J. Symbolic Computation (1993) **16**, 355–376

On Computing Closed Forms for Indefinite Summations

YIU-KWONG MAN

School of Mathematical Sciences, QMW, University of London, U.K.

(Received 2 July 1993; Revised 15 November 1993)

A decision procedure for finding closed forms for indefinite summation of polynomials, rational functions, quasipolynomials and quasirational functions is presented. It is also extended to deal with some non-hypergeometric sums with rational inputs, which are not summable by means of Gosper's algorithm. Discussion of its implementation, analysis of degree bounds and some illustrative examples are included.

1. Introduction

Let K be a constant field of characteristic zero.[†] Finding a closed form solution for an indefinite sum $\sum_x f(x)$,[‡] is equivalent to finding a formula $g(x)$ such that $\Delta g(x) = f(x)$ (or simply $\Delta g = f$, if the summation index is understood)[§]. In Karr (1981,1985), he defined a difference field as follows:

DEFINITION 1.1. *A difference field is a field F together with an automorphism σ of F. The constant field $K \in F$ is the fixed field of σ.*

One can think of σ as a shift operator: $\sigma(g(x)) = g(x+1)$ and it is related to Δ by the relation $\Delta = \sigma - 1$. With this definition, we can say more precisely what we meant by solving an indefinite summation problem: – i.e. given an element $f \in F$, compute, if it exists, an element $g \in F$ (or in a difference field extension E of F) such that $\Delta g = f$. Karr (1981) actually presented a decision procedure to solve such a problem. But his algorithm is more theoretical than practical; even a simple sum (e.g. Karr (1981) p335 example 11) may need lengthy calculations to decide it is summable or not. In Karr (1981) p349, he made the following remarks:

"Loosely speaking, if f is summable in E, then part of it is summable in F, and the rest consists of pieces whose formal sums have been adjoined to F in the construction of E. This makes the construction of extension fields in which f is summable somewhat uninteresting and

† Unless stated otherwise, K will be taken as $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$.

‡ Note: If x is used as a summation index variable, it is always assumed that $x \in \mathbb{Z}$.

§ $\Delta g(x)$ means $g(x+1) - g(x)$, where Δ is called the upper difference operator.

0747–7171/93/040347+22 \$08.00/0

justifies the tendency to look for sums of $f \in F$ only in F."

The problem of deciding whether the indefinite sum of a rational function is rational was solved by Abramov (1971). Later, he (Abramov (1975)) also generalized his algorithm to search for rational components¶ of the solution of a first order difference equation of the form $g(x+1) + a\ g(x) = f(x)$, where $g(x)$ is unknown, $f(x)$ is a given rational function and $a \neq 0$ is a constant. But he remarked (Abramov 1975 p220) that his earlier algorithm (1971) is more efficient for indefinite rational summations than his 1975 algorithm when it is applied to the special case $a = -1$. (Similarly, his more recent algorithms on nth order linear difference equations (Abramov (1989,1991)) are not the same as his 1971 algorithm nor the quasirational summation algorithm to be presented in section 4 when they are applied to the case $n = 1$. Further discussion of this point can be found in section 7.) The problem of indefinite hypergeometric summation was solved by Gosper (1978). He succeeded in recognizing that if f is hypergeometric‖, then one can convert the difference equation $\Delta g = f$ into a system of linear equations by change of variables, whose solution gives coefficients for a polynomial that yields a closed form for the indefinite hypergeometric sum g. In other words, Gosper's algorithm decides whether the indefinite sum of a hypergeometric term is hypergeometric. Many computer algebra systems (e.g. Macsyma, Reduce, Maple and Mathematica) have implemented it as a built-in solver for indefinite or definite summations. Based on the Gosper algorithm and the theory of holonomic functions, Zeilberger (1991) developed an algorithm for definite hypergeometric summations – his method can be used to evaluate sums of the form $\sum_{k=1}^{n} F(n,k)$, where $F(n,k)$ is hypergeometric in both n and k. More recently, Petkovsek (1992) also used some of Gosper's techniques (as well as some modifications of Abramov's (1989) algorithm) to find all hypergeometric solutions of linear recurrences with polynomial coefficients. Perhaps one might ask the following question: Is there anyone who attempted to express some non-rational or non-hypergeometric indefinite sums in closed form ? The answer is affirmative. Moenck (1977) had made such an attempt in his work on summation of rational functions. He made use of the properties of polygamma functions (see section 3) to express the transcendental part of an indefinite sum in closed form. His method is a discrete analog of the Hermite method of integration and he used summation by parts in his calculations. Therefore, he can express the indefinite sum of any rational functions in closed form which the Gosper or the Abramov algorithm cannot. For example, Moenck's method can be used to find closed forms for Harmonic numbers, which motivated those who are interested in summation of various types of Harmonic numbers (e.g. Lamagna et al (1989)). The procedure to be presented in this paper also uses polygamma functions, whenever it is necessary, to express the transcendental part of the solution, but it is different from Moenck's approach. Its inspiration came from the method of undetermined coefficients** (see Richardson (1954)), the work of Abramov (1971) and Moenck's (1977) method. It can be applied to summation of quasirational functions which are not summable by means of Moenck's method. It can also express a harmonic number in closed form and solve some sub-classes of rational functions for which the Gosper or the Abramov algorithm fails to give closed forms (see section 2). It

¶ This is an analog to Hermite's method of finding the rational part of the integral of a rational function.

‖ i.e. $f(x)/f(x-1)$ is a rational function of x.

** One can consult Ross (1991) on why such methods work for both differential and difference equations.

is a decision procedure for summation of rational or quasirational functions. Discussion of its implementation, efficiency and degree bound settings can be found in section 6. Illustrative examples are included in various sections as well as the appendix.

2. Non-hypergeometric sums

In this section, some classes of non-hypergeometric sums will be mentioned. The aim is to show that they are not summable by means of Gosper's (1978) or Abramov's (1971, 1989, 1991) algorithms. In a later section, we shall show how to express such classes of sums as closed forms via the use of polygamma functions.

2.1. Gosper decision procedure

For reference purposes, a brief description of the main steps of Gosper's algorithm is given here (see Gosper (1978), Lafon (1983) or Paule (1991)).

Algorithm 2.1. *(Gosper).*

Input: *a summand a_k, where k is an index variable.*

Output: *a hypergeometric indefinite sum in closed form, if it exists, or say "no hypergeometric indefinite sum exists".*

Procedure:

S1: *Convert the ratio a_k/a_{k-1} into the form*

$$\frac{a_k}{a_{k-1}} = \frac{p_k}{p_{k-1}} \times \frac{q_k}{r_k}$$

where p_k, q_k and r_k are polynomials in k such that $\gcd(q_k, r_{k+j}) = 1$ *for all non-negative integers j.*†

S2: *Let $l = deg(q_{k+1} + r_k)$ and $s = deg(q_{k+1} - r_k)$. Obtain a degree bound α as follows:*
(a) If $l \leq s$ then $deg(p_k) - s \mapsto \alpha$ else
(b) Define $\alpha_0 = -2l_1/l_2$, where l_1 is the coefficient of k^{l-1} in $q_{k+1} - r_k$ and l_2 is the leading coefficient of $q_{k+1} + r_k$. If α_0 is an integer then $max(\alpha_0, deg(p_k) - l + 1) \mapsto \alpha$ else $deg(p_k) - l + 1 \mapsto \alpha$.
If $\alpha < 0$ then return and say "no hypergeometric indefinite sum exists".
(Note: Assume the degree of the zero polynomial is -1.)

S3: *Construct a generic polynomial $f(k)$ with degree α and then determine its unknown coefficients by equating equal powers of k in the following equation*

$$p_k = q_{k+1} f(k) - r_k f(k-1).$$

† This can be done by computing the resultant of q_k and r_{k+j}, and performing 'gcd' operations. (see Lafon (1983) for details).

S4: *If the unknown coefficients of $f(k)$ can be solved, then a hypergeometric indefinite sum exists and it is equal to $q_{k+1}f(k)a_k/p_k$, otherwise return and say "no hypergeometric indefinite sum exists".*

2.2. Results

LEMMA 2.1.

$$\sum_k \frac{1}{k^m}$$

is not summable by Gosper's algorithm, where m is a positive integer.

PROOF. Let $a_k = \frac{1}{k^m}$. Then $\frac{a_k}{a_{k-1}} = \frac{(k-1)^m}{k^m}$. Set $p_k = 1, q_k = (k-1)^m, r_k = k^m$. It is obvious that $\gcd(q_k, r_{k+j}) = 1$ for all non-negative integer j. Setting $l = m, s = -1, \alpha_0 = 0$, we find that the degree bound of $f(k)$ is zero. Let $f(k) = c$, where c is a constant. We get $q_{k+1}f(k) - r_k f(k-1) = k^m c - k^m c = 0$, which is not equal to $p_k = 1$. Therefore $\sum_k \frac{1}{k^m}$ is not summable by the Gosper algorithm. □

LEMMA 2.2.

$$\sum_k \frac{p(k)}{\prod_{j=0}^{m-1}(k+a+j)}$$

is not summable by Gosper's algorithm, where $p(k)$ is a non-zero polynomial in k, with deg $p(k) = m-1$, m is a positive integer and a is a constant.

PROOF. Let

$$a_k = \frac{p(k)}{\prod_{j=0}^{m-1}(k+a+j)}.$$

Then

$$\frac{a_k}{a_{k-1}} = \frac{(k+a-1)p(k)}{(k+a+m-1)p(k-1)}.$$

Set $p_k = p(k)$, $q_k = k+a-1$ and $r_k = k+a+m-1$. Obviously, $\gcd(q_k, r_{k+j}) = 1$ for all non-negative integer j and we find that $l = 1, s = 0$ or -1[†]. So, $\alpha_0 = m-1$ and $f(k)$ is a polynomial of degree $m-1$, say $f(k) = c_{m-1}k^{m-1} + c_{m-2}k^{m-2} + \cdots + c_0$. Substituting into $q_{k+1}f(k) - r_k f(k-1)$ and simplifying, we obtain that the coefficient of k^{m-1} is zero. But $deg(p(k)) = m-1$ and hence the lemma holds. □

EXAMPLE 2.1. $\sum_k \frac{2k-1}{(k+1)(k+2)}$ *and* $\sum_k \frac{4k-3}{(k+1)(k+3)}$ *are non-hypergeometric sums. The first one is obvious in view of Lemma 2.2 above. For the second one, if we rewrite it as* $\sum_k \frac{(4k-3)(k+2)}{(k+1)(k+2)(k+3)}$, *then we can apply Lemma 2.2 again.*

Perhaps a remark which should be made here is that Gosper (1978) proved the correctness of his procedure by stipulating the sum is hypergeometric first and deriving that the input summand must be hypergeometric and then obtaining a degree bound for $f(k)$

[†] These 2 values correspond to the cases $m \neq 1$ and $m = 1$.

such that $p_k = q_{k+1}f(k) - r_k f(k-1)$. But he does not assert that a summand which is hypergeometric (e.g. rational or quasirational functions) must have a hypergeometric sum, as clearly shown in the results above.

3. Background

In this section, some definitions and known results are mentioned. They will be used in subsequent sections.

DEFINITION 3.1.
(a) The digamma function is defined as

$$\psi(x) = \frac{d}{dx}\log\Gamma(x),$$

where $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt \quad (x > 0)$ *is the gamma function.*

(b) The polygamma function of order n is defined as

$$\psi(n,x) = \frac{d^n}{dx^n}\psi(x) = \frac{d^{n+1}}{dx^{n+1}}\log\Gamma(x).$$

These definitions follow those in Abramowitz and Stegun(1970). The following recurrence relations[†], which are important in representing harmonic numbers in closed forms, are also mentioned in their book.

LEMMA 3.1.
(a)

$$\Delta\psi(x) = \psi(x+1) - \psi(x) = \frac{1}{x}.$$

(b)

$$\Delta\psi(n-1,x) = \psi(n-1,x+1) - \psi(n-1,x) = \frac{(-1)^{n-1}(n-1)!}{x^n}$$

Summing from $x = 1$ to $m-1$ on both sides, we can obtain the following results.

COROLLARY 3.1.
(a)

$$\sum_{x=1}^{m-1}\frac{1}{x} = \psi(m) - \psi(1) = \psi(m) + \gamma,$$

where $\gamma \approx 0.5772156649$ *is called the Euler constant.*

[†] Proofs can be found in Richardson (1954).

(b)

$$\sum_{x=1}^{m-1} \frac{1}{x^n} = \frac{\psi(n-1,m)-\psi(n-1,1)}{(n-1)!} \times (-1)^{n-1}.$$

One can readily see that harmonic numbers can be put into closed form via the use of polygamma functions. Now let us explain the meaning of 'dispersion',† which was defined in Abramov (1971, 1974).

DEFINITION 3.2. *The dispersion of a polynomial $f(x)$, with deg $f(x) > 0$,‡ is defined as*

$$dis\ f(x) = \max\{\alpha \in \mathbb{Z}^+ \cup \{0\} | deg(\gcd(f(x), f(x+\alpha)) \geq 1\}.$$

Since $f(x)$ has only a finite number of irreducible factors, *dis* $f(x)$ is well-defined. In short, it is a measure of the shift-freeness of a polynomial. e.g. *dis* $(x+1) = 0$ and *dis* $(x+1)(x+2) = 1$. It can be determined by computing the resultant $res(f(x), f(x+\alpha))$ w.r.t. x, which is a polynomial in α, and then solving for the non-negative integer roots of such a polynomial.

4. Decision procedures

In this section, we shall derive the conditions under which an indefinite sum of a quasipolynomial is a quasipolynomial or an indefinite sum of a quasirational function is a quasirational function. Before we describe the proof, let us mention some of the results for summation of polynomials and rational functions first.

4.1. Summation of polynomials

In section 1, we have mentioned that solving an indefinite summation problem $\sum_x f(x)$ is equivalent to solving a first order difference equation $\Delta g(x) = f(x)$. Analogously to the case of differential equations, $\Delta g = (\sigma - 1)g = 0$ is called the associated homogeneous equation of $\Delta g = f$, and $m - 1 = 0$ is called the characteristic equation of $\Delta g = 0$. Since 1 is the root of multiplicity 1 of the characteristic equation, it follows from the general result§ stated in Abramov (1974, p247) that

THEOREM 4.1. *If $f(x)$ is a polynomial of degree m, then $\sum_x f(x)$ is summable and the closed form is a polynomial of degree $m+1$.*

So one only needs to construct a generic polynomial $g(x)$ of degree $m+1$ and use the method of undetermined coefficients to determine the coefficients.

† Karr (1981, p342) also gave a very similar definition, although he did not call it 'dispersion'.

‡ If $f(x)$ is a constant, then we define *dis* $f(x)$ as zero.

§ For a linear recurrence relation with constant coefficients in the left hand side and a polynomial right hand side, say $F(x)$, if λ is the multiplicity of the root 1 of the characteristic equation, then the degree of the polynomial solution is of $deg\ F(x) + \lambda$.

4.2. Summation of rational functions

Without loss of generality, we assume all the rational functions considered here are proper rational functions, i.e. the degree of the numerator is strictly less than the degree of the denominator.¶ As stated in section 1 before, K denotes a constant field of characteristic zero. Below, $K(x)$ and $K[x]$ denote the field of rational functions and the ring of polynomials in x over K respectively. The following results and proofs can be found in Abramov (1971).

THEOREM 4.2. *Let $f(x)$, $g(x) \in K(x)$ such that $\Delta g(x) = f(x)$ and their irreducible forms*† *be $\frac{f_1(x)}{f_2(x)}$ and $\frac{g_1(x)}{g_2(x)}$ respectively, where $f_i(x)$, $g_i(x) \in K[x]$ $(i = 1, 2)$. Then*

$$dis\ g_2(x) = dis\ f_2(x) - 1.$$

This theorem provides us some information about the denominator of the prospective indefinite sum $g(x)$ once the dispersion of the denominator of the summand, say α is known. It also tells us that if $f(x)$ is a proper rational function with $\alpha = 0$, then $\sum_x f(x)$ is not summable in $K(x)$, otherwise $dis\ g_2(x) = -1$, which would contradict the definition of 'dispersion' being non-negative.

EXAMPLE 4.1. *If $f(x) = \frac{1}{x(x+3)}$, then it is easy to see that $dis\ x(x+3) = 3$ and the denominator of $g(x)$ is $x(x+1)(x+2)$ which has dispersion equal to 2.*

The following algorithm tells us more precisely how to construct both the numerator and the denominator of $g(x)$, if a rational indefinite sum exists. It is a summary of the descriptions in Abramov (1971).

ALGORITHM 4.1. *(Abramov).*

Input: *$f(x) = \frac{f_1(x)}{f_2(x)}$ in irreducible form, where $f_i(x) \in K[x]$ $(i = 1, 2)$ such that $deg\ f_1(x) < deg\ f_2(x)$ and $f_1(x)$ is a non-zero polynomial.*

Output: *a rational sum in closed form, if it exists, or say "not summable in $K(x)$".*

Procedure:

S1: *Compute the dispersion $\alpha = dis\ f_2(x)$. If $\alpha = 0$, then return and say "not summable in $K(x)$".*

S2: *Determine*

$$\frac{s(x)}{t(x)} = \sum_{i=0}^{\alpha-1} \frac{f_1(x+i)}{f_2(x+i)},$$

¶ This will imply that proper rational functions are not polynomials.

† It means the numerator and the denominator are coprime.

where $s(x)$ and $t(x) \in K[x]$. Let $g_2(x)$ be the numerator of the irreducible form of $t(x)/t(x+\alpha)$. Find

$$p(x) = f(x)g_2(x+1)g_2(x).^{‡}$$

Let deg $p(x) = l$ and deg $g_2(x) = m$. Define a degree bound n as follows: if $l > m-2$, then $l - m + 1 \mapsto n$ else $m \mapsto n$.

S3: *Construct a generic polynomial $g_1(x)$ with degree n and determine its unknown coefficients by the following relation*

$$p(x) = g_1(x+1)g_2(x) - g_2(x+1)g_1(x).$$

S4: *If the unknown coefficients of $g_1(x)$ are solvable, then a rational indefinite sum exists and it is equal to $g_1(x)/g_2(x)$, otherwise return and say "not summable in $K(x)$".*

Note: If $f_2(x)$ is a monic polynomial, then a quicker way to compute $g_2(x)$ in **S2** is as follows: Let $h(x) = \gcd(f_2(x), f_2(x+\alpha)) \in K[x]$. Then

$$g_2(x) = \prod_{i=0}^{\alpha-1} h(x - \alpha + i).$$

The following examples illustrate how this algorithm works.

EXAMPLE 4.2. *$\sum_k \frac{1}{k^m}$ is not a rational indefinite sum (cf. Lemma 2.1) because dis $k^m = 0$, where m is a positive integer.*[†]

The following example is taken from Karr (1981, p344), and also appeared in Lafon (1983, p74).[‡]

EXAMPLE 4.3. *To determine the definite sum $\sum_{x=1}^{n} \frac{1}{x(x+2)}$, let $f(x) = \frac{1}{x(x+2)}$ and compute $\alpha =$ dis $x(x+2) = 2$. So $h(x) = x+2$ and $g_2(x) = x(x+1)$. Hence, $p(x) = g_2(x)g_2(x+1)/x(x+2) = (x+1)^2$, $l = m = 2$, and $n = 1$. Let $g_2(x) = c_1x + c_0$. Substituting into $p(x) = g_1(x+1)g_2(x) - g_2(x+1)g_1(x)$ and solving, we obtain $c_1 = -1$, and $c_0 = -1/2$. Hence $g(x) = (-x-1/2)/x(x+1)$ and*

$$\sum_{x=1}^{n} \frac{1}{x(x+2)} = g(n+1) - g(1) = \frac{3n^2+5n}{4(n+1)(n+2)}.$$

Compared with Moenck's method, Abramov's algorithm has some advantages if the indefinite sum exists in $K(x)$ – we do not need to use shift-free decomposition (see Moenck (1977)) and summation by parts in the calculations. Similar to Gosper's algorithm, this algorithm needs to compute a resultant (for the dispersion α) and solve a system of linear equations.

‡ This must be a polynomial since $f_2(x)$ is a factor of $t(x)$ and hence a factor of $g_2(x)$.

† In Karr (1981, p343), he gave a more lengthy proof for the special case $\sum_k \frac{1}{k}$ being non-rational.

‡ Unfortunately, the answers in both papers were misprinted.

4.3. Summation of quasipolynomials

A quasipolynomial is a function of the form $p(x)\lambda^x$ in some difference field F over a constant field K, where $p(x) \in K[x]$ and λ is a non-zero constant and $|\lambda| \neq 1$.

Definition 4.1. *The degree of a quasipolynomial $p(x)\lambda^x$ is defined as the degree of the polynomial $p(x)$.*

The following theorem indicates that a quasipolynomial is always summable in F.

Theorem 4.3. *If $f(x)\lambda^x$ is a quasipolynomial of degree m, then $\sum_x f(x)\lambda^x$ is summable and the closed form is also a quasipolynomial of degree m.*

Proof. Since Σ is a linear operator on the summand, so without loss of generality, it suffices to prove for a quasipolynomial term only. Let $f(x) = x^m\lambda^x$ be the summand, with $m \in \mathbb{Z}$. Suppose $g(x) = (c_n x^n + c_{n-1}x^{n-1} + \cdots + c_0)\lambda^x$, where $c_i \in K$ and $c_n \neq 0$. Consider $g(x+1) - g(x) = f(x)$; we have

$$\lambda(c_n(x+1)^n + c_{n-1}(x+1)^{n-1} + \cdots + c_0)) - (c_n x^n + c_{n-1}x^{n-1} + \cdots + c_0) = x^m.$$

For this equality to hold, we must have $n \geq m$. But $|\lambda| \neq 0$ or 1, and the leading term of the left hand side is $(\lambda - 1)c_n \neq 0$, so n must be equal to m. It is easy to see that $c_m = 1/(\lambda - 1)$. If we consider the coefficient of x^{m-1} on both sides, then we obtain

$$\lambda(c_m m + c_{m-1}) = 0 \text{ and so } c_{m-1} = \frac{m}{\lambda - 1}.$$

Proceeding successively, we can compute each c_i for $0 \leq i \leq m$. In fact, the unknown coefficients satisfy the following recurrence relation

$$c_m = \frac{1}{\lambda - 1}$$

$$c_m \frac{m!}{k!(m-k)!} + c_{m-1}\frac{(m-1)!}{(k-1)!(m-k)!} + \cdots + c_{m-k} = 0(1 \leq k \leq m),$$

for $0 \leq i \leq m$. Therefore, it implies that $\sum_x x^m\lambda^x$ is summable in F and the indefinite sum is a quasipolynomial of degree m. □

Similar to the case of summation of polynomials, one only needs to construct a generic quasipolynomial of degree m and then use the method of undetermined coefficients or the recurrence relation in the proof to determine the coefficients.

4.4. Summation of quasirational functions

A quasirational function is a function of the form $f(x)\lambda^x \in F$ (some difference field over a constant field K), where $f(x) \in K(x)$, λ is a non-zero constant and $|\lambda| \neq 1$. $f(x)$ and λ^x are called the rational part and transcendental part of the quasirational function respectively. Without loss of generality, assume the rational part is a proper fraction (see section 4.2). The following theorem is similar to Theorem 4.2.

THEOREM 4.4. *Let $f(x)\lambda^x$, $g(x)\lambda^x$ be quasirational functions, such that $f(x)$, $g(x) \in K(x)$, $\lambda \in K$ and $\Delta g(x)\lambda^x = f(x)\lambda^x$. If $f_1(x)/f_2(x)$ and $g_1(x)/g_2(x)$ are the irreducible forms of $f(x)$ and $g(x)$ respectively, where $f_i(x)$, $g_i(x) \in K[x]$ for $i = 1, 2$, then $dis\ g_2(x) = dis\ f_2(x) - 1$.*

PROOF. By partial fraction decomposition, we obtain

$$g(x) = \sum_{i=1}^{m} \frac{q_i(x)}{p_i(x)^{n_i}} + r(x),$$

where $p_i(x)$, $q_i(x)$, $r(x) \in K[x]$, $deg\ q_i(x) < deg\ p_i(x)^{n_i}$, $n_i \in \mathbb{Z}$ and $p_i(x)$'s are irreducible in $K[x]$ for $1 \le i \le m$. Let $\alpha = dis\ g_2(x)$ and $p_1(x+\alpha) = p_2(x)$.† Now $\Delta g(x)\lambda^x = f(x)\lambda^x$ implies $\lambda g(x+1) - g(x) = f(x)$, so

$$f(x) = \sum_{i=1}^{m} \frac{\lambda q_i(x+1)}{p_i(x+1)^{n_i}} + \lambda r(x+1) - \sum_{i=1}^{m} \frac{q_i(x)}{p_i(x)^{n_i}} - r(x)$$

$$= \lambda r(x+1) - r(x) + \frac{\lambda q_1(x+1)}{p_1(x+1)^{n_1}} - \frac{q_1(x)}{p_1(x)^{n_1}} +$$

$$\frac{\lambda q_2(x+1)}{p_1(x+\alpha+1)^{n_2}} - \frac{q_2(x)}{p_1(x+\alpha)^{n_2}} + \sum_{i=3}^{m} \left(\frac{\lambda q_i(x+1)}{p_i(x+1)^{n_i}} - \frac{q_i(x)}{p_i(x)^{n_i}} \right).$$

Since the partial fraction decomposition (**pfd**) is unique and the shift from x to $x+1$ transforms an irreducible polynomial into an irreducible polynomial, so the last expression is the **pfd** of $f(x)$. By observing the denominators, no powers of $p_1(x+\alpha+1)$ and $p_1(x)$ can be found under the summation sign in the **pfd** of $f(x)$, hence $dis\ f_2(x) = \alpha + 1$. □

EXAMPLE 4.4. *$\sum_x \frac{2^x}{x+1}$ is not summable in F because $dis\ (x+1) = 0$; otherwise Theorem 4.4 would imply $dis\ g_2(x) = -1$, which contradicts the definition of dispersion being non-negative.*

The following algorithm tells us how to decide if a quasirational function is summable in F or not. The proof is an extension of Abramov's (1971) proof for summation of rational functions.

ALGORITHM 4.2.

Input: *$f(x) = f_1(x)\lambda^x / f_2(x)$ in irreducible form, where $f_i(x) \in K[x]$ for $i = 1, 2$, such that $deg\ f_1(x) < deg\ f_2(x)$, $f_1(x)$ is a non-zero polynomial, λ is a non-zero constant and $|\lambda| \neq 1$.*

Output: *a quasirational sum in closed form, if it exists, or say "not summable in F".*

Procedure:

† If not, it can be done by rearranging the partial fractions and then relabelling each $p_i(x)$.

S1: *Compute the dispersion $\alpha = dis\ f_2(x)$. If $\alpha = 0$, then return and say "not summable in F".*

S2: *Determine*

$$\frac{s(x)}{t(x)} = \sum_{i=0}^{\alpha-1} \frac{f_1(x+i)\lambda^i}{f_2(x+i)},$$

where $s(x),\ t(x) \in K[x]$. Let $g_2(x)$ be the numerator of the irreducible form of $\frac{t(x)}{t(x+\alpha)}$. Determine

$$p(x) = f_1(x)g_2(x)g_2(x+1)/f_2(x).$$

Let $deg\ p(x) = l$ and $deg\ g_2(x) = m$. Define a degree bound $n = l - m$. If $n < 0$, then return and say "not summable in F".

S3: *Construct a generic polynomial $g_1(x)$ with degree n and determine its unknown coefficients by the following relation*

$$p(x) = \lambda g_1(x+1)g_2(x) - g_2(x+1)g_1(x).$$

S4: *If the unknown coefficients of $g_1(x)$ are solvable, then a quasirational indefinite sum exists and it is equal to $g_1(x)\lambda^x/g_2(x)$, otherwise return and say "not summable in F".*

PROOF. Assume $f(x) = f_1(x)\lambda^x/f_2(x)$ is in irreducible form. By Theorem 4.4, if the dispersion $\alpha = dis\ f_2(x) = 0$, then $g(x)$ does not exist. So we can assume $\alpha \geq 1$ and $g(x) = g_1(x)\lambda^x/g_2(x)$ is in irreducible form. Consider the relation $g(x+1) - g(x) = f(x)$. Summing from $i = 0$ to $\alpha - 1$ on both sides, we can obtain

$$g(x+\alpha) - g(x) = \sum_{i=0}^{\alpha-1} \frac{f_1(x+i)\lambda^{x+i}}{f_2(x+i)},$$

which implies

$$\frac{g_1(x+\alpha)g_2(x)\lambda^\alpha - g_1(x)g_2(x+\alpha)}{g_2(x+\alpha)g_2(x)} = \sum_{i=0}^{\alpha-1} \frac{f_1(x+i)\lambda^i}{f_2(x+i)}$$

Suppose $s(x)/t(x)$ is the irreducible form of the last expression. Then we have

$$\frac{t(x)}{t(x+\alpha)} = \frac{g_2(x)}{g_2(x+2\alpha)}.$$

Since $\gcd(g_2(x), g_2(x+2\alpha)) = 1$, so $g_2(x)$ is equal to the numerator of the irreducible form of $t(x)/t(x+\alpha)$. Substituting into the relation $\Delta g(x) = f(x)$, we get

$$\frac{\lambda g_1(x+1)g_2(x) - g_1(x)g_2(x+1)}{g_2(x+1)g_2(x)} = \frac{f_1(x)}{f_2(x)}.$$

Hence, if we let $p(x) = f_1(x)g_2(x)g_2(x+1)/f_2(x)$,† then $p(x) = \lambda g_1(x+1)g_2(x) - g_1(x)g_2(x+1)$. Assume $deg\ p(x) = l$, $deg\ g_2(x) = m$ and $deg\ g_1(x) = n$. Let

$$g_1(x) = c_n x^n + c_{n-1}x^{n-1} + \cdots + c_0$$

† It must be a polynomial following the same argument as in the description of Abramov's algorithm in section 4.2.

and

$$g_2(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_0.$$

where $c_n \neq 0$ and $b_m \neq 0$. Since the leading term of the left hand side of

$$p(x) = \lambda g_1(x+1)g_2(x) - g_1(x)g_2(x+1)$$

is $(\lambda - 1)c_n b_m x^{n+m} \neq 0$, so n must be equal to $deg\ p(x) - m$, i.e. $n = l - m$, provided $l \geq m$. Once the possible degree of $g_1(x)$ is found, the method of undetermined coefficients can be used to solve for the coefficients in $g_1(x)$. If the coefficients are solvable, then an indefinite sum exists and is equal to $g_1(x)\lambda^x/g_2(x)$, otherwise no quasirational indefinite sum exists. □

From this result, one can see that the occurrence of λ makes the determination of the degree bound n much simpler than the case of summation of rational functions, although we have the same way for finding the polynomial $p(x)$ in step 2.

EXAMPLE 4.5. *(Davies (1962)) To determine* $\sum_x 2^x(x^2 - 2x - 1)/(x^2(x+1)^2)$, *first compute the dispersion* $\alpha = dis\ x^2(x+1)^2 = 1$. *Then consider*

$$\frac{s(x)}{t(x)} = \frac{x^2 - 2x - 1}{x^2(x+1)^2}$$

and

$$\frac{t(x)}{t(x+1)} = \frac{x^2}{(x+2)^2}.$$

So $g_2(x) = x^2$ *and* $p(x) = x^2 - 2x - 1$, *which implies* $l = m = 2$ *and* $n = 0$. *Let* $g_1(x) = c_0$, *where* c_0 *is a constant. Substituting into the equation* $p(x) = 2g_1(x+1)g_2(x) - g_1(x)g_2(x+1)$ *and solving, we obtain* $c_0 = 1$ *and hence an indefinite sum* $2^x/x^2$.

5. Summation in a difference field extension

Given a difference field F with an associated automorphism σ, we have already investigated the decision procedures for summation of polynomials, rational functions, quasipolynomials and quasirational functions in section 4. At that stage, we are only interested in computing closed forms in F itself, which certainly cannot guarantee every indefinite sum can be put into closed form. In this section, we attempt to extend a little bit – to allow a closed form to be expressed in a difference field extension of F, which contains the polygamma functions.† The basic facts we use are the recurrence relations mentioned in section 3, which tell us the relationships between the polygamma functions and the harmonic numbers. This will enable us to express the two classes of non-hypergeometric sums mentioned in section 2 in closed forms. Besides, we can also obtain a rational component, if it exists, in addition to the polygamma parts, which is an alternative way of finding a rational component other than the Hermite method of reduction mentioned in Abramov (1975).

† *It is analogous to extending a differential field by adding logarithmic functions to get the closed forms of integrals in indefinite integration theory.*

EXAMPLE 5.1. *In order to express* $\sum_x (x+1)/((x+2)(x+3))$ *in closed form, first we decompose it into* $\sum_x -1/(x+2)$ *and* $\sum_x 2/(x+3)$*, and then apply the recurrence relation* $\Delta\psi(x) = 1/x$ *as well as the inverse recurrence relation* $\Delta^{-1}1/x = \psi(x)$ *to obtain*

$$-\psi(x+2) + 2\psi(x+3) = -\psi(x+2) + 2(\psi(x+2) + \frac{1}{x+2})$$

$$= \psi(x+2) + \frac{2}{x+2},$$

where $\frac{2}{x+2}$ *is called the rational component*† *of the closed form.*

But the drawback in applying this technique is that we need to perform a complete partial fraction decomposition (**cpfd**) because the denominators on the right hand sides of Lemma 3.1‡ must be single linear factors (with multiplicity ≥ 1). It is computationally expensive and may not be easy to do in the case where we have to perform a **cpfd** in an algebraic extension of $\mathbb{Q}(x)$, e.g. decomposing $1/(x^2+\sqrt{5}x-1)$ into $\frac{1}{3(x-(3-\sqrt{5})/2)} - \frac{1}{3(x+(3+\sqrt{5})/2)}$, we will need to work on $\mathbb{Q}(\sqrt{5})(x)$. Besides, it will be somewhat 'overkill' to use polygamma functions if an indefinite sum can be expressed in closed form in F itself.

EXAMPLE 5.2.

$$\sum_x \frac{1}{x(x+2)} = \frac{1}{2}\sum_x \frac{1}{x} - \frac{1}{2}\sum_x \frac{1}{x+2} = \frac{1}{2}(\psi(x) - \psi(x+2))$$

$$= \frac{1}{2}(\psi(x) - \frac{1}{x+1} - \psi(x+1)) = \frac{1}{2}(\psi(x) - \frac{1}{x+1} - \frac{1}{x} - \psi(x)) = \frac{-2x-1}{2x(x+1)}.$$

Therefore, compromising between the pros and cons, it is sensible to work on the summand's defining difference field first; if no closed form can be found and the summand is a rational function, then we can use polygamma functions to find a closed form. Perhaps one remark should be made here is that M. Bronstein and B. Salvy (1993) recently presented an algorithm for full partial fraction decomposition of rational functions, using only gcd operations over the initial coefficient field. For instance, if $P + \sum_\alpha \sum_{i=1}^{n_\alpha} \frac{b_{\alpha,i}}{(x-\alpha)^i}$ is the **cpfd** of a rational function A/D, where P, A, D are polynomials in $K[x]$ and n_α is the multiplicity of the pole α, then it is possible to compute $b_{\alpha,i}$ without any factorization, and using only the operations of the field K. But of course, part of the **cpfd** can only be expressed as a formal sum in the case where α belongs to the algebraic closure of K. For example, $\frac{1}{x^2-2}$ can only be expressed as $\sum_\alpha \frac{\alpha}{4(x-\alpha)}$, where α is the root of $x^2 - 2 = 0$; which means we still need to solve for α if we want to use the polygamma functions. Nevertheless, this partial fraction decomposition algorithm is still quite useful and requires less computational effort since it works on K first and its algebraic closure (if we need to solve for all roots of α) afterwards.

† Note: It is not unique because it depends on how we express the polygamma part. In our case, we can further reduce $\psi(x+2)$ into $\psi(x+1)$ or $\psi(x)$.

‡ This lemma is only applicable to rational summands, and we cannot exploit the recurrence relations of polygamma functions to express a quasirational function into closed form.

6. A single procedure for the 4 cases

For ease of implementation, we combine the 4 separate procedures for computing closed forms for summation of polynomials, rational functions, quasipolynomials and quasirational functions into one, together with the technique of summing non-hygeometric functions via the use of polygamma functions. No further justification of this single procedure is provided as the theoretical backgrounds have been given in sections 3 and 4. The implementation of this general procedure in Reduce has been done and some illustrative examples will be provided in an appendix to show how the program works.

6.1. Algorithm

Assume the input summand belongs to a difference field F, and let E be the difference field extension of F obtained by adjoining one or several polygamma functions and K be the constant field of F or E. In practice, K can be taken as $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ or an algebraic extension of one of them, whenever it is necessary.

Algorithm 6.1. **(Fsum).**

Input: *$f(x) = \frac{f_1(x)\lambda^x}{f_2(x)} \in F$ in irreducible form*†*, where $\lambda \in K$, $f_i(x) \in K[x]$ $(i = 1, 2)$ and $f_2(x)$ is a non-zero polynomial.*

Output: *a closed form for $\sum_x f(x)$ in F or E, if it exists, otherwise say "not summable in E or F".*

Procedure:

S1: *(preprocessing) If $\lambda = 0$, then return a zero sum else if deg $f_1(x) \geq$ deg $f_2(x)$ then use division to split $f(x)$ into 2 parts: $h(x)\lambda^x$ and $r(x)\lambda^x/f_2(x)$, such that $f_1(x) = f_2(x)h(x) + r(x)$, where $r(x)$,$h(x) \in K[x]$ and deg $r(x) <$ deg $f_2(x)$.*
Apply the algorithm **Fsum** *to these two parts separately and then combine their outputs finally.*

S2: *(dispersion) If deg $f_2(x) = 0$,*‡ *then $0 \mapsto \alpha$ else dis $f_2(x) \mapsto \alpha$.*
If deg $f_2(x) > 0$ and $\alpha = 0$, then
(a) go to **S6** *if $|\lambda| = 1$ (b) return and say "not summable in F and E" otherwise.*

S3: *(degree bound)*
(a) If $\alpha = 0$ then
(i) deg $f_1(x) \mapsto n$ if $|\lambda| \neq 1$
(ii) deg $f_1(x) + 1 \mapsto n$ otherwise
(b) If $\alpha > 1$ then determine

$$\frac{s(x)}{t(x)} = \sum_{i=0}^{\alpha-1} \frac{f_1(x+i)\lambda^i}{f_2(x+i)},$$

† $|\lambda|$ can be 0 or 1 in this procedure.

‡ i.e. the summand is either a polynomial or a quasipolynomial.

where $s(x)$, $t(x) \in K[x]$. *Let* $g_2(x)$ *be the numerator of the irreducible form of* $\frac{t(x)}{t(x+\alpha)}$. *Determine*

$$p(x) = \frac{f_1(x)g_2(x)g_2(x+1)}{f_2(x)}.$$

Let deg $p(x) = l$ *and deg* $g_2(x) = m$.
If $|\lambda| = 1$ *then*
(i) $l - m + 1 \mapsto n$ *if* $l > m - 2$
(ii) $m \mapsto n$ *otherwise.*
else $l - m \mapsto n$

S4: *(determining relation) Construct a generic polynomial* $g_1(x)$ *with degree* n *and determine its unknown coefficients by one of the following relations:*†
(a) $\lambda g_1(x+1) - g_1(x) = f(x)$ *if* $\alpha = 0$
(b) $p(x) = \lambda g_1(x+1)g_2(x) - g_2(x+1)g_1(x)$ *otherwise.*

S5: *(closed form in* F *or not existent in* F*) If the unknown coefficients of* $g_1(x)$ *are solvable, then an indefinite sum exists in* F *and it is equal to*
(a) $g_1(x)\lambda^x$ *if* $\alpha = 0$
(b) $g_1(x)\lambda^x/g_2(x)$ *otherwise.*
else if $|\lambda| \neq 1$ *then return and say "not summable in* F *or* E*".*

S6: *(closed form in* E*) Perform a* **cpfd** *on* $f(x)$ *such that*

$$f(x) = f_0(x) + \sum_{i=1}^{k}\sum_{j=1}^{n_i} \frac{f_{ij}}{(x+t_i)^j},$$

where $k, n_i \in \mathbb{Z}$, $f_{ij}, t_i \in K$ *and* $f_0(x) \in K[x]$. *Let* $g(x)$ *be the polynomial obtained by applying FSUM to* $f_0(x)$. *Then return an indefinite sum*

$$g(x) + \sum_{j=1}^{n_i}\sum_{i=1}^{k} \frac{f_{ij}(-1)^{j-1}}{(j-1)!}\psi(j-1, x+t_i) \quad \text{mod(Lemma 3.1(a))}.^{\ddagger}$$

6.2. Implementation

The algorithm in section 6.1 has been implemented in RLISP in Reduce (version 3.4.1). The main operations such as computing 'resultant', solving a system of linear equations and performing partial fraction decompositions heavily rely on the top level symbolic procedures defined in Reduce. The implemented program is called **Fsum** and it has the following input format:

$$fsum(summand, n\ [, lower, upper]),$$

† In fact, the two relations can be combined as one — (a) can be obtained if we replace $p(x)$ by $f(x)$, and $g_2(x)$ by 1 in (b). That means (b) is a general relation.

‡ It means to use the recurrence relations in Lemma 3.1(a) to combine the polygamma functions with the same order into a single polygamma function, if it is possible, as well as to obtain a rational component, if it exists (see Example 5.1).

where *summand* is self-explanatory, n is the index variable of the summand, *lower* and *upper* refers to the lower and upper summation index respectively. The latter two arguments are optional, which means if they are not specified explicitly, the program will assume an indefinite sum to be found,§ otherwise the program will compute an indefinite sum first, say g and then return $sub(n = upper + 1, g) - sub(n = lower, g)$ as the answer (i.e. a definite sum). If the program fails to find a closed form, then the input expression is returned, e.g. $fsum(2^x/x, x)$ will be returned since it is not a quasirational sum. The program also provides a trace command which will show most of the key steps of the algorithm if the switch "trfsum" is turned on. As mentioned before, a complete partial fraction decomposition is needed when we use polygamma functions to express the closed form of a rational input¶, but we cannot guarantee to achieve this if the constant field is required to extend from $\mathbb{Q}$ to an algebraic extension of $\mathbb{Q}$ or even the complex field $\mathbb{C}$. In such a case, the program will return failure in the form $f\_sum(summand, n\ [, lower, upper])$ to indicate the problem encountered. Tests have been done by running **Fsum** on a SPARC station II for more than 100 examples chosen from Jolley (1961) and Davis (1962). **Fsum** can solve all of them provided the summand belongs to one of the following types: polynomials, quasipolynomials, rational functions or quasirational functions. Table 1 below shows some examples solved by **Fsum** (see also the appendix). Each example was run for 10 times and the average time was taken. All timings|| are in terms of milliseconds (garbage collection time is excluded) and the sources of the examples are shown in the column called *Ref.*

0acci

Table 1. Test examples

Number	Example	Answer	Time	Ref
1	$\sum_x \frac{2^x(x^2-2x-1)}{x^2(x+1)^2}$	$\frac{2^x}{x^2}$	40.8	Davis (1962)
2	$\sum_x \frac{3x^2+3x+1}{x^3(x+1)^3}$	$\frac{-1}{x^3}$	365.5	Davis (1962)
3	$\sum_{x=1}^m \frac{6x+3}{4x^4+8x^3+8x^2+4x+3}$	$\frac{m^2+2m}{2m^2+4m+3}$	64.6	Abramov (1971)
4	$\sum_{x=1}^m \frac{-(x^2+3x+3)}{x^4+2x^3-3x^2-4x+2}$	$\frac{m(2m+5)}{m^2+2m-1}$	79.9	Moenck (1977)
5	$\sum_{x=1}^m \frac{x^2 4^x}{(x+1)(x+2)}$	$\frac{2}{3} + \frac{4^{m+1}(m-1)}{3(m+2)}$	25.5	Jolley (1961)
6	$\sum_{x=1}^m (2x-1)^3$	$m^2(2m^2-1)$	28.9	Jolley (1961)
7	$\sum_x 3x^2+3x+1$	x^3	22.1	Davis (1962)
8	$\sum_x (x^2+4x+2)2^x$	$x^2 2^x$	22.1	Davis (1962)
9	$\sum_x \frac{2^x(x^3-3x^2-3x-1)}{x^3(x+1)^3}$	$\frac{2^x}{x^3}$	372.3	Davis (1962)
10	$\sum_{x=0}^m xk^x$	$\frac{mk^{m+2}-(m+1)k^{m+1}+k}{(k-1)^2}$	15.3	Lafon (1983)

§ i.e. a formula g which satisfies the relation $\Delta g = f$.

¶ It is done after the searching for a rational indefinite sum has failed.

|| The test is based on the first version of **Fsum**, which can be further optimized if more low level procedures in Reduce's source codes are used.

6.3. How to improve its efficiency

In general, the computations in step **S2** (determination of the dispersion of $f_2(x)$) and step **S4** (solving the unknown coefficients in $g_1(x)$) are relatively more expensive than other steps in **Fsum**. One way to improve the efficiency in **S2** is to perform a square-free operation to $f_2(x)$ first before computing the resultant $res(f_2(x), f_2(x+\alpha))$ w.r.t. x (see the short paragraph after *definition 3.2*) – doing so will decrease the dimension of the Sylvester matrix and hence the cost of computation of the resultant. For example, the dimension of the associated Sylvester matrix for $res(x^3(x+3)^3, (x+\alpha)^3(x+\alpha+3)^3)$ is 12×12, while the dimension of the Sylvester matrix for $res(x(x+3), (x+\alpha)(x+\alpha+3))$ w.r.t x is only 4×4. To illustrate this fact, five examples are chosen and the comparison of efficiencies are shown in table 2, which also include the timings for Reduce built-in indefinite summation solver (based on Gosper's algorithm). As mentioned before, each example was run for 10 times and the average tims was taken (excluding the garbage collection time) and the timings were in terms of milliseconds. (Note: The efficiency in

Oacci

Table 2. Comparison of efficiencies

Number	Example	Before improvement	After improvement	Reduce solver
1	$\sum_x \frac{3x^2+3x+1}{x^3(x+1)^3}$	384.2	20.4	280.5
2	$\sum_x \frac{2^x(x^3-3x^2-3x-1)}{x^3(x+1)^3}$	391.2	34.2	1159.4
3	$\sum_x \frac{3^x(2x^3+x^2+3x+6)}{(x^2+2)(x^2+2x+3)}$	107.1	117.3	1856.4
4	$\sum_x \frac{4(1-x)(x^2-2x-1)}{x^2(x+1)^2(x-2)^2(x-3)^2}$	20425.5	180.2	7429.2
5	$\sum_x \frac{2^x(x^4-14x^2-24x-9)}{x^2(x+1)^2(x+2)^2(x+3)^2}$	12172.3	152.5	6596.4

example 3 had not been improved due to the fact that the denominator of the summand is already square-free, so a small amount of time for performing the square-free operation was wasted in such case.)

6.4. Analysis of degree bounds

In this section, a comparison of degree bounds between **Fsum** and Gosper's algorithm is given. The two procedures have some similarities and both at a certain step require one to determine a degree bound for a polynomial ($f(n)$ in Gosper's case and $g_1(x)$ in Fsum's case) which will yield a closed form for an indefinite sum if the unknown coefficients can be solved. Although the derivations of the two algorithms are different, it is interesting to note that the degree bound chosen by **Fsum** agrees with Lisonek's (1991) analysis of the performance of Gosper's algorithm on rational function inputs, which says: for a proper rational input, we always have to consider the α_0 case (see section 2.1) and also if α_0 is an integer, we can take either $max(\alpha_0, deg\ p_k - l + 1)$ or $deg\ p_k - l + 1$ as our degree bound. Both will lead to the same answer if the indefinte sum is hypergeometric, but the latter bound is always smaller than or equal to α_0, which means we can improve the efficiency of Gosper's algorithm by choosing $deg\ p_k - l + 1$ as the degree bound. To illustrate this fact, let us reconsider the test examples in Table 1 above and tabulate the results in Table 3.

0acci

Table 3. Analysis of degree bounds

Example number	Type of input	**Fsum** n	Gosper α	Lisonek $deg\ p_k - l + 1$
1	quasirational	0	0	NA
2	rational	0	3	0
3	rational	0	2	0
4	rational	1	2	1
5	quasirational	0	0	NA
6	polynomial	4	4	NA
7	polynomial	3	3	NA
8	quasipolynomial	2	2	NA
9	quasirational	0	0	NA
10	quasipolynomial	1	1	NA

6.5. Further analysis

Table 3 shows the degree bound chosen by **Fsum** is smaller than or equal to that chosen by Gosper's algorithm. In fact it is true in general because the degree bound for **Fsum** is based on Theorem 4.1, Theorem 4.3, Algorithm 4.1 and Algorithm 4.2, which corresponds to polynomial, quasipolynomial, rational and quasirational input respectively. On the other hand, the degree bound for Gosper's algorithm is not specific to any one of the above types and so it is not surprising that it will overestimate the degree bound in some cases.[†] The next thing worthwhile to look at is the summability conditions of the two procedures – there are some relationships between $p_k = q_{k+1}f(k) - r_k f(k-1)$ (S3 of Gosper's algorithm) and $p(x) = \lambda g_1(x+1)g_2(x) - g_2(x+1)g_1(x)$[‡] (S4 of **Fsum**). In order to clarify the situation, we reconsider the ten examples in table 1 and tabulate the values of each quantity that appears in each determining conditions in Tables 4 and 5. Notations are almost the same as those used in section 2.1 and section 6.1, except the variable k is changed to x in Gosper's case. It is easy to see that q_{x+1} is equal to $\lambda g_2(x)$, r_x is equal to $g_2(x+1)$ and p_x is equal to $p(x)$ (up to a certain constant).[§] If such relations are true in general, then it may economize the computationally expensive step S1 of Gosper's algorithm (see Lafon (1983), Zeilberger (1991)) because we have a priori knowledge that the ratio a_x/a_{x-1} can be represented as $\frac{p(x)}{p(x-1)} \times \frac{g_2(x-1)}{g_2(x+1)}$ up to some constant multiple, which means that the two fractional parts are actually produced by shifting the two polynomials $p(x)$ and $g_2(x)$ independently.

† But for polynomial and quasipolynomial inputs, it is easy to check that Gosper's degree bound is always the same as **Fsum**.

‡ As mentioned before, it is a general determining condition for **Fsum**.

§ It is not surprising that p_x and $p(x)$ differ by a constant because the first step of Gosper's algorithm is a_x/a_{x-1}, which will cancel out any constant multiples which appear in the summand; but on the other hand, **Fsum** does not change the numerator in the input.

Oacci

Table 4. Fsum

Example	$p(x)$	$g_2(x+1)$	$g_2(x)$	λ	n
1	x^2-2x-1	$(x+1)^2$	x^2	2	0
2	$3x^2+3x+1$	$(x+1)^3$	x^3	1	0
3	$3(2x+1)$	$2x^2+4x+3$	$2x^2+1$	1	0
4	$-x^2-3x-3$	x^2+2x-1	x^2-2	1	1
5	$-3x-2$	$x+2$	$x+1$	4	0
6	$(2x-1)^3$	1	1	1	4
7	$3x^2+3x+1$	1	1	1	3
8	x^2+4x+2	1	1	2	2
9	x^3-3x^2-3x-1	$(x+1)^3$	x^3	2	0
10	x	1	1	x	1

Oacci

Table 5. Gosper's algorithm

Example	p_x	r_x	q_{x+1}	α
1	x^2-2x-1	$(x+1)^2$	$2x^2$	0
2	$3x^2+3x+1$	$(x+1)^3$	x^3	3
3	$2x+1$	$2x^2+4x+3$	$2x^2+1$	2
4	x^2+3x+3	x^2+2x-1	x^2-2	2
5	$3x+2$	$x+2$	$4(x+1)$	0
6	$(2x-1)^3$	1	1	4
7	$3x^2+3x+1$	1	1	3
8	x^2+4x+2	1	2	2
9	x^3-3x^2-3x-1	$(x+1)^3$	$2x^3$	0
10	x	1	x	1

7. Conclusion

We have presented a general procedure called **Fsum** for handling polynomials, rational functions, quasipolynomials and quasirational functions, The last 2 cases are based on the extension of Abramov's (1971) work on summation of rational functions. Extension to deal with non-hypergeometric sums with rational input is done via the use of polygamma functions, which is an alternative to Moenck's (1977) approach but **Fsum** does not require summation by parts nor Hermite method of reduction. Besides, it is also an alternative to Abramov's (1975) method for finding rational components for rational inputs. Analysis of degree bounds and the determining condition for summability indicate that **Fsum** has connections with Gosper's algorithm which may provide some a priori knowledge how to perform step one of Gosper's algorithm more efficiently. In addition, we have seen from some examples that the degree bound of **Fsum** is always smaller than or equal to that chosen by Gosper's algorithm. So one may make use of such information to improve the efficiency of Gosper's algorithm if the input summand belongs to polynomial, quasipolynomial, rational or quasirational type; especially Gosper's algorithm is a major building block for more involved algorithms (such as Zeilberger (1991) and Petkovsek (1992)). Finally, it should be noted that the adaptation of Abramov's algorithms (1989, 1991) for nth order difference equations to indefinite summation problems is not the same as his algorithm in 1971 nor the **Fsum** algorithm presented in this paper – for instance, if $f(x)$ is a given summand and we want to find the closed form for $\sum_x f(x)$, we have to solve for the rational solution of a first order difference equation of the form $p(x)g(x+1) - q(x)g(x) = q(x)$,¶ where $g(x)$ is unknown, $p(x)$ and $q(x)$ are polynomials which are the numerator and denominator of the irreducible form of $f(x+1)/f(x)$ respectively, by using Abramov's 1989 algorithm. In some occasions, the degree bound setting for the numerator of $g(x)$ can be higher|| and the condition for summability can be more complicated than **Fsum**, Gosper's algorithm or his own 1971 algorithm. Therefore, if we are heading for a more general and efficient algorithm for indefinite summation (or at least hypergeometric summation), then one needs to do further research, and perform more comparisons of efficiencies for the existing algorithms to find out what the optimal strategies should be and hence improve the efficiencies of the existing methods.

Acknowledgement

I would like to thank Prof. M.A.H. MacCallum and Dr. F.J. Wright for their valuable suggestions and help during the preparation of this paper. I also thank the referees' useful comments to me. Special thanks to Dr. M. Petkovsek for some very helpful discussions on hypergeometric summations and Dr. M. Bronstein for sending me a copy of the Maple implementation of his full partial fraction decomposition algorithm. Finally, I have to thank the ORS awards committee in U.K. for partial financial support of my current research in QMW, University of London.

Appendix

The following examples were tested by using the improved version of **Fsum**.

```
REDUCE 3.4.1, 15-Jul-92 ...
1: load fsum$

2: on time$

3: fsum((3x^2+3x+1)/(x^3*(x+1)^3),x);

  - 1
------
   3
  X

Time: 34 ms

4: fsum((6n+3)/(4n^4+8n^3+8n^2+4n+3),n,1,m);

   M*(M + 2)
----------------
     2
 2*M  + 4*M + 3
```

¶ based on Petkovsek's (1992) proposition 5.1 and discussion with M. Petkovsek.

|| In fact, Abramov had pointed out this in his 1989a paper p65.

```
Time: 68 ms

5: fsum(2^x*(x^3-3x^2-3x-1)/(x^3*(x+1)^3),x);

  X
 2
----
  3
 X

Time: 34 ms

6: fsum(n*x^n,n,0,m);

     M         M      M
 X*(X *M*X - X *M - X  + 1)
----------------------------
          2
         X  - 2*X + 1

Time: 34 ms

7: fsum((2x+3)/((x+1)*(x+2)),x);

     2
 2*X *PSI(X) + 2*X*PSI(X) + 3*X + 2
-----------------------------------
              X*(X + 1)

Time: 17 ms
```

References

Abramov, S. A. (1971). On the summation of rational functions. *USSR Comp. Maths. Math. Phy.* **11**, 324–330.

Abramov, S.A. (1974). Solution of linear finite-difference equations with constant coefficients in the field of rational functions. *USSR Comp. Maths. Math. Phy.* **14**, 247–251.

Abramov, S.A. (1975). The rational component of the solution of a first order linear recurrence relation with a rational right side. *USSR Comp. Maths. Math. Phy.* **15**, 216–221.

Abramov, S.A. (1989a). Problems of computer algebra involved in the search for polynomial solutions of linear differential and difference equations. *Moscow University Comp. Math. Cyber.* **3**, 63–68.

Abramov, S.A. (1989b). Rational solutions of linear differential and difference equations with polynomial coefficients. *USSR Comp. Maths. Math. Phy.* **29**, 7–12.

Abramov, S.A. (1991). An algorithm for constructing quasirational solutions to linear differential and finite-difference equations with polynomial coefficients. *Moscow University Comp. Math. Cyber.* **1**, 41–44.

Abramowitz, M., Stegun, I.A. (1970). *Handbook of Mathematical Functions.* New York: Dover.

Boole, G. (1880). *A Treatise on the Calculus of Finite Differences.* London: Macmillan.

Bronstein, M., Salvy, B. (1993). Full partial fraction decomposition of rational functions. *Proc. ISSAC'93 (M. Bronstein eds.), July 6-8, Kiev, Ukraine.* , 157–160.

Cohen, J., Katcoff, J. (1978). Symbolic solution of finite difference equation. *ACM Trans. Math. Soft.* **3**, 261–271.

Davis, H.T. (1962). *The Summation of Series.* San Antonio, Texas: Principia Press of Trinity University.

Gosper, R.W. (1978). Decision procedure for indefinite hypergeometric summation. *Proc. Natl. Acad. Sci.* **USA 75**, 40–42.

Jolley, L.B.W. (1961). *Summation of Series.* New York: Dover.

Karr, M. (1981). Summation in finite terms. *J. ACM* **28**, 305–350.

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

Lafon, J.C. (1983). Summation in finite terms. In: *Computer Algebra– Symbolic and Algebraic Computation (edited by Buchberger et al).* Springer Verlag.

Lamagna, E.A., Liu, S.M., Savio, D.Y. (1989). Summation of harmonic numbers. In: *Computers and Mathematics (edited by Kaltofen et al)* . Springer Verlag.

MacCallum, M.A.H., Wright, F.J. (1991). *Algebraic Computing with Reduce.* Oxford: Clarendon Press.

Lisonek, P. (1991). The performance of Gosper's algorithm on rational function inputs. *Technical Report, July, RISC-Linz 91-31* .

Moenck, R. (1977). On computing closed forms for summation. *Proc. Macsyma Users Conference* , 225–236.

Paule, P., Strehl, V. (1991). A remark on an instance of Gosper's algorithm. *Technical Report, March, RISC-Linz 91-14* .

Petkovsek, M. (1992). Hypergeometric solutions of linear recurrences with polynomial coefficients. *J. Symbolic Comp.* **14**, 243–264.

Richardson, C.H. (1954). *An Introduction to the Calculus of Finite Differences.* New York: Van Nostrand.

Ross, C.C. (1991). Why the method of undetermined coefficients works. *Amer. Math. Month.* **98**, 8, 747–749.

Wilf, H.S., Zeilberger, D. (1990). Towards computerized proofs of identities. *Bull. AMS.* **23**, 1, 77–83.

Wilf, H.S., Zeilberger, D. (1992). Rational function certification on multisum/integral/"q" identities. *Bull. AMS.* **27**, 1, 148–153.

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