Efficient Rational Creative Telescoping

Mark Giesbrecht

Symbolic Computation Group, Cheriton School of Computer Science, University of Waterloo,
Waterloo, ON, N2L 3G1, Canada

Hui Huang

School of Mathematical Sciences, Dalian University of Technology,
Dalian, Liaoning, 116024, China

George Labahn

Symbolic Computation Group, Cheriton School of Computer Science, University of Waterloo,
Waterloo, ON, N2L 3G1, Canada

Eugene Zima

Physics and Computer Science, Wilfrid Laurier University,
Waterloo, ON, N2L 3C5, Canada

Abstract

We present a new algorithm to compute minimal telescopers for rational functions in two discrete variables. As with recent reduction-based approaches, our algorithm has the important feature that the computation of a telescoper is independent of its certificate. In addition, our algorithm uses a compact representation of the certificate, which allows it to be easily manipulated and analyzed without knowing the precise expanded form. This representation hides potential expression swell until the final (and optional) expansion, which can be accomplished in time polynomial in the size of the expanded certificate. A complexity analysis, along with a Maple implementation, indicates that our algorithm has better theoretical and practical performance than the reduction-based approach in the rational case.

Keywords: Rational function, GGSZ reduction, Left scalar division with remainder, Telescoper

1. Introduction

Creative telescoping is a powerful method pioneered by Zeilberger (1990a,b, 1991) in the 1990s and has now become the cornerstone for finding closed forms for definite sums and definite integrals in computer algebra. The method mainly constructs a recurrence (resp. differential) equation admitting the prescribed definite sum (resp. integral) as a solution. Employing other

algorithms applicable to the resulting recurrence or differential equation, it is then possible to find closed form solutions or prove that there is no such solution. In the latter case, one can still make use of creative telescoping for such operations as determining asymptotic expansions about the sum or integral under investigation.

In the case of summation, in order to compute a sum of the form $\sum_{y=a}^{b} f(x, y)$, the main task of creative telescoping consists of constructing polynomials $c_0, \ldots, c_\rho$ in x, not all zero, and another function g in the same domain as f such that

$$c_\rho(x)f(x + \rho, y) + \cdots + c_1(x)f(x + 1, y) + c_0(x)f(x, y) = g(x, y + 1) - g(x, y). \qquad (1.1)$$

The number ρ may or may not be part of the input. If $c_0, \ldots, c_\rho$ and g are as above, then we say that $L = c_\rho S_x^\rho + \cdots + c_1 S_x + c_0$ with S_x being the shift operator in x is a *telescoper* for f and g is a *certificate* for L. If $c_\rho \neq 0$ then the integer ρ is the *order* of L. Finally, the maximum degree in x among the polynomials c_ℓ is the *degree* of L.

The technique of creative telescoping has seen various generalizations and improvements over the past two decades. At the present time, the so-called reduction-based approach originating from (Bostan et al., 2010) has drawn the most attention, as it is both efficient in practice and equipped with the useful feature that it allows one to find a telescoper without necessarily also computing the corresponding certificate. In other words, the computation of the c_ℓ in (1.1) is separated from the computation of g. In a typical situation where the size of the c_ℓ is much smaller than the size of g and the right-hand side of (1.1) collapses to zero when summing over the defining interval, this approach enables one to merely compute the c_ℓ avoiding the costly yet unnecessary computation of the certificate g. In applications where a certificate is required, the approach also allows one to express the certificate as an unnormalized sum so that the summands are concatenated symbolically without actually calculating the sum. These summands are often of much smaller sizes than the original certificate. So far, the reduction-based approach has been worked out for many special functions. We refer to (Chen, 2019) for an excellent exposition of all these algorithms.

However, it is also the case that the unnormalized expression for the certificate returned by the reduction-based approach can introduce superfluous terms which eventually cancel out when normalized. These terms will not contribute to the final output but will increase sizes of intermediate results and thus deteriorate the performance of the approach in these applications. In order to illustrate this issue, let us consider a simple discrete rational function of the form

$$f(x, y) = \frac{x}{x + 3y + 3m} - \frac{x}{x + 3y + 3} + \frac{x}{x + 3y}, \qquad (1.2)$$

where m is an integer greater than one. Applying a reduction method, for example, in (Abramov, 1975), to the given rational function f yields

$$f(x, y) = g_0(x, y + 1) - g_0(x, y) + r_0 \quad \text{with } g_0(x, y) = \sum_{k=1}^{m-1} \frac{x}{x + 3y + 3k} \quad \text{and} \quad r_0 = \frac{x}{x + 3y}, \quad (1.3)$$

where r_0 has the denominator of lowest possible degree in y. Based on the form (1.3), iteratively applying the chosen reduction method to each $f(x + \ell, y)$ for $\ell \geq 0$ gives

$$f(x + \ell, y) = g_\ell(x, y + 1) - g_\ell(x, y) + r_\ell \quad \text{with } r_\ell = \frac{x + \ell}{x + 3y + \bar\ell},$$

2

where $\bar{\ell} \in \{0, 1, 2\}$ is ℓ reduced modulo 3 and

$$g_\ell(x, y) = g_0(x + \ell, y) + \sum_{k=1}^{\lfloor \ell/3 \rfloor} \frac{x + \ell}{x + 3y + 3(k - 1) + \bar{\ell}}.$$

Finding a linear dependency amongst the r_ℓ reduces to solving the following linear system

$$\begin{pmatrix} 9x & 9x + 9 & 9x + 18 & 9x + 27 \\ 6x^2 + 9x & 6x^2 + 12x + 6 & 6x^2 + 15x + 6 & 6x^2 + 27x + 27 \\ x^3 + 3x^2 + 2x & x^3 + 3x^2 + 2x & x^3 + 3x^2 + 2x & x^3 + 6x^2 + 11x + 6 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \qquad (1.4)$$

A nontrivial polynomial solution $(c_0, c_1, c_2, c_3) = (-(x + 3), 0, 0, x)$ then gives

$$L = x\, S_x^3 - (x + 3), \qquad (1.5)$$

a telescoper for f of minimal order with a corresponding certificate

$$g(x, y) = x \cdot g_3(x, y) - (x + 3) \cdot g_0(x, y) = \frac{x(x + 3)}{x + 3y + 3m} - \frac{x(x + 3)}{x + 3y + 3} + \frac{x(x + 3)}{x + 3y} \qquad (1.6)$$

obtained by canceling out the common $m - 2$ terms in the summation. As the m increases, the size of each g_ℓ grows rapidly, whereas the expanded certificate g may still be small. In this particular example, it is actually more reasonable to use the decomposition

$$f(x, y) = g_0(x, y + 1) - g_0(x, y) + r_0, \quad \text{with } g_0(x, y) = -\frac{x}{x + 3y} \text{ and } r_0 = \frac{x}{x + 3y + 3m},$$

instead of (1.3). This leads to an alternate choice of r_ℓ for each $f(x + \ell, y)$, with the corresponding g_ℓ having the denominator of much smaller degree in y. With this choice one gets the same telescoper L and the same certificate g as before, but this time there is no cancellation happening in (1.6). That is, the unnormalized sum gives the final size of the certificate. This suggests a solution to the above issue. Namely, find an initial decomposition (1.3) with both r_0 and g_0 having denominators of lowest possible degrees in y using the method proposed in (Polyakov, 2011; Zima, 2011) to initiate the iterative process of the reduction-based approach. However this process requires a full irreducible factorization of a polynomial.

Separate from the previously mentioned work, there is an alternate method developed by Le (2003) which constructs telescopers in a direct fashion. This method was later used by Chen and Kauers (2012) to obtain the best order-degree curve known so far for telescopers of bivariate rational functions. Currently, the method has only been worked out for bivariate rational functions in the $(q$-$)$shift case. Nevertheless, the method is still interesting because it also has the feature that the computation of a telescoper does not depend on its certificate. In order to demonstrate its main idea, consider again the rational function f given in (1.2). As with the reduction-based approach, this method first decomposes f as in (1.3). The difference is that it later decomposes r_0 as the sum of several simple fractions of numerators in x only, which in our example is merely $x\left(\frac{1}{x+3y}\right)$. By viewing $x = x\, S_x^0$ as a recurrence operator of order zero and using the fact that $S_x^3 - 1$ is a minimal telescoper for $\frac{1}{x+3y}$ with a corresponding certificate $\frac{1}{x+3y}$, Le's method then computes the least common left multiple of x and $S_x^3 - 1$ with the left cofactor of x (resp. $S_x^3 - 1$) giving rise to the same telescoper L as in (1.5) (resp. its certificate $\frac{x(x+3)}{x+3y}$) for the simple fraction

3

$x\left(\frac{1}{x+3y}\right) = r_0$. In the more general case where there is more than one simple fraction in r_0, one finds a telescoper of minimal order for r_0 by calculating the least common left multiple of all telescopers for individual simple fractions. Together with (1.3), the method yields a telescoper of minimal order for f, namely L, as well as its (optional) certificate of the form

$$g = L(g_0) + \frac{x(x+3)}{x+3y}.$$

Rather than leaving the certificate as a (potentially large) unnormalized sum as done by the reduction-based approach, this method represents the certificate by recurrence operators. This representation enables one to more easily manipulate the certificate or analyze its various properties such as the singularities without knowing its expanded form. However, the intermediate expression swell which happens in the certificate is still unavoidable due to (1.3). A second disadvantage is that this method requires the numerator of each simple fraction appearing in the decomposition to be independent of y, often requiring one to work in algebraic extensions of the base field.

1.1. Proposed new approach

Our new algorithm constructs a telescoper for a rational function in a similar fashion as the reduction-based approach, but incorporating the idea from the method of Le (2003). As a result, our algorithm completely avoids algebraic extensions of the base field and intermediate expression swell in the certificate. In order to describe the main idea of our algorithm, let us continue the example (1.2). Unlike the reduction-based approach and the method of Le, we first find a recurrence operator M allowing us to rewrite f in the form

$$f = \underbrace{(x\,S_x^{3m} - x\,S_x^3 + x)}_{M}\left(\frac{1}{x+3y}\right).$$

Assume that we want to find a telescoper for f of order no more than $\rho \in \mathbb{N}$, say $\rho = 3$. We then make an ansatz $L = c_3 S_x^3 + c_2 S_x^2 + c_1 S_x + c_0$ with c_0, c_1, c_2, c_3 to be determined. Using the relation $S_x^3(x+3y) = S_y(x+3y)$ with S_y being the shift operator in y, we calculate the left scalar remainder

$$R = (x+2)c_2 S_x^2 + (x+1)c_1 S_x + ((x+3)c_3 + xc_0)$$

from a so-called left scalar division of $L \odot M$ by $S_y - 1$, where $L \odot M$ is the multiplication of M by L from the left-hand side modulo the left ideal generated by $S_y - S_x^3$ (see Section 3 for a precise definition). We show that L is a telescoper if and only if $R = 0$. The problem is then reduced to solving the following linear system

$$\begin{pmatrix} x & 0 & 0 & x+3 \\ 0 & 0 & x+2 & 0 \\ 0 & x+1 & 0 & 0 \end{pmatrix}\begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{1.7}$$

One immediately reads a nontrivial polynomial solution $(c_0, c_1, c_2, c_3) = (-(x+3), 0, 0, x)$, which yields the telescoper L given by (1.5). In terms of certificate, we either follow the idea from (Gerhard et al., 2003) and use the compact representation

$$g = \mathrm{LSQ}(L \odot M, S_y - 1)\left(\frac{1}{x+3y}\right),$$

or expand it as (1.6) by noticing $\mathrm{LSQ}(L \odot M, \mathrm{S}_y - 1) = x(x+3)\mathrm{S}_x^{3m} - x(x+3)\mathrm{S}_x^3 + x(x+3)$, where LSQ denotes the left scalar quotient obtained from the left scalar division.

In the case where the induced linear system admits no nontrivial solutions, we then have shown that there does not exist any telescoper of order no more than ρ for the given rational function. In order to find a telescoper of minimal order, one can execute the above process incrementally by letting $\rho = 0, 1, 2, \ldots$. The termination of the new algorithm is guaranteed by the existence criterion for telescopers of rational functions given in (Abramov and Le, 2002, Theorem 1), which essentially boils down to the integer-linearity of polynomials. In the general case, the operator S_x in M is replaced by a special recurrence operator acting particularly on integer-linear rational functions of one type, and the given rational function is initially separated into several simple fractions according to integer-linear types.

In summary, our main contribution is a new algorithm for computing minimal telescopers for rational functions. As with the reduction-based approach and the method of Le, our algorithm separates the computation of the telescoper from that of the certificate. When the certificate is needed our algorithm computes it in a compact form, hiding potential expression swell until a final, optional expansion. Compared to Le's method, our algorithm avoids the need for algebraic extensions. In addition, if an expanded form for the certificate is desired then it can be computed easily in time polynomial in the size of the expanded certificate. Moreover, comparing (1.7) with (1.4) suggests that our algorithm also has better control for the size of intermediate expressions involved in the computation of the telescoper.

The arithmetic cost of our new algorithm, as well as that of the reduction-based approach in the rational case, is analyzed in this paper. We note that, until recently, most complexity analyses were done for the differential case (Bostan et al., 2010, 2013, 2018; van der Hoeven, 2020) whereas little has been known for the shift case. The complexity analysis shows that our new algorithm is at least one order of magnitude faster than the reduction-based approach in the rational case when the certificate is not expanded. A Maple implementation further confirms that our approach outperforms the reduction-based approach when restricted to the rational case. In addition, the new algorithm is easy to analyze and leads to a tight order-degree curve for telescopers, a property shared with the method of Le.

The remainder of the paper proceeds as follows. Some basic notions and results are recalled in the next section for later use. In particular, two important decompositions of polynomials in the bivariate setting are reviewed. A kind of recurrence operators specifically working on integer-linear rational functions of one type is introduced in Section 3. Based on basic arithmetic for operators of this kind, Section 4 describes a new algorithm to construct a telescoper of minimal order for bivariate rational functions. Section 5 provides a cost analysis of our new algorithm, followed in Section 6 by a brief summary and a cost analysis of the reduction-based approach in the rational case. Section 7 contains some experimental comparison among all above-mentioned approaches. The paper ends with some topics for future research.

2. Preliminaries

Throughout the paper $\mathbb{K}$ denotes a field of characteristic zero with $\mathbb{K}(x, y)$ the field of rational functions in x, y over $\mathbb{K}$. We let σ_x and σ_y be the automorphisms over $\mathbb{K}(x, y)$, which, for any $f \in \mathbb{K}(x, y)$, are defined by

$$\sigma_x(f(x, y)) = f(x+1, y) \quad \text{and} \quad \sigma_y(f(x, y)) = f(x, y+1).$$

A rational function $f \in \mathbb{K}(x, y)$ is called *summable* with respect to y (or σ_y-*summable* for short) if $f = \sigma_y(g) - g$ for some $g \in \mathbb{K}(x, y)$. A nonzero polynomial $f \in \mathbb{K}[x, y]$ is called *shift-free* with respect to y (or σ_y-*free* for short) if $\gcd(f, \sigma_y^{\ell}(f)) \in \mathbb{K}[x]$ for all nonzero integers ℓ.

Let f be a polynomial in $\mathbb{K}[x, y]$. Throughout this paper, we will order terms using a pure lexicographic order with $x \prec y$. For this order, we let $\mathrm{lc}_{x,y}(f)$ denote the leading coefficient of f over $\mathbb{K}$ with respect to x, y. We say that f is *monic* with respect to x, y if $\mathrm{lc}_{x,y}(f) = 1$. In the sequel, unless there is a danger of confusion, we will just say that f is monic, omitting the variables. We also denote by $\deg_x(f)$ and $\deg_y(f)$ the degrees of f with respect to x and y, respectively, following the convention that $\deg_x(0) = \deg_y(0) = -\infty$.

Let $\mathbb{K}(x, y)[\mathrm{S}_x, \mathrm{S}_y]$ be the ring of linear recurrence operators in x, y over $\mathbb{K}(x, y)$, in which the following commutation rules hold: $\mathrm{S}_x\mathrm{S}_y = \mathrm{S}_y\mathrm{S}_x$ and $\mathrm{S}_x f = \sigma_x(f)\mathrm{S}_x$, $\mathrm{S}_y f = \sigma_y(f)\mathrm{S}_y$ for any $f \in \mathbb{K}(x, y)$. The application of an operator $L = \sum_{i,j \geq 0} a_{ij}\mathrm{S}_x^i\mathrm{S}_y^j$ in $\mathbb{K}(x, y)[\mathrm{S}_x, \mathrm{S}_y]$ to a rational function $f \in \mathbb{K}(x, y)$ is then defined as $L(f) = \sum_{i,j \geq 0} a_{ij}\sigma_x^i\sigma_y^j(f)$.

Definition 2.1. *Let f be a rational function in $\mathbb{K}(x, y)$. A nonzero operator $L \in \mathbb{K}[x][\mathrm{S}_x]$ is called a* telescoper *for f if $L(f)$ is σ_y-summable, or equivalently, there exists a rational function $g \in \mathbb{K}(x, y)$ such that*

$$L(f) = (\mathrm{S}_y - 1)(g),$$

where 1 denotes the identity map of $\mathbb{K}(x, y)$. We call g a corresponding certificate *for L. The* order *and* degree *of L are defined to be its degree in S_x and the maximum degree in x of its coefficients with respect to S_x, respectively. A telescoper of minimal order is also called a* minimal telescoper.

In the rest of this section, we introduce two important decompositions of polynomials, both of which will play crucial roles in our later algorithms.

2.1. Shift-homogeneous decomposition and GGSZ reduction

Recall that two polynomials $f, g \in \mathbb{K}[x, y]$ are called *shift-equivalent* with respect to y (or σ_y-*equivalent* for short), denoted by $f \sim_y g$, if $f = \sigma_y^m(g)$ for some $m \in \mathbb{Z}$. Clearly, $\sim_y$ is an equivalence relation. The σ_y-equivalence of two polynomials can be easily recognized by comparing coefficients.

By grouping together its σ_y-equivalent irreducible factors, any polynomial $g \in \mathbb{K}[x, y]$ can be written in the form

$$g = c \prod_{i=1}^{m} \prod_{j=1}^{n_i} \sigma_y^{\nu_{ij}}(g_i)^{e_{ij}}, \tag{2.1}$$

where $c \in \mathbb{K}[x]$, $m, n_i, \nu_{ij}, e_{ij} \in \mathbb{N}$ with $0 = \nu_{i1} < \nu_{i2} < \cdots < \nu_{in_i}$ and $e_{ij} > 0$, $g_i \in \mathbb{K}[x, y]$ is monic, irreducible and of positive degree in y, and the g_i are pairwise σ_y-inequivalent. Since $\mathbb{K}[x, y]$ is a unique factorization domain, the decomposition (2.1) is unique up to the order of factors. In view of this, we call (2.1) the *shift-homogeneous decomposition* of g with respect to y.

We note that in the context of univariate polynomials, the shift-homogeneous decomposition is equivalent to the most refined shiftless decomposition defined in (Gerhard et al., 2003). In the same paper, based on shiftless decompositions, a reduction algorithm for univariate rational functions, named **RatSum**, was developed. This algorithm can be carried over to the case of bivariate rational functions in a straightforward manner, to which we will refer as the GGSZ reduction later for convenience, named after the authors. The input and output of the GGSZ reduction are given below.

GGSZReduction. Given a rational function $f \in \mathbb{K}(x, y)$, compute two rational functions h, r in $\mathbb{K}(x, y)$ with $r = a/b$, $a, b \in \mathbb{K}[x, y]$, $\deg_y(a) < \deg_y(b)$ and b being σ_y-free such that

$$f = (S_y - 1)(h) + r. \tag{2.2}$$

Such a reduction algorithm is vital for many creative telescoping approaches, including the reduction-based one in (Chen et al., 2015), the method of Le (2003) and the algorithm introduced in this paper. Unlike previous reduction algorithms as given in (Abramov, 1975; Paule, 1995), the GGSZ reduction uses a compact representation of h in (2.2) in terms of left quotients (see Example 2.2 for an illustration), and hence works in polynomial-time of the size of the input without the final expansion.

Example 2.2. *Let g be a polynomial of the form*

$$(xy+1)(x(y+1)+1)(x(y+29)+1)(x(y+30)+1)((-5x+2y)^2+1)((-5x+2y+1)^2+1)((3x+10y)^3+1).$$

Then by grouping together σ_y-equivalent irreducible factors, we obtain

$$g = g_0 \sigma_y(g_0) \sigma_y^{29}(g_0) \sigma_y^{30}(g_0) g_1 g_2 g_3 g_4,$$

where $g_0 = xy + 1$, $g_1 = (-5x + 2y)^2 + 1$, $g_2 = (-5x + 2y + 1)^2 + 1$, $g_3 = (3x + 10y) + 1$ and $g_4 = (3x + 10y)^2 - (3x + 10y) + 1$. Up to making g_1, g_2, g_3, g_4 monic, the above equation gives the shift-homogeneous decomposition of g with respect to y.

Let f be a rational function of denominator g and admit the following decomposition

$$\frac{2x + 3}{\sigma_y^{30}(g_0)} - \frac{2x + 3}{\sigma_y^{29}(g_0)} - \frac{1}{\sigma_y(g_0)} + \frac{1}{g_0} + \frac{2x^2 + 1}{(-5x + 2y)^2 + 1} + \frac{x - 1}{(-5x + 2y + 1)^2 + 1} + \frac{xy + 1}{(3x + 10y)^3 + 1}.$$

We remark that all decomposed forms given in our examples are for readability only. Applying the GGSZ reduction to f then yields (2.2) with

$$h = \mathrm{LQ}((2x + 3)S_y^{30} - (2x + 3)S_y^{29} - S_y + 1, S_y - 1)\left(\frac{1}{g_0}\right) = ((2x + 3)S_y^{29} - 1)\left(\frac{1}{g_0}\right)$$

$$\textit{and} \quad r = \frac{2x^2 + 1}{(-5x + 2y)^2 + 1} + \frac{x - 1}{(-5x + 2y + 1)^2 + 1} + \frac{xy + 1}{(3x + 10y)^3 + 1}, \tag{2.3}$$

where LQ denotes the left quotient in the ring $\mathbb{Q}(x, y)[S_y]$. Note that, in this example, the left quotient in h is a sparse operator although it is of relatively high order 29. Hence the expanded form of h is small. Since $r \neq 0$, then f is not σ_y-summable by (Gerhard et al., 2003, Theorem 12). We will use f as a running example in this paper.

2.2. Integer-linear decomposition and its refinement

Recall that an irreducible polynomial $g \in \mathbb{K}[x, y]$ is called *integer-linear* (over $\mathbb{K}$) if it is of the form $p(\lambda x + \mu y)$ for some integers λ, μ and a univariate polynomial $p \in \mathbb{K}[z]$. Note that λ, μ cannot both be zero since g is irreducible and thus nonunit. By pulling out a common factor and absorbing it into p, one may assume without loss of generality that λ, μ are coprime and that $\mu \geq 0$. Such a pair (λ, μ) is unique and is called the *integer-linear type* of g. For the sake of completeness, we let a constant polynomial be integer-linear of type $(0, 0)$. A polynomial in $\mathbb{K}[x, y]$ is then called *integer-linear* (over $\mathbb{K}$) if all its irreducible factors are integer-linear, possibly with different integer-linear types. A rational function in $\mathbb{K}(x, y)$ is called *integer-linear* (over $\mathbb{K}$) if its denominator and numerator are both integer-linear.

Definition 2.3. *Let $g \in \mathbb{K}[x, y]$ be a polynomial admitting the decomposition*

$$g = p_0(x, y) \prod_{i=1}^{m} p_i(\lambda_i x + \mu_i y), \qquad (2.4)$$

where $p_0 \in \mathbb{K}[x, y]$, $m \in \mathbb{N}$, $\lambda_i, \mu_i \in \mathbb{Z}$ and $p_i \in \mathbb{K}[z]$ for $1 \leq i \leq m$. Then (2.4) *is called the* integer-linear decomposition *of g if*

- *none of irreducible factors of p_0 is integer-linear;*

- *$p_1, \ldots, p_m$ are monic and of positive degrees in z;*

- *each (λ_i, μ_i) satisfies $\gcd(\lambda_i, \mu_i) = 1$ and $\mu_i \geq 0$;*

- *any two pairs of the (λ_i, μ_i) are distinct.*

The (λ_i, μ_i) are called integer-linear types *of g. If g is clear from the context, we will simply say that the (λ_i, μ_i) are integer-linear types.*

Clearly, g is integer-linear if and only if $p_0 \in \mathbb{K}$ in (2.4). By the uniqueness of full factorization and integer-linear types, we see that every polynomial admits a unique integer-linear decomposition up to the order of the factors.

In terms of computation, an efficient algorithm for finding integer-linear decompositions of general multivariate polynomials was recently proposed by authors (Giesbrecht et al., 2019). Compared with previous known approaches (Abramov and Le, 2002; Li and Zhang, 2013), this algorithm performs better both in theory and in practice.

Recall that two polynomials $f, g \in \mathbb{K}[x, y]$ are called *shift-equivalent* with respect to x, y (or (σ_x, σ_y)-*equivalent* for short), denoted by $f \sim_{x,y} g$, if there exist $\ell, m \in \mathbb{Z}$ such that $f = \sigma_x^\ell \sigma_y^m(g)$. Clearly, $\sim_{x,y}$ is an equivalence relation and contains the relation $\sim_y$. Suppose that f, g are integer-linear of the forms $f(x, y) = p_1(\lambda_1 x + \mu_1 y)$ and $g(x, y) = p_2(\lambda_2 x + \mu_2 y)$ for $p_i \in \mathbb{K}[z]$ and $\lambda_i, \mu_i \in \mathbb{Z}$ with $\mu_i \geq 0$ and $\gcd(\lambda_i, \mu_i) = 1$. Then $f \sim_{x,y} g$ implies that $(\lambda_1, \mu_1) = (\lambda_2, \mu_2)$ and $p_1(z) = p_2(z + \ell)$ for some $\ell \in \mathbb{Z}$, and conversely. This indicates that for any two integer-linear polynomials of single types, testing their (σ_x, σ_y)-equivalence amounts to checking the equality of the integer-linear types and the shift-equivalence of univariate polynomials.

Let $p \in \mathbb{K}[z]$ be a monic polynomial of positive degree in z, and let (λ, μ) be an integer-linear type with $\mu > 0$. By computing the shift-homogeneous decomposition of p with respect to z, we obtain $p = \prod_{i=1}^{m} \prod_{j=1}^{n_i} \sigma_y^{v_{ij}}(p_i)^{e_{ij}}$, where $m, n_i, v_{ij}, e_{ij} \in \mathbb{N}$ with $0 = v_{i1} < v_{i2} < \cdots < v_{in_i}$ and $e_{ij} > 0$, $p_i \in \mathbb{K}[z]$ is monic and irreducible, and the p_i are pairwise shift-inequivalent with respect to z. It then follows that

$$p(\lambda x + \mu y) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} p_i(\lambda x + \mu y + v_{ij})^{e_{ij}},$$

where the $p_i(\lambda x + \mu y)$ are pairwise (σ_x, σ_y)-inequivalent.

Consider now a polynomial $g \in \mathbb{K}[x, y]$ with the integer-linear decomposition (2.4). For each factor $p_i(\lambda_i x + \mu_i y)$ with $1 \leq i \leq m$ in (2.4), if $\mu_i = 0$ we then absorb it into p_0; otherwise we further split it into distinct (σ_x, σ_y)-equivalence classes using the procedure described in

the preceding paragraph. By relabeling all the resulting factors, we finally derive the following decomposition (with a slight abuse of notation)

$$g = p_0(x, y) \prod_{i=1}^{m} \prod_{j=1}^{n_i} p_i(\lambda_i x + \mu_i y + \nu_{ij})^{e_{ij}}, \qquad (2.5)$$

where $p_0 \in \mathbb{K}[x, y]$, $m, n_i, \nu_{ij}, e_{ij} \in \mathbb{N}$, $\lambda_i, \mu_i \in \mathbb{Z}$ and $p_1, \ldots, p_m \in \mathbb{K}[z]$ satisfying

- none of irreducible factors of p_0 of positive degree in y is integer-linear;

- $p_1, \ldots, p_m$ are monic and irreducible;

- each (λ_i, μ_i) is an integer-linear type with $\mu_i > 0$;

- $p_i(\lambda_i x + \mu_i y) \not\sim_{x,y} p_j(\lambda_j x + \mu_j y)$ for any two integers i, j with $1 \leq i < j \leq m$; or equivalently, either $(\lambda_i, \mu_i) \neq (\lambda_j, \mu_j)$ or p_i is shift-inequivalent with p_j with respect to z;

- $0 = \nu_{i1} < \cdots < \nu_{in_i}$ and $e_{ij} > 0$.

Evidently, the above decomposition is unique up to the order of factors. We will call (2.5) the *refined integer-linear decomposition* of the polynomial g.

Example 2.4. *Let g be the same polynomial as given in Example 2.2. By definition, it is easy to see that g possesses the integer-linear decomposition*

$$g = p_0(x, y)\tilde{p}_1(-5x + 2y)\tilde{p}_2(3x + 10y),$$

where $p_0 = g_0 \sigma_y(g_0)\sigma_y^{29}(g_0)\sigma_y^{30}(g_0)$ with $g_0 = xy + 1$, $\tilde{p}_1(z) = (z^2 + 1)((z + 1)^2 + 1)$ and $\tilde{p}_2(z) = z^3 + 1$. Computing the shift-homogeneous decompositions of $\tilde{p}_1$ and $\tilde{p}_2$ with respect to z then yields the refined integer-linear decomposition

$$g = p_0(x, y)p_1(-5x + 2y)p_1(-5x + 2y + 1)p_2(3x + 10y)p_3(3x + 10y) \qquad (2.6)$$

with $p_1(z) = z^2 + 1$, $p_2(z) = z + 1$ and $p_3(z) = z^2 - z + 1$.

3. Integer-linear operators

In this section, we introduce another vital ingredient of our algorithms, in this case a special recurrence operator specifically acting on integer-linear rational functions of a single type.

By a standard localization at a left Ore set (see (Cohn, 1985, §0.9) or (Rowen, 1988, §3.1)), the ring $\mathbb{K}(x, y)[\mathrm{S}_x, \mathrm{S}_y]$ can be extended to

$$\mathcal{A} := \mathbb{K}(x, y)[\mathrm{S}_x, \mathrm{S}_y, \mathrm{S}_x^{-1}, \mathrm{S}_y^{-1}].$$

Here $\mathrm{S}_x^{-1} f = \sigma_x^{-1}(f)\mathrm{S}_x^{-1}$ and $\mathrm{S}_y^{-1} f = \sigma_y^{-1}(f)\mathrm{S}_y^{-1}$ for all $f \in \mathbb{K}(x, y)$ with $\sigma_x^{-1}, \sigma_y^{-1}$ denoting the inverse maps of the automorphisms σ_x, σ_y, respectively. For an operator $L \in \mathcal{A}$, there exist unique rational functions $a_{ij} \in \mathbb{K}(x, y)$, finitely many nonzero, such that $L = \sum_{i,j \in \mathbb{Z}} a_{ij} \mathrm{S}_x^i \mathrm{S}_y^j$.

In the rest of this section, we fix a pair (λ, μ) of coprime integers with $\mu > 0$. Then there exist unique integers α, β such that

$$\alpha\lambda + \beta\mu = 1, \qquad (3.1)$$

9

with the constraints $0 \leq \alpha < \mu$ and $|\beta| \leq |\lambda|$ if $\lambda \neq 0$, or $\alpha = 0$ and $\beta = 1$ otherwise. Set $S_{\lambda,\mu}$ to be the product $S_x^\alpha S_y^\beta$. Then

$$\mathcal{A}_{\lambda,\mu} := \mathbb{K}(x,y)[S_{\lambda,\mu}, S_{\lambda,\mu}^{-1}]$$

is a subring of $\mathcal{A}$, which consists of all integer-linear operators of type (λ, μ).

We can view $\mathcal{A}_{\lambda,\mu}$ as a left module over $\mathcal{A}$ as follows. Define the left $\mathbb{K}(x,y)$-linear map

$$\phi_{\lambda,\mu}: \begin{array}{ccc} \mathcal{A} & \longrightarrow & \mathcal{A}_{\lambda,\mu} \\ \sum_{i,j \in \mathbb{Z}} a_{ij} S_x^i S_y^j & \mapsto & \sum_{i,j \in \mathbb{Z}} a_{ij} S_{\lambda,\mu}^{i\lambda+j\mu}. \end{array}$$

The image and kernel of $\phi_{\lambda,\mu}$ are determined below.

Proposition 3.1. *The restriction of $\phi_{\lambda,\mu}$ on $\mathcal{A}_{\lambda,\mu}$ is the identity. Consequently, $\phi_{\lambda,\mu}$ is surjective.*

Proof. By (3.1), $\phi_{\lambda,\mu}(S_{\lambda,\mu}^i) = S_{\lambda,\mu}^i$ for all $i \in \mathbb{Z}$, which, together with the definition of $\phi_{\lambda,\mu}$, implies the assertion. $\qquad\square$

Lemma 3.2. *For every $L \in \mathcal{A}$ and $k, \ell \in \mathbb{Z}$, we have $\phi_{\lambda,\mu}(LS_x^k S_y^\ell) = \phi_{\lambda,\mu}(L)\phi_{\lambda,\mu}(S_x^k S_y^\ell)$.*

Proof. A straightforward calculation based on the definition of $\phi_{\lambda,\mu}$ implies that

$$\phi_{\lambda,\mu}((S_x^i S_y^j)(S_x^k S_y^\ell)) = \phi_{\lambda,\mu}(S_x^i S_y^j)\phi_{\lambda,\mu}(S_x^k S_y^\ell) \quad \text{for all } i, j \in \mathbb{Z}.$$

The lemma then follows from the linearity of $\phi_{\lambda,\mu}$. $\qquad\square$

The above lemma does not imply that $\phi_{\lambda,\mu}$ is a ring homomorphism. In fact, one can easily verify that $\phi_{\lambda,\mu}(S_y y) \neq \phi_{\lambda,\mu}(S_y)\phi_{\lambda,\mu}(y)$ provided that $\beta\mu \neq 1$.

Proposition 3.3. *The kernel of $\phi_{\lambda,\mu}$ is the left ideal generated by $S_x - S_{\lambda,\mu}^\lambda$ and $S_y - S_{\lambda,\mu}^\mu$ in $\mathcal{A}$.*

Proof. Let I be the left ideal generated by $S_x - S_{\lambda,\mu}^\lambda$ and $S_y - S_{\lambda,\mu}^\mu$ in $\mathcal{A}$. For any $L \in I$, there are $P, Q \in \mathcal{A}$ such that $L = P(S_x - S_{\lambda,\mu}^\lambda) + Q(S_y - S_{\lambda,\mu}^\mu)$. By Lemma 3.2, $\phi_{\lambda,\mu}(L) = \phi_{\lambda,\mu}(P)\phi_{\lambda,\mu}(S_x - S_{\lambda,\mu}^\lambda) + \phi_{\lambda,\mu}(Q)\phi_{\lambda,\mu}(S_y - S_{\lambda,\mu}^\mu)$. It follows from the definition of $\phi_{\lambda,\mu}$ and Proposition 3.1 that $\phi_{\lambda,\mu}(S_x - S_{\lambda,\mu}^\lambda) = \phi_{\lambda,\mu}(S_y - S_{\lambda,\mu}^\mu) = 0$, and so also $\phi_{\lambda,\mu}(L) = 0$. We have that $I \subset \ker(\phi_{\lambda,\mu})$.

Conversely, we first observe that every $L \in \mathcal{A}$ can be decomposed as $L = M + R$ for some $M \in I$ and $R \in \mathcal{A}_{\lambda,\mu}$. This is because every monomial $S_x^i S_y^j$ in L with $i, j \in \mathbb{Z}$ can be rewritten as $(S_x - S_{\lambda,\mu}^\lambda + S_{\lambda,\mu}^\lambda)^i (S_y - S_{\lambda,\mu}^\mu + S_{\lambda,\mu}^\mu)^j$ and $(S_x - S_{\lambda,\mu}^\lambda), S_{\lambda,\mu}^\lambda, (S_y - S_{\lambda,\mu}^\mu), S_{\lambda,\mu}^\mu$ multiplicatively commute with each other, so expanding the powers yields the desired result. Then $\phi_{\lambda,\mu}(L) = \phi_{\lambda,\mu}(R)$ since $M \in I \subset \ker(\phi_{\lambda,\mu})$. Moreover, $\phi_{\lambda,\mu}(L) = R$ by Proposition 3.1. We see that $L \in \ker(\phi_{\lambda,\mu})$ implies $R = 0$. Hence $\ker(\phi_{\lambda,\mu}) \subset I$. $\qquad\square$

According to Proposition 3.1, $\mathcal{A}/\ker(\phi_{\lambda,\mu})$ is isomorphic to $\mathcal{A}_{\lambda,\mu}$ as additive groups. Furthermore, $\mathcal{A}/\ker(\phi_{\lambda,\mu})$ is a left module over $\mathcal{A}$ by Proposition 3.3. Hence, $\mathcal{A}_{\lambda,\mu}$ can be viewed as a left module over $\mathcal{A}$ as well. Its left scalar multiplication is defined via $\phi_{\lambda,\mu}$ as follows. For all $L \in \mathcal{A}$ and $M \in \mathcal{A}_{\lambda,\mu}$, the result obtained by multiplying L from the left-hand side to M is $\phi_{\lambda,\mu}(LM)$, which is denoted by $L \odot M$ when the pair (λ, μ) is clear from context.

Using the scalar multiplication, we introduce a left division, which will allows us to characterize telescopers and represent certificates in a compact form. To this end, we need to define the notion of orders in $\mathcal{A}_{\lambda,\mu}$. Let $M = \sum_{i=m}^n a_i S_{\lambda,\mu}^i \in \mathcal{A}_{\lambda,\mu}$, where $m, n \in \mathbb{Z}$ with $m \leq n$ and $a_i \in \mathbb{K}(x,y)$ with $a_m a_n \neq 0$. We say that m and n are the lowest and highest orders of M, and denote them by $\text{lord}(M)$ and $\text{hord}(M)$, respectively.

10

Lemma 3.4. *Let $L \in \mathbb{K}(x, y)[S_y, S_y^{-1}]$ and $M \in \mathcal{A}_{\lambda,\mu}$ be two nonzero operators. Then $L \odot M$ is nonzero. Moreover,*

$$\mathrm{lord}(L \odot M) = \mathrm{lord}(\phi_{\lambda,\mu}(L)) + \mathrm{lord}(M) \quad \text{and} \quad \mathrm{hord}(L \odot M) = \mathrm{hord}(\phi_{\lambda,\mu}(L)) + \mathrm{hord}(M).$$

Proof. Let

$$L = \sum_{i=k}^{\ell} a_i S_y^i \in \mathbb{K}(x, y)[S_y, S_y^{-1}] \quad \text{and} \quad M = \sum_{j=m}^{n} b_j S_{\lambda,\mu}^j \in \mathcal{A}_{\lambda,\mu}, \tag{3.2}$$

where $k, \ell, m, n \in \mathbb{Z}$ with $k \leq \ell$ and $m \leq n$, and $a_i, b_j \in \mathbb{K}(x, y)$ with $a_k a_\ell b_m b_n \neq 0$. Then $\phi_{\lambda,\mu}(L) = \sum_{i=k}^{\ell} a_i S_{\lambda,\mu}^{i\mu}$, which is nonzero. Hence, $\mathrm{lord}(\phi_{\lambda,\mu}(L)) = k\mu$ and $\mathrm{hord}(\phi_{\lambda,\mu}(L)) = \ell\mu$. Observe that $S_y^i \odot (f S_{\lambda,\mu}^j) = \sigma_y^i(f) S_{\lambda,\mu}^{j+i\mu}$ for all $f \in \mathbb{K}(x, y)$ and $i, j \in \mathbb{Z}$. Since $a_k a_\ell b_m b_n \neq 0$, then $L \odot M \neq 0$, and thus $\mathrm{lord}(L \odot M) = k\mu + m$ and $\mathrm{hord}(L \odot M) = \ell\mu + n$. $\qquad\square$

Lemma 3.5. *Let $L \in \mathbb{K}(x, y)[S_y]$ with $L \neq 0$ and $M \in \mathcal{A}_{\lambda,\mu}$. Then there exist $Q, R \in \mathcal{A}_{\lambda,\mu}$ such that $M = L \odot Q + R$, and R is either zero or satisfies $0 \leq \mathrm{lord}(R) \leq \mathrm{hord}(R) < \mathrm{hord}(\phi_{\lambda,\mu}(L))$.*

Proof. If $M = 0$, then we set $Q = 0$ and $R = 0$. Otherwise, let L and M be the same as in (3.2) with $k \geq 0$. Then $\mathrm{hord}(\phi_{\lambda,\mu}(L)) = \ell\mu$.

Case 1. Assume that $m \geq 0$. If $n < \ell\mu$, then we set $Q = 0$ and $R = M$. Otherwise, let $f = \sigma_y^{-\ell}(b_n/a_\ell)$. By Lemma 3.4,

$$N := M - L \odot (f S_{\lambda,\mu}^{n-\ell\mu}) = M - (b_n S_{\lambda,\mu}^n + \text{lower terms in } S_{\lambda,\mu}).$$

Thus either $N = 0$ or $0 \leq \mathrm{lord}(N) \leq \mathrm{hord}(N) < n$. If $N = 0$ or $\mathrm{hord}(N) < \ell\mu$, then we are done. Otherwise, we recursively apply the same reduction on N. The conclusion will be reached in a finite number of steps.

Case 2. Assume that $m < 0$. We reduce M to an integer-linear operator which is either zero or of nonnegative lowest order. Let $g = \sigma_y^{-k}(b_m/a_k)$. Again, by Lemma 3.4, $M - L \odot (g S_{\lambda,\mu}^{m-k\mu}) = M - (b_m S_{\lambda,\mu}^m + \text{higher terms in } S_{\lambda,\mu})$, which is either zero or of lower order higher than m. Repeating the above reduction finitely many times, we will obtain $Q_1, R_1 \in \mathcal{A}_{\lambda,\mu}$ such that $M = L \odot Q_1 + R_1$ and either $R_1 = 0$ or $\mathrm{lord}(R_1) \geq 0$. If $R_1 = 0$, then we are done. Otherwise, applying the argument in the first case to R_1 yields the lemma. $\qquad\square$

Theorem 3.6. *Let $L \in \mathbb{K}(x, y)[S_y, S_y^{-1}]$ with $L \neq 0$ and $M \in \mathcal{A}_{\lambda,\mu}$. Then there exist unique $Q, R \in \mathcal{A}_{\lambda,\mu}$ such that $M = L \odot Q + R$, and R is either zero or satisfies*

$$\mathrm{lord}(\phi_{\lambda,\mu}(L)) \leq \mathrm{lord}(R) \leq \mathrm{hord}(R) < \mathrm{hord}(\phi_{\lambda,\mu}(L)).$$

Proof. Let L be given as in (3.2). If $k \geq 0$, then the existence of Q and R follows from Lemma 3.5. Assume that $k < 0$. The same lemma implies that there exist $\tilde{Q}, \tilde{R} \in \mathcal{A}_{\lambda,\mu}$ such that $S_y^{-k} \odot M = (S_y^{-k}L) \odot \tilde{Q} + \tilde{R}$. In addition, either $\tilde{R} = 0$ or $0 \leq \mathrm{lord}(\tilde{R}) \leq \mathrm{hord}(\tilde{R}) < \mathrm{hord}(\phi_{\lambda,\mu}(S_y^{-k}L))$. It follows that $M = L \odot \tilde{Q} + S_y^k \odot \tilde{R}$. Assume that $\tilde{R}$ is nonzero. Then $\mathrm{lord}(S_y^k \odot \tilde{R}) \geq \mathrm{lord}(\phi_{\lambda,\mu}(L))$ by Lemma 3.4 and the fact that $\mathrm{lord}(\tilde{R}) \geq 0$. Moreover, $\mathrm{hord}(S_y^k \odot \tilde{R}) < \mathrm{hord}(\phi_{\lambda,\mu}(L))$ by Lemma 3.4 and the fact that $\mathrm{hord}(\tilde{R}) < \mathrm{hord}(\phi_{\lambda,\mu}(S_y^{-k}L))$. Setting $Q = \tilde{Q}$ and $R = S_y^k \odot \tilde{R}$ establishes the existence of Q and R.

To show the uniqueness, we let $\tilde{Q}, \tilde{R} \in \mathcal{A}_{\lambda,\mu}$ be such that $M = L \odot \tilde{Q} + \tilde{R}$, and $\tilde{R}$ is either zero or satisfies $\mathrm{lord}(\phi_{\lambda,\mu}(L)) \leq \mathrm{lord}(\tilde{R}) \leq \mathrm{hord}(\tilde{R}) < \mathrm{hord}(\phi_{\lambda,\mu}(L))$. Then $L \odot (Q - \tilde{Q}) = \tilde{R} - R$. Suppose

11

that $Q \neq \tilde{Q}$. Then $\tilde{R} \neq R$ by Lemma 3.4. Suppose that $\mathrm{hord}(Q - \tilde{Q}) \geq 0$. By Lemma 3.4 and the fact that $\mathrm{hord}(\phi_{\lambda,\mu}(L)) > \mathrm{hord}(\tilde{R} - R)$, we have $\mathrm{hord}(L \odot (Q - \tilde{Q})) > \mathrm{hord}(\tilde{R} - R)$, a contradiction. Otherwise, a similar argument yields $\mathrm{lord}(L \odot (Q - \tilde{Q})) < \mathrm{lord}(\tilde{R} - R)$, a contradiction. Hence $Q = \tilde{Q}$ and then $R = \tilde{R}$. $\qquad\square$

In view of the above theorem, we call Q the *left scalar quotient* and R the *left scalar remainder* of M by L, and denote them by $\mathrm{LSQ}(M, L)$ and $\mathrm{LSR}(M, L)$, respectively.

Remark 3.7. *It is possible to extend Theorem 3.6 to the general case when the scalar divisor L is an arbitrary nonzero operator in $\mathcal{A}$. However, as doing this extension is somewhat tedious and as this extension is not used in the paper we do not investigate this aspect further.*

Remark 3.8. *We are particularly interested in the case where the difference operator $\mathrm{S}_y - 1$ plays the part of a scalar divisor. For later reference, we collect below explicit formulas for left scalar remainders, as well as for left scalar quotients, in this case.*

Let $M = \sum_{i=m}^{n} a_i \mathrm{S}_{\lambda,\mu}^i \in \mathcal{A}_{\lambda,\mu}$, where $m, n \in \mathbb{Z}$ with $m \leq n$ and $a_i \in \mathbb{K}(x, y)$. Then

$$\mathrm{LSR}(M, \mathrm{S}_y - 1) = \sum_{r=0}^{\mu-1} \left(\sum_{i_r} \sigma_y^{-q_{i_r}}(a_{i_r}) \right) S_{\lambda,\mu}^r, \tag{3.3}$$

where the inner summation runs over all integers i_r with $m \leq i_r \leq n$ such that $i_r = \mu q_{i_r} + r$ for some integer q_{i_r}, and

$$\mathrm{LSQ}(M, \mathrm{S}_y - 1) = - \sum_{j=m}^{-1} \left(\sum_{i_j} \sigma_y^{-q_{i_j}}(a_{i_j}) \right) \mathrm{S}_{\lambda,\mu}^j + \sum_{j=0}^{n-\mu} \left(\sum_{i_j} \sigma_y^{-q_{i_j}}(a_{i_j}) \right) \mathrm{S}_{\lambda,\mu}^j,$$

where the first inner summation runs over all integers i_j with $m \leq i_j \leq n$ such that $i_j = \mu q_{i_j} + j$ for some nonpositive integer q_{i_j}, while the second inner summation runs over all integers i_j with $m \leq i_j \leq n$ such that $i_j = \mu q_{i_j} + j$ for some positive integer q_{i_j}.

Example 3.9. *Let $M = (x - 1)\mathrm{S}_{-5,2} + (2x^2 + 1)$ with $\mathrm{S}_{-5,2} = \mathrm{S}_x \mathrm{S}_y^3$. Let L be an operator in $\mathbb{Q}[x][\mathrm{S}_x]$ of the form $L = c_2 \mathrm{S}_x^2 + c_1 \mathrm{S}_x + c_0$ for some $c_0, c_1, c_2 \in \mathbb{Q}[x]$. Multiplying L from the left-hand side to M yields*

$$L \odot M = c_0(x-1)\mathrm{S}_{-5,2} + c_0(2x^2+1) + c_1 \sigma_x(x-1)\mathrm{S}_{-5,2}^{-4} + c_1 \sigma_x(2x^2+1)\mathrm{S}_{-5,2}^{-5}$$
$$+ c_2 \sigma_x^2(x-1)\mathrm{S}_{-5,2}^{-9} + c_2 \sigma_x^2(2x^2+1)\mathrm{S}_{-5,2}^{-10}. \tag{3.4}$$

A direct calculation based on Remark 3.7 then delivers

$$\mathrm{LSR}(L \odot M, \mathrm{S}_y - 1) = a_1 \mathrm{S}_{-5,2} + a_0, \tag{3.5}$$

where

$$a_1 = c_0(x-1) + c_1(\sigma_y^3 \sigma_x(2x^2+1)) + c_2(\sigma_y^5 \sigma_x^2(x-1))$$
$$\text{and} \quad a_0 = c_0(2x^2+1) + c_1(\sigma_y^2 \sigma_x(x-1)) + c_2(\sigma_y^5 \sigma_x^2(2x^2+1)).$$

We note that $L \odot M$ is a sparse operator by (3.4); the left scalar quotient $\mathrm{LSQ}(L \odot M, \mathrm{S}_y - 1)$, however, is a dense operator with exponents in $\mathrm{S}_{-5,2}$ ranging consecutively from -10 to -1.

12

4. Telescoping with compact certificates

In this section, we demonstrate how to construct a telescoper for a given rational function, along with its certificate in a compact form, using left scalar divisions of integer-linear operators introduced in the preceding section.

For an operator $L = \sum_{i,j \in \mathbb{Z}} a_{ij} S_x^i S_y^j \in \mathcal{A}$ and a rational function $f \in \mathbb{K}(x, y)$, the application of L to f is defined to be

$$L(f) = \sum_{i,j \in \mathbb{Z}} a_{ij} \sigma_x^i(f) \sigma_y^j(f).$$

Let (λ, μ) be a pair of coprime integers with $\mu > 0$, and $g \in \mathbb{K}(x, y)$ of the form $g = p(\lambda x + \mu y)$ with $p \in \mathbb{K}(z)$. Then

$$S_{\lambda,\mu}^i(g) = p(\lambda x + \mu y + i) \quad \text{for all } i \in \mathbb{Z}.$$

It follows that $S_x(g) = S_{\lambda,\mu}^\lambda(g)$ and $S_y(g) = S_{\lambda,\mu}^\mu(g)$. Thus, for all $L \in \mathcal{A}$, we have that $L(g) = \phi_{\lambda,\mu}(L)(g)$. Assume further that $M \in \mathcal{A}_{\lambda,\mu}$. Then

$$LM(g) = (L \odot M)(g), \tag{4.1}$$

which allows us to describe telescopers and their corresponding certificates in terms of module-theoretic language.

Let $f \in \mathbb{K}(x, y)$ be a rational function with denominator $g \in \mathbb{K}[x, y]$. Based on the refined integer-linear decomposition (2.5) of g, there is a unique partial fraction decomposition of f with respect to y, that is, there exist unique $a_0, a_{ijk} \in \mathbb{K}(x)[y]$ with $\deg_y(a_{ijk}) < \deg_z(p_i)$ such that

$$f = \frac{a_0}{p_0} + \sum_{i=1}^m \sum_{j=1}^{n_i} \sum_{k=1}^{e_{ij}} \frac{a_{ijk}}{p_i(\lambda_i x + \mu_i y + \nu_{ij})^k}. \tag{4.2}$$

Let $d_i = \max_{1 \le j \le n_i}\{e_{ij}\}$ and specify that $a_{ijk} = 0$ in case $k > e_{ij}$. Interchanging the order of summations in (4.2) and introducing the operator $M_{ik} = \sum_{j=1}^{n_i} a_{ijk} S_{\lambda_i,\mu_i}^{\nu_{ij}}$ then gives

$$f = \frac{a_0}{p_0} + \sum_{i=1}^m \sum_{k=1}^{d_i} M_{ik}\left(\frac{1}{p_i(\lambda_i x + \mu_i y)^k}\right). \tag{4.3}$$

Note that $M_{ik} \in \mathbb{K}(x)[y, S_{\lambda_i,\mu_i}]$ and $\deg_y(M_{ik}) < \deg_z(p_i)$ for all $i = 1, \ldots, m$ and $k = 1, \ldots, d_i$. Using the above argument in the opposite direction, one can easily derive the partial fraction decomposition (4.2) from (4.3). It thus follows from the uniqueness of (4.2) that (4.3) is unique. In particular, the operators M_{ik} are uniquely determined by the given rational function f. We will refer to (4.3) as the *RILD-based partial fraction decomposition* of f.

4.1. The basic case

In order to illustrate the main idea of our algorithm in a concise way, we first focus on the simpler yet important subcase when the given rational function $f \in \mathbb{K}(x, y)$ possesses the form

$$f = \sum_{i \in \mathbb{Z}} \frac{a_i}{p(\lambda x + \mu y + i)^k} = M\left(\frac{1}{p(\lambda x + \mu y)^k}\right), \tag{4.4}$$

where $\lambda, \mu, k \in \mathbb{Z}$ with $\gcd(\lambda, \mu) = 1$ and $\mu, k > 0$, $p \in \mathbb{K}[z]$ is monic and irreducible, $a_i \in \mathbb{K}(x)[y]$, finitely many nonzero, with $\deg_y(a_i) < \deg_z(p)$, and $M = \sum_{i \in \mathbb{Z}} a_i S_{\lambda,\mu}^i \in \mathbb{K}(x)[y, S_{\lambda,\mu}, S_{\lambda,\mu}^{-1}]$. Note that such a function f has a telescoper by the criterion of Abramov and Le (2002, Theorem 1).

13

Proposition 4.1. *Let $f \in \mathbb{K}(x, y)$ be of the form* (4.4)*, and let $L \in \mathbb{K}[x][S_x]$ be a nonzero operator. Then L is a telescoper for f if and only if there exists $Q \in \mathcal{A}_{\lambda,\mu}$ such that $L \odot M = (S_y - 1) \odot Q$. When this is the case, a certificate corresponding to L is given by*

$$Q\left(\frac{1}{p(\lambda x + \mu y)^k}\right).$$

Proof. Assume that L is a telescoper for f. Then there exists $h \in \mathbb{K}(x, y)$ such that

$$L(f) = (L \odot M)\left(\frac{1}{p(\lambda x + \mu y)^k}\right) = (S_y - 1)(h),$$

where the first equality follows by (4.1). From the uniqueness of the RILD-based partial fraction decomposition we have that there exists $Q \in \mathbb{K}(x)[y, S_{\lambda,\mu}, S_{\lambda,\mu}^{-1}]$ with $\deg_y(Q) < \deg_z(p)$ such that

$$h = Q\left(\frac{1}{p(\lambda x + \mu y)^k}\right).$$

It thus follows that

$$(L \odot M)\left(\frac{1}{p(\lambda x + \mu y)^k}\right) = ((S_y - 1) \odot Q)\left(\frac{1}{p(\lambda x + \mu y)^k}\right).$$

Again, by the uniqueness of the RILD-based partial fraction decomposition, we find that $L \odot M = (S_y - 1) \odot Q$. Applying both sides of this equality to $1/p(\lambda x + \mu y)^k$ proves the converse. $\qquad\square$

Based on Proposition 4.1, we can compute a telescoper for a rational function $f \in \mathbb{K}(x, y)$ of the form (4.4) as follows. Assume that we want to find a telescoper of order no more than $\rho \in \mathbb{N}$. Making an ansatz $L = c_\rho S_x^\rho + \cdots + c_0$ with $c_0, \ldots, c_\rho \in \mathbb{K}[x]$ to be determined, we first compute the left scalar remainder R of $L \odot M$ by $S_y - 1$. Note that $R \in \mathbb{K}(x)[y, S_{\lambda,\mu}, S_{\lambda,\mu}^{-1}]$. Sending R to zero, we then obtain a linear homogeneous system in $c_0, \ldots, c_\rho$ over $\mathbb{K}(x)$. Any nontrivial solution of this system over $\mathbb{K}[x]$ will give rise to a desired telescoper. If no such solutions can be found, then there does not exist any telescoper of order no more than ρ. In order to compute a minimal telescoper for f, one may apply the above procedure incrementally with $\rho = 0, 1, \ldots$. The termination of this process is assured by the existence of telescopers for such an f. With a telescoper L for the given rational function f at hand, by Proposition 4.1, a corresponding certificate can be formally represented as $\mathrm{LSQ}(L \odot M, S_y - 1)\left(\frac{1}{p(\lambda x + \mu y)^k}\right)$.

Example 4.2. *Consider the rational function f of the form*

$$f = \frac{2x^2 + 1}{(-5x + 2y)^2 + 1} + \frac{x - 1}{(-5x + 2y + 1)^2 + 1}.$$

A simple calculation shows that the RILD-based partial fraction decomposition of f is given by

$$f = \underbrace{((x - 1)S_{-5,2} + 2x^2 + 1)}_{M}\left(\frac{1}{p(-5x + 2y)}\right) \quad \text{with } S_{-5,2} = S_x S_y^3 \text{ and } p(z) = z^2 + 1.$$

Let $L = c_2 S_x^2 + c_1 S_x + c_0$ with $c_0, c_1, c_2 \in \mathbb{Q}[x]$ to be determined. By Example 3.9, we know that the left scalar remainder R of $L \odot M$ by $S_y - 1$ is given by (3.5)*. Sending R to zero then delivers*

14

the following linear homogeneous system

$$
\begin{pmatrix} x - 1 & \sigma_y^3 \sigma_x (2x^2 + 1) & \sigma_y^5 \sigma_x^2 (x - 1) \\ 2x^2 + 1 & \sigma_y^2 \sigma_x (x - 1) & \sigma_y^5 \sigma_x^2 (2x^2 + 1) \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{4.5}
$$

Solving this system over $\mathbb{Q}[x]$ gives a telescoper

$$
L = (4x^4 + 8x^3 + 7x^2 + 5x + 3)S_x^2 + 2(2x^2 - 5)S_x - (4x^4 + 24x^3 + 55x^2 + 59x + 27)
$$

and then a corresponding certificate

$$
h = \mathrm{LSQ}(L \odot M, S_y - 1)\left(\frac{1}{p(-5x + 2y)} \right).
$$

We note that L is actually a telescoper for f of minimal order.

4.2. The general case

We now turn our attention to the general case, namely the case when the input is an arbitrary rational function in $\mathbb{K}(x, y)$. Let $f \in \mathbb{K}(x, y)$ be a rational function admitting the RILD-based partial fraction decomposition (4.3). By (Abramov and Le, 2002, Theorem 1), f has a telescoper if and only if a_0/p_0 in (4.3) is σ_y-summable. Thus it suffices to construct a telescoper for $r := f - a_0/p_0$, which possesses the following form

$$
r = \sum_{i=1}^{m} \sum_{k=1}^{d_i} M_{ik}\left(\frac{1}{p_i(\lambda_i x + \mu_i y)^k} \right), \tag{4.6}
$$

where each (λ_i, μ_i) is a pair of coprime integers with $\mu_i > 0$, each $p_i \in \mathbb{K}[z]$ is monic and irreducible, the $p_i(\lambda_i x + \mu_i y)$ are pairwise (σ_x, σ_y)-inequivalent, and each $M_{ik} \in \mathbb{K}(x)[y, S_{\lambda_i, \mu_i}]$ with $\deg_y(M_{ik}) < \deg_z(p_i)$.

There are two natural ways to proceed. The first method separately takes each simple fraction $M_{ik}\left(\frac{1}{p_i(\lambda_i x + \mu_i y)^k} \right)$ in (4.6) as the basic case and computes its own minimal telescoper $L_{ik} \in \mathbb{K}[x][S_x]$ using the approach presented in the preceding subsection, and then returns the least common left multiple of all these L_{ik} as the output. By taking use of (Le, 2003, Theorem 2), one can show that this least common left multiple gives a minimal telescoper for r (and thus for f). Preliminary experiments, however, suggest that in practice this method does not perform as well as expected. In fact, it is often less efficient than the second method which we are going to explore shortly.

This second method shares exactly the same spirit as the basic case given in the preceding subsection, in the sense that it also reduces the problem of constructing a telescoper to the problem of computing left scalar remainders of integer-linear operators.

Theorem 4.3. *Let $r \in \mathbb{K}(x, y)$ be a rational function possessing the decomposition* (4.6)*, and let $L \in \mathbb{K}[x][S_y]$ be a nonzero operator. Then L is a telescoper for r if and only if, for all $i = 1, \ldots, m$ and $k = 1, \ldots, d_i$, there exist operators $Q_{ik} \in \mathcal{A}_{\lambda_i, \mu_i}$ such that $L \odot M_{ik} = (S_y - 1) \odot Q_{ik}$. When this is the case, a corresponding certificate of L is given by*

$$
\sum_{i=1}^{m} \sum_{k=1}^{d_i} Q_{ik}\left(\frac{1}{p_i(\lambda_i x + \mu_i y)^k} \right). \tag{4.7}
$$

15

Proof. Assume that L is a telescoper for r. Then there exists $h \in \mathbb{K}(x, y)$ such that $L(r) = (S_y - 1)(h)$. By (4.1) and (4.6), we have

$$\sum_{i=1}^{m} \sum_{k=1}^{d_i} (L \odot M_{ik}) \left(\frac{1}{p_i(\lambda_i x + \mu_i y)^k} \right) = (S_y - 1)(h).$$

The RILD-based partial fraction decomposition of the left-hand side in the above equality implies that the same decomposition of h is of the form (4.7), in which $Q_{ik} \in \mathbb{K}(x)[y, S_{\lambda_i, \mu_i}, S_{\lambda_i, \mu_i}^{-1}]$ with $\deg_y(Q_{ik}) < \deg_z(p_i)$. The uniqueness of the RILD-based partial fraction decomposition then forces

$$L \odot M_{ik} = (S_y - 1) \odot Q_{ik} \quad \text{for all } i = 1, \ldots, m \text{ and } k = 1, \ldots, d_i.$$

Conversely, we apply L to r. By (4.1) and (4.6),

$$L(r) = \sum_{i=1}^{m} \sum_{k=1}^{d_i} (L \odot M_{ik}) \left(\frac{1}{p_i(\lambda_i x + \mu_i y)^k} \right) = \sum_{i=1}^{m} \sum_{k=1}^{d_i} ((S_y - 1) \odot Q_{ik}) \left(\frac{1}{p_i(\lambda_i x + \mu_i y)^k} \right).$$

It follows that $L(r) = (S_y - 1)(h)$, where h is given by the formula (4.7). $\square$

In analogy to the basic case, the above theorem induces an iterative strategy to compute a telescoper for a given rational function.

Putting this all together, we obtain a new creative telescoping algorithm for rational functions.

RationalCT. Given a rational function $f \in \mathbb{K}(x, y)$, compute a minimal telescoper $L \in \mathbb{K}[x][S_x]$ for f and a corresponding certificate $h \in \mathbb{K}(x, y)$ if telescopers exist. The steps are:

1. Compute the RILD-based partial fraction decomposition of f to get (4.3).
2. Apply the GGSZ reduction to a_0/p_0 in (4.3) to find $h, r \in \mathbb{K}(x, y)$ with h being of a compact form such that
$$\frac{a_0}{p_0} = (S_y - 1)(h) + r. \tag{4.8}$$
3. If $r \neq 0$ then return "No telescoper exists!".
4. For $i = 1, \ldots, m$ and $k = 1, \ldots, d_i$ set $R_{ik} = 0$.

   For $\ell = 0, 1, 2, \ldots$ do

   4.1 For $i = 1, \ldots, m$ and $k = 1, \ldots, d_i$ do

       4.1.1 Compute the left scalar remainder $\bar{R}$ of $S_x^\ell \odot M_{ik}$ by $S_y - 1$.

       4.1.2 Update R_{ik} to be $R_{ik} + c_\ell \bar{R}$, where c_ℓ is an indeterminate.

   4.2 Find $c_0, \ldots, c_\ell \in \mathbb{K}[x]$ such that $R_{ik} = 0$ for all $i = 1, \ldots, m$ and $k = 1, \ldots, d_i$, by solving a linear system in $c_0, \ldots, c_\ell$ over $\mathbb{K}[x]$. If there is a nontrivial solution, set $L = \sum_{j=0}^{\ell} c_j S_x^j$ and return

$$\left(L, L(h) + \sum_{i=1}^{m} \sum_{k=1}^{d_i} \text{LSQ}(L \odot M_{ik}, S_y - 1) \left(\frac{1}{p_i(\lambda_i x + \mu_i y)^k} \right) \right).$$

Theorem 4.4. *Let f be a rational function in $\mathbb{K}(x, y)$. Then the algorithm* **RationalCT** *terminates and correctly finds a minimal telescoper for f and a corresponding certificate in a compact form when such telescopers exist.*

16

Proof. By (Abramov and Le, 2002, Theorem 1), f has a telescoper if and only if a_0/p_0 in (4.3) is σ_y-summable, which, according to (Gerhard et al., 2003, Theorem 12), is equivalent to the condition that $r = 0$ in (4.8). Thus steps 1-3 are correct.

For $\ell = 0$, It is evident that R_{ik} obtained in step 4.1 is equal to $\mathrm{LSR}(c_0 \odot M_{ik}, S_y - 1)$ for all $i = 1, \ldots, m$ and $k = 1, \ldots, d_i$. By a direct induction on ℓ, we see that in the outer loop of step 4, $R_{ik} = \mathrm{LSR}((c_\ell S_x^\ell + \cdots + c_0) \odot M_{ik}, S_y - 1)$ holds for all $i = 1, \ldots, m$ and $k = 1, \ldots, d_i$ every time the algorithm passes through step 4.1.

Assume that $L = \sum_{\ell=0}^{\rho} \tilde{c}_\ell S_x^\ell$ with $\tilde{c}_\ell \in \mathbb{K}[x]$ and $\tilde{c}_\rho \neq 0$ is a minimal telescoper for f. By Theorem 4.3, the left scalar remainders $\tilde{R}_{ik}$ of the $L \odot M_{ik}$ by $S_y - 1$ are all zero. Thus, the linear homogeneous system over $\mathbb{K}[x]$ obtained by equating all the R_{ik} at the ρth iteration of the outer loop in step 4 to zero has a nontrivial solution, which gives rise to a telescoper of minimal order. The compact representation for a corresponding certificate follows by Theorem 4.3. $\qquad\square$

Example 4.5. *Consider the same rational function f as in Example 2.2. By Example 2.4, the refined integer-linear decomposition of the denominator g is given by (2.6). Then in step 1, we obtain the RILD-based partial fraction decomposition*

$$f = \frac{a_0}{p_0} + \underbrace{((x-1)S_{-5,2} + 2x^2 + 1)}_{M_1}\left(\frac{1}{p_1(-5x+2y)}\right) + \underbrace{\tfrac{1}{30}(-3x^2 - x + 10)S_{3,10}^0}_{M_2}\left(\frac{1}{p_2(3x+10y)}\right),$$

$$+ \underbrace{\tfrac{1}{30}(9x^3 + 30x^2y - 3x^2 + 10xy - 29x - 100y + 20)S_{3,10}^0}_{M_3}\left(\frac{1}{p_3(3x+10y)}\right),$$

where $S_{-5,2} = S_x S_y^3$, $S_{3,10} = S_x^7 S_y^{-2}$ and

$$\frac{a_0}{p_0} = \frac{2x+3}{\sigma_y^{30}(g_0)} - \frac{2x+3}{\sigma_y^{29}(g_0)} - \frac{1}{\sigma_y(g_0)} + \frac{1}{g_0}, \tag{4.9}$$

to which subsequently applying the GGSZ reduction in step 2 yields (4.8) with h represented by the compact form given in (2.3) and $r = 0$. In step 4, we execute the outer loop for $\ell = 0, 1, \ldots, 22$ and iteratively compute the left scalar remainder R_i of $(c_{22}S_x^{22} + \cdots + c_0) \odot M_i$ by $S_y - 1$ for $i = 1, 2, 3$, where $c_0, \ldots, c_{22}$ are indeterminates. By equating R_1, R_2, R_3 to zero, we set up an overdetermined system of 32 linear equations in unknowns $c_0, \ldots, c_{22}$ over $\mathbb{Q}[x]$, in which each linear equation is of degree in x at most 3. Solving this linear system over $\mathbb{Q}[x]$ gives a minimal telescoper

$$L = (3x^2 + 42x + 82)S_x^{22} - (3x^2 + 30x + 10)S_x^{20} - 2(3x^2 + 72x + 142)S_x^{12}$$
$$+ 2(3x^2 + 60x + 10)S_x^{10} + (3x^2 + 102x + 802)S_x^2 - (3x^2 + 90x + 610), \tag{4.10}$$

along with a corresponding certificate in the compact expression

$$L(h) + \mathrm{LSQ}(L \odot M_1, S_y - 1)\left(\frac{1}{p_1(-5x+2y)}\right)$$

$$+ \mathrm{LSQ}(L \odot M_2, S_y - 1)\left(\frac{1}{p_2(3x+10y)}\right) + \mathrm{LSQ}(L \odot M_3, S_y - 1)\left(\frac{1}{p_3(3x+10y)}\right).$$

4.3. Efficiency considerations

The efficiency of the algorithm **RationalCT** can be enhanced by incorporating the following modifications in the algorithm.

(i) *Modification in step 1.*

In step 1, we employ the shift-homogeneous decomposition to obtain the refined integer-linear decomposition of the denominator of f, which leads to the RILD-based partial fraction decomposition of f. In fact, the role of the shift-homogeneous decomposition can be played by any shiftless decomposition introduced in (Gerhard et al., 2003, Definition 1). In particular, the coarsest shiftless decomposition, namely the one which groups all irreducible factors g_i having the same tuples $(v_{i1}, \ldots, v_{in_i})$ and $(e_{i1}, \ldots, e_{in_i})$ in (2.1), can be used. Such a decomposition can be computed via GCD computation (see (Gerhard et al., 2003, §3)). In this way, we avoid the need of full factorization while maintaining the uniqueness of the induced RILD-based partial fraction decomposition, which in turn ensures the correctness of the algorithm.

(ii) *Modification in step 2.*

In step 2, with $\tilde{a}_0 \in \mathbb{K}[x, y]$ and $u_0 \in \mathbb{K}[x]$ denoting the numerator and denominator of a_0, respectively, it actually suffices to apply the GGSZ reduction to $\tilde{a}_0/p_0$ (instead of a_0/p_0) since $\tilde{a}_0/p_0 = (S_y - 1)(hu_0) + ru_0$ and $ru_0 = 0$ if and only if $r = 0$. This reduces the cost of this step.

(iii) *Modification in step 4.*

Let $u \in \mathbb{K}[x]$ be the common denominator of the M_{ik} and write each M_{ik} as $M_{ik} = \frac{1}{u}\tilde{M}_{ik}$ for some $\tilde{M}_{ik} \in \mathbb{K}[x, y, S_{\lambda_i, \mu_i}]$. Inspired by the proof of (Chen and Kauers, 2012, Theorem 10), it actually amounts to looking for a telescoper of the form $L = \sum_{\ell=0}^{\rho} c_\ell \sigma_x^\ell(u) S_x^\ell$. As such, for all $i = 1, \ldots, m$ and $k = 1, \ldots, d_i$, we have

$$L \odot M_{ik} = \left(\sum_{\ell=0}^{\rho} c_\ell \sigma_x^\ell(u) S_x^\ell \frac{1}{u} \right) \odot \tilde{M}_{ik} = \left(\sum_{\ell=0}^{\rho} c_\ell S_x^\ell \right) \odot \tilde{M}_{ik} = \sum_{\ell=0}^{\rho} c_\ell \left(S_x^\ell \odot \tilde{M}_{ik} \right), \qquad (4.11)$$

and thus $L \odot M_{ik} \in \mathbb{K}[x, y, S_{\lambda_i, \mu_i}, S_{\lambda_i, \mu_i}^{-1}]$, so that operations in step 4 only induce arithmetic with polynomial coefficients. For doing so, we compute in step 4.1.1 the left scalar remainder of $\tilde{R}$ of $S_x^\ell \odot \tilde{M}_{ik}$ (instead of $S_x^\ell \odot M_{ik}$) by $S_y - 1$ and return in step 4.2 a telescoper of the form $L = \sum_{j=0}^{\ell} c_j \sigma_x^j(u) S_x^j$ once a nontrivial solution is found.

We note that looking for a telescoper of the specified form in fact does not lose any generality because, for a telescoper $\tilde{L} = \sum_{\ell=0}^{\rho} \tilde{c}_\ell S_x^\ell \in \mathbb{K}[x][S_x]$, multiplying from the left the least common multiple u_ρ of $u, \sigma_x(u), \ldots, \sigma_x^\rho(u)$ gives

$$u_\rho \tilde{L} = \sum_{\ell=0}^{\rho} c_\ell \, \sigma_x^\ell(u) S_x^\ell \quad \text{with } c_\ell = \frac{\tilde{c}_\ell u_\rho}{\sigma_x^\ell(u)},$$

which is again a telescoper with the same order as $\tilde{L}$ and of the required form. On the other hand, it is often observed in experiments that taking such a special form for telescopers actually helps to decrease sizes of the c_ℓ to be determined, which might deserve further investigation.

(iv) *Further modification in step 4.*

Following the preceding modification, before executing the outer loop of step 4, we can first compute the left scalar remainder N_{ik} of each $\tilde{M}_{ik}$ by $S_y - 1$; then we let N_{ik} play the role of $\tilde{M}_{ik}$ in step 4.1.1. This is because any operator in $\mathbb{K}[x][S_x]$ commutes with $S_y - 1$ and then $\mathrm{LSR}(S_x^\ell \odot \tilde{M}_{ik}, S_y - 1) = \mathrm{LSR}(S_x^\ell \odot N_{ik}, S_y - 1)$ for any $\ell \in \mathbb{N}$. Note that every nonzero N_{ik} has highest order no more than $\mu_i - 1$ and typically can be handled more easily than $\tilde{M}_{ik}$.

Let us now reconsider Example 4.5 in the light of the above modifications.

18

Example 4.6. *Consider the same rational function f as Example 2.2. Using the coarsest shiftless decomposition instead of the shift-homogeneous decomposition as described in modification (i), we obtain the following refined integer-linear decomposition*

$$g = p_0(x, y)p_1(-5x + 2y)p_1(-5x + 2y + 1)p_2(3x + 10y), \tag{4.12}$$

where $p_0 = g_0\sigma_y(g_0)\sigma_y^{29}(g_0)\sigma_y^{30}(g_0)$ with $g_0 = xy + 1$, $p_1(z) = z^2 + 1$ and $p_2(z) = z^3 + 1$. Based on (4.12), we find in step 1 the partial fraction decomposition

$$f = \frac{a_0}{p_0} + \underbrace{((x-1)\mathrm{S}_{-5,2} + 2x^2 + 1)}_{M_1}\left(\frac{1}{p_1(-5x + 2y)}\right) + \underbrace{(xy + 1)\mathrm{S}_{3,10}^0}_{M_2}\left(\frac{1}{p_2(3x + 10y)}\right),$$

where a_0/p_0 is given by (4.9), $\mathrm{S}_{-5,2} = \mathrm{S}_x\mathrm{S}_y^3$ and $\mathrm{S}_{3,10} = \mathrm{S}_x^7\mathrm{S}_y^{-2}$. Again, in step 2, we apply the GGSZ reduction to a_0/p_0 which yields (4.8) with h represented by the compact form given in (2.3) and r = 0. In step 4, the loop will be executed for $\ell = 0, \dots, 22$. The final, induced linear system contains 22 equations in unknowns $c_0, \dots, c_{22}$ over $\mathbb{Q}[x]$ and each equation has degree in x at most 2. This compares to Example 4.5 which involves a linear system of 32 linear equations of degree in x at most 2. The basis to the nullspace of the linear system over $\mathbb{Q}(x)$ gives rise to the same minimal telescoper L in (4.10). Note that modifications (ii)-(iv) are trivial in this example.

5. Arithmetic cost for the new algorithm

In this section, we give a complexity analysis of the new algorithm described in the preceding section. For this purpose, we first collect some classical complexity notations and facts needed in this paper. More background on these can be found in (von zur Gathen and Gerhard, 2013).

5.1. Complexity background

In this paper, costs of algorithms will be counted by the number of arithmetic operations in the field $\mathbb{K}$. All costs are analyzed in terms of O-estimates for classical arithmetic and O˜-estimates for fast arithmetic, where the *soft-Oh notation* "O˜" is basically "O" but suppressing logarithmic factors (see (von zur Gathen and Gerhard, 2013, Definition 25.8) for a precise definition).

We summarize the facts needed for our analysis below and will freely use them later. For proofs, we refer to (von zur Gathen and Gerhard, 2013), (Gerhard, 2004, §3 and §5) and (Zhou et al., 2012, Theorem 4.1).

The first fact gives sharp degree bounds for two basic arithmetic operations – division with remainder and partial fraction decomposition. This turns out to be very useful in estimating degree sizes. The proofs are mainly based on Cramer's rule and determinant expansions and will be skipped.

Fact 5.1 (Degree bounds)**.** *Let f, g be two nonzero polynomials in $\mathbb{K}[x, y]$.*

(i) Assume that $\deg_y(f) \geq \deg_y(g)$. Then there exist unique $q, r \in \mathbb{K}[x, y]$ with

$$(\deg_x(q), \deg_y(q)) \leq ((\deg_y(f) - \deg_y(g))\deg_x(g) + \deg_x(f), \deg_y(f) - \deg_y(g))$$

and $\quad (\deg_x(r), \deg_y(r)) \leq ((\deg_y(f) - \deg_y(g) + 1)\deg_x(g) + \deg_x(f), \deg_y(g) - 1)$

such that $\mathrm{lc}_y(g)^{\deg_y(f)-\deg_y(g)+1} f = qg + r$.

19

(ii) *Assume that* $\deg_y(f) < \deg_y(g)$ *and* $g = g_1^{e_1} \ldots g_m^{e_m}$ *with* $e_i \in \mathbb{N} \setminus \{0\}$ *and* $g_i \in \mathbb{K}[x, y]$ *being pairwise coprime. Then there exists* $u \in \mathbb{K}[x]$ *and* $\{f_{ij}\}_{1 \leq i \leq m, 1 \leq j \leq e_i} \subseteq \mathbb{K}[x, y]$ *with*

$$\deg_x(u) \leq \deg_x(g) \deg_y(g) - \sum_{i=1}^{m} \frac{e_i(1 + e_i)}{2} \deg_x(g_i) \deg_y(g_i) \quad \textit{and}$$

$$(\deg_x(f_{ij}), \deg_y(f_{ij})) \leq (\deg_x(g) \deg_y(g) + \deg_x(f) - \deg_x(g) + j \deg_x(g_i), \deg_y(g_i) - 1)$$

such that

$$\frac{f}{g} = \frac{1}{u} \left(\frac{f_{11}}{g_1} + \cdots + \frac{f_{1e_1}}{g_1^{e_1}} + \cdots + \frac{f_{m1}}{g_m} + \cdots + \frac{f_{me_m}}{g_m^{e_m}} \right).$$

The next fact contains the cost of some basic arithmetics for univariate polynomials.

Fact 5.2 (Arithmetic of univariate polynomials). *Let* $f, g \in \mathbb{K}[x]$ *with* $\deg_x(f), \deg_x(g) \leq d_x$. *Then the following operations can be performed at most in* $\mathrm{O}(d_x^2)$ *arithmetic operations in* $\mathbb{K}$ *with classical arithmetic and* $\mathrm{O}^\sim(d_x)$ *with fast arithmetic.*

   (i) *Addition, multiplication, division with remainder, GCD computation of* f *and* g;

  (ii) *Evaluation* f *at* $d_x + 1$ *distinct points in* $\mathbb{K}$ *or interpolation in* $\mathbb{K}[x]$ *at these points;*

  (iii) *Partial fraction decomposition of* f/g *with respect to a given factorization of* g, *provided that* f, g *are nonzero coprime polynomials with* $\deg_x(f) < \deg_x(g)$.

In order to analyze the cost for operations on bivariate polynomials, a general (although not optimal) technique is to use evaluation and interpolation on polynomials and to perform operations on univariate polynomials based on the above fact. We will frequently use this technique without explicitly pointing it out.

As mentioned in the introduction, most of recent creative telescoping algorithms, including our new one presented in Section 4, eventually reduce the problem of finding telescopers to the problem of solving linear systems, which can be accomplished efficiently.

Fact 5.3 (Solving linear systems). *Let* M *be a polynomial matrix in* $\mathbb{K}[x]^{m \times n}$ *with entries being polynomials in* $\mathbb{K}[x]$ *of degree in* x *less than* d_x. *Assume that* $n \in \mathrm{O}(m)$. *Then a basis of the null space of* M *in* $\mathbb{K}[x]$ *can be computed using* $\mathrm{O}(m^3 d_x^2)$ *arithmetic operations in* $\mathbb{K}$ *with classical arithmetic (Gaussian elimination) and* $\mathrm{O}^\sim(m^{\omega-1} n d_x)$ *with fast arithmetic, where* $\omega \in \mathbb{R}$ *with* $2 < \omega \leq 3$ *is the exponent of matrix multiplication over* $\mathbb{K}$.

5.2. Output size estimates

We define the degree of a rational function in $\mathbb{K}(x, y)$ with respect to x (resp. y) to be the maximum of the degrees of its numerator and denominator with respect to x (resp. y). Using Fact 5.1, we are now able to estimate sizes of intermediate results.

Lemma 5.4. *Let* $f \in \mathbb{K}(x, y)$ *be a rational function with* $\deg_x(f) = d_x$ *and* $\deg_y(f) = d_y$. *Assume that the RILD-based partial fraction decomposition of* f *takes the form* (4.3). *Let* $\tilde{a}_0 \in \mathbb{K}[x, y]$ *be the numerator of* a_0. *Let* $u \in \mathbb{K}[x]$ *be the common denominator of the* M_{ik} *and write each* M_{ik} *as* $M_{ik} = \frac{1}{u} \tilde{M}_{ik}$ *for some* $\tilde{M}_{ik} \in \mathbb{K}[x, y, \mathrm{S}_{\lambda_i, \mu_i}]$. *Then*

$$(\deg_x(\tilde{a}_0), \deg_y(\tilde{a}_0)) \in \mathrm{O}(d_x d_y) \times \mathrm{O}(d_y), \quad \deg_x(u) \in \mathrm{O}(d_x d_y)$$

and $\quad (\deg_x(\tilde{M}_{ijk}), \deg_y(\tilde{M}_{ijk})) \in \mathrm{O}(d_x d_y) \times \mathrm{O}(\deg_z(p_i))$ *for all* $i = 1, \ldots, m$ *and* $k = 1, \ldots, d_i$.

Proof. We know from definition that (4.3) gives the partial fraction decomposition of f with respect to y, based on the refined integer-linear decomposition of its denominator. The degree bounds thus follow directly by Fact 5.1. $\qquad\square$

Lemma 5.5. *Let $r \in \mathbb{K}(x, y)$ be a rational function of the form (4.6). Let $u \in \mathbb{K}[x]$ be the common denominator of the M_{ik} and write each M_{ik} as $M_{ik} = \frac{1}{u}\tilde{M}_{ik}$ for some $\tilde{M}_{ik} \in \mathbb{K}[x, y, S_{\lambda_i, \mu_i}]$. Let $L = \sum_{\ell=0}^{\rho} c_\ell \sigma_x^\ell(u) S_x^\ell \in \mathbb{K}[x][S_x]$ with $\rho \in \mathbb{N}$ and $c_\ell \in \mathbb{K}[x]$. Then for each integer pair (i, k) with $1 \le i \le m$ and $1 \le k \le d_i$, the left scalar remainder R_{ik} of $L \odot M_{ik}$ by $S_y - 1$ can be written as*

$$R_{ik} = c_\rho \tilde{R}_{ik\rho} + \cdots + c_0 \tilde{R}_{ik0}, \tag{5.1}$$

where $\tilde{R}_{ik\ell} \in \mathbb{K}[x, y, S_{\lambda_i, \mu_i}]$ with

$$(\deg_x(\tilde{R}_{ik\ell}), \deg_y(\tilde{R}_{ik\ell})) \le (\deg_x(\tilde{M}_{ik}), \deg_y(\tilde{M}_{ik})) \quad and \quad \deg_{x,y}(\tilde{R}_{ik\ell}) \le \deg_{x,y}(\tilde{M}_{ik}).$$

Here $\deg_{x,y}(\cdot)$ denotes the total degree of the argument with respect to x, y.

Proof. For each integer pair (i, k) with $1 \le i \le m$ and $1 \le k \le d_i$, it follows from (4.11) that letting $\tilde{R}_{ik\ell} = \mathrm{LSR}(S_x^\ell \odot \tilde{M}_{ik}, S_y - 1)$ for all $\ell = 0, \ldots, \rho$ gives the decomposition (5.1). It remains to check the degree estimates of $\tilde{R}_{ik\ell}$, which in turn is an immediate result of (3.3). $\qquad\square$

The following depicts an order-degree curve of telescopers for bivariate rational functions.

Lemma 5.6. *Let $r \in \mathbb{K}(x, y)$ be a rational function of the form (4.6). Let $u \in \mathbb{K}[x]$ be the common denominator of the M_{ik} and write each M_{ik} as $M_{ik} = \frac{1}{u}\tilde{M}_{ik}$ for some $\tilde{M}_{ik} \in \mathbb{K}[x, y, S_{\lambda_i, \mu_i}]$. For each integer pair (i, k) with $1 \le i \le m$ and $1 \le k \le d_i$, define $\alpha_{ik} = \max\{-1, \deg_{x,y}(\tilde{M}_{ik})\}$ and $\beta_{ik} = \max\{-1, \deg_y(\tilde{M}_{ik})\}$, and let*

$$\rho_0 = \sum_{i=1}^{m} \sum_{k=1}^{d_i} \mu_i(\beta_{ik} + 1). \tag{5.2}$$

Then for any nonnegative integer pair (ρ, τ) with $\rho \ge \rho_0$ and

$$\tau > \deg_x(u) - 1 + \frac{\sum_{i=1}^{m} \sum_{k=1}^{d_i} \mu_i(\alpha_{ik} - \frac{1}{2}\beta_{ik})(\beta_{ik} + 1)}{\rho + 1 - \rho_0}, \tag{5.3}$$

there exists a telescoper for r of order at most ρ and degree at most τ.

Proof. Let $\rho, \tau \in \mathbb{N}$ with $\rho \ge \rho_0$ and τ satisfying (5.3). To prove the lemma, it is sufficient to show that there exist $c_0, \ldots, c_\rho \in \mathbb{K}[x]$, not all zero, with $\deg_x(c_\ell) \le \tau - \deg_x(u)$ such that

$$\mathrm{LSR}\left(\left(\sum_{\ell=0}^{\rho} c_\ell \sigma_x^\ell(u) S_x^\ell\right) \odot M_{ik}, S_y - 1\right) = 0 \quad \text{for all } i = 1, \ldots, m \text{ and } k = 1, \ldots, d_i, \tag{5.4}$$

because then Theorem 4.3 asserts that $\sum_{\ell=0}^{\rho} c_\ell \sigma_x^\ell(u) S_x^\ell$ gives a desired telescoper for r. Now we consider the linear system over $\mathbb{K}$ (rather than $\mathbb{K}[x]$) obtained by vanishing coefficients of like powers of x and y in (5.4). In other words, we view the coefficients of the c_ℓ with respect to x, not the c_ℓ themselves, as unknowns. This then gives us $(\tau - \deg_x(u) + 1)(\rho + 1)$ unknowns in total. On the other hand, we derive from Lemma 5.5 that each equation in (5.4) has total degree in x, y

21

at most $\tau - \deg_x(u) + \alpha_{ik}$ and degree in y at most β_{ik}. It follows that the induced linear system contains at most

$$(\tau - \deg_x(u) + 1)\rho_0 + \sum_{i=1}^{m} \sum_{k=1}^{d_i} \mu_i(\alpha_{ik} - \frac{1}{2}\beta_{ik})(\beta_{ik} + 1)$$

equations over $\mathbb{K}$. Since $\rho \geq \rho_0$, one concludes from (5.3) that the linear system over $\mathbb{K}$ resulting from (5.4) have more unknowns than equations, assuring such a nontrivial solution. $\qquad\square$

We note that the left scalar remainders of the $\tilde{M}_{ik}$ by $S_y - 1$ can be employed to further refine the bounds given by (5.2) and (5.3).

Remark 5.7. *Under the assumptions of the above lemma, in the context of (Chen and Kauers, 2012, §4), all $\tilde{M}_{ik}$ are actually in $\mathbb{K}[x, S_{\lambda_i,\mu_i}]$, yielding $\alpha_{ik} = \max\{-1, \deg_x(\tilde{M}_{ik})\}$ and $\beta_{ik} = 0$. Then $\rho_0 = \sum_{i=1}^{m} \sum_{k=1}^{d_i} \mu_i$ by (5.2), and (5.3) becomes*

$$\tau > \deg_x(u) - 1 + \frac{\sum_{i=1}^{m} \sum_{k=1}^{d_i} \mu_i \alpha_{ik}}{\rho + 1 - \rho_0},$$

which coincides with the order-degree curve given in (Chen and Kauers, 2012, Theorem 10) (after correcting the typos in the formula of the lower bound for d there).

5.3. Cost analysis of algorithm

Recall that the *auto-dispersion set* of a polynomial $g \in \mathbb{K}[x, y]$ with respect to y consists of all integers ℓ such that $\deg_y(\gcd(g, \sigma_y^\ell(g))) > 0$.

Lemma 5.8. *Let $a_0, p_0 \in \mathbb{K}[x, y]$ be two coprime polynomials with $p_0 \neq 0$ and $\deg_y(a_0/p_0) = d_y$. Then the GGSZ reduction computes $h, r \in \mathbb{K}(x, y)$ with h in a compact form such that (4.8) holds, using $O(\deg_x(p_0)d_y^4 + \deg_x(p_0)^2 d_y^3 + \deg_x(a_0)\deg_x(p_0)d_y^2 + \deg_x(a_0)^2 d_y)$ arithmetic operations in $\mathbb{K}$ with classical arithmetic and $\tilde{O}(\deg_x(p_0)d_y^3 + \deg_x(a_0)d_y)$ with fast arithmetic, plus the cost of computing the auto-dispersion set of p_0 with respect to y.*

Proof. By (Gerhard et al., 2003, Theorem 13), the cost of the GGSZ reduction is dominated by computing the shiftless decomposition of p_0 and the subsequent partial fraction decomposition of a_0/p_0. By (Gerhard et al., 2003, Theorem 10) and making use of the evaluation-interpolation technique, one obtains that the former operation takes $O(\deg_x(p_0)d_y^4 + \deg_x(p_0)^2 d_y)$ arithmetic operations in $\mathbb{K}$ with classical arithmetic and $\tilde{O}(\deg_x(p_0)d_y^3)$ with fast arithmetic, plus the cost of computing the auto-dispersion set of p_0 with respect to y. While the latter operation takes $O(\deg_x(p_0)^2 d_y^3 + \deg_x(a_0)\deg_x(p_0)d_y^2 + \deg_x(a_0)^2 d_y)$ with classical arithmetic and $\tilde{O}(\deg_x(p_0)d_y^2 + \deg_x(a_0)d_y)$ with fast arithmetic. Combining these two costs concludes the lemma. $\qquad\square$

Now we are ready to study the cost of the algorithm **RationalCT**, in which we shall assume that the four enhancements discussed in Section 4.3 have been taken into account.

Theorem 5.9. *Let $f \in \mathbb{K}(x, y)$ be a rational function with $\deg_x(f) = d_x$ and $\deg_y(f) = d_y$. Assume that f has a telescoper and let ρ be the actual order of its minimal telescopers. Further assume (4.3) and (4.6) hold, and define ρ_0 by (5.2). Then the algorithm **RationalCT** finds a minimal telescoper for f and a certificate in a compact form using $O(d_x d_y^4 + \rho d_x^2 d_y^3 + \rho \rho_0^3 d_x^2 d_y^2)$ arithmetic operations in $\mathbb{K}$ with classical arithmetic and $\tilde{O}(d_x d_y^3 + \rho d_x d_y^2 + \rho^2 \rho_0^{\omega-1} d_x d_y)$ with fast arithmetic, plus the cost of computing auto-dispersion sets and finding rational roots.*

Proof. Based on modification (i) in Section 4.3, in step 1, we incorporate the coarsest shiftless decomposition into the integer-linear decomposition to obtain the refined one of the denominator of f, which, by (Giesbrecht et al., 2019, Theorem 3.5) and (Gerhard et al., 2003, Theorem 10), takes $\mathrm{O}(d_x^2 d_y + d_x d_y^3 + d_y^4)$ arithmetic operations with classical arithmetic and $\tilde{\mathrm{O}}(d_x d_y^2 + d_y^3)$ with fast arithmetic, plus the cost of finding rational roots. Therefore, the RILD-based integer-linear decomposition of f can be obtained using $\mathrm{O}(d_x^2 d_y^3 + d_y^4)$ with classical arithmetic and $\tilde{\mathrm{O}}(d_x d_y^2 + d_y^3)$ with fast arithmetic in total. Regardless of the cost of computing auto-dispersion sets, one concludes from modification (ii) and Lemmas 5.4, 5.8 that step 2 takes $\mathrm{O}(d_x d_y^4 + d_x^2 d_y^3)$ with classical arithmetic and $\tilde{\mathrm{O}}(d_x d_y^3)$ with fast arithmetic. By assumption, $r = 0$ in (4.8) and thus the algorithm continues after step 3.

Based on modifications (iii)-(iv), we proceed to find the common denominator $u \in \mathbb{K}[x]$ of the operators M_{ik}, reformulate each of them as $M_{ik} = \frac{1}{u}\tilde{M}_{ik}$ for $\tilde{M}_{ik} \in \mathbb{K}[x, y, \mathrm{S}_{\lambda_i, \mu_i}]$, and compute the left scalar remainders N_{ik} of the $\tilde{M}_{ik}$ by $\mathrm{S}_y - 1$. By Lemma 5.4, $\deg_x(\tilde{M}_{ik}) \in \mathrm{O}(d_x d_y)$ and $\deg_y(\tilde{M}_{ik}) \in \mathrm{O}(\deg_z(p_i))$. It then follows from (3.3) that computing all the N_{ik} in total requires $\mathrm{O}(d_x^2 d_y^3)$ with classical arithmetic and $\tilde{\mathrm{O}}(d_x d_y^2)$ with fast arithmetic. Since $\deg_x(N_{ik}) \leq \deg_x(\tilde{M}_{ik})$ and $\deg_y(N_{ik}) \leq \deg_y(\tilde{M}_{ik})$, for each iteration of the outer loop of step 4, the same cost applies to step 4.1 with M_{ik} replaced by N_{ik} as discussed in modifications (iii)-(iv).

Since ρ is the actual order of minimal telescopers for f, the outer loop of step 4 runs exactly ρ iterations. Thus the total cost of step 4.1 in the whole loop is $\mathrm{O}(\rho d_x^2 d_y^3)$ with classical arithmetic and $\tilde{\mathrm{O}}(\rho d_x d_y^2)$ with fast arithmetic. For the ℓ-th iteration with $0 \leq \ell \leq \rho$, Lemmas 5.4 and 5.5 assert that the coefficient matrix over $\mathbb{K}[x]$ attached to the linear system obtained in step 4.2 has at most ρ_0 rows and $\ell + 1$ columns, and each of its nonzero entries has degree in x in $\mathrm{O}(d_x d_y)$. Thus Fact 5.3 implies that finding a solution needs $\mathrm{O}(\rho_0^3 d_x^2 d_y^2)$ with classical arithmetic and $\tilde{\mathrm{O}}(\ell \rho_0^{\omega-1} d_x d_y)$ with fast arithmetic. This yields the total cost of $\mathrm{O}(\rho \rho_0^3 d_x^2 d_y^2)$ with classical arithmetic and $\tilde{\mathrm{O}}(\rho^2 \rho_0^{\omega-1} d_x d_y)$ with fast arithmetic for solving linear systems in step 4.2 in the whole loop, as there are ρ iterations.

When a nontrivial solution is found, it virtually takes no arithmetic operations for returning the certificate in such a compact representation. By modification (iii), we eventually construct a minimal telescoper of the form $L = \sum_{j=0}^{\ell} c_j \sigma_x^j(u) \mathrm{S}_x^j$. Computing the $\sigma_x^j(u)$ in the telescoper L requires $\mathrm{O}(\rho d_x^2 d_y^2)$ with classical arithmetic and $\tilde{\mathrm{O}}(\rho d_x d_y)$ with fast arithmetic. In addition, by Lemma 5.6, $\deg_x(c_j) \in \mathrm{O}(\rho_0 d_x d_y)$. Therefore, expanding the telescoper L takes $\mathrm{O}(\rho_0 d_x^2 d_y^2)$ with classical arithmetic and $\tilde{\mathrm{O}}(\rho_0 d_x d_y)$ with fast arithmetic. The announced cost follows. $\qquad\square$

Corollary 5.10. *With the assumptions of Theorem 5.9, further let $\mu = \max\{\mu_1, \ldots, \mu_m\}$. Then $\rho_0 \in \mathrm{O}(\mu d_y)$, and the algorithm* **RationalCT** *takes $\mathrm{O}(\mu^4 d_x^2 d_y^6)$ arithmetic operations in $\mathbb{K}$ with classical arithmetic and $\tilde{\mathrm{O}}(\mu^{\omega+1} d_x d_y^{\omega+2})$ with fast arithmetic, plus the cost of computing auto-dispersion sets and finding rational roots.*

Proof. By assumption, with $u \in \mathbb{K}[x]$ denoting the common denominator of the M_{ik} in (4.3), each operator M_{ik} has the form $M_{ik} = \frac{1}{u}\tilde{M}_{ik}$ for $\tilde{M}_{ik} \in \mathbb{K}[x, y, \mathrm{S}_{\lambda_i, \mu_i}]$ with $\deg_y(\tilde{M}_{ik}) < \deg_z(p_i)$. It follows from (5.2) that $\rho_0 \in \mathrm{O}(\mu d_y)$. Since ρ is the actual order of minimal telescopers for f, we conclude from Lemma 5.6 that $\rho \leq \rho_0$. The announced cost is then evident by Theorem 5.9. $\qquad\square$

Remark 5.11. *Under the assumptions of the above corollary, according to Lemma 5.6, there exists a minimal telescoper for f of total size in $\mathrm{O}(\mu^2 d_x d_y^3)$.*

Remark 5.12. *In the case of $\mathbb{K} = \mathbb{Q}$, by incorporating the cost of computing the auto-dispersion set of an integer polynomial (cf. (Gerhard et al., 2003, Theorem 14)) and the cost of finding rational roots of an integer polynomial (cf. (von zur Gathen and Gerhard, 2013, Theorem 15.21)), one sees from the above corollary that the algorithm* **RationalCT** *has the total running time bounded by $(\mu + d_x + d_y + \log \|f\|_\infty)^{O(1)}$ word operations, where the max-norm $\|f\|_\infty$ of $f \in \mathbb{Q}(x, y)$ is defined as the maximal absolute value of the integer coefficients appearing in the numerator and denominator of f with respect to x, y. See (Gerhard, 2004; von zur Gathen and Gerhard, 2013) for more information on word operations.*

6. Arithmetic cost for the reduction-based approach

In this section, we review the reduction-based creative telescoping algorithm developed in (Chen et al., 2015) in the context of bivariate rational functions and further analyze its cost in this setting. As indicated by the name of the algorithm, a reduction method plays a fundamental role. The original reduction method employed by (Chen et al., 2015) in the rational case was developed by Abramov (1975). In order to highlight more significant discrepancies between this creative telescoping algorithm and the one developed in Section 4, we instead use the GGSZ reduction recalled in Section 2.1 to carry out all the reduction steps in the algorithm.

Before discussing the concrete algorithm, let us recall some notions. As a generalization of auto-dispersion sets, the *dispersion set* of a polynomial $f \in \mathbb{K}[x, y]$ with respect to another polynomial $g \in \mathbb{K}[x, y]$ is defined to be the integer set

$$\mathrm{DS}_y(f, g) = \{\ell \in \mathbb{Z} \mid \deg_y(\gcd(f, \sigma_y^\ell(g))) > 0\}.$$

Such a dispersion set can be achieved by the algorithm of Man and Wright (1994) or by the procedure **pDispersionSet** from (Gerhard et al., 2003, §6) in the particular case where $\mathbb{K} = \mathbb{Q}$.

A polynomial in $\mathbb{K}[x, y]$ is called *primitive* with respect y (or y-primitive for short) if the greatest common divisor over $\mathbb{K}[x]$ of all its coefficients with respect to y is equal to one. A rational function in $\mathbb{K}(x, y)$ is called *proper* with respect y (or y-proper for short) if the degree of its numerator with respect to y is less than that of its denominator. For a rational function $f \in \mathbb{K}(x, y)$, another rational function $r \in \mathbb{K}(x, y)$ is called a *shift-remainder* with respect to y (or σ_y-*remainder* for short) of f if $f - r$ is σ_y-summable and r is y-proper with denominator being σ_y-free. For brevity, we just say that r is a σ_y-remainder if f is clear from the context. Clearly, any integer shift of a σ_y-remainder with respect to x is again a σ_y-remainder. By (2.2), we see that the GGSZ reduction reduces a rational function to a σ_y-remainder modulo σ_y-summable rational functions.

A rational function in $\mathbb{K}(x, y)$ usually has more than one σ_y-remainder and any two of them differ by a σ_y-summable rational function. The following proposition implies that zero is the only σ_y-remainder in the case of a σ_y-summable rational function.

Proposition 6.1 (Abramov 1975, Proposition 7). *A rational function in $\mathbb{K}(x, y)$ is σ_y-summable if and only if any of its σ_y-remainders is zero.*

We summarize below the main idea of the reduction-based algorithm in (Chen et al., 2015).

Let f be a rational function in $\mathbb{K}(x, y)$. Applying the GGSZ reduction to f yields (2.2). If the denominator of r in (2.2) is not integer-linear, then by (Abramov and Le, 2002, Theorem 1), f does not have any telescoper. Otherwise, the existence of telescopers for f is guaranteed.

Assume now that we want to find a telescoper for f of order no more than $\rho \in \mathbb{N}$. In this respect, we make an ansatz

$$L = c_\rho S_x^\rho + \cdots + c_1 S_x + c_0 \quad \text{with } c_0, \ldots, c_\rho \in \mathbb{K}[x] \text{ to be determined.}$$

For $\ell = 0, \ldots, \rho$, compute a rational function $h_\ell \in \mathbb{K}(x, y)$ and a σ_y-remainder r_ℓ such that

$$\sigma_x^\ell(f) = (S_y - 1)(h_\ell) + r_\ell \quad \text{and} \quad \sum_{i=0}^\ell c_i r_i \text{ is a } \sigma_y\text{-remainder.} \tag{6.1}$$

A direct calculation then shows that

$$L(f) = (S_y - 1)\left(\sum_{\ell=0}^\rho c_\ell h_\ell\right) + \sum_{\ell=0}^\rho c_\ell r_\ell.$$

Therefore, $\sum_{\ell=0}^\rho c_\ell r_\ell$ is a σ_y-remainder of $L(f)$. By Proposition 6.1, L is a telescoper for f if and only if $\sum_{\ell=0}^\rho c_\ell r_\ell = 0$. This reduces the problem of finding telescopers to the simple task of solving a linear system over $\mathbb{K}[x]$. In other words, we obtain a linear homogeneous system in unknowns $c_0, \ldots, c_\rho$ by equating $\sum_{\ell=0}^\rho c_\ell r_\ell$ to zero, whose any nontrivial solution over $\mathbb{K}[x]$ gives rise to a desired telescoper for f. Failing to find such a solution implies that no required telescopers exist.

In order to compute a minimal telescoper for f, the reduction-based algorithm then applies the above process incrementally with $\rho = 0, 1, \ldots$, with the termination assured by the existence of telescopers.

The proof of (Chen et al., 2015, Theorem 5.6) contains an algorithm for computing such a σ_y-remainder r_ℓ that satisfies (6.1). The key tool is the so-called shift-coprime decompositions of σ_y-free polynomials. Let $b, b_0 \in \mathbb{K}[x, y]$ be two nonzero σ_y-free polynomials. The σ_y-*coprime decomposition* of b with respect to b_0 is defined as

$$b = p_0 \sigma_y^{\ell_1}(p_1) \cdots \sigma_y^{\ell_m}(p_m), \tag{6.2}$$

where $p_0 \in \mathbb{K}[x, y]$ with $\deg_y(\gcd(b_0, \sigma_y^i(p_0))) = 0$ for any nonzero integer i, $p_1, \ldots, p_m \in \mathbb{K}[x, y]$ are monic and y-primitive factors of b_0 of positive degrees in y, and $\ell_1, \ldots, \ell_m$ are distinct nonzero integers. Note that the factors $p_0, \sigma_y^{\ell_1}(p_1), \ldots, \sigma_y^{\ell_m}(p_m), p_1, \ldots, p_m$ are pairwise coprime, since b and b_0 are both σ_y-free. Such a decomposition (6.2) is clearly unique up to the order of factors. It is evident from (6.2) and the σ_y-freeness of b that $\mathrm{DS}_y(b, b_0) = \{0, \ell_1, \ldots, \ell_m\}$ and $p_i = \gcd(\sigma_y^{-\ell_i}(b), b_0)$ for all $i = 1, \ldots, m$. Thus the decomposition (6.2) can be obtained using GCD computation, provided that the dispersion set $\mathrm{DS}_y(b, b_0)$ is known.

Let $r, r_0 \in \mathbb{K}(x, y)$ be two nonzero σ_y-remainders of respective denominators $b, b_0 \in \mathbb{K}[x, y]$. By partial fraction decomposition, based on the σ_y-coprime decomposition (6.2) of b with respect to b_0, there exist unique $f_0, f_1, \ldots, f_m \in \mathbb{K}(x)[y]$ with $\deg_y(f_i) < \deg_y(p_i)$ such that

$$r = \frac{f_0}{p_0} + \frac{f_1}{\sigma_y^{\ell_1}(p_1)} + \cdots + \frac{f_m}{\sigma_y^{\ell_m}(p_m)}. \tag{6.3}$$

We will refer to (6.3) as the *SCD-based* partial fraction decomposition of r with respect to r_0.

The following result can be read from the proof of (Chen et al., 2015, Theorem 5.6).

Proposition 6.2. *Let $r, r_0 \in \mathbb{K}(x, y)$ be two nonzero σ_y-remainders. Assume that the SCD-based partial fraction decomposition of r with respect to r_0 is given by (6.3). Let*

$$\tilde{r} = \frac{f_0}{p_0} + \frac{\sigma_y^{-\ell_1}(f_1)}{p_1} + \cdots + \frac{\sigma_y^{-\ell_m}(f_m)}{p_m}. \tag{6.4}$$

Then $\tilde{r}$ is a σ_y-remainder of r and $c_0 r_0 + c_1 \tilde{r}$ is a σ_y-remainder for any $c_0, c_1 \in \mathbb{K}[x]$.

In view of the above proposition, we call $\tilde{r}$ the *adjusted σ_y-remainder* of r by r_0. It then follows from Proposition 6.2 that (6.1) naturally holds by letting r_ℓ be the adjusted σ_y-remainder of $\sigma_x(r_{\ell-1})$ with respect to $\sum_{i=1}^{\ell-1} c_i r_i$. With all these adjusted σ_y-remainders at hand, the reduction-based algorithm works smoothly in an iterative manner as described before.

Remark 6.3. *As already pointed out in (Chen et al., 2019, §5.2), it is actually sufficient to let each r_ℓ be the adjusted σ_y-remainder of $\sigma_x(r_{\ell-1})$ with respect to r_0 (rather than $\sum_{i=0}^{\ell-1} c_i r_i$) so as to insure the property (6.1). This may reduce the total cost for computing adjusted σ_y-remainders.*

Let us return to the two examples from Section 4. We will use the above reduction-based algorithm in order to illustrate the difference between the two approaches.

Example 6.4. *Let f be the rational function given in Example 4.2. We know from Example 4.2 that f has a minimal telescoper of order two. With $\rho = 2$, the reduction-based algorithm finds the additive decompositions*

$$\sigma_x^\ell(f) = (S_y - 1)(h_\ell) + \frac{a_\ell}{b_\ell} \quad \text{for } \ell = 0, 1, 2,$$

where $h_\ell \in \mathbb{Q}(x, y)$, $a_\ell \in \mathbb{Z}[x, y]$, $b_\ell = ((-5x + 2y)^2 + 1)((-5x + 2y + 1)^2 + 1)$ and all a_ℓ / b_ℓ, as well as their $\mathbb{Q}[x]$-linear combination, are σ_y-remainders. Note that the h_ℓ and a_ℓ are not displayed here for space reasons. In order to find a $\mathbb{Q}[x]$-linear dependency among the a_ℓ / b_ℓ, we set up a linear system attached by the coefficient matrix

$$\begin{pmatrix} 8x^2 + 4x & 8x^2 + 20x + 12 & 8x^2 + 36x + 40 \\ -40x^3 - 12x^2 + 4 & -40x^3 - 100x^2 - 56x & -40x^3 - 172x^2 - 168x + 36 \\ 50x^4 + 5x^3 + 4x^2 - 9x + 1 & 50x^4 + 125x^3 + 67x^2 + 6x + 3 & 50x^4 + 205x^3 + 174x^2 - 73x + 19 \end{pmatrix}.$$

This linear system admits the same solutions as (4.5), in other words, it leads to the same minimal telescoper as Example 4.2. The corresponding certificate is left as an unnormalized dense sum.

Example 6.5. *Consider the same rational function f as Example 2.2. From the same example, we see that f satisfies (2.2) with h, r given by (2.3). Moreover, there exist telescopers for f since the denominator of r is integer-linear. Let $h_0 = h$ and $r_0 = r$. Then for $\ell = 1, \ldots, 22$, the reduction-based algorithm iteratively finds rational functions $h_\ell \in \mathbb{Q}(x, y)$ and adjusted σ_y-remainders r_ℓ such that (6.1) holds. Finding a $\mathbb{Q}[x]$-linear dependency among the r_ℓ yields a linear system with the coefficient matrix of 33 rows and 23 columns and having entries of degree in x at most 34, which yields the same minimal telescoper given by (4.10) as Example 4.6, yet leaving the corresponding certificate as a large, unnormalized dense sum. This compares to Example 4.6 where the induced coefficient matrix has 22 rows and 23 columns with entries of degree in x at most 2.*

26

6.1. Output size estimates

Lemma 6.6. *Let $f \in \mathbb{K}(x, y)$ be a rational function with $\deg_x(f) = d_x$ and $\deg_y(f) = d_y$. Let $r \in \mathbb{K}(x, y)$ be the σ_y-remainder obtained by applying the GGSZ reduction to f. Write $r = a/(ub)$, where $u \in \mathbb{K}[x]$ and $a, b \in \mathbb{K}[x, y]$ with $\deg_y(a) < \deg_y(b)$, $\gcd(a, ub) = 1$ and b being y-primitive and σ_y-free. Then*

$$\deg_x(u) \in \mathrm{O}(d_x d_y), \quad (\deg_x(b), \deg_y(b)) \in \mathrm{O}(d_x) \times \mathrm{O}(d_y)$$

$$and \quad (\deg_x(a), \deg_y(a)) \in \mathrm{O}(d_x d_y) \times \mathrm{O}(d_y).$$

Proof. Assume that the denominator g of f admits the shift-homogeneous decomposition of the form (2.1). With respect to this, we obtain the unique partial fraction decomposition

$$f = p + \frac{1}{\tilde{u}} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{k=1}^{e_{ij}} \frac{f_{ijk}}{\sigma_y^{v_{ij}}(g_i)^k},$$

where $p \in \mathbb{K}(x)[y]$, $\tilde{u} \in \mathbb{K}[x]$ and $f_{ijk} \in \mathbb{K}[x, y]$ with $\deg_y(f_{ijk}) < \deg_y(g_i)$. Applying Fact 5.1 to the above decomposition yields $\deg_x(\tilde{u}) \in \mathrm{O}(d_x d_y)$ and $\deg_x(f_{ijk}) \in \mathrm{O}(d_x d_y)$. Let $d_i = \max_{1 \le j \le n_i}\{e_{ij}\}$ and specify that $f_{ijk} = 0$ in case $k > e_{ij}$. By (Gerhard et al., 2003, Theorem 12),

$$r = \frac{a}{ub} = \frac{1}{\tilde{u}} \sum_{i=1}^{m} \sum_{k=1}^{d_i} \frac{\sum_{j=1}^{n_i} \sigma_y^{-v_{ij}}(f_{ijk})}{g_i^k}.$$

Since b is y-primitive, u divides $\tilde{u}$ in $\mathbb{K}[x]$ and thus $\deg_x(u) \in \mathrm{O}(d_x d_y)$. Notice that $d_i \le \sum_{j=1}^{n_i} e_{ij}$ for all $i = 1, \dots, m$, so $\deg_y(b) \le \sum_{i=1}^{m} d_i \deg_y(g_i) \le d_y$ and similarly, $\deg_x(b) \le d_x$. Moreover, $\deg_x(a) \le \max_{ijk}\{\deg_x(f_{ijk})\} + d_x$, implying $\deg_x(a) \in \mathrm{O}(d_x d_y)$. The lemma follows. $\qquad\square$

Lemma 6.7. *Let $r = a/(ub) \in \mathbb{K}(x, y)$ be a σ_y-remainder, where $u \in \mathbb{K}[x]$ and $a, b \in \mathbb{K}[x, y]$ with $\deg_y(a) < \deg_y(b)$, $\gcd(a, ub) = 1$ and b being y-primitive and σ_y-free. Let $\ell \in \mathbb{N}$ and assume that $r_\ell \in \mathbb{K}(x, y)$ is a σ_y-remainder of $\sigma_x^\ell(r)$. Write $r_\ell = a_\ell/(u_\ell b_\ell)$, where $u_\ell \in \mathbb{K}[x]$ and $a_\ell, b_\ell \in \mathbb{K}[x, y]$ with $\deg_y(a_\ell) < \deg_y(b_\ell)$, $\gcd(a_\ell, u_\ell b_\ell) = 1$ and b_ℓ being y-primitive and σ_y-free. Then*

$$\deg_x(u_\ell) \le \deg_x(u) + \deg_x(b)\deg_y(b), \quad (\deg_x(b_\ell), \deg_y(b_\ell)) = (\deg_x(b), \deg_y(b)),$$

$$and \quad (\deg_x(a_\ell), \deg_y(a_\ell)) \le (\deg_x(a) + \deg_x(b)\deg_y(b), \deg_y(b) - 1).$$

Proof. Since b_ℓ is y-primitive, it admits the full factorization of the form $b_\ell = c_\ell p_1^{e_1} \dots p_m^{e_m}$, where $c_\ell \in \mathbb{K}$ and $p_1, \dots, p_m \in \mathbb{K}[x, y] \setminus \mathbb{K}[x]$ are distinct, monic and irreducible factors of b_ℓ of multiplicities $e_1, \dots, e_m$, respectively. Then by (Huang, 2016, Proposition 5.2), $\sigma_x^\ell(b)$ must have the form

$$\sigma_x^\ell(b) = c\, \sigma_y^{k_1}(p_1)^{e_1} \cdots \sigma_y^{k_m}(p_m)^{e_m} \quad \text{for some } c \in \mathbb{K} \text{ and } k_1, \dots, k_m \in \mathbb{Z}. \tag{6.5}$$

Consequently, $(\deg_x(b_\ell), \deg_y(b_\ell)) = (\deg_x(b), \deg_y(b))$.

On the other hand, notice that b_ℓ is σ_y-free, so $\sigma_y^{k_1}(p_1), \dots, \sigma_y^{k_m}(p_m)$ are pairwise coprime. Based on the factorization (6.5) of $\sigma_x^\ell(b)$, we then find unique polynomials $\tilde{u} \in \mathbb{K}[x]$ and $f_1, \dots, f_m \in \mathbb{K}[x, y]$ with $\deg_y(f_i) < e_i \deg_y(p_i)$ such that

$$\sigma_x^\ell\left(\frac{a}{b}\right) = \frac{1}{\tilde{u}}\left(\frac{f_1}{\sigma_y^{k_1}(p_1)^{e_1}} + \cdots + \frac{f_m}{\sigma_y^{k_m}(p_m)^{e_m}}\right). \tag{6.6}$$

27

Since r_ℓ is a σ_y-remainder of $\sigma_x^\ell(r)$, then $r_\ell - \sigma_x^\ell(r)$ is σ_y-summable. Notice that each $f_i/\sigma_y^{k_i}(p_i)^{e_i}$ differs from $\sigma_y^{-k_i}(f_i)/p_i^{e_i}$ by a σ_y-summable rational function. We conclude from (6.6) that

$$\frac{a_\ell}{u_\ell b_\ell} - \frac{1}{\sigma_x^\ell(u)\tilde{u}}\left(\frac{\sigma_y^{-k_1}(f_1)}{p_1^{e_1}} + \cdots + \frac{\sigma_y^{-k_m}(f_m)}{p_m^{e_m}}\right) \tag{6.7}$$

is σ_y-summable. Observe that the denominator of the above rational function divides b_ℓ over $\mathbb{K}(x)$, so it is σ_y-free. Since the rational function (6.7) is evidently y-proper, it is a σ_y-remainder by definition. It thus follows from Proposition 6.1 that (6.7) is equal to zero, that is, $a_\ell/(u_\ell b_\ell) = 1/(\sigma_x^\ell(u)\tilde{u})\sum_{i=1}^m \sigma_y^{-k_i}(f_i)/p_i^{e_i}$. Since $b_\ell = c_\ell p_1^{e_1}\ldots p_m^{e_m}$ is y-primitive, u_ℓ divides $\sigma_x^\ell(u)\tilde{u}$ in $\mathbb{K}[x]$ and then $\deg_x(a_\ell) \le \max_{1\le i\le m}\{\deg_x(f_i) + \deg_x(b_\ell) - e_i\deg_x(p_i)\}$. The degree estimates for u_ℓ and a_ℓ thus follow by one application of Fact 5.1 (ii) to (6.6). $\qquad\square$

The reduction-based approach also provides us an order-degree curve of telescopers for bivariate rational functions.

Lemma 6.8. *Let $r = a/(ub) \in \mathbb{K}(x, y)$ be a σ_y-remainder, where $u \in \mathbb{K}[x]$ and $a, b \in \mathbb{K}[x, y]$ with $\deg_y(a) < \deg_y(b)$, $\gcd(a, ub) = 1$ and b being y-primitive, σ_y-free and integer-linear. Assume that b admits the refined integer-linear decomposition of the form given by the right-hand side of (2.5). Define $\rho_0 = \sum_{i=1}^m \mu_i \deg_z(p_i)\max\{e_{i1},\ldots,e_{in_i}\}$. Then for any nonnegative integer pair (ρ, τ) with $\rho \ge \rho_0$ and*

$$\tau > \frac{\left((\rho + 1)\deg_x(b)\deg_y(b) + \rho\deg_x(u) + \deg_x(a) + \rho_0\right)\rho_0 - \frac{1}{2}\rho_0(\rho_0 - 1) - (\rho + 1)}{\rho + 1 - \rho_0}, \tag{6.8}$$

there exists a telescoper for r of order at most ρ and degree at most τ.

Proof. Let $\rho, \tau \in \mathbb{N}$ with $\rho \ge \rho_0$ and τ satisfying (6.8). In order to show the lemma, it amounts to proving that there exist $c_0,\ldots,c_\rho \in \mathbb{K}[x]$, not all zero, with $\deg_x(c_\ell) \le \tau$ such that

$$c_\rho r_\rho + \cdots + c_0 r_0 = 0, \tag{6.9}$$

where $r_0 = r$ and r_ℓ is the adjusted remainder of $\sigma_x(r_{\ell-1})$ by r_0 for $\ell = 1,\ldots,\rho$, because then, by Proposition 6.1 and Remark 6.3, the operator $\sum_{\ell=0}^\rho c_\ell S_x^\ell$ gives a desired telescoper for r. This then suffices to verify that, for the linear homogeneous system over $\mathbb{K}$ induced by (6.9), the number of unknowns, namely $(\tau + 1)(\rho + 1)$ in this case, is greater than the number of equations over $\mathbb{K}$. By (Huang, 2016, Theorem 5.5) and Lemma 6.7, the denominator of the left-hand side of (6.9) in $\mathbb{K}[x, y]$ has total degree in x, y at most $(\rho + 1)(\deg_x(u) + \deg_x(b)\deg_y(b)) + \rho_0$. By separately applying Lemma 6.7 to $r_0,\ldots,r_\rho$, one then calculates that there are at most

$$\left(\tau + (\rho + 1)\deg_x(b)\deg_y(b) + \rho\deg_x(u) + \deg_x(a) + \rho_0\right)\rho_0 - \frac{1}{2}\rho_0(\rho_0 - 1)$$

equations over $\mathbb{K}$. Since $\rho \ge \rho_0$ and (6.8) holds, a direct comparison between the number of unknowns and the above number completes the proof. $\qquad\square$

We remark that for "generic" rational functions, ρ_0 defined in the above lemma coincides with the one given by (5.2), although there are cases in which the latter is smaller. Let $f \in \mathbb{K}(x, y)$ with $\deg_x(f) = d_x$ and $\deg_y(f) = d_y$ be a rational function admitting r as a σ_y-remainder. Lemma 6.8 then asserts that there exists a minimal telescoper for f of degree in $O(\rho_0^2 d_x d_y)$. This compares to Lemma 5.6 which tells us that f can actually have a minimal telescoper of degree in $O(\rho_0 d_x d_y)$.

28

6.2. Cost analysis of algorithm

Lemma 6.9. *Let $r, r_0 \in \mathbb{K}(x,y)$ be two nonzero σ_y-remainders. Write $r = a/(ub)$ with $u \in \mathbb{K}[x]$, $a, b \in \mathbb{K}[x,y]$, $\deg_y(a) < \deg_y(b)$, $\gcd(a, ub) = 1$ and b being y-primitive and σ_y-free. Let $b_0 \in \mathbb{K}[x,y]$ be the y-primitive denominator of r_0. Assume that $\deg_x(b), \deg_x(b_0) \leq d_x$ and $\deg_y(b), \deg_y(b_0) \leq d_y$. Then the adjusted σ_y-remainder $\tilde{r}$ of r by r_0 can be computed using $O(\deg_x(a)^2 d_y + d_x^2 d_y^3 + \deg_x(u) d_x d_y)$ arithmetic operations in $\mathbb{K}$ with classical arithmetic and $\tilde{O}(\deg_x(a) d_y + d_x d_y^2 + \deg_x(u))$ with fast arithmetic, plus the cost of computing the dispersion set of b with respect to b_0.*

Proof. By Proposition 6.2, the adjusted σ_y-remainder $\tilde{r}$ of r by r_0 is obtained by computing the SCD-based partial fraction decomposition (6.3) of r with respect to r_0, along with a subsequent normalization based on (6.4). Notice that with the dispersion set $DS_y(b, b_0)$ at hand, computing the shift-coprime decomposition of b with respect to b_0 merely involves GCD computations with arguments of degree in x no more than d_x and degree in y no more than d_y. Together with the cost of partial fraction decomposition, deriving (6.3) takes $O(\deg_y(a)^2 d_y + d_x^2 d_y^3)$ arithmetic operations with classical arithmetic and $\tilde{O}(\deg_y(a) d_y + d_x d_y^2)$ with fast arithmetic, plus the cost of computing the dispersion set of b with respect to b_0. Based on Facts 5.1-5.2, the final normalization of (6.4) for $\tilde{r}$ requires $O(\deg_y(a)^2 d_y + d_x^2 d_y^3 + \deg_x(u) d_x d_y)$ arithmetic operations with classical arithmetic and $\tilde{O}(\deg_y(a) d_y + d_x d_y^2 + \deg_x(u))$ with fast arithmetic. The announce cost follows. □

Now we are ready to analyze the cost of the reduction-based creative telescoping algorithm for bivariate rational functions.

Theorem 6.10. *Let $f \in \mathbb{K}(x,y)$ be a rational function with $\deg_x(f) = d_x$ and $\deg_y(f) = d_y$. Assume that f has a telescoper and let ρ be the actual order of its minimal telescopers. Further let $r \in \mathbb{K}(x,y)$ be a σ_y-remainder of f, and define ρ_0 as in Lemma 6.8. Then the reduction-based algorithm in (Chen et al., 2015) finds a minimal telescoper for f and an unnormalized certificate using $O(d_x d_y^4 + \rho d_x^2 d_y^3 + \rho^3 \rho_0^3 d_x^2 d_y^2 + \rho \rho_0^5)$ arithmetic operations in $\mathbb{K}$ with classical arithmetic and $\tilde{O}(d_x d_y^3 + \rho d_x d_y^2 + \rho^3 \rho_0^{\omega-1} d_x d_y + \rho^2 \rho_0^\omega)$ with fast arithmetic, plus the cost of computing the (auto-)dispersion sets and finding rational roots.*

Proof. By Lemma 5.8, the GGSZ reduction step takes $O(d_x d_y^4 + d_x^2 d_y^3)$ arithmetic operations with classical arithmetic and $\tilde{O}(d_x d_y^3)$ with fast arithmetic, plus the cost of computing the auto-dispersion set. In addition to the cost of finding rational roots in the integer-linearity detection, the cost of the remaining algorithm is dominated by computing adjusted σ_y-remainders and solving linear homogeneous systems in iteration steps. For the ℓ-th iteration with $0 \leq \ell \leq \rho$, by Lemmas 6.6, 6.7 and 6.9, finding the ℓ-th adjusted σ_y-remainder takes $O(d_x^2 d_y^3)$ with classical arithmetic and $\tilde{O}(d_x d_y^2)$ with fast arithmetic, plus the cost of computing relevant dispersion sets. After this, we need to solve a linear system with the coefficient matrix having at most ρ_0 rows and $\ell + 1$ columns. Moreover, the entries of the matrix are of degrees in x in $O(\ell d_x d_y + \rho_0)$. By Fact 5.3, finding a solution requires $O(\ell^2 \rho_0^3 d_x^2 d_y^2 + \rho_0^5)$ with classical arithmetic and $\tilde{O}(\ell^2 \rho_0^{\omega-1} d_x d_y + \ell \rho_0^\omega)$ with fast arithmetic. Since there are ρ iterations, this step in total takes $O(\rho d_x^2 d_y^3 + \rho^3 \rho_0^3 d_x^2 d_y^2 + \rho \rho_0^5)$ with classical arithmetic and $\tilde{O}(\rho d_x d_y^2 + \rho^3 \rho_0^{\omega-1} d_x d_y + \rho^2 \rho_0^\omega)$ with fast arithmetic, yielding the announced cost. □

In analogy to Corollary 5.10, we obtain the following by the above theorem and Lemma 6.8.

Corollary 6.11. *With the assumptions of Theorem 6.10, further let $\mu = \max\{\mu_1, \ldots, \mu_m\}$. Then, without expanding the certificate, the reduction-based algorithm in (Chen et al., 2015) takes $O(\mu^6 d_x^2 d_y^8)$ arithmetic operations in $\mathbb{K}$ with classical arithmetic and $\tilde{O}(\mu^{\omega+2} d_x d_y^{\omega+3})$ with fast arithmetic, plus the cost of computing (auto-)dispersion sets and finding rational roots.*

Proof. It is evident from the definition of ρ_0 that $\rho_0 \in O(\mu d_y)$. By Lemma 6.8, $\rho \leq \rho_0$ since ρ is the actual order of minimal telescopers for f. Thus $\rho \in O(\mu d_y)$. The announced cost then directly follows by Theorem 6.10. $\qquad\square$

The above result compares to Corollary 5.10 which announces that for the same purpose, the algorithm **RationalCT** takes $O(\mu^4 d_x^2 d_y^6)$ arithmetic operations in $\mathbb{K}$ with classical arithmetic and $\tilde{O}(\mu^{\omega+1} d_x d_y^{\omega+2})$ with fast arithmetic, plus the cost of computing auto-dispersion sets and finding rational roots.

Note that for a polynomial $b \in \mathbb{K}[x, y]$, computing its auto-dispersion set and computing the dispersion set $DS_y(\sigma_x(b), b)$ take almost the same cost. Hence the extra costs for the two algorithms in fact do not differ too much.

7. Implementation and timings

We have implemented our algorithms in the computer algebra system MAPLE 2018. Our implementation includes the four enhancements discussed in Section 4.3. The code is available by email request. In order to get an idea about the efficiency, we compared their running time and memory requirements to the performance of two known algorithms – the one developed by Le (2003) and the reduction-based one reviewed in Section 6. The implementation for the former algorithm uses the built-in Maple procedure **SumTools[Hypergeometric][ZpairDirect]**, while the implementation for the latter algorithm was done in accordance with descriptions of the algorithm **ReductionCT** from (Chen et al., 2015) restricted to the rational case, by embracing the GGSZ reduction and Remark 6.3. All timings are measured in seconds on a Linux computer with 128GB RAM and fifteen 1.2GHz Dual core processors. The computations for the experiments did not use any parallelism.

We take examples of the expanded form of

$$r(x, y) = (S_y - 1)\left(\frac{f_0(x, y)}{g_0(x, y)}\right) + \frac{f(x, y)}{g_1(-\lambda x + \mu y) \cdot g_2(\lambda x + \mu y)}, \tag{7.1}$$

where

- $f_0, f \in \mathbb{Z}[x, y]$ of total degree $m \geq 0$ and max-norm $\|f_0\|_\infty, \|f\|_\infty \leq 20$;

- $g_0 \in \mathbb{Z}[x, y]$ of total degree $n \geq 0$ and max-norm $\|g_0\|_\infty \leq 20$;

- λ, μ are positive integers;

- $g_i \in \mathbb{Z}[z]$ of the form $g_i = p_i(z)p_i(z + \lambda_i)p_i(z + \lambda_i\mu)p_i(z + \lambda_i + \lambda_i\mu)$ for $\lambda_i = (-1)^i\lambda$ and $p_i \in \mathbb{Z}[z]$ of total degree $n > 0$ and max-norm $\|p_i\|_\infty \leq 20$.

Note that in a generic situation, a rational function $r \in \mathbb{Q}(x, y)$ of the form (7.1) admits the following RILD-based partial fraction decomposition

$$r = (S_y - 1)\left(\frac{f_0(x, y)}{g_0(x, y)}\right) + M_1\left(\frac{1}{p_1(-\lambda x + \mu y)}\right) + M_2\left(\frac{1}{p_2(\lambda x + \mu y)}\right),$$

where $M_i = a_{i0} + a_{i1} S_{\lambda_i,\mu}^{\lambda_i} + a_{i2} S_{\lambda_i,\mu}^{\lambda_i\mu} + a_{i3} S_{\lambda_i,\mu}^{\lambda_i+\lambda_i\mu}$ for some $a_{i0}, a_{i1}, a_{i2}, a_{i3} \in \mathbb{Q}(x)[y]$. As such, by modulo some σ_y-summable rational function, it can be further reduced to

$$R_1\left(\frac{1}{p_1(-\lambda x + \mu y)}\right) + R_2\left(\frac{1}{p_2(\lambda x + \mu y)}\right)$$

with $R_i = b_{i0} + b_{i1} S_{\lambda_i,\mu}^{\lambda_i}$ for some $b_{i0}, b_{i1} \in \mathbb{Q}(x)[y]$.

For a selection of random rational functions of this type for different choices of (m, n, λ, μ), Table 1 collects the timings, without expanding the certificate, of the algorithm of Le (DCT), the reduction-based algorithm (RCT) and our algorithm (OCT) developed in Section 4. The column *order* is used to record the actual order of the output minimal telescoper.

(m, n, λ, μ)	DCT	RCT	OCT	order
(1, 1, 1, 1)	0.18	0.17	0.16	2
(1, 1, 4, 1)	0.18	0.20	0.16	2
(1, 1, 16, 1)	0.19	0.21	0.17	2
(5, 1, 4, 1)	0.22	0.23	0.19	3
(10, 1, 4, 1)	0.26	0.27	0.21	3
(15, 1, 4, 1)	0.46	0.40	0.27	4
(15, 1, 4, 5)	10.43	14.63	0.90	10
(15, 1, 4, 7)	46.39	69.64	1.92	14
(15, 1, 4, 9)	181.34	283.65	3.58	18
(15, 1, 4, 11)	456.69	851.72	7.49	22
(15, 1, 4, 13)	892.44	2436.57	13.59	26
(1, 2, 4, 1)	–	15.24	2.48	7
(1, 3, 4, 1)	–	1220.58	49.19	11
(1, 4, 4, 1)	–	30599.21	935.41	15
(10, 2, 4, 1)	–	21.00	3.96	7
(20, 2, 4, 1)	–	27.27	5.92	7
(30, 2, 4, 1)	–	51.82	14.55	8
(30, 2, 4, 3)	–	504.78	51.93	12
(30, 2, 4, 5)	–	6437.51	436.25	20
(30, 2, 4, 7)	–	47763.39	1283.01	28

Table 1: Comparison of three algorithms for a collection of rational functions of the form (7.1).

From the finding we see that our creative telescoping algorithm has comparable timings for random problems of small size. In particular none of the three algorithms have significant set up costs. As m increases our algorithm shows significant improvement over both the direct and reduction-based methods. The dash in the column DCT indicates that the current built-in procedure for DCT in MAPLE 2018 is not applicable for random inputs with this choice of (m, n, λ, μ). The issue in these cases is that the denominator of the input rational function has irreducible factors of degrees greater than one, and then the algorithm of Le (2003) requires recurrence operators with coefficients being polynomials over algebraic numbers, something not yet included in the current implementation of DCT in Maple.

8. Conclusion and future work

A new algorithm of creative telescoping for bivariate rational functions has been developed in this paper. Our algorithm is based on basic arithmetic in the ring of recurrence operators and expresses the certificate part by a compact representation, which, if desired, can be expanded in time polynomial in the size of the final result. In terms of complexity, our algorithm outperforms the reduction-based approach in the case of bivariate rational functions by at least one order of magnitude ignoring the certificate part. In practice, our algorithm is also more efficient according to the experiments.

With the rational case being settled, it is natural to wonder about an analogous algorithm for hypergeometric terms. Recall that a bivariate function $f(x, y)$ is called a *hypergeometric term* if both $f(x + 1, y)/f(x, y)$ and $f(x, y + 1)/f(x, y)$ are rational functions in x, y. The hypergeometric term is a basic and ubiquitous class of special functions appearing in combinatorics (Petkovšek et al., 1996). It is more interesting and also more challenging than the rational case.

In the hypergeometric case, there exists no direct analog of the partial fraction decomposition of rational functions. Thus the method described in this paper will not work directly for this setting. One possible way to proceed is to first compute a multiplicative decomposition of the given hypergeometic term and then reduce the problem to a rational one (cf. (Abramov and Petkovšek, 2001; Chen et al., 2015)). This way, however, may introduce arithmetic operations on recurrence operators over $\mathbb{K}(x, y)$ instead of $\mathbb{K}(x)[y]$, and thus makes it more difficult to derive a hypergeometric telescoping criterion, namely an analog of Theorem 4.3. In the future, we hope to explore this topic further and aim at generalizing our results to the class of hypergeometric terms and beyond.

Acknowledgments

References

Abramov, S. A., 1975. The rational component of the solution of a first-order linear recurrence relation with a rational right side. USSR Comput. Math. Math. Phys. 15 (4), 216–221.
URL https://doi.org/10.1016/0041-5553(75)90181-0

Abramov, S. A., Le, H. Q., 2002. A criterion for the applicability of Zeilberger's algorithm to rational functions. Discrete Math. 259 (1-3), 1–17.
URL http://dx.doi.org/10.1016/S0012-365X(02)00442-9

Abramov, S. A., Petkovšek, M., 2001. Minimal decomposition of indefinite hypergeometric sums. In: Proceedings of ISSAC'01. ACM, New York, pp. 7–14.
URL http://dx.doi.org/10.1145/384101.384103

Bostan, A., Chen, S., Chyzak, F., Li, Z., 2010. Complexity of creative telescoping for bivariate rational functions. In: Proceedings of ISSAC'10. ACM, New York, pp. 203–210.
URL http://dx.doi.org/10.1145/1837934.1837975

Bostan, A., Chyzak, F., Lairez, P., Salvy, B., 2018. Generalized Hermite reduction, creative telescoping and definite integration of d-finite functions. In: Proceedings of ISSAC'18. ACM, New York, pp. 95–102.
URL http://doi.acm.org/10.1145/3208976.3208992

Bostan, A., Lairez, P., Salvy, B., 2013. Creative telescoping for rational functions using the Griffiths-Dwork method. In: Proceedings of ISSAC'13. ACM, New York, pp. 93–100.
URL http://dx.doi.org/10.1145/2465506.2465935

Chen, S., 2019. A reduction approach to creative telescoping. In: Proceedings of ISSAC'19. ACM, New York, pp. 11–14.
URL http://doi.acm.org/10.1145/3326229.3326277

Chen, S., Hou, Q.-H., Huang, H., Labahn, G., Wang, R.-H., 2019. Constructing minimal telescopers for rational functions in three discrete variables. Preprint: arXiv:1904.11614.

Chen, S., Huang, H., Kauers, M., Li, Z., 2015. A modified Abramov-Petkovšek reduction and creative telescoping for hypergeometric terms. In: Proceedings of ISSAC'15. ACM, New York, pp. 117–124.
URL http://doi.acm.org/10.1145/2755996.2756648

Chen, S., Kauers, M., 2012. Order-degree curves for hypergeometric creative telescoping. In: Proceedings of ISSAC'12. ACM, New York, pp. 122–129.
URL http://dx.doi.org/10.1145/2442829.2442850

Cohn, P. M., 1985. Free rings and their relations, 2nd Edition. Vol. 19 of London Mathematical Society Monographs. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers],London.

von zur Gathen, J., Gerhard, J., 2013. Modern Computer Algebra, 3rd Edition. Cambridge University Press, Cambridge.
URL http://dx.doi.org/10.1017/CBO9781139856065

Gerhard, J., 2004. Modular Algorithms in Symbolic Summation and Symbolic Integration (Lecture Notes in Computer Science). Springer-Verlag.

Gerhard, J., Giesbrecht, M., Storjohann, A., Zima, E. V., 2003. Shiftless decomposition and polynomial-time rational summation. In: Proceedings of ISSAC'03. ACM, New York, pp. 119–126.
URL http://dx.doi.org/10.1145/860854.860887

Giesbrecht, M., Huang, H., Labahn, G., Zima, E., 2019. Efficient integer-linear decomposition of multivariate polynomials. In: Proceedings of ISSAC'19. ACM, New York, pp. 171–178.
URL http://doi.acm.org/10.1145/3326229.3326261

van der Hoeven, J., 2020. Constructing reductions for creative telescoping. Appl. Algebra Eng. Commun. Comput.
URL http://dx.doi.org/10.1007/s00200-020-00413-3

Huang, H., 2016. New bounds for hypergeometric creative telescoping. In: Proceedings of ISSAC'16. ACM, New York, pp. 279–286.
URL http://doi.acm.org/10.1145/2930889.2930893

Le, H. Q., 2003. A direct algorithm to construct the minimal Z-pairs for rational functions. Adv. in Appl. Math. 30 (1-2), 137–159, Formal power series and algebraic combinatorics (Scottsdale, AZ, 2001).
URL http://dx.doi.org/10.1016/S0196-8858(02)00529-8

Li, Z., Zhang, Y., 2013. An algorithm for decomposing multivariate hypergeometric terms. A contributed talk in CM'13.

Man, Y.-K., Wright, F. J., 1994. Fast polynomial dispersion computation and its application to indefinite summation. In: Proceedings of ISSAC'94. ACM, New York, pp. 175–180.
URL http://doi.acm.org/10.1145/190347.190413

Paule, P., 1995. Greatest factorial factorization and symbolic summation. J. Symbolic Comput. 20 (3), 235–268.
URL http://dx.doi.org/10.1006/jsco.1995.1049

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

Polyakov, S. P., 2011. Indefinite summation of rational functions with factorization of denominators. Program. Comput. Softw. 37 (6), 322–325, translated from Programmirovanie **3**7 (2011), no. 4.
URL https://doi.org/10.1134/S0361768811060077

Rowen, L. H., 1988. Ring theory. Vol. I. Vol. 127 of Pure and Applied Mathematics. Academic Press, Inc., Boston, MA.

Zeilberger, D., 1990a. A fast algorithm for proving terminating hypergeometric identities. Discrete Math. 80 (2), 207–211.
URL http://dx.doi.org/10.1016/0012-365X(90)90120-7

Zeilberger, D., 1990b. A holonomic systems approach to special functions identities. J. Comput. Appl. Math. 32 (3), 321–368.
URL http://dx.doi.org/10.1016/0377-0427(90)90042-X

Zeilberger, D., 1991. The method of creative telescoping. J. Symbolic Comput. 11 (3), 195–204.
URL http://dx.doi.org/10.1016/S0747-7171(08)80044-2

Zhou, W., Labahn, G., Storjohann, A., 2012. Computing minimal nullspace bases. In: Proceedings of ISSAC'12. ACM, New York, pp. 366–373.
URL https://doi.org/10.1145/2442829.2442881

Zima, E., 2011. Synthetic division in the context of indefinite summation. In: Proceedings of SNC'11. ACM, New York, pp. 151–152.
URL http://doi.acm.org/10.1145/2331684.2331708