Improved Abramov–Petkovšek's Reduction and Creative Telescoping for Hypergeometric Terms*

Shaoshi Chen[†], Hui Huang[†,‡] and Ziming Li[†]
[†]Key Laboratory of Mathematics Mechanization
AMSS, Chinese Academy of Sciences, Beijing (100190), China
[‡]Research Institute for Symbolic Computation,
Johannes Kepler University, Linz A-4040, Austria
{schen,huanghui}@amss.ac.cn,zmli@mmrc.iss.ac.cn

1 Introduction

The ubiquity of hypergeometric terms in enumerative combinatorics is widely recognized, such as binomial coefficients, power functions and factorials, etc. Let C be a field of characteristic zero. A univariate term $T(n)$ is said to be *hypergeometric* if its shift quotient $T(n+1)/T(n)$ is in $C(n)$.

A hypergeometric term $T(n)$ is said to be *hypergeometric-summable* if there exists another hypergeometric term $G(n)$ such that

$$T(n) = G(n+1) - G(n). \tag{1}$$

We abbreviate "hypergeometric-summable" as "summable" in the sequel.

There are two methods for determining whether a hypergeometric term is summable or not. One is Gosper's algorithm [5] that finds a hypergeometric term $G(n)$ such that (1) holds whenever $T(n)$ is summable. The other is Abramov–Petkovšek's reduction [1, 2] that computes an additive decomposition

$$T(n) = (T_1(n+1) - T_1(n)) + T_2(n), \tag{2}$$

where T_1 is hypergeometric, and T_2 is either hypergeometric or zero. It is shown that T is summable if and only if $T_2 = 0$.

Both methods amount to computing polynomial solutions of some auxiliary first-order linear difference equations with polynomial coefficients. A difference between the two methods is that the second one also provides us with a non-summable part when T is not summable.

We improve Abramov–Petkovšek's reduction so that the additive decomposition (2) is computed without solving any auxiliary linear difference equation. Computational experiments illustrate that the improved Abramov–Petkovšek's reduction is superior to the original one, and more efficient than Gosper's algorithm when a given hypergeometric input gets complicated. We will also report an on-going research in computing the minimal telescoper for a bivariate hypergeometric term with the help of the improved Abramov–Petkovšek's reduction.

*S. Chen was supported in part by the National Science Foundation (NSF) grant CCF-1017217, H. Huang by the Austrian Science Fund (FWF) grant W1214-13, H. Huang and Z. Li by two NSFC grants (91118001, 60821002/F02) and a 973 project (2011CB302401).

2 Improved Abramov–Petkovšek's reduction

To describe original and improved Abramov–Petkovšek's reductions concisely, we need some terminologies. A nonzero polynomial in $C[n]$ is said to be *shift-free* if its two distinct roots do not differ by an integer. A nonzero rational function is said to be *shift-reduced* if its numerator is co-prime with any shift of its denominator. For two hypergeometric terms $T(n)$ and $H(n)$, we write $T(n) \equiv H(n)$ if $T(n) - H(n)$ is summable. According to [1, 2], every hypergeometric term $T(n)$ can be written as $S(n)H(n)$, where $S(n)$ is in $C(n)$ and $H(n)$ is another hypergeometric term whose shift quotient is shift-reduced. We call the shift quotient $H(n+1)/H(n)$ a *kernel* of $T(n)$ and S the corresponding *shell*.

From now on, we assume that $T(n)$ is a hypergeometric term whose kernel is K and the corresponding shell is S. Moreover, assume that $K \neq 1$, for otherwise, $T(n)$ would be rational. We set $K = u/v$ with $u, v \in C[n]$ and $\gcd(u, v) = 1$.

Definition 2.1 *An irreducible polynomial f in $C[n]$ is said to be* strongly prime *with K if either $f \in C$ with $f \neq 0$, or $f \nmid uv$, $f(n+i) \nmid u$ and $f(n-i) \nmid v$ for all $i \in \mathbb{Z}^+$. A nonzero polynomial in $C[n]$ is said to be* strongly prime *with K if all its irreducible factors are strongly prime with K.*

The original Abramov–Petkovšek's reduction proceeds as follows. First, decompose $T(n)$ as $S(n)H(n)$, where $H(n)$ is a hypergeometric term with shift quotient K. Second, reduce the shell S to find polynomials $a, b, p \in C[n]$ such that

$$T(n) \equiv \left(\frac{a}{b} + \frac{p}{v}\right) H(n), \tag{3}$$

where $\deg(a) < \deg(b)$, $\gcd(a, b) = 1$, and b is shift-free and strongly prime with K. Moreover, $\deg(b)$ is minimal and $\deg(p)$ is bounded (see [2, Theorem 7]). At last, compute a polynomial solution of an auxiliary first-order linear difference equation. If a polynomial solution is found, the algorithm constructs a hypergeometric term $G(n)$ such that (1) holds. Otherwise, $T(n)$ is not summable.

Based on (3), we further reduce the number of terms of p using the idea in [4]. Define a C-linear map $\phi_K : C[n] \longrightarrow C[n]$ with $f(n) \mapsto u(n)f(n+1) - v(n)f(n)$ for all $f(n) \in C[n]$. It can be shown that ϕ_K is injective. So $\left\{\phi_K(x^i) \mid i \in \mathbb{N}\right\}$ is a C-basis of $\operatorname{im}(\phi_K)$. Set

$$\mathcal{N}_K = \operatorname{span}_C \left\{ n^\ell \mid \ell \in \mathbb{N} \text{ and } \ell \neq \deg(g) \text{ for all } g \in \operatorname{im}(\phi_K) \right\}.$$

Then $C[n] = \operatorname{im}(\phi_K) \oplus \mathcal{N}_K$. With the above C-basis of $\operatorname{im}(\phi_K)$, we can easily find a C-basis of $\mathcal{N}_K$. The two bases enable one to project a polynomial into $\operatorname{im}(\phi_K)$ and $\mathcal{N}_K$ by merely addition and scalar multiplication in the C-linear space $C[n]$.

Definition 2.2 *A rational function $r \in C(n)$ is called a (discrete) residual form w.r.t. K if r can be decomposed as $a/b + q/v$, where $a, b, q \in C[n]$, $\deg(a) < \deg(b)$, $\gcd(a, b) = 1$, b is shift-free and strongly prime with K, and q belongs to $\mathcal{N}_K$.*

Residual forms have the following property.

Proposition 2.3 *If r is a nonzero residual form w.r.t. K, and H is a hypergeometric term with shift quotient K, then rH is not summable.*

We further reduce the polynomial p in (3) by the following steps: (i) Compute the projection q of p in $\mathcal{N}_K$. (ii) Set $H_1 = \phi_K^{-1}(p-q)H$ and $r = a/b + q/v$. Then r is a residual form w.r.t. K. By (3) and the definition of ϕ_K, we find that $T(n) \equiv (a/b + p/v)\, H(n) = H_1(n+1) - H_1(n) + rH(n) \equiv rH(n)$. Thus, $T(n)$ is summable if and only if $r = 0$ by Proposition 2.3. This avoids the step to find a polynomial solution of any auxiliary first-order linear difference equation.

3 Computing telescopers

In [4], the authors use Hermite reduction for univariate hyperexponential functions to compute telescopers for bivariate hyperexponential functions. It allows one to separate the computation of telescopers from that of certificates. We try to translate their idea into the hypergeometric setting.

A bivariate term $T(k,n)$ is said to be *hypergeometric* if its two shift quotients $T(k+1,n)/T(k,n)$ and $T(k,n+1)/T(k,n)$ are in $C(k,n)$. For two bivariate hypergeometric terms $T_1(k,n)$ and $T_2(k,n)$, we write $T_1 \equiv_n T_2$ if $T_1 - T_2$ is summable w.r.t. n.

Let σ_k be the shift operator that maps k to $k+1$. Then $\sigma_k^j(T) = T(k+j,n)$. Applying the improved Abramov–Petkovšek's reduction to $\sigma_k^j(T(k,n))$ w.r.t n, where j ranges from 0 to a nonnegative integer i, we get $\sigma_k^j(T) \equiv_n r_j H$, where H is another bivariate hypergeometric term whose shift quotient K w.r.t. n is shift-reduced w.r.t. n, and r_j is a residual form w.r.t. K. For univariate rational functions $a_0, \ldots, a_i \in C(k)$, not all zero, we have

$$\sum_{j=0}^{i} a_j \sigma_k^j(T) \equiv_n \sum_{j=0}^{i} a_j r_j H.$$

Clearly, $\sum_{j=0}^{i} a_j \sigma_k^j$ is a telescoper for T w.r.t. n if $\sum_{j=0}^{i} a_j r_j = 0$. Unfortunately, the converse is false. This is because $\sum_{j=0}^{i} a_j r_j$ is not necessarily a residual form, although all the r_j's are. So Proposition 2.3 is not applicable to $\sum_{j=0}^{i} a_j r_j$. This situation does not occur in the differential case [4]. To make Proposition 2.3 applicable, we will develop a new reduction by the Ore-Sato theorem [6, 7] and the criterion on the existence of telescopers for hypergeometric terms [3].

References

[1] S.A. Abramov and M. Petkovšek. Minimal decomposition of indefinite hypergeometric sums. *Proc. of ISSAC 2001*, 7–14, New York, ACM, 2001.

[2] S.A. Abramov and M. Petkovšek. Rational normal forms and minimal decompositions of hypergeometric terms. *J. Symbolic Comput.*, 33:521–543, 2002.

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

[4] A. Bostan, S. Chen, F. Chyzak, Z. Li, and G. Xin. Hermite reduction and creative telescoping for hyperexponential functions. *Proc. of ISSAC 2013*, 77–84, New York, ACM, 2013.

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

[6] O. Ore. Sur la forme des fonctions hypergéométriques de plusieurs variables. *J. Math. Pures Appl.*, 9(4):311–326, 1930.

[7] M. Sato. Theory of prehomogeneous vector spaces (algebraic part) – the English translation of Sato's lecture from Shintani's note. *Nagoya Math. J.*, 120:1–34, 1990