# Rational Canonical Forms and Efficient Representations of Hypergeometric Terms 

S. A. Abramov ${ }^{*}$<br>Dorodnicyn Computing Centre Russian Academy of Science<br>Moscow, Russia<br>abramov@ccas.ru

H. Q. Le ${ }^{\dagger}$<br>Symbolic Computation Group<br>University of Waterloo<br>Waterloo, Canada<br>hqle@math.uwaterloo.ca

M. Petkovšek ${ }^{\ddagger}$<br>Department of Mathematics<br>University of Ljubljana<br>Ljubljana, Slovenia<br>marko.petkovsek@fmf.uni-<br>lj.si


#### Abstract

We propose four multiplicative canonical forms that exhibit the shift structure of a given rational function. These forms in particular allow one to represent a hypergeometric term efficiently. Each of these representations is optimal in some sense.


Categories and Subject Descriptors
I.1.2 [Symbolic and Algebraic Manipulation]: Algebraic algorithms

## General Terms

Algorithms, Design

## Keywords

Rational functions, hypergeometric terms, canonical forms, efficient representations

## 1. INTRODUCTION

Let $K$ be a field of characteristic zero. Representations of a rational function $R \in K(x)$ in the form

$$
\begin{equation*}
R(x)=F(x) \cdot \frac{V(x+1)}{V(x)} \tag{1}
\end{equation*}
$$

where $F, V \in K(x)$ satisfy some specific conditions, play a substantial role in various computer algebra algorithms operating on hypergeometric terms. Recall that a sequence

[^0]Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ISSAC'03, August 3-6, 2003, Philadelphia, Pennsylvania, USA.
Copyright 2003 ACM 1-58113-641-2/03/0008 ...\$5.00.
$T(n)$ of elements of $K$ defined for all integers $n \geq n_{0}$ is a hypergeometric term if there are polynomials $p, q \in K[x]$ such that $q(n) T(n+1)=p(n) T(n)$ for all $n \geq n_{0}$. If $T(n)$ is eventually nonzero then the rational function $p / q$ is unique, and is called the certificate of $T$.

The main part of Gosper's algorithm for hypergeometric indefinite summation [5], Zeilberger's algorithm for hypergeometric definite summation [5], the algorithm for finding a minimal multiplicative decomposition of a hypergeometric term [2], and the algorithm for finding a minimal additive decomposition of a hypergeometric term [2] starts with the certificate of a hypergeometric term. Each algorithm then proceeds by representing this certificate in the form (1). Algorithm Hyper [5], and the algorithm to compute the hypergeometric dispersion [1] use the representation of certificates in the form (1) as an auxiliary tool.

The algorithm for finding a minimal multiplicative decomposition of a hypergeometric term can be used to construct an economic representation of a hypergeometric term $T(n)$. Using the certificate $R$ of $T(n)$, we can write

$$
\begin{equation*}
T(n)=c \prod_{k=n_{0}}^{n-1} R(k) \tag{2}
\end{equation*}
$$

where $c$ is determined from some initial conditions. Let $F(x)$ in (1) be written as $r(x) / s(x)$ where $r, s \in K[x]$ and $\operatorname{gcd}(r(x), s(x+k))=1$ for all $k \in \mathbb{Z}$. Then (1) is a rational normal form (RNF) of $R(x)$, and $F(x), V(x)$ are the kernel and the shell of this RNF, respectively. By using any RNF of $R$, we can rewrite (2) in the form

$$
\begin{equation*}
T(n)=c V(n) \prod_{k=n_{0}}^{n-1} F(k) \tag{3}
\end{equation*}
$$

where both the numerator and the denominator of $F$ are of minimal possible degrees [2].

It was shown in [2] that a rational function can have several different RNF's. In Sections 3-4.4, we distinguish four rational canonical forms (RCF's) in the set of all RNF's. Each of these four RCF's minimizes the shell in one sense or another: $\mathrm{RCF}_{1}$ and $\mathrm{RCF}_{2}$ minimize the degree of the denominator and of the numerator of the shell, respectively. $R C F_{1}^{*}$ and $\mathrm{RCF}_{2}^{*}$ both minimize the sum of the degrees of the numerator and of the denominator of the shell, and under this condition, also minimize the degree of the denominator and of the numerator of the shell, respectively. By using
these canonical forms in the problem of representing a hypergeometric term $T(n)$ economically, we can minimize $V$ in (3) (recall that $F$ is minimized by any RNF of $R$ ). As a consequence, we can rewrite (3) in the "optimal" form

$$
\begin{equation*}
\alpha^{n} V(n) Q(n), \tag{4}
\end{equation*}
$$

where $\alpha \in K$, and $Q(n)$ is a product of Gamma-function values (if $K=\mathbb{C}$ ) or Pochhammer symbols (i.e., rising factorial powers) and their reciprocals. Additionally,

- $Q(n)$ has the minimal possible number of factors,
- $V(n)$ is a rational function which is minimal in one sense or another, depending on the particular RCF chosen to represent the certificate of $T(n)$.

Economic representations of hypergeometric terms are useful in the output routines of algorithms which return hypergeometric terms, but compute their certificates first and need to construct the terms themselves before outputting them. Other important problems where these representations can be used to advantage include simplification of hypergeometric terms (algorithms which accept a hypergeometric term $T$ as input, and construct a rational function $R$ such that the output hypergeometric term is $R T$; in this case, a simplification is desirable), and investigation of asymptotics of hypergeometric terms.

The algorithms for constructing the four RCF's of a rational function, and the four economic representations of a hypergeometric term have been implemented in Maple, and are available from
http://www.scg.uwaterloo.ca/~hqle/code/RNF/RNF.html

## 2. PRELIMINARIES

In this section we give definitions of basic notions, and formulate some necessary results from [2].

Throughout the paper, $K$ is a field of characteristic zero, $\mathbb{Z}$ and $\mathbb{N}$ respectively denote the set of integers and nonnegative integers, $E$ denotes the shift operator acting both on rational functions by $E R(x)=R(x+1)$, and on sequences by $E T(n)=T(n+1)$. For $p, q \in K[x]$, we write $p \perp q$ to indicate that $p$ and $q$ are coprime. We denote the leading coefficient of $p$ by $l c(p)$. For every rational function $R \in K(x)$, its numerator num $R$ and denominator den $R$ are uniquely determined by requiring that num $R$, den $R \in K[x]$, $R=\operatorname{num} R / \operatorname{den} R$, num $R \perp \operatorname{den} R$, and $l c(\operatorname{den} R)=1$. The leading coefficient of $R$ is $l c(R)=l c($ num $R)$, and $R$ is monic if $l c(R)=1$.

### 2.1 PNF, RNF and Their Strict Versions

A pair of polynomials $(p, q) \in K[x] \times K[x]$ is shift-reduced if $p \perp E^{k} q$ for all $k \in \mathbb{Z}$. We also call a rational function $R \in K(x)$ shift-reduced if the pair (num $R$, den $R$ ) is shift-reduced. Irreducible polynomials $p, q \in K[x]$ are shiftequivalent if $p \mid E^{k} q$ for some $k \in \mathbb{Z}$. A rational function $R \in K(x)$ is shift-homogeneous if all irreducible factors of num $R$ and den $R$ belong to the same shift-equivalence class. By grouping together shift-equivalent irreducible monic factors of its numerator and denominator, every rational function $R(x) \in K(x)$ can be written in the form

$$
\begin{equation*}
R(x)=z R_{1}(x) R_{2}(x) \cdots R_{k}(x) \tag{5}
\end{equation*}
$$

where $z \in K, k \geq 0$, each $R_{i}$ is a monic shift-homogeneous rational function, and $R_{i} R_{j}$ is not shift-homogeneous unless
$i=j$ or $R_{i}=1$ or $R_{j}=1$. We call (5) a shift-homogeneous factorization of $R$.

Definition 2.1. Let $R \in K(x)$. If there are $z \in K$, and monic polynomials $a, b, c \in K[x]$ such that
(i) $R=z \cdot \frac{a}{b} \cdot \frac{E c}{c}$,
(ii) $a \perp E^{k} b$ for all $k \in \mathbb{N}$,
then ( $z, a, b, c$ ) is a polynomial normal form (PNF) of $R$. If, in addition,
(iii) $a \perp c$ and $b \perp E c$,
then $(z, a, b, c)$ is a strict PNF of $R$.
Every nonzero rational function has a unique strict PNF. For a proof of this, and for an algorithm to compute it, see [5]. We denote the strict PNF of $R \in K(x) \backslash\{0\}$ by $\operatorname{sPNF}(R)$.

Definition 2.2. Let $R \in K(x)$. If there are $z \in K$, and monic polynomials $r, s, u, v \in K[x]$ such that
(i) $R=F \cdot \frac{E V}{V}$ where $F=z \cdot \frac{r}{s}, V=\frac{u}{v}$ and $u \perp v$,
(ii) the pair $(r, s)$ is shift-reduced,
then $(z, r, s, u, v)$ is a rational normal form (RNF) of $R$. We denote the set of all RNF's of $R$ by $\operatorname{RNF}_{x}(R)$.

If in addition,
(iii) $r \perp u \cdot E v$ and $s \perp E u \cdot v$,
then $(z, r, s, u, v)$ is a strict RNF of $R$. We denote the set of all strict RNF's of $R$ by $\operatorname{sRNF}_{x}(R)$.

Every nonzero rational function has a strict RNF. For a proof of this, and for an algorithm to compute it, see [2]. Note that all four rational canonical forms introduced in this paper, $\mathrm{RCF}_{1}, \mathrm{RCF}_{2}, \mathrm{RCF}_{1}^{*}$ and $\mathrm{RCF}_{2}^{*}$, are strict RNF's.

Definition 2.3. The rational functions $F=z r / s$ and $V=u / v$ are called, respectively, the kernel and the shell of the RNF $(z, r, s, u, v)$.

For notational convenience, an RNF of a rational function $R$ is sometimes written in the short form $(F, V)$ instead of in the long form ( $z, r, s, u, v$ ).

We will often use the following result [5, Lemma 5.3.1]:
Lemma 2.1. Let $a, b, c, A, B, C \in K[x]$ be polynomials such that $a \perp c, b \perp E c$, and $a \perp E^{k} b$ for all $k \in \mathbb{N}$. If

$$
\frac{a}{b} \frac{E c}{c}=\frac{A}{B} \frac{E C}{C}
$$

then $c$ divides $C$.

### 2.2 Minimizing the Kernel

Theorem 2.2. Let $\varphi=(z, r, s, u, v)$ be any $R N F$ of $R \in$ $K(x) \backslash\{0\}$. Then
(i) $z$ is unique;
(ii) if $R$ is shift-homogeneous then $r=1$ or $s=1$;
(iii) the degrees of the polynomials $r$ and $s$ are unique, and have minimal possible values in the sense that if

$$
R(x)=\frac{p(x)}{q(x)} \frac{E G(x)}{G(x)}
$$

where $p, q \in K[x]$ and $G \in K(x)$, then $\operatorname{deg} r \leq \operatorname{deg} p$ and $\operatorname{deg} s \leq \operatorname{deg} q$;
(iv) given $F=z r / s$, the $R N F$ of $R$ is uniquely determined;
(v) $\varphi^{-1}:=(1 / z, s, r, v, u)$ is an RNF of $1 / R$. If $\varphi$ is strict then so is $\varphi^{-1}$;
(vi) if $\varphi$ is strict then $r \mid \operatorname{num} R$ and $s \mid$ den $R$;
(vii) the set $\operatorname{sRNF}_{x}(R)$ is finite.

For a proof, see [2].
Proposition 2.3. Let $R \in K(x) \backslash\{0\}$. A strict $R N F$ $(z, r, s, u, v)$ of $R$ is uniquely determined by either $u$ or $v$.

Proof: Let $\left(z_{1}, r_{1}, s_{1}, u_{1}, v_{1}\right),\left(z_{2}, r_{2}, s_{2}, u_{2}, v_{2}\right)$ be two strict RNF's of $R$. This implies

$$
z_{1} \frac{r_{1}}{s_{1}} \frac{E u_{1}}{u_{1}} \frac{v_{1}}{E v_{1}}=z_{2} \frac{r_{2}}{s_{2}} \frac{E u_{2}}{u_{2}} \frac{v_{2}}{E v_{2}} .
$$

If $v_{1}=v_{2}$ then $\left(z_{1}, r_{1}, s_{1}, u_{1}\right)$ and $\left(z_{2}, r_{2}, s_{2}, u_{2}\right)$ are both strict PNF's of $R_{1}=R \cdot\left(E v_{1} / v_{1}\right)$. Similarly, if $u_{1}=u_{2}$ then $\left(1 / z_{1}, s_{1}, r_{1}, v_{1}\right)$ and ( $1 / z_{2}, s_{2}, r_{2}, v_{2}$ ) are both strict PNF's of $R_{2}=(1 / R)\left(E u_{1} / u_{1}\right)$. Since the strict PNF of a rational function is unique, we have proved the assertion.

Example 2.1. Let $R \in K(x) \backslash\{0\}$. While the strict $P N F$ of $R$ is unique, $R$ can have infinitely many distinct PNF's. For instance, for any irreducible monic $p \in K[x]$, the four-tuple ( $1, p, E p, p$ ) is a PNF of $R(x)=1$. Likewise, some rational functions $R$ can have infinitely many distinct RNF's. For instance, for any monic $p \in K[x]$ and $k \in \mathbb{N}$, the five-tuple $\left(1, E^{k} p, 1,1, p E p \cdots E^{k-1} p\right)$ is an $R N F$ of $R(x)=p(x)$.

Property (iii) in Theorem 2.2 shows the minimality of the kernel of any RNF. An interesting question is how to compute an RNF not only with the minimal kernel, but also with a minimal shell (in some sense).

## 3. MINIMIZING THE SHELL: deg num $V$ OR deg den $V$

### 3.1 Definition and Properties of $\mathrm{RCF}_{1}$ and $\mathrm{RCF}_{2}$

Among all possible RNF's of $R$ we distinguish two (not necessarily distinct) forms which are called the first and the second rational canonical forms $\left(\mathrm{RCF}_{1}\right.$ and $\left.\mathrm{RCF}_{2}\right)$ of $R$.

Definition 3.1. Let $R \in K(x) \backslash\{0\}$. A strict $R N F$ $\left(z, r_{1}, s_{1}, u_{1}, v_{1}\right)$ of $R$ is the first rational canonical form $\left(\mathrm{RCF}_{1}\right)$ of $R$ if $v_{1} \mid v$ for every $R N F(z, r, s, u, v)$ of $R$. A strict $R N F\left(z, r_{2}, s_{2}, u_{2}, v_{2}\right)$ of $R$ is the second rational canonical form $\left(\mathrm{RCF}_{2}\right)$ of $R$ if $u_{2} \mid u$ for every $R N F$ $(z, r, s, u, v)$ of $R$.

Theorem 3.1. Every $R \in K(x) \backslash\{0\}$ has a unique $R C F_{1}$ and a unique $R C F_{2}$.

Proof: By Definition 3.1, any two $\mathrm{RCF}_{1}$ 's of $R$ have the same $v$, hence by Proposition 2.3 they are equal. Similarly, any two $\mathrm{RCF}_{2}$ 's of $R$ have the same $u$, hence they are equal. This proves uniqueness of $\mathrm{RCF}_{1}$ and $\mathrm{RCF}_{2}$ (and justifies our use of the word "canonical"). Their existence is established constructively by Algorithms $\mathrm{RCF}_{1}$ and $\mathrm{RCF}_{2}$, respectively, in Section 3.2.

We denote the unique $\mathrm{RCF}_{1}$ and $\mathrm{RCF}_{2}$ of $R \in K(x) \backslash\{0\}$ by $\mathrm{RCF}_{1}(R)$ and $\mathrm{RCF}_{2}(R)$, respectively.

From Definition 3.1 it follows that $\mathrm{RCF}_{1}(R)$ (resp. $\mathrm{RCF}_{2}(R)$ ) guarantees minimality of the denominator (resp. of the numerator) of the shell among all RNF's of $R$. Furthermore, it also guarantees minimality of its numerator (resp. of its denominator) among all those RNF's of $R$ that have the same (i.e., minimal) degree of the denominator (resp. of the numerator) as $\mathrm{RCF}_{1}(R)$ (resp. $\mathrm{RCF}_{2}(R)$ ):

Proposition 3.2. Let $R C F_{i}(R)=\left(z, r_{i}, s_{i}, u_{i}, v_{i}\right), i \in$ $\{1,2\}$. Let $(z, r, s, u, v)$ be an RNF of $R$.
(i) If $\operatorname{deg} v=\operatorname{deg} v_{1}$ then $v_{1}=v$ and $u_{1} \mid u$.
(ii) If $\operatorname{deg} u=\operatorname{deg} u_{2}$ then $u_{2}=u$ and $v_{2} \mid v$.

Proof: (i) Let $\operatorname{deg} v=\operatorname{deg} v_{1}$. By definition of $\mathrm{RCF}_{1}$ we have $v_{1} \mid v$, hence $v_{1}=v$. Then

$$
\frac{r_{1}}{s_{1}} \frac{E u_{1}}{u_{1}}=\frac{r}{s} \frac{E u}{u} .
$$

As $\operatorname{RCF}_{1}(R)$ is strict and $r / s$ is shift-reduced, Lemma 2.1 implies that $u_{1} \mid u$. - The proof of (ii) is analogous.

However, as shown by the next proposition, the price for absolute minimality of the denominator (resp. of the numerator) of the shell in $\mathrm{RCF}_{1}$ (resp. in $\mathrm{RCF}_{2}$ ) is maximality of its numerator (resp. of its denominator) among all strict RNF's of $R$.

Proposition 3.3. Let $R C F_{i}(R)=\left(z, r_{i}, s_{i}, u_{i}, v_{i}\right), i \in$ $\{1,2\}$. If $(z, r, s, u, v)$ is a strict $R N F$ of $R$ then $u \mid u_{1}$ and $v \mid v_{2}$.

Proof: By definition of $\mathrm{RCF}_{1}$, there is $w \in K[x]$ such that $v=v_{1} w$. Then

$$
\frac{r_{1}}{s_{1}} \frac{E\left(u_{1} w\right)}{\left(u_{1} w\right)}=\frac{r}{s} \frac{E u}{u} .
$$

As $(z, r, s, u, v)$ is strict and $r_{1} / s_{1}$ is shift-reduced, Lemma 2.1 implies that $u \mid u_{1} w$. From $u \perp v$ it follows that $u \perp w$, so $u \mid u_{1}$ as claimed. - The proof that $v \mid v_{2}$ is analogous.

Corollary 3.4. If $R C F_{1}(R)=R C F_{2}(R)$ then this is the only strict $R N F$ of $R$.

Proof: Let $\varphi=(z, r, s, u, v)$ be any strict RNF of $R$. Write $\operatorname{RCF}_{1}(R)=\operatorname{RCF}_{2}(R)=\left(z, r_{1}, s_{1}, u_{1}, v_{1}\right)$. Then $v_{1} \mid v$ by Definition 3.1 and $v \mid v_{1}$ by Proposition 3.3, hence $v=v_{1}$. By Proposition 2.3, $\varphi=\mathrm{RCF}_{1}(R)=\mathrm{RCF}_{2}(R)$.

### 3.2 Existence and Computation of $\mathrm{RCF}_{1}$ and $\mathrm{RCF}_{2}$

In this section we prove the existence of $\mathrm{RCF}_{1}$ and $\mathrm{RCF}_{2}$ by giving algorithms to construct them.

## Algorithm $\mathbf{R C F}_{1}$

input: $\quad R \in K(x) \backslash\{0\}$
output: $\operatorname{RCF}_{1}(R)$

$$
\begin{aligned}
& (z, a, b, c):=\operatorname{sPNF}(R) ; \\
& \left(1, a_{1}, b_{1}, c_{1}\right):=\operatorname{sPNF}(b / a) ; \\
& g:=\operatorname{gcd}\left(c, c_{1}\right) ; \quad(\operatorname{take} g \text { monic }) \\
& d:=c / g ; d_{1}:=c_{1} / g ; \\
& \text { return }\left(z, b_{1}, a_{1}, d, d_{1}\right) .
\end{aligned}
$$

## Algorithm $\mathbf{R C F}_{\mathbf{2}}$

input: $\quad R \in K(x) \backslash\{0\}$
output: $\mathrm{RCF}_{2}(R)$

$$
\begin{aligned}
& (z, r, s, u, v):=\operatorname{RCF}_{1}(1 / R) ; \\
& \text { return }(1 / z, s, r, v, u) .
\end{aligned}
$$

Now we proceed to prove correctness of these algorithms.
Lemma 3.5. Let $(z, a, b, c)$ be the strict PNF of $R \in$ $K(x) \backslash\{0\}$, and let $(z, r, s, u, v)$ be an RNF of $R$ such that $r \perp u$ and $s \perp E u$. Then $u \mid c$.

Proof: We have

$$
\begin{equation*}
R=z \frac{a}{b} \frac{E c}{c}=z \frac{r}{s} \frac{E(u / v)}{(u / v)} . \tag{6}
\end{equation*}
$$

Set

$$
R_{1}=\frac{a}{b} \frac{E(c v)}{(c v)}
$$

It follows from (6) that

$$
R_{1}=\frac{1}{z} \frac{E v}{v} R=\frac{r}{s} \frac{E u}{u} .
$$

As $r \perp u, s \perp E u$ and $\operatorname{gcd}\left(a, E^{k} b\right)=1$ for all $k \in \mathbb{N}$, it follows from Lemma 2.1 that $u \mid c v$. Hence $u \mid c$.

Lemma 3.6. Let $(z, r, s, u, v)$ be any $R N F$ of $R \in K(x) \backslash$ $\{0\}$. Then there is an $R N F\left(z, r^{\prime}, s^{\prime}, u^{\prime}, v\right)$ of $R$ such that $r^{\prime} \perp u^{\prime}$ and $s^{\prime} \perp E u^{\prime}$.

Proof: Let $\mathcal{R}$ be the set of all pairs of monic polynomials $(\rho, \tau)$ such that the pair $(\rho, s)$ is shift-reduced and $\rho E \tau / \tau=$ $r E u / u$. The set $\mathcal{R}$ contains $(r, u)$, so $\mathcal{R} \neq \emptyset$. Let $\left(r^{\prime}, u_{1}\right) \in$ $\mathcal{R}$ be such that $\operatorname{deg} u_{1}$ is minimal among all pairs in $\mathcal{R}$. Then

$$
\begin{equation*}
r^{\prime} \frac{E u_{1}}{u_{1}}=r \frac{E u}{u} . \tag{7}
\end{equation*}
$$

Denote $g=\operatorname{gcd}\left(r^{\prime}, u_{1}\right), r_{2}=r^{\prime} / g$ and $u_{2}=u_{1} / g$. Then $\operatorname{deg} u_{2} \leq \operatorname{deg} u_{1}$. As $r_{2}\left|r^{\prime}, E g\right| E r^{\prime}$, and ( $r^{\prime}, s$ ) is shiftreduced, so is ( $r_{2} E g, s$ ). As $r_{2} E g E u_{2} / u_{2}=r^{\prime} E u_{1} / u_{1}=$ $r E u / u$, it follows that $\left(r_{2} E g, u_{2}\right) \in \mathcal{R}$. By definition of $u_{1}$ we have $\operatorname{deg} u_{1} \leq \operatorname{deg} u_{2}$, so $\operatorname{deg} u_{1}=\operatorname{deg} u_{2}$ and $\operatorname{deg} g=0$. Hence $r^{\prime} \perp u_{1}$.

Let $\mathcal{S}$ denote the set of all pairs of monic polynomials $(\sigma, \tau)$ such that the pair $\left(r^{\prime}, \sigma\right)$ is shift-reduced and $(1 / \sigma) E \tau / \tau=(1 / s) E u_{1} / u_{1}$. The set $\mathcal{S}$ contains $\left(s, u_{1}\right)$, so $\mathcal{S} \neq \emptyset$. Let $\left(s^{\prime}, u^{\prime}\right) \in \mathcal{S}$ be such that $\operatorname{deg} u^{\prime}$ is minimal among all pairs in $\mathcal{S}$. Then

$$
\begin{equation*}
\frac{1}{s^{\prime}} \frac{E u^{\prime}}{u^{\prime}}=\frac{1}{s} \frac{E u_{1}}{u_{1}} \tag{8}
\end{equation*}
$$

It can be shown that $s^{\prime} \perp E u^{\prime}$ (the proof is analogous to the one showing that $r^{\prime} \perp u_{1}$ given in the preceding paragraph).

Together with (8) and Lemma 2.1 this implies that $u^{\prime} \mid u_{1}$, and so $r^{\prime} \perp u^{\prime}$. Finally, from (7) and (8) we have

$$
\frac{r^{\prime}}{s^{\prime}} \frac{E u^{\prime}}{u^{\prime}}=\frac{r^{\prime}}{s} \frac{E u_{1}}{u_{1}}=\frac{r}{s} \frac{E u}{u}
$$

so $\left(z, r^{\prime}, s^{\prime}, u^{\prime}, v\right)$ is an RNF of $R$ with required properties.
Theorem 3.7. Algorithms $R C F_{1}$ and $R C F_{2}$ are correct.
Proof: Let $z, a, b, c, a_{1}, b_{1}, c_{1}, g, d, d_{1}$ be as in Algorithm $\mathrm{RCF}_{1}$. We claim that $\varphi_{1}=\left(z, b_{1}, a_{1}, d, d_{1}\right)$ is $\operatorname{RCF}_{1}(R)$. It follows from (the proof of) [2, Theorem 1] that $\varphi_{1}$ is a strict RNF of $R$. We need to show that if $\varphi=(z, r, s, u, v)$ is any RNF of $R$ then $d_{1} \mid v$. By Lemma 3.6, there is an RNF $\left(z, r^{\prime}, s^{\prime}, u^{\prime}, v\right)$ of $R$ such that $r^{\prime} \perp u^{\prime}$ and $s^{\prime} \perp E u^{\prime}$. By Lemma 3.5, $u^{\prime} \mid c$, so $c_{2}:=c v / u^{\prime}$ is a polynomial and

$$
\frac{a_{1}}{b_{1}} \frac{E c_{1}}{c_{1}}=\frac{b}{a}=z \frac{1}{R} \frac{E c}{c}=\frac{s^{\prime}}{r^{\prime}} \frac{E c_{2}}{c_{2}}
$$

As $a_{1} \perp c_{1}, b_{1} \perp E c_{1}$ and $s^{\prime} / r^{\prime}$ is shift-reduced, Lemma 2.1 implies that $c_{1} \mid c_{2}$. Let $q_{1}=c_{2} / c_{1} \in K[x]$. Then

$$
d_{1} q_{1} u^{\prime}=\frac{c_{1}}{g} q_{1} u^{\prime}=\frac{c_{2}}{g} u^{\prime}=\frac{c v}{g}=d v
$$

As $d_{1} \perp d$, it follows that $d_{1} \mid v$ which proves the claim.
Let $\varphi_{2}=(1 / z, s, r, v, u)$ be the output of Algorithm $\mathrm{RCF}_{2}$. We claim that $\varphi_{2}$ is $\operatorname{RCF}_{2}(R)$. By Theorem $2.2(\mathrm{v}), \varphi_{2}$ is a strict RNF of $R$. Let $\varphi=\left(1 / z, s^{\prime}, r^{\prime}, v^{\prime}, u^{\prime}\right)$ be any strict RNF of $R$. Then by Theorem $2.2(\mathrm{v}),\left(z, r^{\prime}, s^{\prime}, u^{\prime}, v^{\prime}\right)$ is a strict RNF of $1 / R$. Since $(z, r, s, u, v)=\operatorname{RCF}_{1}(1 / R)$, it follows that $v \mid v^{\prime}$, proving the claim.

Example 3.1. Consider the rational function

$$
R=\frac{x(x+2)(x-4+\sqrt{2})(x-3+\sqrt{2})(x+2+\sqrt{2})(x+11+\sqrt{2})}{(x-3)(x-2)^{2}(x+6)(x+12)(x-1+\sqrt{2})(x+1+\sqrt{2})} .
$$

Following Algorithm $\mathrm{RCF}_{1}, R C F_{1}\left(z, r_{1}, s_{1}, u_{1}, v_{1}\right)$ of $R$ is

$$
\begin{aligned}
& (1,(x-4+\sqrt{2})(x-3+\sqrt{2}),(x-3)(x+6)(x+12), \\
& (x-2)^{2}(x-1)^{2} x(x+1)(x-1+\sqrt{2})(x+\sqrt{2}) \\
& (x+1+\sqrt{2})^{2}(x+2+\sqrt{2})(x+3+\sqrt{2})(x+4+\sqrt{2}) \\
& (x+5+\sqrt{2})(x+6+\sqrt{2})(x+7+\sqrt{2})(x+8+\sqrt{2}) \\
& (x+9+\sqrt{2})(x+10+\sqrt{2}), 1) .
\end{aligned}
$$

Following Algorithm $\mathrm{RCF}_{2}, R C F_{2}\left(z, r_{2}, s_{2}, u_{2}, v_{2}\right)$ of $R$ is

$$
\begin{aligned}
& \left(1,(x+2+\sqrt{2})(x+11+\sqrt{2}),(x-3)(x-2)^{2}, 1,\right. \\
& \quad x(x+1)(x+2)^{2}(x+3)^{2}(x+4)^{2}(x+5)^{2}(x+6) \\
& \quad(x+7)(x+8)(x+9)(x+10)(x+11)(x-4+\sqrt{2}) \\
& \left.\quad(x-3+\sqrt{2})^{2}(x-2+\sqrt{2})^{2}(x-1+\sqrt{2})(x+\sqrt{2})\right) .
\end{aligned}
$$

Notice that $\operatorname{deg} r_{1}=\operatorname{deg} r_{2}, \operatorname{deg} s_{1}=\operatorname{deg} s_{2}, u_{2}\left|u_{1}, v_{1}\right| v_{2}$, as expected.

## 4. MINIMIZING THE SHELL: TOTAL DEGREE

### 4.1 The Multiplicative Structure of the Shell

Let (5) be the shift-homogeneous factorization of a rational function $R$. Then, obviously, there exists a one-to-one
correspondence between $\mathrm{RNF}_{x}(R)$ and $\mathrm{RNF}_{x}\left(R_{1}\right) \times \cdots \times$ $\operatorname{RNF}_{x}\left(R_{k}\right)$ :

$$
(F, V) \leftrightarrow\left(\left(F_{1}, V_{1}\right), \ldots,\left(F_{k}, V_{k}\right)\right)
$$

where $F=z F_{1} \cdots F_{k}$ and $V=V_{1} \cdots V_{k}$ are shifthomogeneous factorizations of $F$ resp. $V$ such that for $1 \leq i \leq k$ the irreducible factors of $R_{i}, F_{i}$ and $V_{i}$ are shiftequivalent, $(F, V)$ is an RNF of $R$, and $\left(F_{i}, V_{i}\right)$ is an RNF of $R_{i}$ for $1 \leq i \leq k$. Therefore we can limit our attention to a monic shift-homogeneous rational function $R$ of the form

$$
\begin{equation*}
\frac{p\left(x+a_{1}\right) p\left(x+a_{2}\right) \cdots p\left(x+a_{m}\right)}{p\left(x+b_{1}\right) p\left(x+b_{2}\right) \cdots p\left(x+b_{n}\right)} \tag{9}
\end{equation*}
$$

where $p(x)$ is an irreducible polynomial while $a_{1} \leq a_{2} \leq$ $\cdots \leq a_{m}$ and $b_{1} \leq b_{2} \leq \cdots \leq b_{n}$ are nonnegative integers such that $a_{i} \neq b_{j}$ for all $i$ and $j$. If $m=n$ then it follows from Theorem 2.2 (ii) and (iv) that (9) has a unique RNF $(z, r, s, u, v)$ such that $z=r=s=1$. Clearly, $u=u_{1} \cdots u_{m}$ and $v=v_{1} \cdots v_{m}$ where

$$
\begin{align*}
& u_{i}(x)= \begin{cases}1, & a_{i}<b_{i} \\
\prod_{k=b_{i}}^{a_{i}-1} p(x+k), & a_{i}>b_{i}\end{cases}  \tag{10}\\
& v_{i}(x)= \begin{cases}\prod_{k=a_{i}}^{b_{i}-1} p(x+k), & a_{i}<b_{i} \\
1, & a_{i}>b_{i}\end{cases} \tag{11}
\end{align*}
$$

Thus

$$
\begin{align*}
& \operatorname{deg} u+\operatorname{deg} v=(\operatorname{deg} p) \sum_{k=1}^{m}\left|a_{k}-b_{k}\right|  \tag{12}\\
& \operatorname{deg} u-\operatorname{deg} v=(\operatorname{deg} p) \sum_{k=1}^{m}\left(a_{k}-b_{k}\right)
\end{align*}
$$

Otherwise $m \neq n$. From now on assume that $m<n$ (the case $m>n$ can be treated similarly, cf. Algorithm mshRCF ${ }_{1}^{*}$ below). It follows from Theorem 2.2 (ii), (iv) and (vi) that any strict RNF ( $1,1, s, u, v$ ) of $R$ arises from an injection

$$
\begin{equation*}
f:\{1,2, \ldots, m\} \rightarrow\{1,2, \ldots, n\} \tag{13}
\end{equation*}
$$

such that $f(1)<f(2)<\cdots<f(m)$, by taking $s(x)=$ $\prod_{k \notin \mathrm{rng} f} p\left(x+b_{k}\right)$, and $(1,1,1, u, v)$ to be the unique RNF of the rational function $\prod_{k=1}^{m} p\left(x+a_{k}\right) / p\left(x+b_{f(k)}\right)$. Here $\operatorname{rng} f=\{f(1), f(2), \ldots, f(m)\}$ is the range of $f$. Similarly to (12) we obtain the following theorem.

Theorem 4.1. Let $R$ be written in the form (9), let $f$ be an injection of the form (13), and let $(F, V)$ be the corresponding RNF of $R$. Then

$$
\begin{aligned}
& \operatorname{deg} \operatorname{num} V+\operatorname{deg} \operatorname{den} V=(\operatorname{deg} p) \sum_{k=1}^{m}\left|a_{k}-b_{f(k)}\right| \\
& \operatorname{deg} \text { num } V-\operatorname{deg} \operatorname{den} V=(\operatorname{deg} p) \sum_{k=1}^{m}\left(a_{k}-b_{f(k)}\right)
\end{aligned}
$$

In general, not all RNF's induced by injections of the form (13) are strict.

Lemma 4.2. An injection $f$ of the form (13) induces a strict RNF of $R$ iff for any $j \notin \operatorname{rng}(f)$ the number of $k$ such that $b_{f(k)}<b_{j}$ is equal to the number of $k$ such that $a_{k}<b_{j}$.

Proof: Suppose that $j \notin \operatorname{rng}(f)$, and $l=\max \{k: f(k)<j\}$. Let $V_{1}, V_{2}$ be such that

$$
\begin{aligned}
\frac{p\left(x+a_{1}\right) \cdots p\left(x+a_{l}\right)}{p\left(x+b_{f(1)}\right) \cdots p\left(x+b_{f(l)}\right)} & =\frac{E V_{1}}{V_{1}} \\
\frac{p\left(x+a_{l+1}\right) \cdots p\left(x+a_{m}\right)}{p\left(x+b_{f(l+1)}\right) \cdots p\left(x+b_{f(m)}\right)} & =\frac{E V_{2}}{V_{2}}
\end{aligned}
$$

Then (11) implies that $p\left(x+b_{j}\right)$ does not divide den $V_{1}$ because $b_{j}>b_{f(l)}$, and (10) implies that $p\left(x+b_{j}\right)$ divides num $E V_{1}$ iff $b_{j} \leq a_{l}$. The case of $V_{2}$ is treated similarly.

### 4.2 RNF*: Forms with Minimal Total Degree of the Shell

Definition 4.1. An $R N F(z, r, s, u, v)$ of $R \in K(x) \backslash\{0\}$ is an RNF* if $\operatorname{deg} u+\operatorname{deg} v$ is minimal among all $R N F$ 's of $R$.

Proposition 4.3. Any $R N F^{*}$ is strict.
Proof: Let $\varphi=(z, r, s, u, v)$ be a non-strict RNF of a rational function $R$. Then $\operatorname{deg} \operatorname{gcd}(r, u) \geq 1$ or $\operatorname{deg} \operatorname{gcd}(r, E v) \geq 1$ or $\operatorname{deg} \operatorname{gcd}(s, E u) \geq 1$ or $\operatorname{deg} \operatorname{gcd}(s, v) \geq 1$.

If $\operatorname{deg} \operatorname{gcd}(r, u) \geq 1$, write $g=\operatorname{gcd}(r, u), r=g r^{\prime}$ and $u=g u^{\prime}$ where $r^{\prime}, u^{\prime} \in K[x]$. Then $\left(z, r^{\prime} E g, s, u^{\prime}, v\right)$ is an RNF of $R$. As $\operatorname{deg} u^{\prime}+\operatorname{deg} v<\operatorname{deg} u+\operatorname{deg} v, \varphi$ is not an RNF*.

If $\operatorname{deg} \operatorname{gcd}(r, E v) \geq 1$, write $g=\operatorname{gcd}(r, E v), r=g r^{\prime}$ and $E v=g E v^{\prime}$ where $r^{\prime}, v^{\prime} \in K[x]$. Then $\left(z, r^{\prime} E^{-1} g, s, u, v^{\prime}\right)$ is an RNF of $R$. As $\operatorname{deg} u+\operatorname{deg} v^{\prime}<\operatorname{deg} u+\operatorname{deg} v, \varphi$ is not an RNF*.

Similarly, one can show that $\varphi$ is not an RNF* in the other two cases.

Thus, by Theorem 2.2 (vi), the problem of finding an RNF* is equivalent to the problem of finding an injection $f$ of the form (13) such that the sum

$$
\begin{equation*}
\sum_{k=1}^{m}\left|a_{k}-b_{f(k)}\right| \tag{14}
\end{equation*}
$$

is minimal.
Example 4.1. Consider the rational function $R$ in Example 3.1. $R$ can be written as $R_{1} \cdot R_{2}$ where $R_{1}, R_{2}$ each is a monic shift-homogeneous rational function:

$$
\begin{aligned}
R_{1} & =\frac{x(x+2)}{(x-3)(x-2)^{2}(x+6)(x+12)} \\
R_{2} & =\frac{(x-4+\sqrt{2})(x-3+\sqrt{2})(x+2+\sqrt{2})(x+11+\sqrt{2})}{(x-1+\sqrt{2})(x+1+\sqrt{2})}
\end{aligned}
$$

For the monic shift-homogeneous factor $R_{1}$, there exist two injections $f_{1}, f_{2}$ such that the sum $\sum_{k=1}^{2}\left|a_{k}-b_{f_{j}(k)}\right|$ is minimal for $1 \leq j \leq 2$.
For the injection $f_{1}$ :

$$
R_{1}=\frac{1}{x-3} \cdot \frac{x}{x-2} \cdot \frac{x+2}{x-2} \cdot \frac{1}{x+6} \cdot \frac{1}{x+12}
$$

the corresponding $\operatorname{RNF}^{*}\left(z, r_{1}, s_{1}, u_{1}, v_{1}\right)$ is
$\left(1,1,(x-3)(x+6)(x+12),(x-2)^{2}(x-1)^{2} x(x+1), 1\right)$.

For the injection $f_{2}$ :

$$
R_{1}=\frac{1}{x-3} \cdot \frac{x}{x-2} \cdot \frac{1}{x-2} \cdot \frac{x+2}{x+6} \cdot \frac{1}{x+12}
$$

the corresponding $\operatorname{RNF}^{*}\left(z, r_{2}, s_{2}, u_{2}, v_{2}\right)$ is

$$
\begin{aligned}
& (1,1,(x-3)(x-2)(x+12),(x-1)(x-2) \\
& \quad(x+2)(x+3)(x+4)(x+5)) .
\end{aligned}
$$

For the monic shift-homogeneous factor $R_{2}$, there exists one injection $f$ such that the sum $\sum_{k=1}^{2}\left|a_{k}-b_{f(k)}\right|$ is minimal:
$R_{2}=(x-4+\sqrt{2}) \cdot \frac{x-3+\sqrt{2}}{x-1+\sqrt{2}} \cdot \frac{x+2+\sqrt{2}}{x+1+\sqrt{2}} \cdot(x+11+\sqrt{2})$,
and the corresponding $\operatorname{RNF}^{*}\left(z_{3}, r_{3}, s_{3}, u_{3}, v_{3}\right)$ is

$$
\begin{aligned}
& (1,(x-4+\sqrt{2})(x+11+\sqrt{2}), 1,(x+1+\sqrt{2}) \\
& \quad(x-2+\sqrt{2})(x-3+\sqrt{2})) .
\end{aligned}
$$

As the result, the two $\mathrm{RNF}^{*}$ 's $\left(z, r_{1}^{*}, s_{1}^{*}, u_{1}^{*}, v_{1}^{*}\right)$ and $\left(z, r_{2}^{*}, s_{2}^{*}, u_{2}^{*}, v_{2}^{*}\right)$ respectively of $R$ are

$$
\begin{aligned}
& (1,(x-4+\sqrt{2})(x+11+\sqrt{2}),(x-3)(x+6)(x+12) \\
& (x-2)^{2}(x-1)^{2} x(x+1)(x+1+\sqrt{2}),(x-3+\sqrt{2}) \\
& (x-2+\sqrt{2})), \text { and } \\
& (1,(x-4+\sqrt{2})(x+11+\sqrt{2}),(x-3)(x-2)(x+12) \\
& (x-1)(x-2)(x+1+\sqrt{2}) \\
& (x+2)(x+3)(x+4)(x+5)(x-3+\sqrt{2})(x-2+\sqrt{2})) .
\end{aligned}
$$

Note that the total degree of the shell in both RNF*'s is 9, while it is 19 for $R C F_{1}(R)$, and 23 for $R C F_{2}(R)$ (see Example 3.1).

### 4.3 Reduction to a Linear Programming Problem

The problem of computing an injection $f$ such that the sum in (14) is minimal can be reduced to a well-known combinatorial problem which can be solved by linear programming techniques. This is the Minimum Weighted Bipartite Matching Problem (MWBM): Given a complete bipartite graph $K_{m, n}$ (where $m \leq n$ ) with rational weights on the edges, find a matching (i.e., a set of pairwise nonadjacent edges) of size $m$ which has minimum total weight. It is well known [4] that MWBM can be solved efficiently (in time polynomial in $\max \{m, n\}$, i.e., avoiding exhaustive search).

To reduce the problem of constructing an RNF* to MWBM (which is also known as the Assignment Problem), construct a complete bipartite graph with vertex sets $\left\{u_{1}, u_{2}, \ldots, u_{m}\right\}$ and $\left\{v_{1}, v_{2}, \ldots, v_{n}\right\}$ where all the $u_{j}$ 's and $v_{k}$ 's are pairwise distinct, and let the weight on the edge connecting $u_{j}$ with $v_{k}$ be $\left|a_{j}-b_{k}\right|$. This special case of MWBM can be solved even in linear time [3]. If the injection $f$ given by the solution to MWBM is not monotonically increasing we replace it by the unique monotonically increasing injection having the same range as $f$. Note that this will not increase the weight of the corresponding matching.

### 4.4 Definition and Properties of $\mathrm{RCF}_{1}^{*}$ and $\mathrm{RCF}_{2}^{*}$

There may exist several $\mathrm{RNF}^{*}$ 's $(z, r, s, u, v)$ of $R \in K(x) \backslash$ $\{0\}$. Among all such forms, we can again distinguish two forms which minimize $\operatorname{deg} v$ resp. $\operatorname{deg} u$ (i.e., maximize or minimize $\operatorname{deg} u-\operatorname{deg} v$, respectively). We denote them by $\mathrm{RCF}_{1}^{*}$ and $\mathrm{RCF}_{2}^{*}$, respectively.

Remark 4.1. If $\varphi$ is an $\mathrm{RCF}_{1}^{*}$ of $1 / R$ then, clearly, $\varphi^{-1}$ is an $\mathrm{RCF}_{2}^{*}$ of $R$.

Theorem 4.4. Every $R \in K(x) \backslash\{0\}$ has a unique $R C F_{1}^{*}$ and a unique $R C F_{2}^{*}$.

Proof: Existence of $\mathrm{RCF}_{1}^{*}$ and $\mathrm{RCF}_{2}^{*}$ follows from the existence of RNF's.

Let us prove, for example, uniqueness of $\mathrm{RCF}_{1}^{*}$. Uniqueness of $\mathrm{RCF}_{2}^{*}$ will then follow from Remark 4.1. Suppose that $R$ is of the form (9), and that there are two injections $f, f^{\prime}$ which induce two different $\mathrm{RCF}_{1}^{*}$ 's of $R$. Let $l, 1 \leq l \leq m$, be the least such that $f(l) \neq f^{\prime}(l)$. W.l.g. assume that $f(l)>f^{\prime}(l)$, and hence $b_{f(l)}>b_{f^{\prime}(l)}$.
(a) $\left(a_{l}-b_{f(l)}\right)\left(a_{l}-b_{f^{\prime}(l)}\right)>0$.
(a1) $a_{l}-b_{f(l)}<0$ and $a_{l}-b_{f^{\prime}(l)}<0$. The injection $f^{\prime \prime}$ such that if $k \neq l$ then $f^{\prime \prime}(k)=f(k)$ while $f^{\prime \prime}(l)=f^{\prime}(l)$ produces a smaller sum (14) than the one produced by $f$.
(a2) $a_{l}-b_{f(l)}>0$ and $a_{l}-b_{f^{\prime}(l)}>0$. This case is similar to (a1).
(b) $\left(a_{l}-b_{f(l)}\right)\left(a_{l}-b_{f^{\prime}(l)}\right)<0$. Since $b_{f(l)}>b_{f^{\prime}(l)}$, we have $a_{l}-b_{f(l)}<0$ and $a_{l}-b_{f^{\prime}(l)}>0$. Consider two cases:
(b1) $\left|a_{l}-b_{f(l)}\right| \neq\left|a_{l}-b_{f^{\prime}(l)}\right|$. Similarly to (a), it is possible to decrease the sum produced by $f$.
(b2) $\left|a_{l}-b_{f(l)}\right|=\left|a_{l}-b_{f^{\prime}(l)}\right|$. By changing $f$ as described in (a), we get $f^{\prime \prime}$ which does not change the sum, but decreases $\operatorname{deg} v$.

Example 4.2. For the rational function $R$ in Example 4.1, the computed $\operatorname{RNF}^{*}\left(z, r_{1}^{*}, s_{1}^{*}, u_{1}^{*}, v_{1}^{*}\right)$ is the $\mathrm{RCF}_{1}^{*}$ of $R$, and the computed $\operatorname{RNF}^{*}\left(z, r_{2}^{*}, s_{2}^{*}, u_{2}^{*}, v_{2}^{*}\right)$ is the $\mathrm{RCF}_{2}^{*}$ of $R\left(\operatorname{deg} u_{1}^{*}=7, \operatorname{deg} v_{1}^{*}=2, \operatorname{deg} u_{2}^{*}=3, \operatorname{deg} v_{2}^{*}=6\right)$.

### 4.5 Computation of RCF $_{1}^{*}$ and RCF $_{2}^{*}$

Suppose again that $m<n$ in (14). Computation of $\operatorname{RCF}_{1}^{*}(R)$ is a special choice of an injection $f$ or, equivalently, of $m$ factors $p\left(x+b_{f(1)}\right) p\left(x+b_{f(2)}\right) \cdots p\left(x+b_{f(m)}\right)$ of the denominator of (9). If we wish to obtain $\operatorname{RCF}_{1}^{*}(R)$ then we should find an RNF* that maximizes the sum $\sum_{k=1}^{m}\left(a_{k}-\right.$ $\left.b_{f(k)}\right)$ or, equivalently, minimizes the sum $b_{f(1)}+\cdots+b_{f(m)}$. For this purpose, we add $n-m$ new vertices $u_{m+1}, \ldots, u_{n}$ to the vertex set $\left\{u_{1}, \ldots, u_{m}\right\}$, and connect each of them with each of $v_{1}, \ldots, v_{n}$. Set $N=b_{1}+\cdots+b_{n}+1$. Let the weight $w_{j k}$ of the edge $\left[u_{j}, v_{k}\right]$ be equal to $\left|a_{j}-b_{k}\right|$ if $j \leq m$, and to $1-b_{k} / N$ otherwise. When MWBM is solved, any vertex $u_{j}, j \leq m$, is connected with a unique vertex $v_{k}$. This gives an injection $f$ of the form (13).

Lemma 4.5. The algorithm described above constructs an injection $f$ such that the sum (14) is minimal. Additionally, among all injections that minimize this sum, the constructed injection minimizes the sum $b_{f(1)}+\cdots+b_{f(m)}$.

Proof: It is easy to see that if we set

$$
w_{\sigma \tau}=1-\varepsilon_{\tau}, \varepsilon_{\tau} \geq 0, \sigma=m+1, \ldots, n, \tau=1, \ldots, n
$$

and $\varepsilon_{1}+\cdots+\varepsilon_{n}<1$, then any solution of MWBM minimizes the sum of the corresponding integer weights (and, thereby, gives us an $\operatorname{RNF}^{*}(R)$ ), and under this condition, maximizes the sum of those $\varepsilon_{\tau}$ for which the solution of MWBM contains an edge $\left[u_{\sigma}, v_{\tau}\right]$ with $\sigma>m$. This means that if we define $\varepsilon_{k}$ as $b_{k} / N$ then we obtain an injection $f$ that gives an RNF* with minimal $\left(b_{f(1)}+\cdots+b_{f(m)}\right) / N$ or, equivalently, with minimal $b_{f(1)}+\cdots+b_{f(m)}$.

THEOREM 4.6. The algorithm described above constructs an injection $f$ which induces $\mathrm{RCF}_{1}^{*}$ of (9).

Proof: The claim follows immediately from Lemma 4.5.
Note that in the case $m>n$ we have to maximize the sum $a_{f(1)}+\cdots+a_{f(n)}$ where $f$ is an injection from $\{1, \ldots, n\}$ to $\{1, \ldots, m\}$. To attain this goal, set $w_{j k}$ to $\left|a_{j}-b_{k}\right|$ if $k \leq n$, and $a_{j} / M$ otherwise, where $M=a_{1}+\cdots+a_{m}+1$.

We conclude this section by giving detailed descriptions of the algorithms to compute $\mathrm{RCF}_{1}^{*}$ and $\mathrm{RCF}_{2}^{*}$. First, we present Algorithm mshRCF ${ }_{1}^{*}$ for computing $\mathrm{RCF}_{1}^{*}$ of a monic shift-homogeneous rational function $R$ of the form (9). Let $\operatorname{MWBM}(m, w)$ be an algorithm for solving the Minimum Weighted Bipartite Matching Problem on the balanced complete bipartite graph $K_{m, m}$ with the $m \times m$ weight matrix $w$. The output of $\operatorname{MWBM}(m, w)$ is the injection $f$ such that the sum (14) is minimal. Then Algorithm mshRCF ${ }_{1}^{*}$ can be described as follows.

```
Algorithm mshRCF \({ }_{1}^{*}\)
input: a monic shift-homogeneous rational function \(R\) of
the form (9)
output: \(\mathrm{RCF}_{1}^{*}(R)\)
if \(m<n\) then
    \(N:=b_{1}+\cdots+b_{n}+1 ;\)
    for \(s\) from 1 to \(n\) do
        for \(r\) from 1 to \(m\) do
            \(w_{r s}:=\left|a_{r}-b_{s}\right| ;\)
        od;
        for \(r\) from \(m+1\) to \(n\) do
            \(w_{r s}:=1-b_{s} / N ;\)
        od;
    od;
    \(f:=\operatorname{MWBM}(n, w)\);
    \((1,1,1, u, v):=\operatorname{RCF}_{1}\left(\prod_{k=1}^{m} p\left(x+a_{k}\right) / p\left(x+b_{f(k)}\right)\right)\);
    return \(\left(1,1, \prod_{k \in\{1, \ldots, n\} \backslash \operatorname{rng}(f)} p\left(x+b_{k}\right), u, v\right)\).
else
    \(M:=a_{1}+\cdots+a_{m}+1 ;\)
    for \(r\) from 1 to \(m\) do
        for \(s\) from 1 to \(n\) do
            \(w_{r s}:=\left|a_{r}-b_{s}\right| ;\)
        od;
        for \(s\) from \(n+1\) to \(m\) do
            \(w_{r s}:=a_{r} / M ;\)
        od;
    od;
    \(f:=\operatorname{MWBM}(m, w)\);
    \((1,1,1, u, v):=\operatorname{RCF}_{1}\left(\prod_{k=1}^{n} p\left(x+a_{f(k)}\right) / p\left(x+b_{k}\right)\right) ;\)
    \(\operatorname{return}\left(1, \prod_{k \in\{1, \ldots, m\} \backslash \operatorname{rng}(f)} p\left(x+a_{k}\right), 1, u, v\right)\).
fi.
```

For $R \in K(x)$, let the output of the function $\operatorname{SHF}(R)$ be the shift-homogeneous factorization in the form (5) of $R$. The following algorithms compute $\mathrm{RCF}_{1}^{*}$ and $\mathrm{RCF}_{2}^{*}$ of $R$, respectively.

```
Algorithm RCF*
input: }R\inK(x)\{0
output: }\mp@subsup{\textrm{RCF}}{1}{*}(R
    (z, R},\mp@subsup{R}{2}{},\ldots,\mp@subsup{R}{k}{}):=\operatorname{SHF}(R)
    for }i\mathrm{ from }1\mathrm{ to }k\mathrm{ do
        (1, ri, si, ui,vi):= mshRCF
    od;
    return (z, \prod
```


## Algorithm RCF ${ }_{\mathbf{2}}^{*}$

input: $\quad R \in K(x) \backslash\{0\}$
output: $\operatorname{RCF}_{2}^{*}(R)$

$$
\begin{aligned}
& (z, r, s, u, v):=\operatorname{RCF}_{1}^{*}(1 / R) ; \\
& \text { return }(1 / z, s, r, v, u) .
\end{aligned}
$$

## 5. REPRESENTING HYPERGEOMETRIC TERMS EFFICIENTLY

A hypergeometric term $T(n)$ is usually represented as

$$
\begin{equation*}
\alpha^{n} P(n) \tag{15}
\end{equation*}
$$

where $\alpha \in K$ and $P(n)$ is a product of Gamma-function values (if $K=\mathbb{C}$ ), or Pochhammer symbols (i.e., rising factorial powers) and their reciprocals. Such representation can be simplified: we can replace (15) by

$$
\begin{equation*}
\alpha^{n} V(n) Q(n) \tag{16}
\end{equation*}
$$

where $V(n)$ is a rational function, and $Q(n)$ is a product that looks like $P(n)$, but has the minimal possible number of factors. This can be achieved by using any RNF of the certificate of $T(n)(V(n)$ is the shell of this RNF). If we use any of the rational canonical forms of the certificate of $T$ as discussed in Sections 3 and 4, then we can additionally minimize $V(n)$ in one sense or another.

### 5.1 Efficient Multiplicative Decompositions using RCF's and RCF*'s

Definition 5.1. Let $T(n)$ be a hypergeometric term. A multiplicative decomposition of $T$ is a triple $\left(F, W, n_{0}\right)$ where $F, W \in K(x)$ and $n_{0} \in \mathbb{Z}$ are such that for all integers $n \geq n_{0}$ :
(i) $T$ is defined at $n, F$ has neither a pole nor a zero at $n, W$ has no pole at $n$,
(ii) $T(n)$ can be written as

$$
\begin{equation*}
T(n)=W(n) \prod_{k=n_{0}}^{n-1} F(k) \tag{17}
\end{equation*}
$$

This decomposition is minimal if for any multiplicative decomposition ( $G, W_{1}, n_{1}$ ) of $T$ we have $\operatorname{deg}$ num $F \leq$ $\operatorname{deg} \operatorname{num} G$ and $\operatorname{deg} \operatorname{den} F \leq \operatorname{deg} \operatorname{den} G$.

Let $T(n)$ be a hypergeometric term with the certificate $R \in K(x)$. Let $n_{0} \in \mathbb{Z}$ be such that $T(n)$ is defined for all integers $n \geq n_{0}$, and $R$ has neither a pole nor a zero at $n$. It is easy to check that the triple $\left(R, T\left(n_{0}\right), n_{0}\right)$ is a multiplicative decomposition of $T$. Let $(F, V)$ be an RNF of $R$. Set $W(n)=V(n) T\left(n_{0}\right) / V\left(n_{0}\right)$. Then it follows from Definition 5.1 and Theorem 2.2 that the multiplicative decomposition ( $F, W, n_{0}$ ) is minimal.

Let $(F, V)$ be one of the four RCF's of $R$ as discussed in Sections 3 and 4, and the hypergeometric term $T(n)$ be written in the form (17). Then in addition to the property that the numerator and the denominator of the kernel $F$ are of minimal possible degrees, the shell $V$ is also minimal in some sense. That is, if we use $\mathrm{RCF}_{1}$, then $\operatorname{den} V$ is of minimal degree; if we use $\mathrm{RCF}_{2}$, then num $V$ is of minimal degree; if we use $\mathrm{RCF}_{1}^{*}$ or $\mathrm{RCF}_{2}^{*}$, then $\operatorname{deg}$ num $V+\operatorname{deg} \operatorname{den} V$ is minimal, and under this condition, $\operatorname{deg} \operatorname{den} V$ is minimal for $\mathrm{RCF}_{1}^{*}$, and deg num $V$ is minimal for $\mathrm{RCF}_{2}^{*}$. In this case the representation of $T(n)$ of the form (17) is called an efficient multiplicative decomposition of $T$, denoted by $\operatorname{EMD}(T)$.

For a hypergeometric term $T(n)$, let $R$ be the certificate of $T$, denoted by $\operatorname{cer}(T)$. Set RCF[3] := RCF ${ }_{1}^{*}$, and RCF[4]:= $\mathrm{RCF}_{2}^{*}$. The following is a description of the algorithm to construct an efficient multiplicative decomposition of $T(n)$.

## Algorithm EMD[i]

input: a hypergeometric term $T(n), i \in\{1,2,3,4\}$
output: an efficient multiplicative decomposition $W(n) \prod_{k=n_{0}}^{n-1} F(k)$ of $T(n)$ where:
If $i=1$ then $\operatorname{deg} \operatorname{den} W$ is minimal.
If $i=2$ then $\operatorname{deg}$ num $W$ is minimal.
If $i=3$ then $\operatorname{deg}$ num $W+\operatorname{deg} \operatorname{den} W$ is minimal, and $\operatorname{deg} \operatorname{den} W$ is minimal. If $i=4$ then $\operatorname{deg}$ num $W+\operatorname{deg} \operatorname{den} W$ is minimal, and deg num $W$ is minimal;

$$
\begin{aligned}
& R:=\operatorname{cer}(T) \\
& (F, V):=\operatorname{RCF}[i](R) ;
\end{aligned}
$$

let $n_{0} \in \mathbb{Z}$ be such that $T(n)$ is defined for all integers $n \geq n_{0}$, and $R$ has neither a pole nor a zero at $n$; $W:=V(n) T\left(n_{0}\right) / V\left(n_{0}\right)$;
return $W(n) \prod_{k=n_{0}}^{n-1} F(k)$.

### 5.2 Gamma-function Values and Pochhammer Symbols

Using Pochhammer symbol we can write

$$
\prod_{k=n_{0}}^{n-1}(k-c)=\left(n_{0}-c\right)_{n-n_{0}}
$$

for any $c \in K, n_{0} \in \mathbb{Z}$. If $K=\mathbb{C}$, then similarly

$$
\prod_{k=n_{0}}^{n-1}(k-c)=\frac{\Gamma(n-c)}{\Gamma\left(n_{0}-c\right)}
$$

Conversely, each expression

$$
\begin{equation*}
(-c)_{n}, \text { or } \Gamma(n-c) \tag{18}
\end{equation*}
$$

can be represented in the form $\delta \prod_{k=n_{0}}^{n-1}(k-c)$, where $\delta$ is a constant. Suppose that a hypergeometric term $T(n)$ is represented in an efficient multiplicative decomposition proposed in Section 5.1 as $\alpha^{n} V(n) \prod_{k=n_{0}}^{n-1} F(n)$ where $\alpha \in K$, and $F(n)$ is a monic rational function. If we factorize the
numerator and the denominator of $F$ over some extension of $K$ into linear factors, then by the above reasoning we can represent $T(n)$ in the form (16) with minimized $V(n)$ and with $Q(n)$ having the minimal possible number of factors of the form (18). Such a form is called an efficient representation of $T$.

Example 5.1. Consider the hypergeometric term $T(n)$ :

$$
24 \prod_{k=1}^{n-1} \frac{1}{2} \frac{\left(3 k^{2}+6 k+4\right)(2 k+3)(4 k+5)(k+1)(4 k+3)}{k(4 k-1)(2 k-1)(4 k-3)(2 k+5)(k+2)\left(3 k^{2}+1\right)}
$$

A multiplicative decomposition $T(n)=T\left(n_{0}\right) \prod_{k=n_{0}}^{n-1} R(k)$ where the product is expressed in terms of a product of Gamma-function values in (18) is:

$$
\begin{equation*}
T_{1}=1536 \sqrt{\pi}\left(\frac{1}{4}\right)^{n} \frac{p}{q} \tag{19}
\end{equation*}
$$

where

$$
\begin{aligned}
p= & \Gamma\left(n+\frac{3}{4}\right) \Gamma(n+1) \Gamma\left(n+\frac{5}{4}\right) \Gamma\left(n+\frac{3}{2}\right) \times \\
& \Gamma\left(n+1-\frac{\sqrt{3}}{3} i\right) \Gamma\left(n+1+\frac{\sqrt{3}}{3} i\right) \\
q= & \Gamma\left(n-\frac{3}{4}\right) \Gamma\left(n-\frac{1}{2}\right) \Gamma\left(n-\frac{1}{4}\right) \Gamma(n) \Gamma(n+2) \times \\
& \Gamma\left(n+\frac{5}{2}\right) \Gamma\left(n-\frac{\sqrt{3}}{3} i\right) \Gamma\left(n+\frac{\sqrt{3}}{3} i\right) .
\end{aligned}
$$

The four efficient representations of the hypergeometric term $T_{1}$ in (19) based on the four RCF's of its certificate are:

$$
\begin{gathered}
1536 \sqrt{\pi}\left(\frac{1}{4}\right)^{n} \frac{\left(n^{2}+\frac{1}{3}\right)\left(n-\frac{3}{4}\right)\left(n-\frac{1}{2}\right)\left(n-\frac{1}{4}\right) n\left(n+\frac{1}{4}\right)\left(n+\frac{1}{2}\right)}{\Gamma(n+2) \Gamma\left(n+\frac{5}{2}\right)}, \\
1536 \sqrt{\pi}\left(\frac{1}{4}\right)^{n} \frac{\left(n^{2}+\frac{1}{3}\right)\left(n-\frac{3}{4}\right)\left(n-\frac{1}{4}\right)\left(n+\frac{1}{4}\right)}{(n+1)\left(n+\frac{3}{2}\right) \Gamma\left(n-\frac{1}{2}\right) \Gamma(n)}, \\
1536 \sqrt{\pi}\left(\frac{1}{4}\right)^{n} \frac{\left(n^{2}+\frac{1}{3}\right)\left(n-\frac{3}{4}\right)\left(n-\frac{1}{4}\right) n\left(n+\frac{1}{4}\right)}{\left(n+\frac{3}{2}\right) \Gamma\left(n-\frac{1}{2}\right) \Gamma(n+2)}, \\
1536 \sqrt{\pi}\left(\frac{1}{4}\right)^{n} \frac{\left(n^{2}+\frac{1}{3}\right)\left(n-\frac{3}{4}\right)\left(n-\frac{1}{4}\right)\left(n+\frac{1}{4}\right)}{(n+1)\left(n+\frac{3}{2}\right) \Gamma\left(n-\frac{1}{2}\right) \Gamma(n)} .
\end{gathered}
$$

## Acknowledgement

The authors wish to thank the anonymous referees for their very careful reading of the paper.

## 6. REFERENCES

[1] S.A. Abramov, M. Bronstein. Hypergeometric dispersion and the orbit problem. Proc. ISSAC 2000, ACM Press (2000) 8-13.
[2] S.A. Abramov, M. Petkovšek. Rational normal forms and minimal representation of hypergeometric terms, J. Symb. Comp. 33 (2002) 521-543.
[3] R.M. Karp, S.-Y.R. Li, Two special cases of the assignment problem. Discrete Math. 13 (1975) 129-142.
[4] C. H. Papadimitriou, K. Steiglitz. Combinatorial Optimization: Algorithms and Complexity. Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1982.
[5] M. Petkovšek, H. S. Wilf, D. Zeilberger. $A=B$. A K Peters, 1996.


[^0]:    *Partially supported by the French-Russian Lyapunov Institute under grant 98-03.
    ${ }^{\dagger}$ Partially supported by Natural Sciences and Engineering Research Council of Canada Grant No. CRD215442-98.
    ${ }^{\ddagger}$ Supported in part by MŠZŠ RS under grant P0-0511-0101.

