CANADIAN THESES

THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ MICROFILMÉE TELLE QUE NOUS L'AVONS REÇUE

The University of Alberta

MODULAR PADE FORMS FOR BIVARIATE POWER SERIES

by

Piotr Kossowski

A thesis
submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree
of Doctor of Philosophy

Department of Computing Science

Edmonton, Alberta
Fall, 1986

ISBN 0-[illegible]-32489-9

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Piotr Kossowski

TITLE OF THESIS: Modular Padé Forms for Bivariate Power Series

DEGREE FOR WHICH THIS THESIS WAS PRESENTED: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: 1986

(Signed) P. Kossowski

Permanent Address:
Department of Computing Science
University of Alberta
Edmonton, Alberta
Canada T6G 2H1

Dated 8 October 1986

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled **Modular Padé Forms for Bivariate Power Series** submitted by **Piotr Kossowski** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

....

Supervisor

....

....

....

Date ....

ABSTRACT

For a bivariate power series $A(z,y)$ with coefficients over a field, the notion of a modular Padé form is introduced. A modular Padé form is a bivariate rational expression $P(z,y)/Q(z,y)$ where $P(z,y)$ is a polynomial of degree k and m in z and y, respectively, and $Q(z,y)$ is a polynomial of degree k and n in z and y, respectively, such that coefficients of the terms $z^i y^j$ for $0 \le i \le k$, $0 \le j \le m+n$ in the power series $A(z,y)Q(z,y) + P(z,y)$ are zero.

Modular Padé forms for a power series $A(z,y)$ always exist, but are never unique. For any given k, m and n, a full characterization of all modular Padé forms is obtained and expressed as a linear combination of one or of k fundamental solutions of a triangular block Hankel system.

An algorithm is developed for constructing modular Padé forms of type (k,m_i,n_i) along an off-diagonal path $m_i - n_i = m - n$. Using classical arithmetic the cost of the algorithm, when $A(0,y)$ is normal, is $O(k^2(m+n)n)$ operations in the field of coefficients. When $A(0,y)$ is abnormal the complexity of the algorithm increases according to the nature of abnormality.

Also developed is a new algorithm for constructing Padé fractions for univariate power series with coefficients over an arbitrary integral domain. The algorithm is built around a generalized notion of power series pseudo-division and power series remainder sequences for formal power series and is $O(m+n)$ faster than fraction-free methods when the power series is normal. When applied to polynomials, rather than to power series, the algorithm, for one specific off-diagonal path, corresponds to Euclid's extended algorithm for computing greatest common divisors of polynomials.

Acknowledgements

I would like to thank my advisor, Stan Cabay for all the time, help and patience that he has given me so freely over the last two years. I am most grateful to him for making me believe that I can do it. He has been both an advisor and a very good friend. Much of this work has been developed in a course of our many lengthy discussions and often heated arguments. I benefited a great deal from his profound knowledge of the area, especially his expertise in the theory of Padé fractions. Many results in this thesis owe its origins to his ideas, in particular, the results of Chapter 4.

I also wish to give thanks to the members of my examining committee; Dr. John Tartar, Dr. Bruce Allison, Dr. Wlodek Dobosiewicz and Dr. Keith Geddes. Their involvement is greatly appreciated. I would especially like to thank Dr. Geddes for providing critical review on such short notice.

My appreciation goes also to Karen Kwiatkowski for being such an excellent typist.

This dissertation is dedicated to my mother who wanted so much for me to obtain this degree.

Table of Contents

Chapter 1

Introduction

The central problem of Padé theory for univariate functions over a field D is that of finding a rational function

$$\frac{P_{mn}(y)}{Q_{mn}(y)} = \frac{\sum_{j=0}^{m} p_j y^j}{\sum_{j=0}^{n} q_j y^j} \tag{1.1}$$

whose Maclaurin expansion agrees with a given power series

$$A(y) = \sum_{j=0}^{\infty} a_j y^j \tag{1.2}$$

as far as possible. The foundations for the Padé theory were laid in Cauchy's (1821) "Cours d'Analyse", and certain algorithmic aspects of the problem were studied (Frobenius 1881) sometime before Padé finally gave it recognition in his doctoral dissertation. Since then the problem has been studied in great depth, and the properties of Padé approximants are now well researched. Indepth surveys of the algebraic and analytical aspects of Padé theory can be found in Gragg [20] and Baker [2], for example.

In a more convenient way, the problem can be restated as one of finding polynomials $P_{mn}(y)$ and $Q_{mn}(y)$ satisfying the order condition

$$A(y)Q_{mn}(y) - P_{mn}(y) = O(y^{m+n+1}), \tag{1.3}$$

where the symbol $O(y^{m+n+1})$ denotes an arbitrary power series beginning with the power y^{m+n+1}.

The pair of polynomials $P_{mn}(y)$, $Q_{mn}(y)$ satisfying (1.3) is often called a Padé form and is known to always exists (Gragg [20]). A Padé fraction refers to the same pair of polynomials with the additional requirement that they are relatively prime; they are unique, but may not exist. Scaled Padé fractions, which always exist and are unique were proposed by Cabay and Choi [9]. The power series $A(y)$ is said to be normal if for any pair (m,n) the order condition is exact. Padé fractions for normal power series always exist and can be conveniently expressed in a determinant form known as resultants, Gragg [20].

In modern times, Padé approximants are a useful tool in many applications. They are used for example, in quantum field theory (Bessis [4]), in electrical engineering (because of their close relationship to continued fractions), in electrical network problems (Sobhy [28]), in signal processing (Chisholm [11]), in digital filtering (Chui et al. [12]) and Brophy and Salazar [5]) and in numerical analysis (Watson [29]), just to name a few. This popularity of Padé approximants in such a wide range of applications inspired attempts to generalize them to multivariate functions and, in particular, to functions of two variables.

Unfortunately, a simple statement of the univariate problem does not have its counterpart for multivariate power series, because agreement "as far as possible" is now ambiguous and the order condition (1.3) must be expressed in a much more elaborate way.

Borrowing with some simplifications the formulation of Levin [26], the order condition can be expressed in terms of finite index sets I_P, I_Q and I_E. Given a bivariate power series

$$A(x,y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} {}_i a_j x^i y^j, \tag{1.4}$$

required are polynomials

$$P(x,y) = \sum_{(i,j)\in I_P} {}_i p_j x^i y^j \tag{1.5}$$

and

$$Q(x,y) = \sum_{(i,j)\in I_Q} {}_i q_j x^i y^j, \tag{1.6}$$

such that

$${}_i r_j = 0 \;\text{ for }\; (i,j) \in I_E, \tag{1.7}$$

where $\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} {}_i r_j \, x^i y^j = A(x,y) \cdot Q(x,y) - P(x,y)$.

Various definitions can be given depending on the explicit description of the index sets. For any such definition, the order condition (1.7) gives rise to linear equations in the coefficients of $Q(x,y)$. Providing that I_P is a subset of I_E, any solution for $Q(x,y)$ uniquely determines the coefficients of $P(x,y)$. It is said that approximants $P(x,y)$, $Q(x,y)$ are non-degenerate if solutions to this system are unique. A power series is normal (with respect to a particular definition of approximants) if all approximants are non-degenerate. Levin [26] has shown that non-degenerate approximants can be expressed as resultants (i.e., in the form of a determinant).

One of the first and best known definitions of a bivariate approximant is given by Chisholm [10]. Much theory based on Chisholm's definition has been developed by a group of researchers at the University of Kent, known as the Canterbury group ([14], [21], [22] and [23]). Chisholm arrived at his definition by requesting that approximants have some convenient properties (e.g., symmetry and invariance under certain transformations) and also that the sets of indices have convenient geometric structures. Chisholm approximants, also known as diagonal approximants, since the degrees of $P(x,y)$ and $Q(x,y)$ in both variables are set equal to some integer m, are such that $I_P = I_Q = \{(i,j): 0 \leq i,j \leq m\}$ and $I_E = \{(i,j): 0 \leq i+j \leq 2m\}$. In addition, in order to arrive at the right number of equations, it is postulated that the terms ${}_i r_{2m+1-i}$ and ${}_{2m+1-i} r_i$ of the remainder power series should add to 0, for $i = 1, \ldots, m$.

Hughes Jones and Makinson [23] described a prong method requiring $O(m^4)$ operations in the field of coefficients, of solving the block system of linear equations generated by a Chisholm approximant. This method solves the system "block by block" under the assumption that each block is non-singular (i.e., under the assumption that the approximants are non-degenerate). In the same paper, they have shown that, in order for Chisholm approximants to be non-degenerate, it is necessary that the univariate approximants to $A(x,0)$ and $A(0,y)$ be non-degenerate.

Graves-Morris, et al. [21] generalize Chisholm's definition to off-diagonal approximants, for which I_P is not necessarily equal to I_Q.

Approximants defined by the sets $I_P = \{(i,j): mn \leq i+j \leq mn+m\}$, $I_Q = \{(i,j): mn \leq i+j \leq mn+n\}$ and $I_E = \{(i,j): mn \leq i+j \leq mn+m+n\}$ are proposed and studied by Cuyt ([15] and [18]). Cuyt's approximants are quite restrictive, because the polynomials $P(x,y)$ and $Q(x,y)$ are not allowed to contain any terms $x^i y^j$ for which $i+j < mn$. She has reformulated the quotient-difference and the epsilon algorithms for her type of approximants, and also has shown their relationship to the theory of multivariate continued fractions (Cuyt [16], [17]). These generalized QD and

ϵ algorithms, as in univariate case, are applicable only to normal power series.

Furthermore, both algorithms introduce extraneous common factors in $P(z,y)$ and $Q(z,y)$ which, without removal, grow exponentially with m and n. The removal of common factors at every step reduces costs significantly (the complexity of the resulting algorithms is $O((m+n)^8)$ operations in D), but not to the extent that the algorithms become practical, except for small m and n. Thus, as in the univariate case, these generalized QD and ϵ algorithms are useful primarily for finding approximants for the power series $A(z,y)$ evaluated specifically at $z=1$ and $y=1$ (i.e., for accelerating the convergence of sequences).

In terms of the index sets, the modular Padé forms introduced in this thesis satisfy (1.5), (1.6) and (1.7) for $I_P = \{(i,j): 0 \le i \le k, 0 \le j \le m\}$, $I_Q = \{(i,j): 0 \le i \le k, 0 \le j \le n\}$ and $I_E = \{(i,j): 0 \le i \le k, 0 \le j \le m+n\}$. They are called modular, because $A(z,y)$, $P(z,y)$ and $Q(z,y)$ in (1.4), (1.5) and (1.6) can be viewed, respectively, as being a univariate power series $A(y)$ and univariate polynomials $P(y)$ and $Q(y)$ in y with coefficients that are polynomials in z. With I_E defined above, the order condition (1.7) for the bivariate problem becomes the order condition (1.3) for the univariate problem modulo z^{k+1}.

Modular Padé forms lack symmetry (i.e., the indeterminate z cannot be exchanged for y), a property which is often desirable in applications and which most but not all of the previous definitions possess. However, in compensation (and primarily because of our definition), we are able to achieve many results which are not available using the other definitions.

For modular Padé forms, the order condition (1.7) leads to a block triangular Hankel system of equations. Because the system is well-structured, in chapter 2, we are able to give a full characterization of all solutions in terms of one or of k fundamental solutions. From these results, in chapter 3, we infer a full characterization of modular Padé forms for a bivariate power series.

For $k=0$, the characterization obtained is the same as the well-known characterization of Padé forms for univariate power series (see, Gragg [20], for example). Such characterizations are not available for other definitions of Padé approximants (typically, results are given for normal power series only).

Modular Padé forms can be computed directly by solving for $Q(x,y)$ the relevant block triangular Hankel system of equations. The methods of Bultheed [8], Hughes Jones and Makinson [13] and Wax and Kailath [30], designed for solving a more general block Hankel system, can each be adapted for solving block triangular Hankel systems in $O(k^2n^2)$ operations in D. The numerator $P(x,y)$ can then be determined by an additional $O(k^2mn)$ operations in D. Unfortunately, all methods impose certain normality conditions on $A(x,y)$.

In chapter 5, the univariate algorithm of Cabay and Choi [9] is generalized to compute modular Padé forms. The algorithm is not restricted by any normality condition on $A(x,y)$. When the power series $A(0,y)$ is normal the cost of the algorithm is $0(k^2(m+n)n)$, assuming classical arithmetic is used. When the power series $A(0,y)$ is abnormal, the cost may increase according to the nature of abnormality, but it does not fail as do the other methods (for finding other bivariate Padé approximants). In addition, should fast methods be applicable the algorithm can achieve a complexity of

$O(k(m+n)\log k \log^2(m+n))$.

To overcome abnormalities, the algorithm MPADE, given in chapter 5, for computing modular Padé forms relies on another algorithm JPADE, which is developed in chapter 4. The algorithm, JPADE, also a generalization of the algorithm of Cabay and Choi, computes scaled Padé fractions for univariate power series over an arbitrary integral domain J, rather than a field. This is achieved by generalizing to power series the notions of polynomial pseudo-division and polynomial remainder sequences. As for polynomial remainder sequences, it becomes necessary to remove common factors from the coefficients of power series remainder sequences in order to avoid exponential growth of coefficients. The factors to be removed are determined by extending many results of Collins [13] and Brown [7] from polynomial resultants and polynomial remainder sequences to power series resultants and power series remainder sequences. When the power series is not too abnormal, the algorithm JPADE is $O(m+n)$ faster than the algorithm of Geddes [19] based on fraction-free methods for computing Padé fractions for power series over J.

Chapter 2

Hankel Systems

In this chapter, the nature of the solutions to a certain class of block Hankel systems is described. Beginning with the statement of the problem in section 2.1, in section 2.2 and section 2.3 conditions for the existence of a family of "fundamental solutions" are given. This family is used in section 2.4 and section 2.5 as a bases for the characterization of solutions to block Hankel systems in general.

2.1 Preliminaries

We begin with some preliminary definitions and the statement of the problem. For $k,t \geq 0$, let

$$ {}_kV_t = \left\{ Q : Q = \begin{bmatrix} {}_kQ \\ \cdot \\ \cdot \\ {}_0Q \end{bmatrix}, \ \textit{where} \quad {}_iQ = \begin{bmatrix} {}_iq_t \\ \cdot \\ \cdot \\ {}_iq_0 \end{bmatrix} \ \text{for} \ i=0,\dots,k \right\} \tag{2.1} $$

be a vector space of dimension $(k+1)(t+1)$. It is assumed that the components ${}_iq_j$, $i=0,\dots,k$, $j=0,\dots,t$ are elements from a field D.

Let l be a non-negative integer. For any $Q \in {}_kV_t$, define

$$ z^l \cdot Q = [\,{}_kQ, \cdots, {}_0Q, \mathbf{0}, \cdots, \mathbf{0}]^* \in {}_{k+l}V_t \tag{2.2} $$

and

$$ z^{l} \cdot Q = [\,{}_{k-l}Q, \cdots, {}_0Q, \mathbf{0}, \cdots, \mathbf{0}]^* \in {}_kV_t, \tag{2.3} $$

where $\mathbf{0}$ is the zero vector of length $t+1$. For any $Q \in {}_kV_t$, also define

$$ y^l \cdot Q = [y^l \cdot {}_kQ, \cdots, y^l \cdot {}_0Q]^* \in {}_kV_{t+l} \tag{2.4} $$

where, for each $i = 0,\ldots,k$,

$$y^l\cdot_i Q = [_iq_t, \cdots, _iq_0, 0, \cdots, 0]^\sim \in {}_0V_{t+l}.$$

Similarly, for any $Q \in {}_kV_t$, define

$$\bar{y}^l\cdot Q = [\bar{y}^l\cdot_k Q, \cdots, \bar{y}^l\cdot_0 Q]^\sim \in {}_kV_{t+l}, \tag{2.5}$$

where, for each $i = 0,\ldots,k$,

$$\bar{y}^l\cdot_i Q = [0, \cdots, 0, _iq_t, \cdots, _iq_0]^\sim \in {}_0V_{t+l}.$$

Associated with definitions (2.2), (2.3), (2.4) and (2.5) are the following mappings defined in the obvious way:

$$\begin{aligned} x^l &: {}_kV_t \to {}_{k+l}V_t \\ \mathbf{x}^l &: {}_kV_t \to {}_kV_t \\ y^l &: {}_kV_t \to {}_kV_{t+l} \\ \bar{y}^l &: {}_kV_t \to {}_kV_{t+l}. \end{aligned} \tag{2.6}$$

It is easy to see that each of the mappings in (2.6) is linear, and that x^l, y^l and $\bar{y}^l$ are injective. For notational convenience, we also adopt the convention that $x = x^1$, $\mathbf{x} = \mathbf{x}^1$, $y = y^1$ and $\bar{y} = \bar{y}^1$. Later, in Lemma 3.1, an isomorphism is established between the mappings x^l and y^l and multiplication of polynomials by the polynomials x^l and y^l, respectively.

Let S be a subspace of ${}_kV_t$. Then the image of S by the transformation x^l is denoted by

$$x^l\cdot S = \{x^l\cdot Q : Q \in S\}. \tag{2.7}$$

Clearly, $x^l\cdot S$ is a subspace of ${}_{k+l}V_t$. A notation similar to (2.7) can be adopted for the other transformations in (2.6), but such is not required in what follows.

For $i=0,\ldots,k$, let

$$ {}_iH_{s,t} = \begin{bmatrix} {}_ih_0 & \cdots & {}_ih_t \\ \vdots & \iddots & \vdots \\ {}_ih_s & \cdots & {}_ih_{s+t} \end{bmatrix} \tag{2.8} $$

be a generalized Hankel matrix with components ${}_ih_j \in D$, $j=0,\ldots,s+t$. The objective of chapter 2 is to characterize the solutions of the block Hankel system

$$ {}_k\mathbf{H}_{s,t}\cdot Q = 0, \tag{2.9} $$

where ${}_k\mathbf{H}_{s,t}$ is of the triangular form

$$ {}_k\mathbf{H}_{s,t} = \begin{bmatrix} & & {}_0H_{s,t} \\ & \iddots & \vdots \\ {}_0H_{s,t} & \cdots & {}_kH_{s,t} \end{bmatrix}. \tag{2.10} $$

Denote the rank of ${}_k\mathbf{H}_{s,t}$ by ${}_kr_{s,t}$ and define ${}_{-1}r_{s,t} = 0$.

The ultimate objective for purposes of chapter 3 is to characterize solutions of the system

$$ {}_k\mathbf{H}_{t-1,t}\cdot Q = 0. \tag{2.11} $$

Consideration is given to the more general system (2.9) primarily to facilitate developments of results for the case $s = t-1$.

2.2 Existence of Solution

Denote the space of solutions of (2.9) by

$$ {}_kS_{s,t} = \left\{ Q : {}_k\mathbf{H}_{s,t}\cdot Q = 0 \right\}. \tag{2.12} $$

Clearly, ${}_kS_{s,t}$ is a subspace of ${}_kV_t$. A sufficient but not a necessary condition for ${}_kS_{s,t}$ to be non-trivial is given by

Theorem 2.1. If $s \le t-1$, then a non-trivial solution Q to equation (2.9) always exists.

Proof: Since ${}_k\mathbf{H}_{s,t}$ has $(k+1)(s+1)$ rows, then

$$ {}_kr_{s,t} \le (k+1)(s+1). $$

Since there are $(k+1)(t+1)$ unknowns in equation (2.9), then

$$ \begin{aligned} dim\ ({}_kS_{s,t}) &= (k+1)(t+1) - {}_kr_s. \\ &\ge (k+1)(t-s) \\ &\ge k+1. \end{aligned} \tag{2.13} $$

The inequality (2.13) provides that for $t > s$ (and in particular for $s = t-1$), equation (2.9) has at least $k+1$ linearly independent solutions. The next few results are concerned with the nature of these solutions. Corresponding to (2.7), for $i = 0,1,\ldots,k$, define

$$ z^i \cdot {}_{k-i}S_{s,t} = \left\{ z^i \cdot Q : Q \in {}_{k-i}S_{s,t} \right\}. \tag{2.14} $$

Lemma 2.2. For $i = 0,\ldots,k$, $z^i \cdot {}_{k-i}S_{s,t}$ is a subspace of ${}_kS_{s,t}$.

Proof: Assume that $Q' \in z^i \cdot {}_{k-i}S_{s,t}$. Then, by the definition of the set $z^i \cdot {}_{k-i}S_{s,t}$, there exists

$$ Q = [\,{}_{k-i}Q, \cdots, {}_0Q]^* \in {}_{k-i}S_{s,t}, $$

such that

$$Q' = z^{i} \cdot Q.$$

Thus,

$$Q' = [\,{}_{k-i}Q, \cdots, {}_{0}Q, 0, \cdots, 0]^{*}$$

and clearly

$${}_{k}\mathbf{H}_{s,t} \cdot Q' = 0.$$

Therefore,

$$z^{i} \cdot {}_{k-i}S_{s,t} \subset {}_{k}S_{s,t}.$$

The result now follows since $z^{i} \cdot {}_{k-i}S_{s,t}$ is a subspace of ${}_{k}V_{t}$. ∎

Definition 2.3. Q is a **fundamental** solution of (2.9) if $Q \in {}_{k}S_{s,t}$ and $Q \notin z \cdot {}_{k-1}S_{s,t}$. ∎

Lemma 2.4. If

$$Q = [\,{}_{k}Q, \cdots, {}_{0}Q]^{*} \in {}_{k}S_{s,t},$$

then Q is a fundamental solution in ${}_{k}S_{s,t}$ if and only if ${}_{0}Q \neq 0$.

Proof: Assume that Q is a fundamental solution in ${}_{k}S_{s,t}$, and suppose that ${}_{0}Q = 0$. Let

$$Q' = [\,{}_{k}Q, \cdots, {}_{1}Q]^{*}.$$

Then $Q = z \cdot Q'$ and $Q' \in {}_{k-1}S_{s,t}$. Thus, $Q \in z \cdot {}_{k-1}S_{s,t}$. This contradicts the assumption that Q is a fundamental solution in ${}_{k}S_{s,t}$.

Conversely, assume that $Q \in {}_{k}S_{s,t}$ and that Q is not a fundamental solution.

Then, by Definition 2.3, $Q \in z \cdot {}_{k-1}S_{s,t}$, and consequently, there exists $Q' \in {}_{k-1}S_{s,t}$ such that $Q = z \cdot Q'$. Thus,

$$[{}_kQ, \cdots, {}_1Q, {}_0Q]^* = [{}_{k-1}Q', \cdots, {}_0Q', 0]^*,$$

from which it follows that ${}_0Q = 0$. Therefore, if ${}_0Q \neq 0$, then Q must be a fundamental solution in ${}_kS_{s,t}$. ∎

Corollary 2.5. If

$$Q = [{}_kQ, \cdots, {}_0Q]^* \in {}_kS_{s,t},$$

then Q is a fundamental solution in ${}_kS_{s,t}$ iff

$$Q' = [{}_{k-1}Q, \cdots, {}_0Q]^*$$

is a fundamental solution in ${}_{k-1}S_{s,t}$.

Proof: Since $Q \in {}_kS_{s,t}$, then Q satisfies equation (2.9). From (2.9), it also follows that $Q' \in {}_{k-1}S_{s,t}$. Thus, by Lemma 2.4, ${}_0Q \neq 0$ iff Q is a fundamental solution in ${}_kS_{s,t}$ and Q' is a fundamental solution in ${}_{k-1}S_{s,t}$. ∎

Theorem 2.6. If $s \leq t-1$, then a fundamental solution in ${}_kS_{s,t}$ of equation (2.9) always exists.

Proof: We shall show that

$$dim\ ({}_kS_{s,t}) > dim\ (z \cdot {}_{k-1}S_{s,t}), \tag{2.15}$$

from which it then follows that there exists at least one $Q \in {}_kS_{s,t}$ such that $Q \notin z \cdot {}_{k-1}S_{s,t}$. Since the mapping z in (2.6) is injective, then

$$dim\ (z \cdot {}_{k-1}S_{s,t}) = dim({}_{k-1}S_{s,t}) = k(t+1) - {}_{k-1}r_{s,t}. \tag{2.16}$$

Thus,

$$\begin{aligned} dim\ ({}_kS_{s,t}) - dim(z \cdot {}_{k-1}S_{s,t}) &= [(k+1)(t+1) - {}_kr_{s,t}] - [k(t+1) - {}_{k-1}r_{s,t}] \\ &= t+1 - ({}_kr_{s,t} - {}_{k-1}r_{s,t}) \\ &\geq t-s. \end{aligned} \tag{2.17}$$

In the last inequality, we have used the fact that

$${}_kr_{s,t} - {}_{k-1}r_{s,t} \leq s+1,$$

since ${}_k\mathbf{H}_{s,t}$ has s+1 more rows than ${}_{k-1}\mathbf{H}_{s,t}$.

2.3 Quotient Spaces

Directly from the definition of a fundamental solution, it follows that the space ${}_kS_{s,t}$ of solutions to Equation (2.9) is composed of (1) fundamental solutions in ${}_kS_{s,t}$ and (2) solutions contained in the subspace $z \cdot {}_{k-1}S_{s,t}$. The same observation can be made about the solution spaces ${}_{k-i}S_{s,t}$, $i = 1,\ldots,k$. Consequently, fundamental solutions in each ${}_{k-i}S_{s,t}$, $i = 0,\ldots,k$ describe the entire solution space ${}_kS_{s,t}$. Unfortunately, fundamental solutions are not a convenient concept to work with, because fundamental solutions in ${}_kS_{s,t}$ form a set and not a vector space.

As a remedy, we introduce the vector space of quotients ${}_k\mathbf{F}_{s,t}$ of ${}_kS_{s,t}$ with respect to the subspace $z \cdot {}_{k-1}S_{s,t}$, namely,

$$\begin{aligned} {}_k\mathbf{F}_{s,t} &= {}_kS_{s,t} \,/\, z \cdot {}_{k-1}S_{s,t} \\ &= \{Q + z \cdot {}_{k-1}S_{s,t} : Q \in {}_kS_{s,t}\},\ k \geq 1. \end{aligned} \tag{2.18}$$

For $k=0$, we define trivially

$$ {}_0\mathbf{F}_{s,t} = {}_0S_{s,t}. \tag{2.19}$$

Clearly, fundamental solutions in ${}_kS_{s,t}$ are representatives of non-zero cosets of ${}_k\mathbf{F}_{s,t}$, and conversely. Thus, representatives of the cosets of any basis for ${}_k\mathbf{F}_{s,t}$ together with a basis for $z \cdot {}_{k-1}S_{s,t}$ constitute a basis for ${}_kS_{s,t}$. Consequently, the problem of constructing a basis for ${}_kS_{s,t}$ (i.e., of characterizing the space of solutions of (2.9)) is reduced to the problem of constructing a basis for each of the quotient spaces ${}_{k-i}\mathbf{F}_{s,t}$, $i=0,\dots,k$. Since

$$ dim\ (z \cdot {}_{k-1}S_{s,t}) = dim\ ({}_{k-1}S_{s,t}), \tag{2.20}$$

then

$$\begin{aligned} dim\ ({}_k\mathbf{F}_{s,t}) &= dim\ ({}_kS_{s,t}) - dim\ ({}_{k-1}S_{s,t}) \\ &= t+1 - ({}_kr_{s,t} - {}_{k-1}r_{s,t}). \end{aligned} \tag{2.21}$$

The next lemma is crucial in the following sections.

Lemma 2.7. For $k \geq 1$, then

$$ dim\ ({}_k\mathbf{F}_{s,t}) \leq dim\ ({}_{k-1}\mathbf{F}_{s,t}). \tag{2.22}$$

Proof: For

$$ Q = [{}_kQ, \cdots, {}_0Q]^* \in {}_kS_{s,t}, $$

define

$$ T(Q) = [{}_{k-1}Q, \cdots, {}_0Q]^* $$

Clearly, T is a linear transformation

$$T : {}_iS_{s,t} \to {}_{i-1}S_{s,t}.$$

By Corollary 2.5, T has the property that $Q + z \cdot {}_{k-1}S_{s,t}$ is the zero coset in ${}_k\mathbf{F}_{s,t}$ iff $T(Q) + z \cdot {}_{k-2}S_{s,t}$ is the zero coset in ${}_{k-1}\mathbf{F}_{s,t}$. Therefore (c.f., Marcus [27]), the mapping $\mathbf{T} : {}_k\mathbf{F}_{s,t} \to {}_{k-1}\mathbf{F}_{s,t}$ induced by T is a monomorphism, that is,

$$\mathbf{T}(Q + z \cdot {}_{k-1}S_{s,t}) = T(Q) + z \cdot {}_{k-2}S_{s,t}.$$

As a consequence,

$$dim\ ({}_k\mathbf{F}_{s,t}) = dim(\mathbf{T}({}_k\mathbf{F}_{s,t})) \leq dim\ ({}_{k-1}\mathbf{F}_{s,t}).$$

■

Corollary 2.8. For $k \geq 1$

$${}_kr_{s,t} - {}_{k-1}r_{s,t} \leq {}_{k+1}r_{s,t} - {}_kr_{s,t}. \tag{2.23}$$

Proof: From Equation (2.21) and Lemma 2.7,

$$t+1 - ({}_{k+1}r_{s,t} - {}_kr_{s,t}) \leq t+1 - ({}_kr_{s,t} - {}_{k-1}r_{s,t}),$$

and (2.23) now follows.

■

From (2.22), it follows that the dimensions of the quotient spaces ${}_k\mathbf{F}_{s,t}$ cannot increase with k, that is,

$$dim\ ({}_0\mathbf{F}_{s,t}) \geq dim\ ({}_1\mathbf{F}_{s,t}) \geq dim\ ({}_2\mathbf{F}_{s,t}) \geq \cdots.$$

Thus, if there are no fundamental solutions in ${}_iS_{s,t}$ (i.e., if $dim\ ({}_i\mathbf{F}_{s,t}) = 0$), then there are no fundamental solutions in ${}_jS_{s,t}$, $i \leq j \leq k$.

When fundamental solutions do exist, we wish to distinguish between two cases by means of

Definition 2.9. For any i, $0 \leq i \leq k$, the matrix ${}_k\mathbf{H}_{s,t}$ is **i-maximal** if $dim\,({}_i\mathbf{F}_{s,t}) = 1$, and **i-nonmaximal** if $dim\,({}_i\mathbf{F}_{s,t}) > 1$. ∎

Thus, from (2.21), ${}_k\mathbf{H}_{s,t}$ is i-maximal if and only if

$$ {}_i r_{s,t} - {}_{i-1} r_{s,t} = t. \tag{2.24}$$

Corollary 2.10. If ${}_k\mathbf{H}_{s,t}$ is i-nonmaximal for some i, $1 \leq i \leq k$, then it is $(i-1)$-nonmaximal.

Proof: The result follows immediately from Lemma 2.7. ∎

In section 2.4, the notion of k-maximality is used to provide a condition for the uniqueness (in a certain sense) of solutions to equation (2.9). When ${}_k\mathbf{H}_{s,t}$ is k-nonmaximal, in section 2.5, we show that solutions of equation (2.9) can be expressed in terms of "unique" solutions of other systems for which the Hankel matrix is k-maximal.

2.4 Uniqueness

If ${}_k\mathbf{H}_{s,t}$ is k-maximal, then by Definition 2.9 there exists a coset

$$ \mathbf{Q}^* = Q^* + z \cdot {}_{k-1}S_{s,t}, \tag{2.25}$$

unique up to a multiplicative constant, which constitutes a basis for ${}_k\mathbf{F}_{s,t}$. The representative vector Q^* in (2.25) is a fundamental solution in ${}_kS_{s,t}$. In addition, if $Q \in {}_kS_{s,t}$ is any other solution of equation (2.9), then there exists a scalar α such that

$$ Q = \alpha Q^* + Q', $$

where

$$Q' \in z \cdot {}_{k-1}S_{s,t}.$$

The same observation applies inductively to the solution spaces ${}_{k-i}S_{s,t}$, $i=1,\ldots,k$, by which a basis for the entire space ${}_kS_{s,t}$ can be built. A stronger result is given by

Theorem 2.11. If ${}_i\mathbf{H}_{s,t}$ is i-maximal for $i=0,\ldots,k$, then there exists $Q^* \in {}_kS_{s,t}$ such that

$$\{Q^*, z\cdot Q^*, \cdots, z^k\cdot Q^*\} \tag{2.26}$$

is a basis for ${}_kS_{s,t}$.

Proof: Let the unique non-zero coset in ${}_k\mathbf{F}_{s,t}$ be given by (2.25), where

$$Q^* = [{}_kQ^*, \cdots, {}_0Q^*]^*$$

is a fundamental solution in ${}_kS_{s,t}$. Then by Corollary 2.5, $[{}_iQ^*, \cdots, {}_0Q^*]^*$ is a fundamental solution in ${}_iS_{s,t}$, $i=0,\ldots,k$. By induction, we now show that a basis for ${}_kS_{s,t}$ is given by

$$\{z^{k-j}\cdot[{}_jQ^*, \cdots, {}_0Q^*]^*\},\ j=0,\ldots,k. \tag{2.27}$$

The theorem shall then follow from (2.27), because

$$z^{k-j}\cdot Q^* = z^{k-j}\cdot[{}_jQ^*, \cdots, {}_0Q^*]^*$$

Since ${}_0Q^*$ is a fundamental solution in ${}_0S_{s,t}$, and

$$dim\ ({}_0S_{s,t}) = dim\ ({}_0\mathbf{F}_{s,t}) = 1,$$

then a basis for ${}_0S_{s,t}$ is composed of the single vector ${}_0Q^*$.

Now assume that a basis for ${}_iS_{s,t}$ is given by

$$\{z^{i-j}\cdot[{}_jQ^*, \cdots {}_0Q^*]^*\},\ j=0,\ldots,i. \tag{2.28}$$

Since ${}_kH_{s,t}$ is $(i+1)$-maximal (i.e., $dim\ ({}_{i+1}F_{s,t}) = 1$) and

$$[{}_{i+1}Q^*, \cdots {}_0Q^*]^* \tag{2.29}$$

is a fundamental solution in ${}_{i+1}S_{s,t}$, then the unique non-zero coset in ${}_{i+1}F_{s,t}$ is

$$[{}_{i+1}Q^*, \cdots {}_0Q^*]^* + z\cdot{}_iS_{s,t}.$$

Therefore, a basis for ${}_{i+1}S_{s,t}$ is obtained by appending the representative vector (2.29) to a basis for $z\cdot{}_iS_{s,t}$. However, using the inductive hypothesis (2.28), a basis for $z\cdot{}_iS_{s,t}$ is given by

$$\{z^{i+1-j}\cdot[{}_jQ^*, \cdots {}_0Q^*]^*\},\ j=0,\ldots,i. \tag{2.30}$$

The vector (2.29), together with (2.30), yield the required basis for ${}_{i+1}S_{s,t}$. ∎

Corollary 2.12. Let ${}_kH_{t-1,t}$ be 0-maximal. If Q is a solution of equation (2.11), then there exist scalars α_i, $i=0,\ldots,k$, such that

$$Q = \sum_{i=0}^{k} \alpha_i z^i \cdot Q^*. \tag{2.31}$$

where Q^* is a fundamental solution in ${}_kS_{t-1,t}$.

Proof: By Theorem 2.6, a fundamental solution in ${}_kS_{t-1,t}$ always exists, and consequently

$$dim\ ({}_kF_{t-1,t}) \geq 1.$$

Furthermore,

$$dim\ ({}_0\mathbf{F}_{t-1,t}) = 1,$$

since ${}_kH_{t-1,t}$ is 0-maximal. Thus, by Lemma 2.7

$$dim\ ({}_i\mathbf{F}_{t-1,t}) = 1,\ i=0,\ldots,k.$$

Equation (2.31) now follows from Theorem 2.11. ∎

2.5 Characterization of Solutions

The construction of a basis for ${}_k\mathbf{F}_{s,t}$ is significantly more complex when ${}_k\mathbf{H}_{s,t}$ is k-nonmaximal, since now $dim\ ({}_k\mathbf{F}_{s,t})>1$. The objective of this section is to construct a representative of a single coset in ${}_i\mathbf{F}_{s,t}$ (only under certain constraints on s and t) which generates a basis for ${}_i\mathbf{F}_{s,t}$. Then, the representatives of the $k+1$ generators of the quotient spaces ${}_i\mathbf{F}_{s,t}$, $i=0,\ldots,k$, yield a basis for the solution space ${}_kS_{s,t}$ of Equation (2.9). We begin with a number of preliminary lemmas.

Lemma 2.13. If $Q \in {}_kS_{s+1,t}$, then $y\cdot Q$, $\overline{y}\cdot Q \in {}_kS_{s,t+1}$.

Proof: The result follows from the definition of y and $\overline{y}$, and a very careful comparison of the matrices ${}_k\mathbf{H}_{s+1,t}$ and ${}_k\mathbf{H}_{s,t+1}$. ∎

Lemma 2.14. Let $Q \in {}_kS_{s+1,t}$. Then the following statements are equivalent:

(1) Q is a fundamental solution in ${}_kS_{s+1,t}$.

(2) $y\cdot Q$ is a fundamental solution in ${}_kS_{s,t+1}$.

(3) $\overline{y}\cdot Q$ is a fundamental solution in ${}_kS_{s,t+1}$.

Proof: Since $Q \in {}_kS_{s+1,t}$, then by Lemma 2.13, $y\cdot Q \in {}_kS_{s,t+1}$. But, ${}_0Q = [{}_0q_t, \cdots, {}_0q_0]^T = 0$ if and only if $y\cdot{}_0Q = [{}_0q_t, \cdots, {}_0q_0, 0]^T = 0$. Thus, by

Lemma 2.4, (1) and (2) are equivalent.

Statements (1) and (3) can be shown to be equivalent in a similar fashion. ■

Lemma 2.15. Let $Q \in {}_kS_{s+1,t}$. Then the following statements are equivalent:

(1) $Q + x \cdot {}_{k-1}S_{s+1,t}$ is the zero coset in ${}_k\mathbf{F}_{s+1,t}$.

(2) $y \cdot Q + x \cdot {}_{k-1}S_{s,t+1}$ is the zero coset in ${}_k\mathbf{F}_{s,t+1}$.

(3) $\bar{y} \cdot Q + x \cdot {}_{k-1}S_{s,t+1}$ is the zero coset in ${}_k\mathbf{F}_{s,t+1}$.

Proof: The [illegible] are a direct consequence of Lemma 2.14 and the definition of a fundamental sol[illegible] ■

Define the mappings

$$\mathbf{y} : {}_k\mathbf{F}_{s+1,t} \rightarrow {}_k\mathbf{F}_{s,t+1} \tag{2.32}$$

$$\bar{\mathbf{y}} : {}_k\mathbf{F}_{s+1,t} \rightarrow {}_k\mathbf{F}_{s,t+1} \tag{2.33}$$

as follows: For an arbitrary coset

$$\mathbf{Q} = Q + x \cdot {}_{k-1}S_{s+1,t} \in {}_k\mathbf{F}_{s+1,t}, \tag{2.34}$$

where $Q \in {}_kS_{s+1,t}$ is a representative, define

$$\mathbf{y} \cdot \mathbf{Q} = y \cdot Q + x \cdot {}_{k-1}S_{s,t+1} \in {}_k\mathbf{F}_{s,t+1} \tag{2.35}$$

$$\bar{\mathbf{y}} \cdot \mathbf{Q} = \bar{y} \cdot Q + x \cdot {}_{k-1}S_{s,t+1} \in {}_k\mathbf{F}_{s,t+1}. \tag{2.36}$$

This definition is shown to be unambiguous in

Lemma 2.16. The mappings $\mathbf{y}$ and $\bar{\mathbf{y}}$ are monomorphisms of ${}_k\mathbf{F}_{s+1,t}$ into ${}_k\mathbf{F}_{s,t+1}$.

Proof: We give a proof for $\mathbf{y}$ only. The result for $\overline{\mathbf{y}}$ follows in a similar fashion.

We first show that $\mathbf{y}$ is well defined (i.e., $\mathbf{y}$ does not depend on the choice of representatives for cosets in ${}_k\mathbf{F}_{s+1,t}$). Suppose that Q, $Q' \in {}_kS_{s+1,t}$ are such that

$$Q + x\cdot_{k-1}S_{s+1,t} = Q' + x\cdot_{k-1}S_{s+1,t}. \tag{2.37}$$

Then, $Q - Q' \in x\cdot_{k-1}S_{s+1,t}$ and consequently $Q - Q' + x\cdot_{k-1}S_{s+1,t}$ is the zero coset in ${}_k\mathbf{F}_{s+1,t}$. Thus, by Corollary 2.15, $y\cdot Q - y\cdot Q' + x\cdot_{k-1}S_{s,t+1}$ is the zero coset in ${}_k\mathbf{F}_{s,t+1}$. That is,

$$y\cdot Q + x\cdot_{k-1}S_{s,t+1} = y\cdot Q' + x\cdot_{k-1}S_{s,t+1}.$$

Thus, we have shown that if (2.37) holds, then

$$\mathbf{y}\cdot(Q + x\cdot_{k-1}S_{s+1,t}) = \mathbf{y}\cdot(Q' + x\cdot_{k-1}S_{s+1,t}).$$

Next, we show that $\mathbf{y}$ is injective. For the coset

$$\mathbf{Q} = Q + x\cdot_{k-1}S_{s+1,t} \in {}_k\mathbf{F}_{s+1,t},$$

where $Q \in {}_kS_{s+1,t}$, suppose that $\mathbf{y}\cdot\mathbf{Q}$ is the zero coset in ${}_k\mathbf{F}_{s,t+1}$. That is, suppose that $y\cdot Q + x\cdot_{k-1}S_{s,t+1}$ is the zero coset in ${}_k\mathbf{F}_{s,t+1}$. Then, by Corollary 2.15, $\mathbf{Q} = Q + x\cdot_{k-1}S_{s,t}$ is the zero coset in ${}_k\mathbf{F}_{s+1,t}$.

Finally, we show that $\mathbf{y}$ is linear. Let $\mathbf{Q}$, $\mathbf{Q}' \in {}_k\mathbf{F}_{s+1,t}$. Then, there exist Q, $Q' \in {}_kS_{s+1,t}$ such that

$$\mathbf{Q} = Q + x\cdot_{k-1}S_{s+1,t}$$

and

$$\mathbf{Q}' = Q' + x\cdot_{k-1}S_{s+1,t}.$$

Then, for any scalars α and α',

$$\begin{aligned}\mathbf{y}\cdot(\alpha\mathbf{Q} + \alpha'\mathbf{Q}') &= \mathbf{y}\cdot(\alpha Q + z\cdot_{k-1}S_{s+1,t} + \alpha' Q' + z\cdot_{k-1}S_{s+1,t}) \\ &= \mathbf{y}\cdot(\alpha Q + \alpha' Q' + z\cdot_{k-1}S_{s+1,t}) \\ &= y\cdot(\alpha Q + \alpha' Q') + z\cdot_{k-1}S_{s,t+1} \\ &= (\alpha y\cdot Q + z\cdot_{k-1}S_{s,t+1}) + (\alpha' y\cdot Q' + z\cdot_{k-1}S_{s,t+1}) \\ &= \alpha(y\cdot Q + z\cdot_{k-1}S_{s,t+1}) + \alpha'(y\cdot Q' + z\cdot_{k-1}S_{s,t+1}) \\ &= \alpha\mathbf{y}\cdot\mathbf{Q} + \alpha'\mathbf{y}\cdot\mathbf{Q}'.\end{aligned}$$

Denote the images of ${}_k\mathbf{F}_{s+1,t}$ under the tranformations $\mathbf{y}$ and $\bar{\mathbf{y}}$ by $\mathbf{y}\cdot{}_k\mathbf{F}_{s+1,t}$ and $\bar{\mathbf{y}}\cdot{}_k\mathbf{F}_{s+1,t}$, respectively. We then have

Lemma 2.17. *If dim $({}_k\mathbf{F}_{s+1,t}) \geq 1$, then*

$$\mathbf{y}\cdot{}_k\mathbf{F}_{s+1,t} \neq \bar{\mathbf{y}}\cdot{}_k\mathbf{F}_{s+1,t}. \tag{2.38}$$

Proof: Define

$$FS = \{Q : {}_0Q \neq 0 \text{ and } Q \in {}_kS_{s+1,t}\} \tag{2.39}$$

to be the set of fundamental solutions in ${}_kS_{s+1,t}$. Then, FS is not empty, because dim $({}_k\mathbf{F}_{s+1,t}) > 0$. Clearly, by the definitions of y and $\bar{y}$ in (2.4) and (2.5), there exists $Q' \in FS$ such that for all $Q \in FS$

$$y\cdot{}_0Q' \neq \bar{y}\cdot{}_0Q. \tag{2.40}$$

We now show that the coset

$$y\cdot Q' + z\cdot_{k-1}S_{s,t+1} \notin \bar{\mathbf{y}}\cdot{}_k\mathbf{F}_{s+1,t}. \tag{2.41}$$

For suppose otherwise. Then, there exists $Q'' \in {}_kS_{s+1,t}$ such that

$$y\cdot Q' + z\cdot_{k-1}S_{s,t+1} = \bar{y}\cdot Q'' + z\cdot_{k-1}S_{s,t+1}.$$

Thus,

$$y \cdot Q' - \bar{y} \cdot Q'' \in z \cdot {}_{k-1}S_{s,t+1};$$

that is, $y \cdot Q' - \bar{y} \cdot Q''$ is not a fundamental solution in ${}_kS_{s,t+1}$. Consequently, by Lemma 2.4,

$$y \cdot {}_0Q' - \bar{y} \cdot {}_0Q'' = 0.$$

But, ${}_0Q' \neq 0$, because $Q' \in FS$. Thus, ${}_0Q'' \neq 0$, which implies Q'' is a fundamental solution in ${}_kS_{s+1,t}$. We have therefore found $Q'' \in FS$ such that

$$y \cdot {}_0Q' = \bar{y} \cdot {}_0Q'' = 0,$$

which violates the definition of Q' in (2.40). Thus, (2.41) is true.

On the other hand, by Lemma 2.14, $y \cdot Q'$ is a fundamental solution in ${}_kS_{s,t+1}$. Therefore,

$$y \cdot Q' + z \cdot {}_{k-1}S_{s,t+1} \in {}_k\mathbf{F}_{s,t+1} \tag{2.42}$$

is a nonzero coset in ${}_k\mathbf{F}_{s,t+1}$.

Corollary 2.18. If $dim\ ({}_k\mathbf{F}_{s+1,t}) \geq 1$, then

$$dim\ ({}_k\mathbf{F}_{s,t+1}) \geq dim\ ({}_k\mathbf{F}_{s+1,t}) + 1. \tag{2.43}$$

Proof: From (2.41) and (2.42), it follows that

$$dim\ (\bar{\mathbf{y}} \cdot {}_k\mathbf{F}_{s+1,t}) < dim\ ({}_k\mathbf{F}_{s,t+1}). \tag{2.44}$$

But, from lemma (2.16), $\bar{\mathbf{y}}$ is a monomorphism so that

$$dim\ ({}_k\mathbf{F}_{s+1,t}) = dim\ (\bar{\mathbf{y}} \cdot {}_k\mathbf{F}_{s+1,t}). \tag{2.45}$$

The result (2.43) now follows from (2.44) and (2.45).

Corollary 2.19. If $dim\ ({}_k\mathbf{F}_{s+1,t}) \geq 1$, then

$$ {}_kr_{s,t+1} - {}_{k-1}r_{s,t+1} \leq {}_kr_{s+1,t} - {}_{k-1}r_{s+1,t}. \tag{2.46}$$

Proof: From Equation (2.21),

$$dim\ ({}_kF_{s,t+1}) = t+2 - ({}_kr_{s,t+1} - {}_{k-1}r_{s,t+1}),$$

and

$$dim\ ({}_kF_{s+1,t}) = t+1 - ({}_kr_{s+1,t} - {}_{k-1}r_{s+1,t}).$$

Thus, by Lemma 2.19,

$$t+2 - ({}_kr_{s,t+1} - {}_{k-1}r_{s,t+1}) \geq t+1 - ({}_kr_{s+1,t} - {}_{k-1}r_{s+1,t}) + 1$$

and (2.46) now follows.

Corollary 2.20. If $dim\ ({}_k\mathbf{F}_{s,t}) \geq 1$, then

$$ {}_kr_{s-i,t+i} - {}_{k-1}r_{s-i,t+i} \leq {}_kr_{s,t} - {}_{k-1}r_{s,t},\ i=0,\ldots,s \tag{2.47}$$

Proof: We proceed by induction to show that (2.47) is true and in addition that

$$dim\ ({}_k\mathbf{F}_{s-i,t+1}) \geq i+1,\ i=0,\ldots,s. \tag{2.48}$$

For $i=0$, (2.47) and (2.48) hold true trivially. Now suppose (2.47) and (2.48) are valid for $i \geq 0$. Then from (2.47) and Corollary 2.19

$$\begin{aligned} {}_kr_{s-i-1,t+i+1} - {}_{k-1}r_{s-i-1,t+i+1} &\leq {}_kr_{s-i,t+i} - {}_{k-1}r_{s-i,t+i} \\ &\leq {}_kr_{s,t} - {}_{k-1}r_{s,t} \end{aligned}$$

and (2.47) is true at $i+1$. Also, (2.48) is true at $(i+1)$, because

$$\begin{aligned} dim\ ({}_k\mathbf{F}_{s-i-1,t+i+1}) &= t+i+2 - ({}_kr_{s-i-1,t+i+1} - {}_{k-1}r_{s-i-1,t+i+1}) \\ &\geq t+i+2 - ({}_kr_{s,t} - {}_{k-1}r_{s,t}) \\ &\geq i+2. \end{aligned}$$

In the last inequality, we have used the fact that

$${}_kr_{s,t} - {}_{k-1}r_{s,t} \leq t,$$

which again follows from (2.21) because

$$1 \leq dim\ ({}_k\mathbf{F}_{s,t}) = t+1 - ({}_kr_{s,t} - {}_{k-1}r_{s,t}).$$

If ${}_k\mathbf{H}_{s,t}$ is k-maximal, then $dim\ ({}_k\mathbf{F}_{s,t}) = 1$. The quotient space ${}_k\mathbf{F}_{s,t}$ is then fully characterized by a single non-zero coset in ${}_k\mathbf{F}_{s,t}$. Characterization of ${}_k\mathbf{F}_{s,t}$ when ${}_k\mathbf{H}_{s,t}$ is k-nonmaximal is accomplished by means of

Theorem 2.21. Let

$$\gamma_k = t - ({}_kr_{s,t} - {}_{k-1}r_{s,t}). \tag{2.49}$$

If

$${}_kr_{s,t} - {}_{k-1}r_{s,t} \leq \min\{s,t\}, \tag{2.50}$$

then there exists a fundamental solution $Q^{(k)} \in {}_kS_{s+\gamma_k,t-\gamma_k}$ such that

$$\{\bar{y}^{\gamma_k - j} \cdot y^j \cdot Q^{(k)} + z \cdot {}_{k-1}S_{s,t}\},\ j=0,\ldots,\gamma_k \tag{2.51}$$

forms a basis for ${}_k\mathbf{F}_{s,t}$.

Proof: From (2.21) and (2.50)

$$dim\ ({}_k\mathbf{F}_{s,t}) = t+1 - ({}_kr_{s,t} - {}_{k-1}r_{s,t}) \geq 1$$

Then, by Corollary 2.20 and equation (2.49), for $i=0,\ldots,s$,

$$ {}_kr_{s-i,t+i} - {}_{k-1}r_{s-i,t+i} \le {}_kr_{s,t} - {}_{k-1}r_{s,t} = t-\gamma_k. \tag{2.52} $$

But, from (2.49) and (2.50),

$$ 0 \le t-\gamma_k \le s. $$

Therefore, in particular, for $i = s-t+\gamma_k$, inequality (2.52) becomes

$$ {}_kr_{t-\gamma_k,s+\gamma_k} - {}_{k-1}r_{t-\gamma_k,s+\gamma_k} \le t-\gamma_k. \tag{2.53} $$

Now consider

$$ {}_kH_{s+\gamma_k,t-\gamma_k} = ({}_kH_{t-\gamma_k,s+\gamma_k})' $$

Clearly,

$$ \begin{aligned} {}_kr_{s+\gamma_k,t-\gamma_k} - {}_{k-1}r_{s+\gamma_k,t-\gamma_k} &= {}_kr_{t-\gamma_k,s+\gamma_k} - {}_{k-1}r_{t-\gamma_k,s+\gamma_k} \\ &\le t-\gamma_k, \end{aligned} \tag{2.54} $$

using (2.53). Thus,

$$ dim\ ({}_kF_{s+\gamma_k,t-\gamma_k}) = t-\gamma_k+1 - ({}_kr_{s+\gamma_k,t-\gamma_k} - {}_{k-1}r_{s+\gamma_k,t-\gamma_k}) \ge 1. $$

Corollary 2.20 can therefore be applied once again to yield

$$ \begin{aligned} t-\gamma_k &= {}_kr_{s,t} - {}_{k-1}r_s \\ &\le {}_kr_{s+\gamma_k,t-\gamma_k} - \end{aligned} \tag{2.55} $$

From (2.54), (2.55) and Coroll[illegible] follows that

$$ {}_kr_{s+i,t-i} - {}_{k-1}r_{s+i,t-i} = \text{[illegible]} \tag{2.56} $$

for all i such that $t-s-\gamma_k \le i \le \gamma_k$. Thus, from (2.21) and (2.56),

$$ \begin{aligned} dim\ ({}_kF_{s+i,t-i}) &= t-i+1 - (t-\gamma_k) \\ &= \gamma_k - i + 1, \end{aligned} \tag{2.57} $$

i such that $t-s-\gamma_k \leq i \leq \gamma_k$.

In particular, observe that

$$dim\ ({}_k\mathbf{F}_{s+\gamma_k,t-\gamma_k}) = 1. \tag{2.58}$$

Therefore, in ${}_k\mathbf{F}_{s+\gamma_k,t-\gamma_k}$, there exists a unique non-zero coset

$$\mathbf{Q}^{(k)} = Q^{(k)} + x\cdot{}_{k-1}S_{s+\gamma_k,t-\gamma_k},$$

where $Q^{(k)}$ is a fundamental solution in ${}_kS_{s+\gamma_k,t-\gamma_k}$. We now show that $\mathbf{Q}^{(k)}$ generates a basis for ${}_k\mathbf{F}_{s+i,t-i}$, for $i=\gamma_k, \ldots, t-s-\gamma_k$. That is, we show that a basis for ${}_k\mathbf{F}_{s+i,t-i}$, $i=\gamma_k, \ldots, t-s-\gamma_k$, is given by

$$\{\bar{\mathbf{y}}^{\gamma_k-i-j}\cdot\mathbf{y}^j\cdot\mathbf{Q}^{(k)}\},\ \ j=0,\ldots,\gamma_k-i. \tag{2.59}$$

We proceed by induction for decreasing values of i. For the initial step in the induction, $i=\gamma_k$, we have trivially that $\{\mathbf{Q}^{(k)}\}$ is a basis for ${}_k\mathbf{F}_{s+\gamma_k,t-\gamma_k}$. Assume now that (2.59) provides a basis for ${}_k\mathbf{F}_{s+\gamma_k,t-\gamma_k}, \cdots, {}_k\mathbf{F}_{s+i,t-i}$. It is required that we show (2.59) provides a basis for ${}_k\mathbf{F}_{s+i-1,t-i+1}$. Since

$$\{\bar{\mathbf{y}}^{\gamma_k-i-j}\cdot\mathbf{y}^j\cdot\mathbf{Q}^{(k)}\},\ \ j=0,\ldots,\gamma_k-i,$$

is a basis for ${}_k\mathbf{F}_{s+i,t-i}$, by Lemma 2.16, a basis for each of $\mathbf{y}\cdot{}_k\mathbf{F}_{s+i,t-i}$ and $\bar{\mathbf{y}}\cdot{}_k\mathbf{F}_{s+i,t-i}$ are given by

$$\{\bar{\mathbf{y}}^{\gamma_k-i-j}\cdot\mathbf{y}^{j+1}\cdot\mathbf{Q}^{(k)}\},\ \ j=0,\ldots,\gamma_k-i \tag{2.60}$$

and

$$\{\bar{\mathbf{y}}^{\gamma_k-i-j+1}\cdot\mathbf{y}^j\cdot\mathbf{Q}^{(k)}\},\ \ j=0,\ldots,\gamma_k-i, \tag{2.61}$$

respectively, since $\mathbf{y}$ and $\bar{\mathbf{y}}$ commute. The union of (2.60) and (2.61)

$$\{\mathbf{y}^{\gamma_k - i - j + 1} \cdot \mathbf{y}^j \cdot \mathbf{Q}^{(k)}\}, \quad j = 0, \ldots, \gamma_k - i + 1 \tag{2.62}$$

are all linearly independent in ${}_k\mathbf{F}_{s+i-1,t-i+1}$, according to Lemma 2.17. By (2.57),

$$dim\ ({}_k\mathbf{F}_{s+i-1,t-i+1}) = \gamma_k - i + 2,$$

and since there are $\gamma_k - i + 2$ cosets in (2.62), then (2.62) forms a basis for ${}_k\mathbf{F}_{s+i-1,t-i+1}$. The induction is therefore complete.

The theorem now follows by setting $i = 0$ in (2.59) ∎

For a given matrix ${}_k\mathbf{H}_{s,t}$, let

$${}_kr_{s,t} - {}_{k-1}r_{s,t} \leq \min\{s,t\}.$$

Then, by Corollary 2.8, it follows that

$${}_ir_{s,t} - {}_{i-1}r_{s,t} \leq \min\{s,t\}, \ i = 0, \ldots, k.$$

Thus, Theorem 2.21 is valid for each submatrix ${}_i\mathbf{H}_{s,t}$, $i = 0, \ldots, k$. We then obtain

Corollary 2.22. Let ${}_k\mathbf{H}_{s,t}$ be such that

$${}_kr_{s,t} - {}_{k-1}r_{s,t} \leq \min\{s,t\}.$$

For $i = 0, \ldots, k$, define

$$\gamma_i = t - ({}_ir_{s,t} - {}_{i-1}r_{s,t}) \tag{2.63}$$

and let $Q^{(i)}$ be a fundamental solution in ${}_iS_{s+\gamma_i,t-\gamma_i}$. If $Q \in {}_kS_{s,t}$, then there exists scalars $\alpha_{i,j}$ such that

$$Q = \sum_{i=0}^{k} \sum_{j=0}^{\gamma_i} \alpha_{i,j} z^{k-i} \cdot \bar{y}^{\gamma_i - j} \cdot y^j \cdot Q^{(i)}. \tag{2.64}$$

Proof: From Theorem 2.21, a basis for ${}_kS_{s,t}$ is composed of the union of

$$\{\bar{y}^{\gamma_k - j} \cdot y^j \cdot Q^{(k)}\}, \quad j = 0,\ldots,\gamma_k$$

and an z-shift of a basis for ${}_{k-1}S_{s,t}$. A basis for ${}_{k-1}S_{s,t}$ (and iteratively for the subspaces ${}_{k-2}S_{s,t}, \cdots, {}_0S_{s,t}$) is obtained in a similar fashion by means of Theorem 2.21, and (2.61) now follows. ∎

Theorem 2.23. Let $s \geq t-1$, and define

$$\gamma_k = t - ({}_kr_{s,t} - {}_{k-1}r_{s,t}). \tag{2.65}$$

If $\gamma_k \geq 0$, then there exists a fundamental solution $Q^{(k)} \in {}_kS_{s+\gamma_k, t-\gamma_k}$ such that

$$\{\bar{y}^{\gamma_k - j} \cdot y^j \cdot Q^{(k)} + z \cdot {}_{k-1}S_{s,t}\}, \quad j = 0,\ldots,\gamma_k \tag{2.66}$$

forms a basis for ${}_k\mathbf{F}_{s,t}$.

Proof: If $s \geq t$, then (2.65) and $\gamma_k \geq 0$ imply that

$${}_kr_{s,t} - {}_{k-1}r_{s,t} = t - \gamma_k \leq t = \min\{s,t\},$$

and the theorem follows from Theorem 2.21.

If $s = t-1$, then ${}_k\mathbf{H}_{s,t}$ has t more rows than ${}_{k-1}\mathbf{H}_{s,t}$, and consequently

$${}_kr_{s,t} - {}_{k-1}r_{s,t} \leq t. \tag{2.67}$$

If, in addition,

$${}_kr_{s,t} - {}_{k-1}r_{s,t} \leq t - 1$$

in (2.67), then again condition (2.50) is satisfied and the theorem follows from Theorem 2.21. Finally, if $s = t-1$ and

$$ {}_k r_{s,t} - {}_{k-1} r_{s,t} = t, $$

then ${}_k\mathbf{H}_{s,t}$ is k-maximal. Therefore, $\gamma_k = 0$ in (2.65) and (2.66) and $Q^{(k)}$ is a representative of the unique non-zero coset in ${}_k\mathbf{F}_{s,t}$.

Corollary 2.24. Let $s \geq t-1$, and define

$$ \gamma_i = t - ({}_i r_{s,t} - {}_{i-1} r_{s,t}), \; i = 0,\ldots,k. \tag{2.68} $$

If $\gamma_k \geq 0$, then for any $Q \in {}_k S_{s,t}$ there exists scalars $\alpha_{i,j}$ such that

$$ Q = \sum_{i=0}^{k} \sum_{j=0}^{\gamma_i} \alpha_{i,j} z^{k-i} \cdot \bar{y}^{\gamma_i - j} \cdot y^j Q^{(i)}, \tag{2.69} $$

where, for $i = 0,\ldots,k$, $Q^{(i)}$ is a fundamental solution in ${}_i S_{s+\gamma_i, t-\gamma_i}$.

Proof: By Corollary 2.8,

$$ {}_i r_{s,t} - {}_{i-1} r_{s,t} \leq {}_k r_{s,t} - {}_{k-1} r_{s,t}, \; i = 0,\ldots,k. $$

Thus, in (2.68) $\gamma_i \geq 0$ for $i = 0,\ldots,k$. From Theorem 2.3, a basis for ${}_i\mathbf{F}_{s,t}$, $i = 0,\ldots,k$ is given by (2.66) with k replaced by i. Then, Corollary 2.24 follows by arguments similar to those in the proof of Corollary 2.22.

Chapter 3

Modular Padé Forms

In this chapter bivariate power series and bivariate polynomials with coefficients from a field D are considered. A definition of a modular Padé form for a bivariate power series is given in terms of an order condition (as is usual, in such a case). A modular Padé form is a bivariate rational expression, whose coefficients are determined by solutions of triangular block Hankel systems of the type defined in chapter 2. Results from chapter 2 are applied by means of an isomorphism between vector spaces ${}_kV_l$ and vector spaces ${}_kBP_l$ of bivariate polynomials. We begin with the introduction of a suitable notation.

A bivariate power series $A(x,y)$ is a formal power series in two variables x and y, i.e. a formal expression of the form

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} {}_ia_j x^i y^j, \tag{3.1}$$

where the coefficients ${}_ia_j$ are from a field D.

For $i \geq 0$ and $j \geq 0$, an expression $O(x^i y^j)$ denotes an arbitrary bivariate power series $R(x,y)$ such that there exists a bivariate power series $R'(x,y)$ and $R(x,y) = x^i y^j R'(x,y)$. In this case it is said that $R(x,y)$ is of the order $x^i y^j$. Thus, for $i_0, j_0, \ldots, i_k, j_k \geq 0$, the expression

$$A(x,y) = O(x^{i_0} y^{j_0}) + \cdots + O(x^{i_k} y^{j_k})$$

indicates that there exist power series $R^{(0)}(x,y), \cdots, R^{(k)}(x,y)$ such that

$$A(x,y) = x^{i_0} y^{j_0} R^{(0)}(x,y) + \cdots + x^{i_k} y^{j_k} R^{(k)}(x,y).$$

Bivariate power series with a finite number of non-zero coefficients are bivariate polynomials. If $P(x,y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} {}_i p_j x^i y^j$ is a polynomial, then a minimal $k \geq 0$ and a minimal $m \geq 0$ such that

$$P(x,y) = \sum_{i=0}^{k} \sum_{j=0}^{m} {}_i p_j x^i y^j$$

are called, respectively, the degree of $P(x,y)$ in x and the degree of $P(x,y)$ in y (in symbols, $\partial_x P(x,y) = k$ and $\partial_y P(x,y) = m$). A pair (k,m) is called simply the degree of $P(x,y)$ (in symbols, $\partial P(x,y) = (k,m)$). Also, the expression $\partial P(x,y) \leq (k,m)$ is used to indicate that $\partial_x P(x,y) \leq k$ and $\partial_y P(x,y) \leq m$.

A vector space of all bivariate polynomials $Q(x,y)$ with $\partial Q(x,y) \leq (k,l)$ is denoted by ${}_kBP_l$. It is an easy observation that the vector space ${}_kV_l$ from chapter 2 and vector space ${}_kBP_l$ are isomorphic in a natural way, i.e., if $Q \in {}_kV_l$, where $Q = [{}_kQ, \cdots, {}_0Q]^*$ and ${}_iQ = [{}_iq_l, \cdots, {}_iq_0]^*$, $i = 0, \ldots, k$, then there exists a corresponding polynomial $Q(x,y) \in {}_kBP_l$, namely

$$Q(x,y) = \sum_{i=0}^{k} \sum_{j=0}^{l} {}_iq_j x^i y^j . \tag{3.2}$$

This isomorphism is denoted by $Pol_{k,l}$ and its inverse isomorphism by $Vec_{k,l}$. Thus,

$$Pol_{k,l} : {}_kV_l \to {}_kBP_l$$

$$Vec_{k,l} : {}_kBP_l \to {}_kV_l ,$$

and with Q and $Q(x,y)$ above, $Q(x,y) = Pol_{k,l}(Q)$ and $Q = Vec_{k,l}(Q(x,y))$. Moreover, the shift transformations x, $\underline{x}$, y and $\overline{y}$ from chapter 2 can be easily translated into operations on polynomials, which are given by the following

Lemma 3.1: Let $Q \in {}_kV_t$. Then

1. $Pol_{k+l,t}(z^l \cdot Q) = z^l \cdot Pol_{k,t}(Q)$,

2. $Pol_{k,t}(\underline{z}^l \cdot Q) = (z^l \cdot Pol_{k,t}(Q)) mod\ z^{k+1}$,

3. $Pol_{k,t+l}(y^l \cdot Q) = y^l \cdot Pol_{k,t}(Q)$, and

4. $Pol_{k,t+l}(\overline{y}^l \cdot Q) = Pol_{k,t}(Q)$.

Proof:

Only case 2 is not trivial, and the proof is given for this case only. Let $Q \in {}_kV_t$.

Then

$$Pol_{k,t}(Q) = \sum_{i=0}^{k} \sum_{j=0}^{t} {}_iq_j z^i y^j.$$

Thus,

$$z^l \cdot Pol_{k,t}(Q) = \sum_{i=0}^{k} \sum_{j=0}^{t} {}_iq_j z^{i+l} y^j$$

and

$$(z^l \cdot Pol_{k,t}(Q)) mod z^{k+1} = \sum_{i=0}^{k-l} \sum_{j=0}^{t} {}_iq_j z^{i+l} y^j.$$

On the other hand, by the definition of the tranformation $\underline{z}$,

$$\underline{z}^l \cdot Q = [{}_{k-l}Q, \ldots, {}_0Q, \mathbf{0}, \ldots, \mathbf{0}]^\frown = [{}_kQ', \ldots, {}_0Q']^\frown,$$

where ${}_iQ' = [{}_{i-l}q_t, \ldots, {}_{i-l}q_t]^\frown$ for $i = l, \ldots, k$, and ${}_iQ' = 0$ for $i = 0, \ldots, l-1$. Thus,

$$Pol_{k,t}(\underline{z}^l \cdot Q) = \sum_{i=l}^{k} \sum_{j=0}^{t} {}_{i-l}q_j z^i y^j = \sum_{i=0}^{k-l} \sum_{j=0}^{t} {}_iq_j z^{i+l} y^j. \qquad \blacksquare$$

Let the bivariate power series $A(z,y)$ and non-negative integers k, m and n be given.

Definition 3.2. A bivariate rational expression $P(z,y)/Q(z,y)$ is called a modular Padé (k,m,n)-form for $A(z,y)$ if $\partial P(z,y) \le (k,m)$, $\partial Q(z,y) \le (k,n)$, $Q(z,y) \ne 0$ and the following order condition is satisfied

$$A(z,y)\cdot Q(z,y) + P(z,y) = O(y^{m+n+1}) + O(z^{k+1}). \tag{3.3}$$

By equating appropriate powers of z and y, it is easy to see that the polynomials

$$P(z,y) = \sum_{i=0}^{k}\sum_{j=0}^{m} {}_ip_j z^i y^j \quad \text{and} \quad Q(z,y) = \sum_{i=0}^{k}\sum_{j=0}^{n} {}_iq_j z^i y^j$$

satisfy the order condition (3.3) if and only if its coefficients satisfy the following systems of linear equations:

$$\sum_{i=0}^{k}\sum_{j=0}^{n} {}_{s-i}a_{t-j}\,{}_iq_j = 0, \qquad 0 \le s \le k,\ m+1 \le t \le m+n, \tag{3.4}$$

and

$$\sum_{i=0}^{k}\sum_{j=0}^{n} {}_{s-i}a_{t-j}\,{}_iq_j + {}_sp_t = 0, \qquad 0 \le s \le k,\ 0 \le t \le m, \tag{3.5}$$

where ${}_ia_j = 0$ if $i<0$ or $j<0$.

The systems of equations (3.4) and (3.5) can be expressed in matrix form as follows: Let ${}_k\mathbf{H}_{n-1,n}$ be the triangular block Hankel matrix defined in (2.8), with components ${}_ih_j$ determined by the coefficients of the power series $A(z,y)$, namely, for $i=0,\ldots,k$ and $j=0,\ldots,2n-1$,

$${}_ih_j = \begin{cases} {}_ia_{m-n+1+j}, & \text{if } m-n+1+j \ge 0, \\ 0, & \textit{otherwise}. \end{cases} \tag{3.6}$$

Let ${}_k\mathbf{G}_{m,n}$ be a triangular block Toeplitz matrix, such that

$$ {}_k\mathbf{G}_{m,n} = \begin{bmatrix} {}_0G & \cdots & {}_kG \\ & \ddots & \vdots \\ & & {}_0G \end{bmatrix} $$

where

$$ {}_iG = \begin{bmatrix} {}_ia_{m-n} & \cdots & {}_ia_m \\ \vdots & & \vdots \\ {}_ia_{-n} & \cdots & {}_ia_0 \end{bmatrix} \tag{3.7} $$

for $i = 0,\ldots,k$, where ${}_ia_j = 0$ if $j<0$. Then, the system (3.4) is equivalent to

$$ {}_k\mathbf{H}_{n-1,n} \cdot Q = 0, \tag{3.8} $$

and the system (3.5) is equivalent to

$$ {}_k\mathbf{G}_{m,n} \cdot Q + P = 0, \tag{3.9} $$

where $Q \in {}_kV_n$ and $P \in {}_kV_m$.

Clearly, the polynomials $P(z,y) \in {}_kBP_m$ and $Q(z,y) \in {}_kBP_n$ satisfy the order condition (3.3) if and only if the corresponding vectors $P = Vec_{k,m}(P(z,y))$ and $Q = Vec_{k,n}(Q(z,y))$ satisfy equations (3.8) and (3.9). Thus, there is a one to one correspondence between modular Padé forms and solutions to block Hankel systems which can be stated as

Lemma 3.3. $P(z,y)/Q(z,y)$ is a modular Padé (k,m,n)-form for $A(z,y)$ if and only if the corresponding vectors P and Q are solutions of systems (3.8) and (3.9). ∎

Solutions of (3.8) and (3.9) are uniquely determined by solutions of the system (3.8) alone (i.e. any solution of (3.8) can be substituted in (3.9) to calculate the vector

P). This observation together with Lemma 3.2., gives a procedure for the characterization of all modular Padé (k,m,n)-forms for $A(z,y)$. First, the family of solutions Q to the system (3.8) is determined, and then the family of solutions P is given by solving system (3.9) for P. Then $P(z,y)/Q(z,y)$, such that $P(z,y) = Pol_{k,n}(P)$ and $Q(z,y) = Pol_{k,m}(Q)$, are all modular Padé (k,m,n)-forms for $A(z,y)$.

General results from chapter 2 are applied to obtain the solution of equation (3.8), which as it was shown in Corollary 2.24 can be expressed as a linear combination of shifts of solutions of a smaller system. It will be shown that the solution P corresponding to equation (3.9) can be expressed as the same linear combination of the same shifts of solutions P corresponding to solutions of these smaller systems.

It should be clear, from the definition of matrices ${}_k\mathbf{G}_{m,n}$, that if $Q \in {}_{k-i}V_n$, $0 \le i \le k$,

$$ {}_k\mathbf{G}_{m,n}\cdot(z^i\cdot Q) = z^i\cdot({}_{k-i}\mathbf{G}_{m,n}\cdot Q), \tag{3.10} $$

and also, that if $Q \in {}_kV_n$, then

$$ {}_k\mathbf{G}_{m,n}\cdot(\underline{z}^i\cdot Q) = \underline{z}^i\cdot({}_k\mathbf{G}_{m,n}\cdot Q). \tag{3.11} $$

Solutions involving shifts y^i and $\bar{y}^i$ are more complex, and are addressed in

Lemma 3.4. If $Q \in {}_kV_{n-\gamma}$, for some γ, $0 \le \gamma \le n$, is such that $Q \in {}_kS_{n-1+\gamma,n-\gamma}$, then

$$ {}_k\mathbf{G}_{m,n}\cdot\bar{y}^{\gamma-l}\, y^l\cdot Q = \bar{y}^{\gamma-l}\cdot y^l\cdot {}_k\mathbf{G}_{m-\gamma,n-\gamma}Q, \tag{3.12} $$

for $0 \le l \le \gamma$.

Proof: Let

$$Q = [{}_kQ, \cdots, {}_0Q], \quad {}_iQ = [{}_iq_{n-\gamma}, \cdots, {}_iq_0], \; i=0,\ldots,k,$$

and define

$$P = {}_k\mathbf{G}_{m-\gamma,n-\gamma}\cdot Q,$$

where

$$P = [{}_kP, \cdots, {}_0P], \quad {}_iP = [{}_ip_{m-\gamma}, \cdots, {}_ip_0], \; i=0,\ldots,k.$$

If we set

$$P'' = \bar{y}^{\gamma-l}\cdot y^l\cdot P,$$

where

$$P'' = [{}_kP'', \cdots, {}_0P''], \quad {}_iP'' = [{}_ip''_m, \cdots, {}_ip''_0], \; i=0,\ldots,k,$$

then the r.h.s. of equality (3.12) is equal to P''.

Similarly, define

$$Q' = \bar{y}^l\cdot y^l\cdot Q,$$

where

$$Q' = [{}_kQ', \cdots, {}_0Q], \quad {}_iQ = [{}_iq'_n, \cdots, {}_iq'_0], \; i=0,\ldots,k.$$

Then the l.h.s. of equality (3.12) become

$$P' = {}_k\mathbf{G}_{m,n}\cdot Q',$$

where

$$P' = [{}_kP', \cdots, {}_0P'], \quad {}_iP' = [{}_ip'_m, \cdots, {}_ip'_0], \; i=0,\ldots,k.$$

It will be shown that $P' = P''$, by proving that ${}_{i_0}p'_{j_0} = {}_{i_0}p''_{j_0}$, for $0 \le i_0 \le k$ and $0 \le j_0 \le m$. For $0 \le i_0 \le k$, $0 \le j_0 \le j-\gamma$

$$ {}_{i_0}p_{j_0} = \sum_{i=0}^{k} \sum_{j=0}^{n-\gamma} {}_{i_0-i}a_{j_0-j}\,{}_iq_j . $$

Thus, for $0 \le i_0 \le k$, $0 \le j_0 \le m$

$$ {}_{i_0}p''_{j_0} = \begin{cases} {}_{i_0}p_{j_0-l}, & l \le j_0 \le m-\gamma-l, \\ 0, & \textit{otherwise}. \end{cases} $$

Then, for $0 \le i_0 \le k$,

$$ {}_{i_0}p''_{j_0} = \begin{cases} \displaystyle\sum_{i=0}^{k} \sum_{j=l}^{n-\gamma+1} {}_{i_0-i}a_{j_0-j}\,{}_iq_{j-l}, & l \le j_0 \le m-\gamma+l, \\ 0, & 0 \le j_0 < l, \quad \textit{or} \quad m-\gamma+l+1 \le j_0 \le m. \end{cases} $$

On the other hand, for $0 \le i_0 \le k$, $0 \le j_0 \le m$,

$$ {}_{i_0}p'_{j_0} = \sum_{i=0}^{k} \sum_{j=0}^{n} {}_{i_0-i}a_{j_0-j}\,{}_iq'_j, $$

where

$$ {}_iq'_j = \begin{cases} {}_iq_{j-l}, & l \le j \le n-\gamma+l, \\ 0, & \textit{otherwise}. \end{cases} $$

Therefore, for $0 \le i_0 \le k$, $0 \le j_0 \le m$,

$$ {}_{i_0}p'_{j_0} = \begin{cases} \displaystyle\sum_{i=0}^{k} \sum_{j=l}^{n-\gamma+l} {}_{i_0-i}a_{j_0-j}\,{}_iq_{j-l}, & l \le j_0 \le m, \\ 0, & 0 \le j_0 < l. \end{cases} $$

It remains to show that for $0 \le i_0 \le k$, $m-\gamma+l+1 \le j_0 \le m$

$$\sum_{i=0}^{k} \sum_{j=l}^{n-\gamma+l} {}_{i_0-i}a_{j_0-j,i}q_{j-l} = 0.$$

From the assumption that $Q \in {}_kS_{n-1+\gamma,n-\gamma}$, it follows that for $0 \le i_0 \le k$, $m-\gamma+1 \le j_0 \le m+n$,

$$\sum_{i=0}^{k} \sum_{j=0}^{n-\gamma} {}_{i_0-i}a_{j_0-j,i}q_j = 0,$$

which is equivalent to

$$\sum_{i=0}^{k} \sum_{j=l}^{n-\gamma+l} {}_{i_0-i}a_{j_0-j,i}q_{j+l} = 0,$$

for $0 \le i_0 \le k$ and $m-\gamma+1+l \le j_0 \le m+n+l$.

Definition 3.5 The power series $A(z,y)$ is (i,m,n)-maximal if the matrix ${}_k\mathbf{H}_{n-1,n}$ is i-maximal.

By definition, the matrix ${}_k\mathbf{H}_{n-1,n}$ for a $(0,m,n)$-maximal power series, is 0-maximal. Thus, by Corollary 2.12, there exists a single fundamental solution Q^*. Let $P^* = - {}_k\mathbf{G}_{m,n} Q^*$, be the corresponding solution to the system (3.9), and let $P^*(z,y)/Q^*(z,y)$ be the corresponding modular Padé (k,m,n)-form.

The next theorem shows that if $A(z,y)$ is $(0,m,n)$-maximal, then all modular Padé (k,m,n)-forms can be characterized in terms of a single modular Padé (k,m,n)-form.

Theorem 3.6. All modular Padé (k,m,n)-forms for a $(0,m,n)$-maximal power series $A(z,y)$ are of the form $P(z,y)$, $Q(z,y)$, where

$$P(z,y) = (U(z)P^*(z,y)) \bmod z^{k+1}$$
$$Q(z,y) = (U(z)Q^*(z,y)) \bmod z^{k+1},$$

and $U(z)$ is an arbitrary polynomial in z.

Proof: Let $P(z,y) = Pol_{k,m}(P)$ and $Q(z,y) = Pol_{k,n}(Q)$ where P and Q are solutions to (3.8) and (3.9). By Corollary 2.12

$$Q = \sum_{i=0}^{k} \alpha_i \mathbf{z}^i \cdot Q^*,$$

for some $\alpha_0, \cdots, \alpha_k \in D$. Thus

$$P = -{}_k\mathbf{G}_{m,n} \cdot Q = \sum_{i=0}^{k} \alpha_i ({}_k\mathbf{G}_{m,n} \mathbf{z}^i \cdot Q^*)).$$

But, from (3.11),

$${}_k\mathbf{G}_{m,n} \cdot (\mathbf{z}^i \cdot Q^*) = \mathbf{z}^i \cdot ({}_k\mathbf{G}_{m,n} \cdot Q^*),$$

and therefore,

$$P = \sum_{i=0}^{k} \alpha_i \mathbf{z}^i \cdot P^*.$$

Let $U(z) = \sum_{i=0}^{k} \alpha_i z^i$. From Lemma 3.1, it follows that

$$Pol_{k,m}(P) = (U(z)\, P^*(z,y)) \bmod z^{k+1} \text{ and}$$
$$Pol_{k,n}(Q) = (U(z)\, Q^*(z,y)) \bmod z^{k+1}.$$

■

The above theorem characterizes modular Padé (k,m,n)-forms in a special case, when $A(z,y)$ is $(0,m,n)$-maximal. Full characterization is given below for the general case.

Given $A(z,y)$, k, m and n, let $Q^{(i)}$, $i=0,\ldots,k$, be the fundamental solutions as given in Corollary 2.24. Thus, the $Q^{(i)}$'s are solutions to a system

$$ {}_i\mathbf{H}_{n-1+\gamma_i,n-\gamma_i} \cdot Q = 0. $$

Therefore, they are also solutions to a smaller system of the form (3.8), i.e.,

$$ {}_i\mathbf{H}_{n-1-\gamma_i,n-\gamma_i} \cdot Q = 0. $$

For $i=0,\ldots,k$, let $P^{(i)}$ be defined as the corresponding solution of the form (3.9), i.e.,

$$ P^{(i)} = -{}_i\mathbf{G}_{m-\gamma_i,n-\gamma_i} \cdot Q. $$

For $i=0,\ldots,k$, let

$$ P^{(i)}(z,y) = Pol_{i,m-\gamma_i}(P^{(i)}) $$

and

$$ Q^{(i)}(z,y) = Pol_{i,n-\gamma_i}(Q^{(i)}). $$

Thus, by Lemma 3.3, $P^{(i)}(z,y)/Q^{(i)}(z,y)$ are modular Padé $(i,m-\gamma_i,n-\gamma_i)$-forms for $A(z,y)$ where $i=0,\ldots,k$.

The following theorem shows that any modular Padé (k,m,n)-form can be expressed as a function of modular Padé $(i,m-\gamma_i,n-\gamma_i)$-forms, $i=0,\ldots,k$.

Theorem 3.7. All modular Padé (k,m,n)-forms for $A(z,y)$ are of the form $P(z,y)/Q(z,y)$, where

$$ P(z,y) = \sum_{i=0}^{k} \sum_{j=0}^{\gamma_i} \alpha_{i,j} z^{k-i} y^j P^{(i)}(z,y), \quad Q(z,y) = \sum_{i=0}^{k} \sum_{j=0}^{\gamma_i} \alpha_{i,j} z^{k-i} y^j \, Q^{(i)}(z,y) $$

and $\alpha_{i,j}$, $i=0,\ldots,k$, $j=0,\ldots,\gamma_i$ are arbitrary scalars.

Proof: Let $P(z,y) = Pol_{k,m}(P)$ and $Q(z,y) = Pol_{k,n}(Q)$, where P and Q are arbitrary solutions to (3.8) and (3.9). By Corollary 2.24,

$$Q = \sum_{i=0}^{k} \sum_{j=0}^{\gamma_i} \alpha_{i,j} z^{k-i} \cdot \overline{y}^{\gamma_i - j} \cdot y^j \cdot Q^{(i)}.$$

Thus,

$$P = -\sum_{i=0}^{k} \sum_{j=0}^{\gamma_i} \alpha_{i,j\,k}\mathbf{G}_{m,n} \cdot (z^{k-i} \cdot \overline{y}^{\gamma_i - j} \cdot y^j \cdot Q^{(i)}).$$

From (3.10), it follows that

$${}_k\mathbf{G}_{m,n} \cdot (z^{k-i} \cdot \overline{y}^{\gamma_i - j} \cdot y^j \cdot Q^{(i)}) = z^{k-i} \cdot {}_i\mathbf{G}_{m,n} \cdot (\overline{y}^{\gamma_i - j} \cdot y^j \cdot Q^{(i)}).$$

Since $Q^{(i)} \in {}_kS_{n-1+\gamma_i, n-\gamma_i}$, $i = 0, \ldots, k$, by Lemma 3.4, it follows that

$${}_i\mathbf{G}_{m,n} \overline{y}^{\gamma_i - j} \cdot y^j \cdot Q^{(i)} = \overline{y}^{\gamma_i - j} \cdot y^j \cdot {}_i\mathbf{G}_{m-\gamma_i, n-\gamma_i} Q^{(i)}.$$

But, ${}_iG_{m-\gamma_i, n-\gamma_i} Q^{(i)} = -P^{(i)}$, and consequently,

$$P = \sum_{i=0}^{k} \sum_{j=0}^{\gamma_i} \alpha_{i,j} z^{k-i} \cdot \overline{y}^{\gamma_i - j} \cdot y^j \cdot P^{(i)}.$$

An application of Lemma 3.1 to the vectors P and Q gives

$$Pol_{k,m}(P) = \sum_{i=0}^{k} \sum_{j=0}^{\gamma_i} \alpha_{i,j} z^{k-i} y^j Pol_{i,m-\gamma_i}(P^{(i)})$$

and

$$Pol_{k,n}(Q) = \sum_{i=0}^{k} \sum_{j=0}^{\gamma_i} \alpha_{i,j} z^{k-i} y^j Pol_{i,n-\gamma_i}(Q^{(i)}).$$

Thus,

$$P(z,y) = \sum_{i=0}^{k} \sum_{j=0}^{\gamma_i} \alpha_{i,j} z^{k-i} y^j P^{(i)}(z,y)$$

and

$$Q(x,y) = \sum_{i=0}^{k} \sum_{j=0}^{\gamma_i} \alpha_{i,j} x^{k-i} y^j Q^{(i)}(x,y).$$

■

Chapter 4

Padé Fractions over an Integral Domain

In this chapter, we turn our attention to the construction of Padé fractions for univariate power series with coefficients over an arbitrary integral domain. The algorithm developed for this construction is used in Chapter 5 to obtain modular Padé forms for a bivariate power series.

4.1 Preliminaries

Let $J[[y]]$ denote the set of all power series in y with coefficients lying in an integral domain J. We are interested primarily in the case $J = D[[x]]$, the set of all power series in x with coefficients in the field D. The set $J[[y]]$ itself comprises an integral domain. Denote the integral domain of all polynomials in $J[[y]]$ by $J[y]$, and the quotient field of $J[y]$ by $J(y)$ (rational functions in y over J). Some of the development in this chapter requires an embedding of $J[[y]]$ in a larger domain $F_J[[y]]$, where F_J is the quotient field of J. We then introduce also the domains $F_J[y]$ and $F_J(y)$.

Proceeding as in Geddes [19], and borrowing from Cabay and Choi [9], we introduce

Definition 4.1. Given two power series

$$A(y) = \sum_{j=0}^{\infty} a_j y^j \in J[[y]] \tag{4.1}$$

and

$$B(y) = \sum_{j=0}^{\infty} b_j y^j \in J[[y]] \tag{4.2}$$

the rational function $U_m(y)/V_n(y) \in J(y)$, where $V_n(y) \neq 0$, is the scaled Padé fraction of type (m,n) for the pair $\{A(y), B(y)\}$ if

(1) $\min\{m - \partial U_m(y),\ n - \partial V_n(y)\} = 0$ (4.3)

(2) $GCD\ (U_m(y),\ V_n(y)) = hy^{\lambda}$, for *some integer* $\lambda \geq 0$, and $h \in J$, and (4.4)

(3) $A(y) \cdot V_n(y) + B(y) \cdot U_m(y) = 0(y^{m+n+1})$ (4.5)

Theorem 4.2. Scaled Padé fractions exist and are unique up to a multiplicative constant in J.

Proof: Cabay and Choi show existence and uniqueness of scaled Padé fractions when J is a field. By embedding $J[[y]]$ in $F_J[[y]]$, it follows that there exists uniquely, up to a multiplicative constant from F_J, a scaled Padé fraction $U''_m(y)/V''_y(y) \in F_J(y)$. For appropriate $f, g \in J$,

$$U''(y) = fU''(y) \in J[y]$$

$$V''(y) = gV''(y) \in J[y]$$

and $U''(y)$, $V''(y)$ still satisfy (4.3), (4.5) and (4.4) for some $h \in J$. ■

Since there are numerous good algorithms for computing Padé fractions when the domain of coefficients is a field, these algorithms can be applied also to power series in $J[[y]]$ by first embedding them in $F_J[[y]]$. The results obtained in $F_J(y)$ can then be converted to results in $J(y)$ in a manner indicated in the proof of Theorem 4.2

Unfortunately, operations in F_J result in costly algorithms since large intermediate growth of coefficients prevail, unless common factors are removed at each step, a costly process. Geddes provides a method for finding Padé fractions where all operations are in J. In this chapter, the algorithm of Cabay and Choi for computation of scaled Padé fractions is united with the algorithms of Collins [13] and Brown [7] for computing greatest common divisors of polynomials in $J[y]$. The algorithm developed performs all operations in J and is superior to the fraction-free algorithm described by Geddes. The new algorithm is used in a fundamental way in Chapter 5 to design a fast method for computing modular Padé forms for bivariate power series.

4.2 Power Series Pseudo-Division

Let J be an integral domain and consider the power series

$$A(y) = \sum_{j=0}^{\infty} a_j y^j, \; a_0 \neq 0, \; a_j \in J, \tag{4.6}$$

and

$$B(y) = \sum_{j=0}^{\infty} b_j y^j, \; b_0 \neq 0, \; b_j \in J. \tag{4.7}$$

We are interested primarily in $J = D[[z]]$, the domain of univariate power series in z with coefficients in the field D, and in $J = D[z]$, the domain of polynomials in z with coefficients also in D.

For a given integer $s \geq 0$, consider the system of equations

$$\begin{bmatrix} & & b_0 \\ & \cdot & \vdots \\ b_0 & \cdots & b_s \end{bmatrix} \begin{bmatrix} \omega_s \\ \vdots \\ \omega_0 \end{bmatrix} = -b_0^{s+1} \begin{bmatrix} a_0 \\ \vdots \\ a_s \end{bmatrix}, \tag{4.8}$$

Because the right-hand-side of (4.8) includes the term b_0^{s+1}, Cramer's rule can be used to show the existence of a unique solution $[\omega_0, \ldots, \omega_s]^{\hat{}}$, where $\omega_j \in J$, $j = 0, \ldots, s$.

Let

$$\Omega(y) = \sum_{j=0}^{s} \omega_j y^j,$$

where $[\omega_0, \ldots, \omega_s]^{\hat{}}$ is the unique solution of (4.8), and let

$$R'(y) = b_0^{s+1} A(y) + \Omega(y) \cdot B(y). \tag{4.9}$$

Then, from (4.8), it follows that either $R'(y) = 0$, or there exists an integer t, $t > s$, such that

$$R'(y) = y^t R(y) = y^t \sum_{j=0}^{\infty} r_j y^j, \tag{4.10}$$

where $r_0 \neq 0$. Adopting the convention that $t = \infty$ in (4.10) when $R'(y) = 0$, we have shown

Lemma 4.3. Let $A(y)$ and $B(y)$ be given by (4.6) and (4.7), respectively. For any integer $s \geq 0$, there exists a unique polynomial

$$\Omega(y) = \sum_{j=0}^{s} \omega_j y^j \tag{4.11}$$

such that

$$b_0^{s+1} A(y) + \Omega(y) \cdot B(y) = y^t R(y), \tag{4.12}$$

where $t > s$ and

$$R(y) = \sum_{j=0}^{\infty} r_j y^j, \quad r_0 \neq 0. \tag{4.13}$$

Definition 4.4: $\Omega(y)$ is called the **power series pseudo-quotient** and $R(y)$ the **power series pseudo-remainder** on "division" of $A(y)$ by $B(y)$ relative to s. A procedure for constructing $\Omega(y)$ satisfying (4.8) and $R(y)$ satisfying (4.12) is called **power series pseudo-division**.

When $A(y)$ and $B(y)$ are polynomials in y, power series pseudo-division corresponds exactly to polynomial pseudo-division (see, for example, Knuth [25]). To see this, it is only necessary to set

$$s = \partial A(y) - \partial B(y)$$

in (4.12), and take reciprocals with respect to $y^{\partial A(y)}$.

Note that (4.12) is not the only way to define power series pseudo division. It may be possible to obtain a unique solution $\omega_j \in J$, $j = 0, \ldots, s$ of (4.8) if we multiply $[a_0, \ldots, a_s]^{-}$ on the r.h.s. of (4.8) by $-b_0^l$, $l \leq s$, rather than by $-b_0^{s+1}$. If l is chosen to be minimal, the resulting power series pseudo-division has a direct anology to the polynomial pseudo-division defined by Collins [13]. We choose to define power series pseudo-division in the form (4.12) because it allows us to extend directly many of the results of Brown and Traub [6] and of Brown [7] on polynomial remainder sequences, results which are adequate for our purposes.

4.3 Power Series Remainder Sequences

Let $A(y)$ and $B(y)$ be power series defined by (4.6) and (4.7), and let m and n be nonnegative integers such that $m \geq n$. Later, m and n shall correspond to the degrees of the scaled Padé fraction $U_m(y)/V_m(y)$ in Definition 4.1. Initially set

$$\begin{aligned} s_{-1} &= 1 \qquad R_{-1}(y) = A(y) \\ s_0 &= n-m \qquad R_0(y) = B(y). \end{aligned} \tag{4.14}$$

We then have

Definition 4.5. A **power series remainder sequence (PSRS)** is the sequence

$$\{s_i, R_i(y)\}_{i=-1,0,} \tag{4.15}$$

such that

$$\alpha_{i+1} R_{i-1}(y) + \Omega_{i+1}(y) \cdot R_i(y) = y^{s_i + s_{i+1}} \beta_{i+1} R_{i+1}(y), \quad i=0,\ldots \tag{4.16}$$

To ensure that a PSRS is well defined, the following additional constraints are imposed. Let

$$c_i = r_{i,0} \tag{4.17}$$

be the leading coefficient of

$$R_{i+1}(y) = \sum_{j=0}^{\infty} r_{i,j} y^j \tag{4.18}$$

in (4.16). Assume inductively that $c_{i+1} \neq 0$, and let

$$\alpha_{i+1} = c_i^{s_i+1}. \tag{4.19}$$

Then, from Lemma 4.3, there exists a unique $\Omega_{i+1}(y)$ such that

$$\alpha_{i+1} R_{i-1}(y) + \Omega_{i+1}(y) \cdot R_i(y) = y^{s_i + s_{i+1}} R(y), \tag{4.20}$$

where $s_{i+1} \geq 1$. If $R(y)=0$ in (4.20), then $s_{i+1}=\infty$ and the PSRS (4.15) terminates. If

$R(y) \neq 0$, let $R_{i+1}(y)$ be defined by

$$R(y) = \beta_{i+1} R_{i+1}(y) = \beta_{i+1} \sum_{j=0}^{\infty} r_{i+1,j} y^j, \quad r_{i+1,0} \neq 0, \tag{4.21}$$

where $\beta_{i+1} \in J$ is some factor which can be removed from $R(y)$. In this case

$$\alpha_{i+2} = c_{i+1}^{s_{i+1}+1} = r_{i+1,0}^{s_{i+1}+1} \neq 0,$$

and the term $\{s_{i+2}, R_{i+2}(y)\}$ in the sequence (4.15) can again be computed by (4.16).

When a method for choosing β_{i+1} is given, and a termination condition is specified, equation (4.16) constitutes an algorithm for constructing a PSRS.

Initially set

$$n_{-1} = n - m - 1, \qquad m_{-1} = -1, \tag{4.22}$$

and define

$$n_{i+1} = n_i + s_i, \qquad m_{i+1} = m_i + s_i, \quad i = -1, \cdots. \tag{4.23}$$

Thus,

$$n_{i+1} = \sum_{l=-1}^{i} s_l, \qquad m_{i+1} = \sum_{l=0}^{i} s_l, \quad i = 0, \cdots. \tag{4.24}$$

We terminate the iteration (4.16) at $i = e$ (i.e., after computing α_{e+1} and $\Omega_{e+1}(y)$), whenever it is determined that

$$n_{e+1} + s_{e+1} > n, \tag{4.25}$$

or equivalently, because $s_0 = m - n$, whenever

$$m_{e+1} + s_{i+1} > m. \tag{4.26}$$

Note that the termination condition is also valid in the case that $R_{e+1}(y) = 0$ in

(4.16), since now $s_{e+1} = n_{e+2} = m_{e+2} = \infty$. Thus, the algorithm terminates with the finite PSRS

$$\{s_i, R_i\}_{i=-1,0,\ldots,e},$$

where e is the smallest integer for which

$$n_{e+2} = \sum_{l=1}^{e+1} s_l > n.$$

Initially also set

$$\begin{aligned} U_{-1}(y) &= 0, \quad V_{-1}(y) = y^{n-m-1}, \\ U_0(y) &= 1, \quad V_0(y) = 0, \end{aligned} \tag{4.27}$$

Corresponding to the PSRS (4.15), define the sequences

$$\{U_i(y)\}_{i=-1,0,\ldots}, \quad \{V_i(y)\}_{i=-1,0,\ldots}, \tag{4.28}$$

such that

$$\beta_{i+1} U_{i+1}(y) = y^{s_i + s_{i-1}} \alpha_{i+1} U_{i-1}(y) + \Omega_{i+1}(y) U_i(y) \tag{4.29}$$

and

$$\beta_{i+1} V_{i+1}(y) = y^{s_i + s_{i-1}} \alpha_{i+1} V_{i-1}(y) + \Omega_{i+1}(y) V_i(y), \tag{4.30}$$

where α_{i+1}, β_{i+1} and $\Omega_{i+1}(y)$ are determined by (4.16). Strictly speaking, the initialization (4.27) is not justified because of the negative powers of y. However, the initialization is acceptable algorithmically, since by using (4.27), $U_i(y)$ and $V_i(y)$ for $i = 1,2$ can be computed directly from (4.29) and (4.30).

As for the PSRS, we terminate the iterations (4.29) and (4.30) at $i = e$, that is, after having computed the sequences

$\{U_i(y)\}_{i=-1,0,\ldots,e+1}$, $\{V_i(y)\}_{i=-1,0,\ldots,e+1}$.

Note that $U_{e+1}(y)$ and $V_{e+1}(y)$ can be computed by means of (4.29) and (4.30) without explicit knowledge of $R_{e+1}(y)$, because α_{e+1} and $\Omega_{e+1}(y)$ in (4.16) depend only on $R_{e-1}(y)$, $R_e(y)$ and s_e, and also, in Theorem 4.11, β_{i+1} is chosen independently of s_{e+1} and $R_{e+1}(y)$.

The fractions $U_i(y)/V_i(y)$, $i \geq 1$, are shown to be scaled Padé fractions in

Theorem 4.6. The sequences (4.16), (4.29) and (4.30) satisfy

$$A(y)\cdot V_i(y) + B(y)\cdot U_i(y) = y^{m_i+n_i+s_i}R_i(y), \quad i=1,\ldots, \tag{4.31}$$

where

$$\min\{m_i - \partial U_i(y), n_i - \partial V_i(y)\} = 0. \tag{4.32}$$

In addition, GCD $\{U_i(y), V_i(y)\} = h$, for some $h \in J$.

Proof: The validity of the theorem for $i=1,2$ can be shown in the obvious manner. Inductively, assume the theorem is true for $l=1,\ldots,i$. Then

$$A(y)\cdot V_{i-1}(y) + B(y)\cdot U_{i-1}(y) = y^{m_{i-1}+n_{i-1}+s_{i-1}}R_{i-1}(y), \tag{4.33}$$

where $\partial U_{i-1}(y) \leq m_{i-1}$, $\partial V_{i-1}(y) \leq n_{i-1}$ and the leading coefficient $r_{i-1,0}$ of $R_{i-1}(y)$ is nonzero (otherwise, the iteration would terminate here). Also, we have

$$A(y)\cdot V_i(y) + B(y)\cdot U_i(y) = y^{m_i+n_i+s_i} R_i(y), \tag{4.34}$$

where $\partial U_i(y) \leq m_i$, $\partial V_i(y) \leq n_i$ and where again $r_{i,0} \neq 0$.

We now show that (4.31) and (4.32) are valid at $i+1$. First, using (4.29) and (4.30), and next the inductive hypotheses (4.33) and (4.34), and finally (4.16), we obtain

$$
\begin{aligned}
\beta_{i+1}[A(y)\cdot V_{i+1}(y) + B(y)\cdot U_{i+1}(y)] \\
&= y^{s_{i-1}+s_i}\,\alpha_{i+1}\,[A(y)\cdot V_{i-1}(y) + B(y)\cdot U_{i-1}(y)] \\
&+ \Omega_{i+1}(y)\cdot[A(y)\cdot V_i(y) + B(y)\cdot U_i(y)] \qquad (4.35) \\
&= y^{m_i+n_i+s_i}\,[\alpha_{i+1}R_{i-1}(y) + \Omega_{i+1}(y)\cdot R_i(y)] \\
&= y^{m_{i+1}+n_{i+1}+s_{i+1}}\beta_{i+1}R_{i+1}(y)
\end{aligned}
$$

The relationship (4.31) now follows after dividing (4.35) by β_{i+1}.

To show $\partial U_{i+1}(y) \leq m_{i+1}$, from (4.29), it follows that

$$
\begin{aligned}
\partial U_{i+1}(y) &\leq \max\{\partial U_{i-1}(y) + s_{i-1} + s_i,\ \partial\Omega_{i+1}(y) + \partial U_i(y)\} \\
&\leq \max\{m_{i-1} + s_{i-1} + s_i,\ s_i + m_i\} \\
&= m_{i+1}.
\end{aligned}
$$

In a similar fashion, it can be shown that $\partial V_{i+1}(y) \leq n_{i+1}$.

Let

$$d = \min\{m_{i+1} - \partial U_{i+1}(y),\ n_{i+1} - \partial V_{i+1}(y)\}$$

and

$$G(y) = GCD(y^d U_{i+1}(y),\ y^d V_{i+1}(y)).$$

We will show that $\partial G(y) = 0$. It will then follow that

$$\min\{m_{i+1} - \partial U_{i+1}(y),\ n_{i+1} - \partial V_{i+1}(y)\} = 0,$$

and that

$$GCD(U_{i+1}(y),\ V_{i+1}(y)) = h,$$

for some $h \in J$.

Let $\lambda = \partial G(y)$ and suppose $\lambda > 0$. Define

$$\overline{U}_{i+1}(y) = y^{d+\lambda-1} U_{i+1}(y)/G(y)$$

and

$$\overline{V}_{i+1}(y) = y^{d+\lambda-1} V_{i+1}(y)/G(y).$$

It is easy to show that $\overline{U}_{i+1}(y)/\overline{V}_{i+1}(y)$ is a scaled Padé fraction of type $(m_{i+1}^{*}-1, n_{i+1}-1)$ for the pair $(A(y),B(y))$. But, by the induction hypothesis, $(y^{s_i-1}U_i(y), y^{s_i-1}V_i(y))$ is also a scaled Padé fraction of type $(m_{i+1}-1, n_{i+1}-1)$ for the pair $(A(y),B(y))$. From Theorem 4.12, by the uniqueness of scaled Padé fractions, it follows that

$$\overline{U}_{i+1}(y)/\overline{V}_{i+1}(y) = U_i(y)/V_i(y),$$

or equivalently, that

$$U_{i+1}(y)V_i(y) - U_i(y)V_{i+1}(y) = 0.$$

Replacing $U_{i+1}(y)$ and $V_{i+1}(y)$ by their definitions given in (4.29) and (4.30), respectively, we obtain from the two equations above that

$$U_i(y)V_{i-1}(y) - U_{i-1}(y)V_i(y) = 0,$$

which violates the inductive hypothesis that $U_{i-1}(y)/V_{i-1}(y)$ and $U_i(y)/V_i(y)$ are distinct scaled Padé fractions. Thus, $\lambda = 0$. ■

Each scaled Padé fraction $U_i(y)/V_i(y)$, $i \geq 1$, in the sequence (4.28) lies along the off-diagonal path

$$m_i - n_i = m - n,$$

according to (4.14) and (4.24). Let $(\overline{m},\overline{n})$, where

$$\overline{m} = m_i + j$$
$$\overline{n} = n_i + j$$

for some j, $0<j<s_i$, be a point along this off-diagonal path between the points (m_i,n_i) and (m_{i+1},n_{i+1}). Then, the scaled Padé fraction $U_{\overline{m}}(y)/V_{\overline{n}}(y)$ of type $(\overline{m},\overline{n})$ for the pair $(A(y),B(y))$ is given by

$$U_{\overline{m}}(y) = y^j U_{m_i}(y)$$
$$V_{\overline{n}}(y) = y^j V_{n_i}(y).$$

In particular, to obtain the scaled Padé fraction $U_m(y)/V_m(y)$ of type (m,n) for the pair $(A(y),B(y))$, set $i = e+1$ and $j = n - n_{e+1}$. Then,

$$\overline{m} = m_i + j = m_{e+1} + n - n_{e+1} = m$$
$$\overline{n} = n_i + j = n,$$

and

$$U_m(y) = U_{\overline{m}}(y) = y^j U_{m_i}(y) = y^{m-m_{e+1}} U_{m_{e+1}}(y)$$
$$V_n(y) = V_{\overline{n}}(y) = y^j V_{n_i}(y) = y^{n-n_{e+1}} V_{n_{e+1}}(y).$$

Then, using (4.26), it follows that

$$\begin{aligned} A(y)V_n(y) &+ B(y)U_m(y) \\ &= y^{n-n_{e+1}}[A(y)V_{n_{e+1}}(y) + B(y)U_{m_{e+1}}(y)] \\ &= y^{n-n_{e+1}}[y^{m_{e+1}+n_{e+1}+s_{e+1}}R_{e+1}(y)] \\ &= 0(y^{m+n+1}). \end{aligned}$$

When J is a field, we may set $\alpha_{i+1} = 1$ and $\beta_{i+1} = 1$ in (4.16), (4.29) and (4.30). The computational steps that result correspond exactly to those of Algorithm 3 in

Cabay and Choi [9]. That is, Theorem 4.6 provides that the fraction $U_i(y)/V_i(y)$ is a scaled Padé fraction of type (m_i, n_i) for the quotient power series $A(y)/B(y)$.

When J is a unique integral domain, we have yet to specify β_{i+1} in (4.16), which in addition must divide the right-hand-sides of (4.29) and (4.30). For algorithm effectiveness, the choice of β_{i+1} is extremely critical. The obvious choice $\beta_l = 1$, $l = 1, \ldots$, typically leads to exponential growth of coefficients in the power series $R_i(y)$ and in the polynomials $U_i(y)$ and $V_i(y)$. Borrowing Brown's [7] example, let

$$R_{-1}(y) = 1 + y^2 - 3y^4 - 3y^5 + 8y^6 + 2y^7 - 5y^8 + \cdots$$
$$R_0(y) = 3 + 5y^2 - 4y^4 - 9y^5 + 21y^6 + \cdots,$$

where J is the domain of integers. Setting $s_0 = 2$ and $\beta_l = 1$, $l = 1, \ldots$, in (4.16) yields

$$R_1(y) = -15 + 3y^2 - 9y^4 + \cdots \tag{4.36}$$
$$R_2(y) = 15795 + 30375y - 59535y^2 + \cdots$$
$$R_3(y) = 1254542875143750 - 1654608338437500y + \cdots$$

When J is the domain of polynomials in x, exponential growth is encountered in terms of degrees; whereas, if J is the domain of power series in x, exponential growth is encountered in terms of the order of the power series.

Fortunately, large factors can be easily removed from the remainder sequence $R_i(y)$ and from the polynomial sequences $U_i(y)$ and $V_i(y)$, $i = 1, \ldots$. Following Brown [7], in section 4.4, we prescribe a choice of β_i, $i = 1, \ldots$ which keeps the growth linear.

4.4 Subresultants

For the power series $A(y)$ and $B(y)$ given by (4.6) and (4.7) with $b_0 \neq 0$, and for integers $j, s \geq 0$, define the resultants (determinants)

$$\hat{U}_{j,s}(A,B) = \begin{vmatrix} a_0 & & & b_0 & & \\ & \ddots & & & \ddots & \\ \vdots & & a_0 & \vdots & & b_0 \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{2j+s} & \cdots & a_{j+s} & b_{2j+s} & \cdots & b_j \\ 0 & \cdots & 0 & 1 & \cdots & y^{j+s} \end{vmatrix}, \tag{4.37}$$

$$\hat{V}_{j,s}(A,B) = \begin{vmatrix} a_0 & & & b_0 & & \\ & \ddots & & & \ddots & \\ \vdots & & a_0 & \vdots & & b_0 \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{2j+s} & \cdots & a_{j+s} & b_{2j+s} & \cdots & b_j \\ 1 & \cdots & y^j & 0 & \cdots & 0 \end{vmatrix}, \tag{4.38}$$

and

$$\hat{R}_{j,s}(A,B) = \begin{vmatrix} a_0 & & & b_0 & & \\ & \ddots & & & \ddots & \\ \vdots & & a_0 & \vdots & & b_0 \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{2j+s} & \cdots & a_{j+s} & b_{2j+s} & \cdots & b_j \\ A & \cdots & y^j A & B & \cdots & y^{j+s} B \end{vmatrix}. \tag{4.39}$$

We have the following well-known relationship on resultants.

Theorem 4.7.

$$A(y)\cdot \hat{V}_{j,s}(A,B) + B(y)\cdot \hat{U}_{j,s}(A,B)$$
$$= \hat{R}_{j,s}(A,B)$$

$$= y^{2j+s}\sum_{l=1}^{\infty} y^l \begin{vmatrix} a_0 & & & & & \\ \vdots & \ddots & & & & \\ \vdots & & a_0 & & & b_0 \\ \vdots & & \vdots & & & \vdots \\ a_{2j+s} & \cdots & a_{j+s} & b_{2j+s} & \cdots & b_j \\ a_{2j+s+l} & \cdots & a_{j+s+l} & b_{2j+s+l} & \cdots & b_{j+l} \end{vmatrix} \tag{4.40}$$

Proof: See Gragg [20], for example. ∎

The primary goal of this section is to show that the resultants $U_{j,s}(A,B)$, $V_{j,s}(A,B)$ and $R_{j,s}(A,B)$, $j \geq 0$, are either all identically zero, or for a correct choice of β_i in (4.16), (4.29) and (4.30) correspond, respectively, to $U_i(y)$, $V_i(y)$ and $R_i(y)$, for some i. To prove the main results given in Theorem 4.11 and Theorem 4.12, we use a number of lemmas.

Proceeding as in Brown and Traub [6], the first lemma considers a single step in the PSRS (4.16). To simplify notation, in (4.16) we make the substitutions

$$\begin{aligned} s_i &= s, \quad s_{i+1} = t, \\ \alpha_{i+1}R_{i-1}(y) &= A(y) = \sum_{l=0}^{\infty} a_l y^l, \quad a_0 \neq 0, \\ R_i(y) &= B(y) = \sum_{l=0}^{\infty} b_l y^l, \quad b_0 \neq 0, \\ \Omega_{i+1}(y) &= \Omega(y) = \sum_{l=0}^{s} \omega_l y^l, \\ \beta_{i+1}R_{i+1}(y) &= G(y) = \sum_{l=0}^{\infty} g_l y^l, \quad g_0 \neq 0. \end{aligned} \tag{4.41}$$

Lemma 4.8. If

$$A(y) + \Omega(y)\cdot B(y) = y^{s+t}G(y), \tag{4.42}$$

where $t \geq 1$, then

$$\hat{R}_{j,t}(A,B) = \begin{cases} (-1)^{s+1} y^{s+t} b_0^{s+1} G(y), & j=0, \\ 0, & 0<j<t-1, \\ (-1)^{(s+1)t} y^{s+2t-1} b_0^{s+t} g_0^{t-1} \; G(y), & j=t-1, \\ (-1)^{(j+1)(j+s+1)} y^{s+t} b_0^{s+t} \hat{R}_{j-t,t}(B,G), & j \geq t. \end{cases} \tag{4.43}$$

Proof: Equating coefficients in powers of y in (4.42), we obtain

$$\begin{bmatrix} a_0 & b_0 & & \\ \vdots & \vdots & \ddots & \\ & & & b_0 \\ & & & \vdots \\ a_{s+t} & b_{s+t} & & b_t \\ \vdots & & & \vdots \end{bmatrix} \begin{bmatrix} 1 \\ w_0 \\ \vdots \\ w_s \end{bmatrix} = \begin{bmatrix} g_{-s-t} \\ \vdots \\ g_{-1} \\ g_0 \\ \vdots \end{bmatrix},$$

where $g_l = 0$, if $l<0$. Thus,

$$\begin{bmatrix} \mathbf{0} & \mathbf{0} & & \mathbf{0} \\ a_0 & b_0 & & \\ & & \ddots & b_0 \\ & & & \vdots \\ a_{2j-l} & b_{2j+s-l} & \cdots & b_{2j-l} \\ y^l A & y^l B & \cdots & y^{l+s} B \end{bmatrix} \begin{bmatrix} 1 \\ w_0 \\ \vdots \\ w_s \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ g_{-s-t} \\ \vdots \\ g_{2j-t-l} \\ y^{s+t+l} G \end{bmatrix}, \tag{4.44}$$

where $\mathbf{0}$ is a column vector of length l, and where the last row follows directly from (4.42). Finally, for all l, $0 \leq l \leq j$, in (4.39), replace column $l+1$ by the r.h.s. of (4.44). Since this corresponds to adding multiples of certain columns in the matrix given in (4.39) to column l, $0 \leq l \leq j$, the determinant does not change. Consequently, (4.39) can be written as

$$\hat{R}_{j,s}(A,B) = \begin{vmatrix} g_{-s-t} & & & & b_0 & & \\ & \ddots & & & & \ddots & \\ \vdots & & g_{-s-t} & & \vdots & & b_0 \\ \vdots & & \vdots & & & & \vdots \\ g_{2j-t} & \cdots & g_{j-t} & & b_{2j+s} & \cdots & b_j \\ y^{s+t}G & \cdots & y^{s+t+j}G & & B & \cdots & y^{j+s}B \end{vmatrix}, \tag{4.45}$$

$$= (-1)^{(j+1)(j+s+1)} \begin{vmatrix} b_0 & & & g_{-s-t} & & \\ & \ddots & & & \ddots & \\ & & b_0 & g_{j-t} & & g_{-s-t} \\ \vdots & & \vdots & & & \vdots \\ b_{2j+s} & \cdots & b_j & g_{2j-t} & \cdots & g_{j-t} \\ B & \cdots & y^{j+s}B & y^{s+t}G & \cdots & y^{s+t+j}G \end{vmatrix}.$$

Setting $j = 0$ in (4.45), we obtain

$$\hat{R}_{0,s}(A,B) = (-1)^{s+1} \begin{vmatrix} b_0 & & & \\ \vdots & \ddots & & \\ b_s & \cdots & b_0 & \\ B & & y^sB & y^{s+t}G \end{vmatrix}$$

$$= (-1)^{s+1} b_0^{s+1} y^{s+t} G(y).$$

For $0 < j < t-1$ let $l = t-j-1 \geq 0$. Then (4.45) becomes are

$$\hat{R}_{j,s}(A,B) = (-1)^{(j+1)(j+s+1)} \left| \begin{array}{cccccccc} b_0 & & & & & & & \\ & \ddots & & & & & & \\ & & b_0 & & & & & \\ \vdots & & & g_{-t} & & & & \\ & & \vdots & & \ddots & & & \\ b_{2j+s} & \cdots & b_j & g_{j-t-1} & \cdots & g_{-t} & 0 & \\ B & \cdots & y^{j+s}B & y^{s+t}G & \cdots & & y^{j+s+t}G & \end{array} \right| = 0 .$$

For $j = t-1$, (4.45) becomes

$$\hat{R}_{j,s}(A,B) = (-1)^{(s+1)t} \left| \begin{array}{cccccccc} b_0 & & & & & & & \\ & \ddots & & & & & & \\ \vdots & & b_0 & & & & & \\ \vdots & & & g_0 & & & & \\ & & \vdots & & \ddots & & & \\ b_{s+2t-2} & \cdots & b_{t-1} & g_{t-2} & \cdots & g_0 & 0 & \\ B & \cdots & y^{s+t-1}B & y^{s+t}G & \cdots & & y^{s+2t-1}G & \end{array} \right|$$

$$= (-1)^{(s+1)t} y^{s+2t-1} b_0^{s+t} g_0^{t-1} G(y).$$

Finally, for $j \geq t$, minor expansions about the first $s+t$ rows in (4.45) yield

$$\hat{R}_{j,s}(A,B) = (-1)^{(j+1)(j+s+1)} b_0^{s+t} \left| \begin{array}{cccccc} b_0 & & & g_0 & & \\ & \ddots & & & \ddots & \\ \vdots & & b_0 & \vdots & & g_0 \\ \vdots & & \vdots & \vdots & & \vdots \\ b_{2j-t} & \cdots & b_j & g_{2j-t} & \cdots & g_{j-t} \\ y^{s+t}B & \cdots & y^{j+s}B & y^{s+t}G & \cdots & y^{j+s+t}G \end{array} \right|$$

$$= (-1)^{(j+1)(j+s+1)} y^{s+t} b_0^{s+t} \hat{R}_{j-t,t}(B,G).$$

Corollary 4.9. Let

$$\{s_i,\ R_i(y)\}_{i=-1,0,\dots}$$

be a power series remainder sequence determined by (4.16). Then, for $i=0,\ldots,$

$$\alpha_{i+1}\hat{R}_{0,s_i}(R_{i-1},R_i) = (-1)^{s_i+1}y^{s_i+s_{i+1}}c_i^{s_i+1}\beta_{i+1}R_{i+1}(y), \tag{4.46}$$

$$\hat{R}_{j,s_i}(R_{i-1},R_i) = 0, \quad 0<j<s_{i+1}-1, \tag{4.47}$$

$$\alpha_{i+1}^{s_{i+1}}\hat{R}_{s_{i+1}-1,s_i}(R_{i-1},R_i)$$
$$= (-1)^{s_{i+1}(s_i+s_{i+1})}y^{s_i+2s_{i+1}-1}c_i^{s_i+s_{i+1}}c_{i+1}^{s_{i+1}-1}\beta_{i+1}^{s_{i+1}}R_{i+1}(y), \tag{4.48}$$

$$\alpha_{i+1}^{j+1}\hat{R}_{j,s_i}(R_{i-1},R_i)$$
$$= (-1)^{(j+1)(j+s_i+1)}y^{s_i+s_{i+1}}c_i^{s_i+s_{i+1}}\beta_{i+1}^{j+1}\hat{R}_{j-s_{i+1}+1,s_{i+1}}(R_i,R_{i+1}), \quad j\geq s_{i+1}. \tag{4.49}$$

Proof: The corollary follows by substituting (4.41) back into (4.43) and using the identity

$$\hat{R}_{j,s}(aA,bB) = a^{j+1}b^{j+s+1}\hat{R}_{j,s}(A,B).$$

Lemma 4.10. For $i=0,\ldots,$

$$\hat{R}_{n_{i+1},s_0}(A,B)\prod_{j=1}^{i+1}\alpha_j^{n_{i+1}-n_j+1}$$
$$= y^{m_{i+1}+n_{i+1}+s_{i+1}}(-1)^{\sigma_{i+1}}R_{i+1}(y)c_i^{1-s_{i+1}}\prod_{j=1}^{i+1}c_{j-1}^{s_{j-1}+s_j}\beta_j^{n_{i+1}-n_j+1}, \tag{4.50}$$

$$\hat{R}_{j,s_0}(A,B) = 0, \quad n_{i+1} < j < n_{i+2}-1, \tag{4.51}$$

$$\hat{R}_{n_{i+2}-1,s_0}(A,B)\prod_{j=1}^{i+1}\alpha_j^{n_{i+2}-n_j}$$
$$= y^{m_{i+2}+n_{i+2}-1}(-1)^{\tau_{i+1}}R_{i+1}(y)c_{i+1}^{s_{i+1}-1}\prod_{j=1}^{i+1}c_{j-1}^{s_{j-1}+s_j}\beta_j^{n_{i+2}-n_j}, \tag{4.52}$$

where

$$\sigma_{i+1} = \sum_{j=1}^{i+1}(n_{i+1}-n_j+1)(m_{i+1}-m_{j-1}+1)$$

$$\tau_{i+1} = \sum_{j=1}^{i+1} (n_{i+2} - n_j)\ (m_{i+2} - m_{j-1}).$$

Proof: For $l \geq n_{i+1}$, an application of (4.49) i times yields

$$\begin{aligned} \hat{R}_{l,s_0}(R_{-1},R_0)\prod_{j=1}^{i} \alpha_j^{\,l-n_j+1}+1 \\ = (-1)^{\rho_i} y^{m_i+n_i+s_i} \prod_{j=1}^{i} [c_{j-1}^{s_{j-1}+s_j}\ \beta_j^{\,l-n_j+1}]\ \hat{R}_{l-n_{i+1},s_i}(R_{i-1},R_i), \end{aligned} \tag{4.53}$$

where

$$\rho_i = \sum_{j=1}^{i} (l-n_j+1)(l+s_0-m_{j-1}+1).$$

The identity (4.50) is obtained by setting $l = n_{i+1}$ in (4.53) and applying (4.46) to the result. The identity (4.51) is an immediate consequence of (4.47) and (4.53), whereas, (4.52) is obtained by setting $l = n_{i+2} - 1$ in (4.53) and using (4.48). ■

Theorem 4.11. Let

$$\begin{aligned} \beta_1 &= (-1)^{s_0+1} \\ \beta_{i+1} &= (-1)^{s_i+1}\ c_{i-1} h_{i-1}^{s_i},\quad i \geq 1, \end{aligned} \tag{4.54}$$

where

$$\begin{aligned} h_0 &= c_0^{s_0}, \\ h_i &= c_i^{s_i} h_{i-1}^{1-s_i},\quad i \geq 1. \end{aligned} \tag{4.55}$$

Then, $h_i, \beta_{i+1} \in J$, for $i = 0, \ldots,$ and

$$\hat{R}_{n_{i+1},s_0}(A,B) = y^{m_{i+1}+n_{i+1}+s_{i+1}} R_{i+1}(y), \tag{4.56}$$

$$\hat{R}_{l,s_0}(A,B) = 0,\quad n_{i+1} < l < n_{i+2} - 1, \tag{4.57}$$

$$c_{i+1}\hat{R}_{n_{i+2}-1,s_0}(A,B) = y^{m_{i+2}+n_{i+2}-1}h_{i+1}R_{i+1}(y). \tag{4.58}$$

Proof: By induction, we first show that

$$h_i = (-1)^{s_i+1}c_i^{-1}\prod_{j=1}^{i+1}(\alpha_j/\beta_j). \tag{4.59}$$

Initially,

$$h_0 = (-1)^{s_0+1}c_0^{-1}\ (\alpha_1/\beta_1) = c_0^{s_0}.$$

Assume that (4.59) is true for $l=0,\ldots,i$. Then, using (4.55),

$$\begin{aligned} h_{i+1} &= (-1)^{s_{i+1}+1}c_{i+1}^{-1}\prod_{j=1}^{i+2}(\alpha_j/\beta_j) \\ &= (-1)^{s_{i+1}+1}c_{i+1}^{-1}(\alpha_{i+2}/\beta_{i+2})c_i\left[c_i^{-1}\prod_{j=1}^{i+1}(\alpha_j/\beta_j)\right] \\ &= (-1)^{s_{i+1}+1}c_{i+1}^{s_{i+1}}\ c_i/\beta_{i+2}\ h_i \\ &= c_{i+1}^{s_{i+1}}h_i^{1-s_{i+1}}, \end{aligned}$$

and, from (4.55), the induction is complete. Observe also that by using (4.50) and (4.55)

$$\begin{aligned} h_{i-1}c_{i-1}\prod_{j=1}^{i}(\beta_i/\alpha_i) &= c_{i-1}^{s_{i-1}+1}h_{i-2}^{1-s_{i-1}}\ (\beta_i/\alpha_i)\prod_{j=1}^{i-1}(\beta_j/\alpha_j) \\ &= (-1)^{s_{i-1}+1}\left[h_{i-2}c_{i-2}\prod_{j=1}^{i-1}(\beta_j/\alpha_j)\right]. \end{aligned}$$

Repeating the above process inductively, we obtain

$$\begin{aligned} h_{i-1}c_{i-1}\prod_{j=1}^{i}(\beta_j/\alpha_j) &= \prod_{j=0}^{i-1}(-1)^{s_j+1} \\ &= (-1)^{m_i+1}. \end{aligned} \tag{4.60}$$

Now, let

$$\theta_{i+1} = (-1)^{\sigma_{i+1}} c_i^{1-\epsilon_{i+1}} \prod_{j=1}^{i+1} c_{j-1}^{\epsilon_{j-1}+\epsilon_j} \cdot \prod_{j=1}^{i+1} (\beta_j/\alpha_j)^{n_{i+1}-n_j+1}, \tag{4.61}$$

so that the identity (4.50) becomes

$$\hat{R}_{n_{i+1},\epsilon_0}(A,B) = y^{m_{i+1}+n_{i+1}+\epsilon_{i+1}}\ \theta_{i+1} R_{i+1}(y). \tag{4.62}$$

Using the definition of β_{i+1} given in (4.54), we will show that $\theta_{i+1} = 1$, $i=0, \ldots$. Trivially, for $i=0$,

$$\begin{aligned}
\theta_1 &= (-1)^{\sigma_1} c_0^{\epsilon_0+1} \beta_1/\alpha_1 \\
&= (-1)^{\sigma_1+\epsilon_0+1} \\
&= 1.
\end{aligned}$$

Suppose that $\theta_j = 1$, $j=1, \ldots, i$. We now show by induction that $\theta_{i+1} = 1$. From (4.50), (4.60) and (4.61),

$$\begin{aligned}
\theta_{i+1}/\theta_i &= (-1)^{\sigma_{i+1}-\sigma_i} c_{i-1}^{\epsilon_i-1}\ \beta_{i+1} \left[\prod_{j=1}^{i} (\beta_j/\alpha_j)\right]^{\epsilon_i} \\
&= (-1)^{\sigma_{i+1}-\sigma_i+\epsilon_i+1} \left[h_{i-1} c_{i-1} \prod_{j=1}^{i} (\beta_j/\alpha_j)\right]^{\epsilon_i} \\
&= (-1)^{\sigma_{i+1}-\sigma_i+\epsilon_i(m_i+i+1)+1} \\
&= 1.
\end{aligned}$$

Therefore, $\theta_{i+1} = 1$, and using (4.62), we have proved (4.56).

To prove (4.58), let

$$\phi_{i+1} = (-1)^{\tau_{i+1}} c_{i+1}^{\epsilon_{i+1}-1} \prod_{j=1}^{i+1} c_{j-1}^{\epsilon_{j-1}+\epsilon_j}\ (\beta_j/\alpha_j)^{n_{i+2}-n_j}, \tag{4.63}$$

so that (4.52) becomes

$$\hat{R}_{n_{i+2}-1,s_0}(A,B) = y^{m_{i+2}+n_{i+2}-1}\phi_{i+1}\, R_{i+1}(y). \tag{4.64}$$

Then, since $\theta_{i+2} = 1$, and using (4.59), (4.61) and (4.63),

$$\begin{aligned} c_{i+1}\phi_{i+1} &= c_{i+1}\phi_{i+1}/\theta_{i+2} \\ &= (-1)^{\tau_{i+1}-\sigma_{i+2}} c_{i-1}^{-1} \prod_{j=1}^{i+2} (\alpha_j/\beta_j) \\ &= h_{i+1}. \end{aligned} \tag{4.65}$$

Thus, (4.58) follows from (4.64) and (4.65).

Observe also that h_{i+1} is the leading coefficient of the r.h.s. of (4.64). Therefore, h_{i+1} is also the first nonzero coefficient of $\hat{R}_{n_{i+1}-1,s_0}(A,B)$, and consequently from (4.39), $h_{i+1} \in J$, $i = -1, 0, \cdots$. In addition, from (4.54), we then get that $\beta_{i+1} \in J$, $i = -1, 0, \cdots$.

To complete the proof of the theorem, it remains to show (4.57), which follows trivially from (4.51). ■

Theorem 4.12. Let the polynomial sequences

$$\{U_i(y)\}_{i=-1,0,\ldots}\ ,\ \{V_i(y)\}_{i=-1,0,\ldots}$$

be given by (4.27), (4.29) and (4.30), and let β_{i+1} and h_i be defined according to (4.54) and (4.55). Then, for $i = 1, 2, \cdots$ the resultants (4.37) and (4.38) satisfy

$$\hat{U}_{n_{i+1},s_0}(A,B) = U_{i+1}(y),$$

$$\hat{V}_{n_{i+1},s_0}(A,B) = V_{i+1}(y),$$

[illegible] $(A,B) = 0, \qquad n_{i+1} < l < n_{i+2},$

[illegible] $(A,B) = 0, \qquad n_{i+1} < l < n_{i+2},$

$$c_{i+1}\hat{U}_{n_{i+2}-1,s_0}(A,B) = h_{i+1}U_{i+1}(y),$$
$$c_{i+1}\hat{V}_{n_{i+2}-1,s_0}(A,B) = h_{i+1}V_{i+1}(y).$$

Proof: The proof is similar to the proof of the Theorem 4.9.

4.5 Algorithm JPADE

The algorithm JPADE given below is a generalization of Algorithm 3 given by Cabay and Choi [9]. The generalization consists of replacing division in their algorithm by pseudo-division (Step 6 of JPADE) and removing the common factor β_{i+1} identified in Theorem 4.11 from the coefficients of the Padé fraction U_{i+1}/V_{i+1} (Step 8 of JPADE). Given non-negative integers m and n and two power series $A(y)$ and $B(y)$, the algorithm computes all the Padé fractions $U_i(y)/V_i(y)$ of type (m_i,n_i) for the pair ($A(y)$ and $B(y)$) along the off-diagonal path $m_i - n_i = m - n$, $i = 1, \cdots$. Prior to exit, in Step 4 of JPADE, the two most recent Padé fractions are scaled by an appropriate power of y. Thus, in particular, the most recent Padé fraction becomes the unique (up to a multiplicative factor in J, scaled Padé fraction of type (m,n) for the pair $(A(y), B(y))$.

When $A(y)$ and $B(y)$ are polynomials of degrees m and n, respectively, it can be shown (as in Cabay and Choi [9]) that JPADE yields exactly the same intermediate and final results as does Brown's extended Euclidean algorithm [7] for computing the greatest common divisors of the reciprocals of $A(y)$ and $B(y)$. This correspondence is made under the assumption tha[illegible] operations in Steps 2 and 5 in JPADE are dropped. Dropping these mod[illegible] however, significantly increases the cost of operations in JPADE.

The initial conditions (Step 1) of JPADE, which involve negative powers of y, are chosen to correspond exactly, after taking reciprocals, to the usual initial conditions given for the extended Euclidean algorithm. However, after two iterations of the algorithm all powers of y become nonnegative, and instead these could have been used as initial conditions.

Algorithm JPADE $\left(\begin{bmatrix} A \\ B \end{bmatrix}, \begin{bmatrix} m \\ n \end{bmatrix}\right)$:

Input: A, B, m, n, where

(1) m and n are nonnegative integers, and

(2) the truncated power series

$$A = \sum_{j=0}^{m+n} a_j y^j, \; a_0 \neq 0,$$

and

$$B = \sum_{j=0}^{m+n} b_j y^j, \; b_0 \neq 0$$

with coefficients $a_j, b_j \in J$, an integral domain.

Output: $\left(\begin{bmatrix} U_i & U_{i-1} \\ V_i & V_{i-1} \end{bmatrix}, \begin{bmatrix} m & m_i - 1 \\ n & n_i - 1 \end{bmatrix}\right)$, where

(1) $U_i/V_i \in J(y)$ is the scaled Padé fraction of type (m,n) for the pair $(A(y),B(y))$

and

(2) $U_{i-1}/V_{i-1} \in J(y)$ is the scaled Padé fraction of type (m_i-1, n_i-1) for the pair $(A(y), B(y))$.

Step 1: #Initialization#

$$c_{-1} \leftarrow 1$$
$$s_{-1} \leftarrow 1$$
$$h_{-1} \leftarrow 1$$
$$i \leftarrow 0$$

If $m \geq n$, then

$$\begin{bmatrix} U_0 & U_{-1} \\ V_0 & V_{-1} \end{bmatrix} \leftarrow \begin{bmatrix} 1 & 0 \\ 0 & y^{n-m-1} \end{bmatrix}$$

$$\begin{bmatrix} m_0 & m_{-1} \\ n_0 & n_{-1} \end{bmatrix} \leftarrow \begin{bmatrix} 0 & -1 \\ n-m & n-m-1 \end{bmatrix},$$

else

$$\begin{bmatrix} U_0 & U_{-1} \\ V_0 & V_{-1} \end{bmatrix} \leftarrow \begin{bmatrix} 0 & y^{m-n-1} \\ 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} m_0 & m_{-1} \\ n_0 & n_{-1} \end{bmatrix} \leftarrow \begin{bmatrix} m-n & m-n-1 \\ 0 & -1 \end{bmatrix}.$$

Step 2: #Compute the residual for U_i/V_i#

Find s_i and R_i, $\partial R_i \leq s_i$, such that

$$(AV_i + BU_i) \bmod y^{m_i+n_i+2s_i+1} = y^{m_i+n_i+s_i} R_i,$$

where $s_i \leq m - n_i$ and $R_i(0) \neq 0$, otherwise set $s_i = n - n_i + 1$.

Set $q_i \leftarrow R_i(0)$.

Step 3: #Calculation of degrees of next Padé fraction #

$$m_{i+1} \leftarrow m_i + s_i$$
$$n_{i+1} \leftarrow n_i + s_i$$

Step 4: #Termination#

If $n_{i+1} > n$ then #Scale the Padé fractions#

$$\begin{bmatrix} U_i & U_{i-1} \\ V_i & V_{i-1} \end{bmatrix} \leftarrow \begin{bmatrix} U_i & U_{i-1} \\ V_i & V_{i-1} \end{bmatrix} \begin{bmatrix} y^{s_i-1} & 0 \\ 0 & y^{s_{i-1}-1} \end{bmatrix}$$

Return $\left(\begin{bmatrix} U_i & U_{i-1} \\ V_i & V_{i-1} \end{bmatrix}, \begin{bmatrix} m & m_i - 1 \\ n & n_i - 1 \end{bmatrix} \right)$,

and exit.

Step 5: #Compute residual for U_{i-1}/V_{i-1}#

Compute R_{i-1}, $\partial R_{i-1} \leq s_i$, such that

$$(AV_{i-1} + BU_{i-1}) \bmod y^{m_{i-1}+n_{i-1}+s_{i-1}+s_i+1} = y^{m_{i-1}+n_{i-1}+s_{i-1}} R_{i-1},$$

where $R_{i-1}(0) \neq 0$.

Step 6: #Pseudo-division#

Compute the pseudo-quotient Ω_{i+1}, $\partial\Omega_{i+1} \leq s_i$, such that

$$(c_i^{s_i+1} R_{i-1} + \Omega_{i+1} R_i) \bmod y^{s_i+1} = 0$$

Step 7: #Compute a common factor#

$$h_i \leftarrow c_i^{s_i} h_{i-1}^{1-s_i}$$
$$\beta_{i+1} \leftarrow (-1)^{s_i+1} c_{i-1} h_{i-1}^{s_i}$$

Step 8: #Advance the computation#

Compute the Padé fraction U_{i+1}/V_{i+1} of type (m_{i+1}, n_{i+1}) by means of

$$\begin{bmatrix} U_{i+1} \\ V_{i+1} \end{bmatrix} \leftarrow \left(\begin{bmatrix} U_i & U_{i-1} \\ V_i & V_{i-1} \end{bmatrix} \begin{bmatrix} \Omega_{i+1} \\ c_i^{e_i+1} y^{e_{i-1}+e_i} \end{bmatrix} \right) / \beta_{i+1}$$

Set $i \leftarrow i+1$ and go to 2.

To illustrate the effect of the removal of the factors β_{i+1} from the power series remainder sequence, we apply the algorithm JPADE to the same example (4.36) to obtain $R_1(y)$, $R_2(y)$, $R_3(y)$ and $R_4(y)$, where

$$\begin{aligned} R_1(y) &= 15 - 3y^2 + 9y^4 + \dots \\ R_2(y) &= 65 + 125y - 245y^2 + \cdots \\ R_3(y) &= 9326 - 12300y + \cdots \\ R_4(y) &= 260708 + \cdots \end{aligned} \tag{4.66}$$

Removed factors β_i are

$$\begin{aligned} \beta_1 &= -1 \\ \beta_2 &= -243 \\ \beta_3 &= -9375 \\ \beta_4 &= 1856465. \end{aligned} \tag{4.67}$$

In assessing the costs of JPADE it is assumed that classical algorithms are used for the multiplication in the integral domain J. Thus, it is assumed that for $a, b \in J$

$$cost(a \cdot b) = size(a) \cdot size(b), \tag{4.68}$$

where the function size measures the total storage space that is required for its argument.

To determine the cost of JPADE, we begin with estimates of sizes of all variables in the algorithm at the i-th pass. Assume that the size of all coefficients of the input power series $A(y)$ and $B(y)$ is bounded by some integer $k>0$. Using theorem 4.12, the size of coefficients of the polynomial $U_i(y)$ is equal to the size of the coefficients of the resultant $\hat{U}_{n_i,j_0}(A,B)$, which is bounded by $k\cdot(m_i+n_i+1)$. In a similar fashion, with the exception of the polynomial Ω_{i+1}, the sizes of all the other variables in JPADE can be determined. To assess the size of coefficients of Ω_{i+1}, notice that it can be obtained by solving a system of equations, similar to the one given in (4.8), by means of Cramer's rule. In this case, the determinants of the matrix in (4.8) will cancel with the multiplier $c_i^{s_i+1}$, and the size of the coefficients in Ω_{i+1} is bounded by the size of the suitable determinant. Bounds for the sizes of the coefficients of the variables in JPADE are summarized in Table 4.1 below.

Variable	Bound on size of coefficients
$U_i(y)$	$k(m_i+n_i+1)$
$V_i(y)$	$k(m_i+n_i+1)$
$R_i(y)$	$k(m_i+n_i+1)$
c_i	$k(m_i+n_i+1)$
h_i	$k(m_i+n_i+1)s_i$
β_{i+1}	$k(m_i+n_i+1)(s_i+1)^2$
$\Omega_{i+1}(y)$	$k(m_i+n_i+1)(s_i+1)$[illegible]

Table 4.1
Bounds on Variable Size

A count of the number of operations required for each step at i-th pass through the algorithm JPADE is now obtained. Since the number of additions at each step is never greater than the number of multiplications, only multiplications are counted.

Step 2 requires the computation of $2s_i$ terms of the power series AV_i+BU_i, namely, the coefficients of y^j, where $j = m_i + n_i + 1, \ldots, m_i + n_i + 2s_i$. The cost of computing one term of AV_i is

$$\begin{aligned} cost\left(\sum_{j=0}^{n_i} a_{i-j} v_{i,j}\right) &= (\partial V_i + 1) \cdot size(A) \cdot size(V_i) \\ &\le (n_i + 1) k^2 (m_i + n_i + 1). \end{aligned}$$

Similarly, the cost of computing one term of BU_i is bounded by $(m_i + 1)k^2(m_i + n_i + 1)$. Thus, the total cost of Step 2 is not larger than

$$2s_i(m_i + n_i + 2)k^2(m_i + n_i + 1).$$

A detailed cost analysis of the other steps in JPADE can be conducted in a similar fashion, and the results are summarized in Table 4.2 below.

Step	# of Multiplications	Bound
2.	$2s_i(m_i + n_i + 2)(m_i + n_i + 1)k^2$	$2s_i(m_i + n_i + 2)^2 k^2$
5.	$(s_i + 1)(m_{i-1} + n_{i-1} + 2)(m_{i-1} + n_{i-1} + 1)k^2$	$(s_i + 1)(m_i + n_i + 2)^2 k^2$
6.	$(s_i + 3)^3(m_i + n_i + 1)^2 k^2$	$(s_i + 3)^3(m_i + n_i + 2)^2 k^2$
7.	$4s_i^2(m_i + n_i + 1)^2 k^2$	$4s_i^2(m_i + n_i + 2)^2 k^2$
8.	$2(s_i + 1)^2(m_i + n_i + 2)^3 k^2$	$2(s_i + 1)^2(m_i + n_i + 2)^3 k^2$

Table 4.2
Bounds on Number of Multiplications

Thus, the total number of multiplications in Steps 2, 5, 6 and 7 is bounded by

$$\begin{aligned} &(s_i^3 + 13s_i^2 + 30s_i + 28)(m_i + n_i + 2)^2 k^2, \\ &\le (s_i + 5)^3 (m_i + n_i + 2)^2 k^2 \end{aligned} \tag{4.69}$$

and number of multiplications in Step 8 is bounded by

$$2(s_i + 1)^2(m_i + n_i + 2)^3 k^2. \tag{4.70}$$

To estimate the total cost of JPADE, two cases are taken into consideration. In the case when the step size is bounded by some constant c (which is the case of normal power series, with $c = 1 \geq s_i$), the number of passes through the algorithm is bounded by m. Then, the total cost is bounded by

$$\sum_{i=0}^{m} [(c+5)^3(m_i+n_i+2)^2k^2 + 2(c+1)^2(m_i+n_i+2)^3k^2] \tag{4.71}$$
$$\leq (c+5)^3(m+1)(m+n+2)^2k^2 + 2(c+1)^2(m+1)(m+n+2)^3k^2$$

Thus, in this case, the algorithm has a cost complexity of $O(k^2(m+n)^4)$.

In the case when the step size is not bounded, the costs of the algorithm are estimated by using the inequality

$$\sum_{i=0}^{l}(m_{i+1})^p s_i^q \leq m^{p+q}, \tag{4.72}$$

where $p,q > 0$ and l is such that $\sum_{i=0}^{l} s_i = m_{l+1} = m$. Since

$$2m_{i+1} \geq m_{i+1} + n_{i+1} = m_i + n_i + 2s_i \geq m_i + n_i + 2,$$

the bounds (4.69) and (4.70), respectively, can be rewritten as

$$(s_i^3 + 13s_i^2 + 30s_i + 28)(2m_{i+1})^2k^2 \tag{4.73}$$
$$= 4(s_i^3 + 13s_i^2 + 30s_i + 28)(m_{i+1})^2k^2$$

and

$$8(s_i+1)^2(m_{i+1})^2k^2. \tag{4.74}$$

Then, by applying inequality (4.72), a bound for the total cost of JPADE is given by

$$\sum_{i=0}^{l}(4(s_i^3 + 13s_i^2 + 30s_i + 28)m_{i+1}^2k^2 + 8(s_i^2+2s_i+1)m_{i+1}^2k^2 \tag{4.75}$$
$$\leq (4m^5 + 60m^4 + 136m^3 + 120m^2)k^2.$$

Thus, in this, the worst case, JPADE has a cost complexity of $O(k^2m^5)$.

These cost estimates concur with the estimates obtained by Brown [7] of the cost of his Euclidean algorithm for finding greatest common divisors of two polynomials with coefficients in J. These costs compare favorably with the costs of the fraction free method proposed by Geddes [19] which according to an analyses done by Bareiss [3] (albeit, on a more general problem) has a cost complexity of $O(k^2(m+n)^5)$, even if the power series is normal.

Chapter 5

Algorithms for Modular Padé Forms

In this chapter, some methods for constructing modular Padé forms for the bivariate power series

$$A(z,y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} {}_ia_j z^i y^j \in D,$$

are considered, and a new algorithm is developed.

The direct approach of constructing modular Padé forms is to solve the block Hankel systems (3.8) and (3.9). A solution of (3.8) yields a vector $Q = [{}_kQ, \cdots, {}_0Q]^T$, where ${}_iQ = [{}_iq_n, \cdots, {}_iq_0]^T$, and then the vector $P = [{}_kP, \cdots, {}_0P]^T$, where ${}_iP = [{}_ip_m, \cdots, {}_ip_0]^T$, can be obtained from (3.9). By Lemma 3.3, the bivariate rational expression $P(z,y)/Q(z,y)$, where

$$P(z,y) = \sum_{i=0}^{k} \sum_{j=0}^{m} {}_ip_j z^i y^j$$

and

$$Q(z,y) = \sum_{i=0}^{k} \sum_{j=0}^{n} {}_iq_j z^i y^j$$

is a modular Padé (k,m,n)-form for $A(z,y)$.

If ${}_0H_{n-1,n-1}$ is nonsingular, a method of solving (3.8) can be described as follows. Partition each Hankel submatrix ${}_iH_{n-1,n}$, $i=0,\ldots,k$, in (3.8), according to

$$_iH_{n-1,n} = \begin{bmatrix} {}_ia_{m-n+1}, \ldots, & {}_ia_m & | & {}_ia_{m+1} \\ \vdots & \vdots & | & \\ {}_ia_m, \ldots, & {}_ia_{m+n-1} & | & {}_ia_{m+n} \end{bmatrix} = [{}_iH_{n-1,n-1} | {}_i\overline{a}],$$

where ${}_i\overline{a} = [{}_ia_{m+1}, \cdots, {}_ia_{m+n}]^T$. By setting ${}_0q_0 = -1$ and ${}_iq_0 = 0$, $i=1,\ldots,k$, the system (3.8) becomes

$$\begin{bmatrix} & & {}_0H_{n-1,n-1} \\ & & \vdots \\ {}_0H_{n-1,n-1}, & \ldots, & {}_kH_{n-1n-1} \end{bmatrix} \begin{bmatrix} {}_k\overline{Q} \\ \vdots \\ {}_0\overline{Q} \end{bmatrix} = \begin{bmatrix} {}_0\overline{a} \\ \vdots \\ {}_k\overline{a} \end{bmatrix}, \tag{5.1}$$

where ${}_i\overline{Q}$ is a subvector of ${}_iQ$ according to

$${}_iQ = [{}_iq_n, \cdots, {}_iq_1 | {}_iq_0] = [{}_i\overline{Q} | {}_iq_0].$$

The system (5.1) can then be written iteratively as

$${}_i\overline{Q} = {}_0H_{n-1,n-1}^{-1} [{}_i\overline{a} - \sum_{j=0}^{i-1} {}_{i-j}H_{n-1,n-1} \cdot {}_j\overline{Q}]^T, \quad i=0,\ldots k. \tag{5.2}$$

The solution (5.2) so obtained provides a nontrivial solution Q (since ${}_0q_0 \neq 0$) of (3.8).

It is an easy matter to show that the cost of the iteration (5.2) is $O(k^2n^2)$ operations in D, using classical arithmetic. Given Q, the vector P can be obtained from (3.9) in $O(k^2mn)$ additional operations in D. Thus, the complexity of this method of constructing one nontrivial modular Padé (k,m,n)-form for $A(z,y)$ is $O(k^2(m+n)n)$ operations in D, using classical arithmetic.

Each of the methods of Bultheel [8], Hughes Jones and Makinson [22] and Wax and Kailath [30] for solving general block Hankel systems, when reduced to the block triangular system (3.8), also require $O(k^2n^2)$ operations in D. Unfortunately, their methods, and the iteration (5.2) as well, require at least that ${}_0H_{n-1,n-1}$ be nonsingular, a restrictive condition that is not necessary for the existance of solutions.

It is an easy observation that modular Padé forms, to a given bivariate power series $A(x,y)$, can be computed also by a straightforward application of the algorithm JPADE. Such an application has no restrictive requirements on the power series $A(x,y)$, but according to the analysis of Section 4.6 is much more costly than the iteration (5.2). As the input power series $A(y)$ to JPADE, we use the power series $A(x,y)$ truncated modulo x^{k+1} (regarded as a univariate power series in y, with polynomial coefficients in $J = D[x]$), and as the input power series $B(y)$, we use $B(y) = 1$. The scaled Padé fraction $U_i(y)/V_i(y)$ obtained from this input is subsequently processed by removing the largest common factor x^l from $U_i(y)$ and $V_i(y)$, and then truncating each modulo x^{k+1}, to obtain a modular Padé form.

The above procedure is the basis for the algorithm MPADE, given later in this chapter. Unfortunately, this procedure, without any modifications, has the same computational complexity as JPADE. To reduce the complexity, we will show that at appropriate steps intermediate results can be truncated modulo x^{k+1}. The steps at which such a truncation can take place are identified using the notion of nonsingularity of a univariate power series.

Definition 5.1: For nonnegative integers m and n, the univariate power series

$$A(y) = \sum_{i=0}^{\infty} a_i y^i \ ,$$

is (m,n)-nonsingular if and only if the following Hankel matrix of its coefficients

$$\begin{bmatrix} a_{m-n+1}, & \dots, & a_m \\ \vdots & & \vdots \\ a_m & \dots, & a_{m+n-1} \end{bmatrix},$$

where $a_j = 0$ for $j<0$, is nonsingular.

If the power series $A(y)$ is (m,n)-nonsingular for all $m,n \geq 0$, then we say that $A(y)$ is normal.

Definition 5.1 is extended to bivariate power series by means of

Definition 5.2. A bivariate power series

$$A(x,y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_{ij} x^i y^j$$

is $(0,m,n)$-nonsingular if and only if the univariate power series $A(0,y)$ is (m,n)-nonsingular.

A direct consequence of Definition 5.2 is the following

Lemma 5.3: If a power series $A(x,y)$ is $(0,m,n)$-nonsingular, then it is $(0,m,n)$-maximal.

Proof: If $A(x,y)$ is $(0,m,n)$-nonsingular, from Definitions 5.1 and 5.2, it follows that the matrix

$$
{}_0H_{n-1,n} = \begin{bmatrix} {}_0a_{m-n+1}, & \ldots, & {}_0a_{m+1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ {}_0a_m & \ldots, & {}_0a_{m+n} \end{bmatrix}
$$

is maximal. Thus, the dimension of ${}_0S_{n-1,n}$, the space of solutions of (2.11), is equal to 1. There[illegible]ions 2.9 and 3.5, $A(x,y)$ is $(0,m,n)$-maximal. ■

W[illegible]h necessary and sufficient conditions for $(0,m,n)$-nonsingularity in terms o[illegible]esultants (4.37), (4.38) and (4.39). Let

$$A(x,y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} {}_ia_j x^i y^j$$

and

$$B(x,y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} {}_ib_j x^i y^j$$

be bivariate power series with coefficients in D. Also, let [illegible](y) and $B(y)$, respectively, be the corresponding univariate power series with coefficients in $D[[x]]$. That is,

$$A(y) = \sum_{j=0}^{\infty} a_j y^j, \; a_j = \sum_{i=0}^{\infty} {}_ia_j x^i,$$

and

$$B(y) = \sum_{j=0}^{\infty} b_j y^j, \; b_j = \sum_{i=0}^{\infty} {}_ib_j x^i.$$

Lemma 5.4: Let ${}_0b_0 \neq 0$, and suppose $m \geq n$. Then, the quotient power series $E(x,y) = A(x,y)/B(x,y)$ is $(0,m,n)$-nonsingular iff the constant term [illegible] the resultant $V_{n,m-n}(A,B)$ is nonzero.

Proof: Let

$$E(z,y) = \sum_{j=0}^{\infty} e_j y^j,$$

where

$$e_j = \sum_{i=0}^{\infty} {}_ie_j z^i.$$

Then, $E(z,y) = A(z,y)/B(z,y)$ implies that

$$a_i = \sum_{l=0}^{i} e_l b_{i-l}, \qquad i = 0, \cdots \tag{5.3}$$

For $i = 0,\ldots,n$ and $j = 0,\ldots,m+n$, define

$${}_if_j = a_{i-j} - \sum_{l=0}^{m-j} e_l b_{i-j-l}. \tag{5.4}$$

Then, using (5.4),

$$\begin{aligned} {}_if_j &= a_{i-j} - \sum_{l=j}^{i} e_{l-j} b_{i-l} + \sum_{l=m+1}^{i} e_{l-j} b_{i-l} \\ &= \sum_{l=m+1}^{i} e_{l-j} b_{i-l}. \end{aligned} \tag{5.5}$$

Now, using (5.4), the resultant (4.38) with $j = n$ and $s = m - n$ becomes

$$\hat{V}_{n,m-n}(A,B) = \begin{vmatrix} a_0 & & & b_0 & & \\ \cdot & & & & & \\ \cdot & a_0 & & \cdot & & b_0 \\ \vdots & a_1 & & \vdots & & \cdot \\ & \cdot & & & & \cdot \\ a_{m+n} & \cdots & a_m & b_{m+n} & \cdots & b_n \\ 1 & \cdots & y^n & 0 & \cdots & 0 \end{vmatrix}, \tag{5.6}$$

$$= \begin{vmatrix} 0 & & & b_0 & & \\ & \cdot & & & \cdot & \\ & & 0 & & & b_0 \\ {}_0f_{m+1} & \cdots & {}_nf_{m+1} & b_{m+1} & \cdots & b_1 \\ \vdots & & \vdots & \vdots & & \vdots \\ {}_0f_{m+n} & \cdots & {}_nf_{m+n} & b_{m+n} & \cdots & b_n \\ 1 & \cdots & y^n & 0 & \cdots & 0 \end{vmatrix}$$

Expanding the second determinant in (5.6) about the last $m+1$ columns, and then adding appropriate multiples of rows to other rows in order to remove all those terms in ${}_if_j$ (expressed by (5.5)) not including the factor b_0, we obtain

$$\hat{V}_{n,m-n}(A,B) = (-1)^{(m+1)(n+1)} b_0^{m+1} \begin{bmatrix} e_{m+1}b_0 & \cdots & e_{m-n+1}b_0 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ e_{m+n}b_0 & \cdots & e_m b_0 \\ 1 & \cdots & y^n \end{bmatrix} \tag{5.7}$$

$$= (-1)^{(m+1)(n+1)} b_0^{m+n+1} \begin{bmatrix} e_{m+1} & \cdots & e_{m-n+1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ e_{m+n} & \cdots & e_m \\ 1 & \cdots & y^n \end{bmatrix}$$

The lemma now follows by expanding the second determinant in (5.7) about the first column and last row and evaluating it at $x=0$.

Given $A(x,y)$, $B(x,y)$ and two nonnegative integers m and n, let the power series remainder sequence

$$\{s_i, R_i(x,y)\}_{i=-1,0,} \tag{5.8}$$

and the cofactor sequences

$$\{U_i(z,y)\}_{i=-1,0,} \quad , \quad \{V_i(z,y)\}_{i=-1,0,} \tag{5.9}$$

be determined by the algorithm JPADE. That is, if $m \geq n$, set

$$R_{-1}(z,y) = A(z,y),\ R_0(z,y) = B(z,y) \text{ and } s_0 = m-n;$$

and, if $m < n$, set

$$R_{-1}(z,y) = B(z,y),\ R_0(z,y) = A(z,y) \text{ and } s_0 = n-m.$$

As a consequence of Lemma 5.4, we have

Theorem 5.5. Let ${}_0b_0 \neq 0$. Then $A(z,y)/B(z,y)$ is $(0,m_i,n_i)$-nonsingular iff $V_i(0,0) \neq 0$. In addition, if $A(z,y)/B(z,y)$ is $(0,m_i,n_i)$-nonsingular, then the leading coefficient $c_{i-1}(z)$ of $R_{i-1}(z,y)$ divides $U_{i-1}(z,y)$, $V_{i-1}(z,y)$ and $R_{i-1}(z,y)$.

Proof: We consider the case $m \geq n$ only. The proof for the case $m < n$ can be shown in a similar fashion, using instead of Lemma 5.4, the result that $E(z,y) = A(z,y)/B(z,y)$ is $(0,m,n)$-nonsingular iff the constant term of the resultant $U_{m,n-m}(A,B)$ is nonzero.

Suppose that $m \geq n$. That $V_i(0,0) \neq 0$ iff $A(z,y)/B(z,y)$ is $(0,m_i,n_i)$-nonsingular is a direct consequence of Theorem 4.12 and Lemma 5.4. To show that $c_{i-1}(z)$ divides $R_{i-1}(z,y)$, from (4.40) and (4.58) with $j = n_i - 1$ and $s_0 = m-n = m_i - n_i$, we obtain

$$y^{m_i+n_i-1} h_{i-1}(z) R_{i-1}(z,y)$$
$$= c_{i-1}(z) \hat{R}_{n_i-1,s_0}(A,B)$$

$$= y^{m_i+n_i+1}c_{i-1}(z)\begin{vmatrix} a_0 & & & b_0 & & \\ & \ddots & & & \ddots & \\ \vdots & & a_0 & \vdots & & b_0 \\ & & \vdots & & & \vdots \\ a_{m_i+n_i-1} & \cdots & a_{m_i} & b_{m_i+n_i-1} & \cdots & b_{n_i} \end{vmatrix} + O(y)^{m_i+n_i+2} \qquad (5.10)$$

But, from (5.6) and Theorem 4.12,

$$V_i(z,y) = \hat{V}_{n_i,m_i-n_i}(A,B),$$

$$= (-1)^{m_i+n_i+1}b_0\begin{vmatrix} a_0 & & & b_0 & & \\ & \ddots & & & \ddots & \\ \vdots & & a_0 & \vdots & & b_0 \\ & & \vdots & & & \vdots \\ a_{m_i+n_i-1} & \cdots & a_{m_i} & b_{m_i+n_i-1} & \cdots & b_{n_i} \end{vmatrix} + O(y) \qquad (5.11)$$

Since $V_i(0,0)\neq 0$ and ${}_0b_0\neq 0$, the determinant on the r.h.s. of (5.11) does not vanish at $z=0$. Thus, the determinant on the r.h.s. of (5.10) also does not vanish at $z=0$, and consequently $h_{i-1}(0)\neq 0$. Now, from (5.10)

$$y^{m_i+n_i-1}R_{i-1}(z,y) = c_{i-1}(z)h_{i-1}^{-1}(z)\hat{R}_{n_i-1,s_0}(A,B),$$

and it is clear that $c_{i-1}(z)$ divides $R_{i-1}(z,y)$, because $h_{i-1}^{-1}(z)\hat{R}_{n_i-1,s_0}(A,B)\in D[[z,y]]$. From Theorem 4.12, it follows also that $c_{i-1}(z)$ divides $U_{i-1}(z,y)$ and $V_{i-1}(z,y)$.

The algorithm MPADE below performs basically the same computations as does the algorithm JPADE. It differs primarily in that MPADE takes advantage of Theorem 5.5 as follows. At any $(0,m_i,n_i)$-nonsingularity, MPADE reduces the sizes (in terms of the degree in the variable z) of the intermediate results $U_i(z,y)$, $V_i(z,y)$,

$U_{i-1}(z,y)$ and $V_{i-1}(z,y)$ to polynomials of degree k in the variable z. After such a reduction, algorithm MPADE continues computations exactly as would JPADE (i.e., MPADE calls JPADE) until either, another nonsingularity is encountered and reduction is once again performed, or it terminates normally having arrived at the ultimate destination $n_{i+1} = n$. In order that it recognize a $(0, m_i, n_i)$-nonsingularity, it is necessary to modify JPADE by replacing Step 4 with

Step 4. #Termination with nonsingularity#

If ($V_i(0,0) \neq 0$ and $i > 0$) or $n_{i+1} > n$, then

Return $\left(\begin{bmatrix} U_i & U_{i-1} \\ V_i & V_{i-1} \end{bmatrix}, \begin{bmatrix} m_i & m_{i-1} \\ n_i & n_{i-1} \end{bmatrix} \right)$ and exit.

The algorithm MPADE below assumes that this modification to JPADE has been made.

Algorithm MPADE $\left(\begin{bmatrix} A \\ B \end{bmatrix}, \begin{bmatrix} m \\ n \end{bmatrix}, k \right)$:

Input: A, B, m, n, and k, where

(1) m, n, and k are nonnegative integers and,

(2) A and B are truncated power series,

$$A = \sum_{j=0}^{m+n} a_j(z) y^j, \quad a_0(0) \neq 0,$$

and

$$B = \sum_{j=0}^{m+n} b_j(z) y^j, \quad b_0(z) \neq 0.$$

with coefficients

$$a_j(z) = \sum_{i=0}^{k} {}_i a_j z^i$$

and

$$b_j(z) = \sum_{i=0}^{k} {}_i b_j z^i, \quad {}_0 b_0 \neq 0$$

in $D[z]$.

Output: $\left(\begin{bmatrix} U_i & U_{i-1} \\ V_i & V_{i-1} \end{bmatrix}, \begin{bmatrix} m_i & m_{i-1} \\ n_i & n_{i-1} \end{bmatrix} \right)$, where

U_i/V_i is a modular Padé (k,m,n)-form for A/B and U_{i-1}/V_{i-1} is a modular Padé (k,m_{i-1},n_{i-1})-form for A/B.

Step 1. #Initialization#

$$\begin{bmatrix} R_0 \\ R_{-1} \end{bmatrix} \leftarrow \begin{bmatrix} A \\ B \end{bmatrix}$$

$$\begin{bmatrix} U_0 & U_{-1} \\ V_0 & V_{-1} \end{bmatrix} \leftarrow \begin{bmatrix} 0 & y^{m-n-1} \\ 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} m_0 & m_{-1} \\ n_0 & n_{-1} \end{bmatrix} \leftarrow \begin{bmatrix} m-n & m-n-1 \\ 0 & [illegible] \end{bmatrix}$$

$s_{-1} \leftarrow 1$
$s_0 \leftarrow n-m$
$i \leftarrow 0$

Step 2. #Call JPADE#

$$\left(\begin{bmatrix}\overline{U}_j & \overline{U}_{j-1}\\ \overline{V}_j & \overline{V}_{j-1}\end{bmatrix}, \begin{bmatrix}\overline{m}_j & \overline{m}_{j-1}\\ \overline{n}_j & \overline{n}_{j-1}\end{bmatrix}\right) \leftarrow JPADE \left(\begin{bmatrix} R_i \\ R_{i-1}\end{bmatrix}, \begin{bmatrix} m-m_i-s_i \\ n-n_i\end{bmatrix}\right)$$

Step 3. #Convert to modular Padé forms#

Determine the largest l_j and l_{j-1} such that z^{l_j} and $z^{l_{j-1}}$, respectively, are common factors of $\overline{U}_j$ and $\overline{V}_j$ and of $\overline{U}_{j-1}$ and $\overline{V}_{j-1}$.

Set

$$\begin{bmatrix}\overline{U}_j & \overline{U}_{j-1}\\ \overline{V}_j & \overline{V}_{j-1}\end{bmatrix} = \left(\begin{bmatrix}\overline{U}_j & \overline{U}_{j-1}\\ \overline{V}_j & \overline{V}_{j-1}\end{bmatrix}\begin{bmatrix} z^{-l_j} & 0\\ 0 & z^{-l_{j-1}}\end{bmatrix}\right) mod z^{k+1}$$

Step 4. #Calculation of degrees#

$$\begin{bmatrix} m_{i+j} & m_{i+j-1}\\ n_{i+j} & n_{i+j-1}\end{bmatrix} = \begin{bmatrix} m_i+s_i & m_i+s_i\\ n_i & n_i\end{bmatrix} + \begin{bmatrix}\overline{m}_j & \overline{m}_{j-1}\\ \overline{n}_j & \overline{n}_{j-1}\end{bmatrix}$$

Step 5. #Combine results#

$$\begin{bmatrix} U_{i+j} & U_{i+j-1}\\ V_{i+j} & V_{i+j-1}\end{bmatrix} = \left(\begin{bmatrix} U_i & y^{s_{i-1}+s_i}U_{i-1}\\ V_i & y^{s_{i-1}+s_i}V_{i-1}\end{bmatrix}\begin{bmatrix}\overline{V}_j & \overline{V}_{j-1}\\ \overline{U}_j & \overline{U}_{j-1}\end{bmatrix}\right) mod z^{k+1}$$

Step 6. #Compute new residuals#

Find s_{i+j}, R_{i+j} and R_{i+j-1} such that

$$y^{m_{i+j}+n_{i+j}+s_{i+j}}R_{i+j} = (AV_{i+j} + BU_{i+j}) mod z^{k+1} mod y^{m+n+1}$$

where ($R_{i+j}(y=0) \neq 0$ and $s_{i+j} \leq n - n_{i+j}$), otherwise set $s_{i+j} = n - n_{i+j} + 1$; and

$$y^{m_{i+j-1}+n_{i+j-1}+s_{i+j-1}}R_{i+j-1} = (AV_{i+j-1} + BU_{i+j-1}) mod z^{k+1} mod y^{m+n+1}$$

Step 7. #Increment i#

Set

$$i \leftarrow i+j$$

Step 8. #Termination#

If $n_i + s_i > n$

then #Scale results and exit#

Set

$$\begin{bmatrix} U_i & U_{i-1} \\ V_i & V_{i-1} \end{bmatrix} \leftarrow \begin{bmatrix} U_i & U_{i-1} \\ V_i & V_{i-1} \end{bmatrix} \begin{bmatrix} y^{s_i-1} & 0 \\ 0 & y^{s_{i-1}-1} \end{bmatrix}$$

and

Return $\left(\begin{bmatrix} U_i & U_{i-1} \\ V_i & V_{i-1} \end{bmatrix}, \begin{bmatrix} m & m_i-1 \\ n & n_i-1 \end{bmatrix} \right)$

else #Continue processing#

Go to Step 2.

Theorem 5.6: The algorithm MPADE terminates. On completion, U_i/V_i is a modular Padé (k,m,n)-form and U_{i-1}/V_{i-1} is a modular (k,m_i-1,n_i-1)-form for A/B.

Proof: We proceed by induction with respect to the number of passes through the algorithm. We show that the following conditions are satisfied just prior to the execution of Step 7:

Condition 1. U_{i-1}/V_{i-1} is a modular Padé (k,m_{i-1},n_{i-1})-form for A/B, where for some s_{i-1} and R_{i-1}

$$AV_{i-1} + BU_{i-1} = y^{m_{i-1}+n_{i-1}+s_{i-1}} R_{i-1} \bmod z^{k+1} \bmod y^{m+n+1} \quad (5.12)$$

Condition 2. U_i/V_i is a modular Padé (k, m_i, n_i)-form for A/B, where for some R_i and for $s_i \geq 1$

$$AV_i + BU_i = y^{m_i+n_i+s_i} R_i \bmod z^{k+1} \bmod y^{m+n+1} \quad (5.13)$$

Condition 3. $R_{i-1}(0,0) \neq 0$ and $V_{i-1}(0,y) \neq 0$ (5.14)

Condition 4. $V_i(0,0) \neq 0$ (5.15)

Condition 5. $s_{i-1} = m_i - m_{i-1} = n_i - n_{i-1} \geq 1$ (5.16)

Condition 6. $m_i - n_i = m_{i-1} - n_{i-1} = m - n$ (5.17)

With these assumptions on U_{i-1}/V_{i-1} and U_i/V_i, we show that U_{i+j-1}/V_{i+j-1} and U_{i+j}/V_{i+j}, when computed by means of Step 2 though Step 6, exist in two distinct states.

Case I: If $\overline{V}_j(0,0) \neq 0$, we then show that U_{i+j-1}/V_{i+j-1} and U_{i+j}/V_{i+j} satisfy all the Conditions 1 through 6 above (with i replaced by $i+j$). Thus, if the algorithm does not terminate at Step 8, the same conditions 1 through 6 will continue to be satisfied at the next pass of MPADE.

Case II: Otherwise, we show that algorithm terminates immediately during the execution of Step 8, and returns U_{i-1}/V_{i-1} and U_i/V_i satisfying all the conditions of the theorem.

First consider Case I.

To show that Conditions 1 through 6 (with i replaced by $i+j$) continue to be satisfied, we trace Step 2 through 6 of MPADE, all of which contribute to the construction of U_{i+j-1}/V_{i+j-1} and U_{i+j}/V_{i+j}.

In Step 2 of MPADE, from the results of Chapter 4, it follows that the fractions $\overline{U}_j/\overline{V}_j$ and $\overline{U}_{j-1}/\overline{V}_{j-1}$ computed by JPADE satisfy

$$R_i\overline{V}_j + R_{i-1}\overline{U}_j = y^{\overline{m}_j+\overline{n}_j+\overline{s}_j}\overline{R}_j \tag{5.18}$$

for some $\overline{R}_j$ and $s_j \geq 1$, and

$$R_i\overline{V}_{j-1} + R_{i-1}\overline{U}_{j-1} = y^{\overline{m}_{j-1}+\overline{n}_{j-1}+\overline{s}_{j-1}}\overline{R}_{j-1} \tag{5.19}$$

for some $\overline{R}_{j-1}$ and $\overline{s}_{j-1}$ such that

$$\overline{s}_{j-1} = \overline{m}_j - \overline{m}_{j-1} = \overline{n}_j - \overline{n}_{j-1}. \tag{5.20}$$

In addition,

$$\overline{n}_j - \overline{m}_j = \overline{n}_{j-1} - \overline{m}_{j-1} = s_i. \tag{5.21}$$

Also because $R_{i-1}(0,0) \neq 0$ (by the inductive hypothesis given by Condition 3), and from the assumption that $\overline{V}_j(0,0) \neq 0$, (i.e., R_i/R_{i-1} is $(0,\overline{m}_j,\overline{n}_j)$-nonsingular), it follows from Theorem 5.5 that the leading coefficient $\overline{c}_{j-1}(z)$ of $\overline{R}_{j-1}(z,y)$ divides $\overline{U}_{j-1}(z,y)$, $\overline{V}_{j-1}(z,y)$ and $\overline{R}_{j-1}(z,y)$. Let l_{j-1} be the largest integer such that $z^{l_{j-1}}$ divides $\overline{c}_{j-1}(z)$. Then $z^{l_{j-1}}$ divides $\overline{U}_{j-1}(z,y)$, $\overline{V}_{j-1}(z,y)$ and $\overline{R}_{j-1}(z,y)$, and in addition $z^{-l_{j-1}}\overline{R}_{j-1}(z,y) \neq 0$ for $z=y=0$. Consequently, the execution of Step 3 of MPADE

$$\begin{bmatrix} \overline{U}_j & \overline{U}_{j-1} \\ \overline{V}_j & \overline{V}_{j-1} \end{bmatrix} \leftarrow \begin{bmatrix} \overline{U}_j & \overline{U}_{j-1} \\ \overline{V}_j & \overline{V}_{j-1} \end{bmatrix} \begin{bmatrix} z^{-l_j} & 0 \\ 0 & z^{-l_{j-1}} \end{bmatrix} \bmod z^{k+1},$$

with $l_j = 0$, yields $\overline{U}_j/\overline{V}_j$ and $\overline{U}_{j-1}/\overline{V}_{j-1}$ which satisfy (5.20) and (5.21), as well as

$$R_i\overline{V}_j + R_{i-1}\overline{U}_j = y^{\overline{m}_j+\overline{n}_j+s}\tilde{R}_j \tag{5.22}$$

for some $\tilde{R}_j$ and $s \geq \overline{s}_j$, and

$$R_i\overline{V}_{j-1} + R_{i-1}\overline{U}_{j-1} = y^{\overline{m}_{j-1}+\overline{n}_{j-1}+\overline{s}_{j-1}}\tilde{R}_{j-1} \tag{5.23}$$

for $\tilde{R}_{j-1}$ such that $\tilde{R}_{j-1}(0,0) \neq 0$.

Then, after Steps 5 and 6, using (5.12), (5.13), (5.15) and (5.22), it follows that, modulo y^{m+n+1}, U_{i+j}/V_{i+j} and U_{i+j-1}/V_{i+j-1} satisfy

$$\begin{aligned} AV_{i+j} + BU_{i+j} &= A(V_i\overline{V}_j + y^{s_{i-1}+s_i}V_{i-1}\overline{U}_j) \\ &\quad + B(U_i\overline{V}_j + y^{s_{i-1}+s_i}U_{i-1}\overline{U}_j) + O(z^{k+1}) \\ &= \overline{V}_j(AV_i + BU_i) + y^{s_{i-1}+s_i}\overline{U}_j(AV_{i-1} + BU_{i-1}) + O(z^{k+1}) \\ &= y^{m_i+n_i+s_i}(\overline{V}_jR_i + \overline{U}_jR_{i-1}) + O(z^{k+1}) \\ &= y^{m_i+n_i+s_i+\overline{m}_j+\overline{n}_j+s}\tilde{R}_j + O(z^{k+1}) \\ &= y^{m_{i+j}+n_{i+j}+s_{i+j}}R_{i+j} \bmod z^{k+1}y^{m+n+1} \end{aligned} \tag{5.24}$$

for some $s_{i+j} \geq s \geq \overline{s}_j \geq 1$ and $R_{i+j}(z,0) \neq 0$ if $n_{i+j} + s_{i+j} < n$. Furthermore, $V_{i+j}(0,0) = V_i(0,0)\overline{V}_j(0,0) \neq 0$. That is, U_{i+j}/V_{i+j} satisfy Conditions 2 and 4.

To prove that U_{i+j-1}/V_{i+j-1} satisfy Conditions 1 and 3, in a fashion similar to (5.24), it can be shown that, modulo y^{m+n+1},

$$AV_{i+j-1} + BU_{i+j-1} = y^{m_{i+j-1}+n_{i+j-1}+\overline{s}_{j-1}}\overline{R}_j + O(z^{k+1}).$$

Thus, $s_{i+j-1} = \overline{s}_{j-1}$ and

$$R_{i+j-1} = \overline{R}_j \bmod z^{k+1} \bmod y^{m-m_{i+j}+n-n_{i+j-1}+1-\overline{s}_{j-1}}.$$

It then follows that $R_{i+j-1}(0,0) \neq 0$, and consequently that $V_{i+j-1}(0,y) \neq 0$ (because $B(0,0) \neq 0$). Finally using (5.16), (5.17), (5.20), (5.21) and the results that $s_{i+j-1} = \bar{s}_j$ and $s_{i+j} \geq \bar{s}_j$, it is trivial to show that

$$s_{i+j-1} = m_{i+j} - m_{i+j-1} = n_{i+j} - n_{i+j-1} \geq 1$$

and

$$m_{i+j} - n_{i+j} = m_{i+j-1} - n_{i+j-1} = m - n,$$

where m_{i+j}, n_{i+j}, m_{i+j-1} and n_{i+j-1} are computed in Step 4 of MPADE. Thus, Conditions 5 and 6 are also satisfied.

Observe again that if $n_{i+j} + s_{i+j} < n$, then the termination condition in Step 8 fails. But then in equation (5.24), $R_{i+j}(z,0) \neq 0$, and with i replaced by $i+j$ (Step 7) all the Conditions 1 through 6 are satisfied during the next pass through MPADE. That is, we have resolved Case I.

Next, we consider Case II. That is, assume that Conditions 1 through 6 are satisfied for U_i/V_i and U_{i-1}/V_{i-1}, and also that, in Step 2 JPADE returns $\bar{U}_j/\bar{V}_j$ and $\bar{U}_{j-1}/\bar{V}_{j-1}$ which still satisfy equations (5.18) and (5.19). However instead of the condition $\bar{V}_j(0,0) \neq 0$ being satisfied as in Case I, now $\bar{U}_j/\bar{V}_j$ satisfies the condition that

$$\bar{n}_j + \bar{s}_j = n - n_i. \tag{5.25}$$

We shall show that, prior to the execution of Step 7, U_{i+j}/V_{i+j} and U_{i+j-1}/V_{i+j-1} are modular Padé (k,m_{i+j},n_{i+j})-form and modular Padé (k,m_{i+j-1},n_{i+j-1})-form, respectively, for A/B. In addition, we shall show that in Step 8 the termination condition is satisfied.

First we discuss U_{i+j}/V_{i+j}. From (5.18), after Step 3, it follows that $\overline{U}_j/\overline{V}_j$ satisfy either $\overline{U}_j(0,y)\neq 0$ or $\overline{V}_j(0,y)\neq 0$, or both. Furthermore, $\overline{V}_j(0,y)\neq 0$, since otherwise (5.22) and the condition that $R_{i-1}(0,0)\neq 0$ would imply that $U_i(0,y)=0$, also.

Similarly, using (5.23), it follows that $\overline{V}_{j-1}(0,y)\neq 0$.

In a fashion similar to (5.24), Steps 5 and 6 yield U_{i+j}/V_{i+j}, s_{i+j} and R_{i+j} such that

$$AV_{i+j} + BU_{i+j} = y^{m_{i+j}+n_{i+j}+s_{i+j}}R_{i+j}\, mod z^{k+1} mod y^{m+n+1}.$$

But, from (5.25) and the definition of h_{i+j} in Step 4, it follows that Step 6 must yield $s_{i+j} = \overline{s}_j$, because

$$n_{i+j} + \overline{s}_j = n_i + \overline{p}_j + \overline{s}_j = n.$$

Thus, the algorithm MPADE must terminate at Step 8.

It is still required to show that U_{i+j}/V_{i+j} is nontrivial. In fact, we shall show that $U_{i+j}(0,y)\neq 0$, or $V_{i+j}(0,y)\neq 0$. For suppose that both $U_{i+j}(0,y)=0$ and $V_{i+j}(0,y)=0$. Then, from the definition of U_{i+j} and V_{i+j} in Step 5, we get

$$V_i(0,y)\overline{V}_j(0,y) = -y^{s_{i-1}+s_i}V_{i-1}(0,y)\overline{U}_j(0,y) \tag{5.26}$$

and

$$U_i(0,y)\overline{V}_j(0,y) = -y^{s_{i-1}+s_i}U_{i-1}(0,y)\overline{U}_j(0,y). \tag{5.27}$$

Since $\overline{V}_j(0,y)\neq 0$, equations (5.26) and (5.27) imply that

$$U_i(0,y)V_{i-1}(0,y) - U_{i-1}(0,y)V_i(0,y) = 0. \tag{5.28}$$

But, $U_i(0,y)/V_i(0,y)$ is a Padé fraction of type (m_i,n_i) for the univariate power series

$A(0,y)/B(0,y)$ which is (m_i,n_i)-nonsingular, and $U_{i-1}(0,y)/V_{i-1}(0,y)$ is its predecessor. Thus, from results on univariate Padé theory (see, for example, Gragg [20], or Cabay and Choi [9]), we have

$$\frac{U_i(0,y)}{V_i(0,y)} = \frac{U_{i-1}(0,y)}{V_{i-1}(0,y)},$$

which contradicts (5.28).

Thus, U_{i+j}/V_{i+j} is a modular Padé (k,m_i,n_i)-form for A/B. In a similar fashion (but, now we use the fact that $\overline{R}_{j-1}(0,y)\neq 0$ to show that U_{i+j-1}/V_{i+j-1} is nontrivial), it follows that U_{i+j-1}/V_{i+j-1} is a modular Padé (k,m_{i+j-1},n_{i+j-1})-form for A/B.

To complete the proof of Case II, it is only necessary to observe that the scaling of U_i/V_i by y^{s_i} and U_{i-1}/V_{i-1} by $y^{s_{i-1}-1}$ in Step 8, just prior to exit, yield the modular Padé forms required in the statement of the theorem.

Finally it is an easy exercise to show that just prior to Step 7 on the first pass through MPADE, with $i=0$, U_{i+j}/V_{i+j} and U_{i+j-1}/V_{i+j-1} fall either under Case I or Case II.

The presentation of the algorithm MPADE above is geared towards the proof of its correctness. Before we consider the costs of using MPADE some modifications on improving its performance are given.

At each pass through the algorithm (in its present form), Step 6 computes the residuals R_{i+j} and R_{i+j-1} modulo y^{m+n+1}. Subsequently in Step 2, JPADE only requires $\overline{m}_j+\overline{n}_j+\overline{s}_j$ terms of the residuals. The modulo y^{m+n+1} operations are therefore wasteful, but they are required because $\overline{m}_j$, $\overline{n}_j$ and $\overline{s}_j$ are not known apriori.

In most implementation environments, by declaring the variables A, B, U_i, V_i, U_{i-1}, and V_{i-1} to be global variables, the computations of the terms of residuals can be performed in JPADE, rather than in MPADE, only as required.

To obtain a cursory cost analysis of MPADE, with the modification indicated as above, we first consider the case when the input power series $E = A/B$ is $(0,m_i,n_i)$-nonsingular for all i such that $n_i \leq n$ (that is, $E(0,y)$ is normal). In this case JPADE always terminates after just one step with a step size of 1. Thus, MPADE performs exactly the same computations as does Algorithm 3 of Cabay and Choi [9] for univariate power series with coefficients over a field D, except that now MPADE performs operations on polynomials in z of degree not larger than k, rather than on elements of D. The algorithm given by Cabay and Choi has been shown to have a cost complexity of $O((m+n)n)$ operations in D. Consequently, for $E(z,y) = A(z,y)/B(z,y)$ for which $E(0,y)$ is normal MPADE has a cost complexity of $O(k^2(m+n)n)$ operations in D if classical arithmetic is used. On the other hand if fast methods (requiring $O(k\log k)$ operations in D for the multiplications and divisions of polynomials of degree k) were used, the complexity of MPADE reduces (again when $E(0,y)$ is normal) to $O(k\log k(m+n)n)$ operations in D.

If $E(0,y)$ is not normal, the complexity of MPADE increases according to the nature of abnormality. If the distances between nonsingularities are bounded by some constant, then the complexity results again hold. Otherwise, the complexity of MPADE can become as large as the complexity of JPADE.

Chapter 6

Suggestions for Further Research

The focal point of the research presented in this thesis is a generalization of univariate Padé theory to bivariate functions. As has been pointed out in Chapter 1, such a generalization is not straightforward and there is no widely accepted, uniform definition of bivariate Padé approximants. The modular Padé forms defined in Chapter 3 are superior to other definitions in the sense that, even though they are never unique, the family of all modular Padé forms at a given point (k,m,n) can be characterized in simple terms for any power series. There are no similar results for other definitions, i.e., characterizations are not provided when approximants are not unique (which is always the case when power series is abnormal). Indeed, it is a trivial observation that for any definition of a Padé-like approximant, which gives the set I_E (c.f., Chapter 1) explicitly, uniqueness of approximants under general conditions, cannot be ensured.

One way of dealing with this problem is to require solutions in a super set of I_E, say I'_E, which in certain sense would be maximal. If the power series is normal (which is an underlying assumption for most of the results in this area), the modular Padé forms can easily be made unique and irreducible. It is sufficient in this case to expand the set I_E to include $\{(i,j)\colon k+1 \le i \le 2k \text{ and } j=0\}$. This is equivalent to the order condition

$$A(x,y)Q(x,y) + P(x,y) = O(y^{m+n+1}) + O(x^{2k+1}) + O(x^{k+1}y).$$

A rational form satisfying such an extended order condition can be obtained first by obtaining a modular Padé (k,m,n)-form for $A(x,y)$, and then by normalizing it so that along $y = 0$, it becomes a univariate Padé fraction of type (k,k) to $A(z,0)$. How to obtain similar results for abnormal power series remains to be investigated.

The results of Chapter 4, even though they are significant on their own, have been obtained primarily in order to facilitate the development of the algorithm MPADE. The specific choices of α_{i+1} and β_{i+1} have been made in order to establish a correspondence between PSRS's and subresultants. Because of this correspondence we were able to prove Theorem 5.5 and consequently justify the correctness of the algorithm MPADE.

Two ways of improving JPADE require further investigation.

(1) The choices of α_{i+1} and β_{i+1} are not optimal ones. The factor α_{i+1} in the pseudodivision (4.8), Step 6 of JPADE, may be made smaller. Also, β_{i+1} (Theorem 4.11) is not the largest constant in J that can be removed from the cofactors and the residual.

(2) The development of a recursive algorithm JPADE in the spirit of Algorithm 2 of Cabay and Choi [9] poses an interesting research problem.

Given the power series $E(z,y) = A(z,y)/B(z,y)$ and the integers k, m and n, the algorithm MPADE always produces a modular Padé (k,m,n)-form. If $E(z,y)$ is (k,m,n)-maximal then the result produced (according to Chapter 3) is unique up to a polynomial in z. Thus, in this case, we obtain all the modular Padé (k,m,n)-forms. If $E(z,y)$ is not (k,m,n)-maximal however, the modular Padé (k,m,n)-form computed by MPADE does not allow us to obtain all other modular Padé (k,m,n)-forms, as given in Theorem 3.7 of Chapter 3. This important problem is left for further investigation.

The efficiency of the algorithm MPADE depends on how often intermediate results can be truncated modulo z^{k+1}. It can be shown, by a counter-example, that if truncation is performed at each step, the algorithm may not always succeed, i.e., the cofactors U_{i+j}, V_{i+j} in Step 5 of MPADE may vanish. We were able to prove that at nonsingularities of the power series $E(0,y)$ truncations of the intermediate results modulo z^{k+1} is justified. It remains an interesting problem to determine:

(1) if less restrictive conditions (that is, conditions other than $(0,m,n)$-nonsingularities) for truncation modulo z^{k+1} can be found, and

(2) if truncations modulo z^{ck+1} can be performed at every step, where c is some minimal integer determined at each step of the algorithm.

Knowing apriori that $E(0,y)$ is normal, Algorithm 2 of Cabay and Choi [9] can be extended in the obvious way by replacing operations in D with operations in $D[z]$. This is possible because in this case all the computed coefficients become units in $D[[z]]$, and truncations modulo z^{k+1} can be performed at each step. Since the complexity of their Algorithm 2 is $O((m+n)\log^2(m+n))$ operations in D, the extended algorithm, using fast polynomial arithmetic on polynomials in $D[z]$ of degree at most k would have a complexity of $O(k(m+n)\log k\log^2(m+n))$ operations in D. It remains a subject for future research to develop an extended algorithm that does not fail at abnormalities of $E(0,y)$.

REFERENCES

[1] H. AKAIKE, *Block Toeplitz matrix inversion*, SIAM J. Appl. Math., 24 (1973), pp. 234-241.

[2] G.A. BAKER, JR., *Essentials of Padé approximants*, Academic Press, New York, 1975

[3] E.H. BAREISS, *Sylvester's identity and multistep integer-preserving Gaussian elimination*, Math. Comp. 22, (1968), pp. 565-578.

[4] D. BESSIS, *Padé approximants in quantum field theory*, in "*Padé approximants and their applications*", ed. P.R. Graves-Morris, Academic Press, London/New York, 1973.

[5] F. BROPHY and A.C. SALAZAR, *Consideration of the Padé approximants technique in synthesis of recursive digital filters*, IEEE Trans. Audio Electroaust., 21 (1973), pp. 500-505.

[6] W.S. BROWN and J.F. TRAUB, *On Euclid's algorithm and the theory of subresultants*, J. Assoc. Comput. Mach., 18 (1971), pp. 505-514.

[7] W.S. BROWN, *The subresultant PRS algorithm*, ACM Trans. Math. Software, 4 (1978), pp. 237-249.

[8] A. BULTHEEL, *Recursive relations for block Hankel and Toeplitz systems, Part I: Direct recursion*, J. Comp. Appl. Math., 10 (1984), pp. 301-328.

[9] S. CABAY and D.K. CHOI, *Algebraic computations of scaled Padé fractions*, SIAM J. Comput., 15 (1986), pp. 243-270.

[10] J.S.R. CHISHOLM, *Rational approximants defined from double power series*, Math. Comp., 27 (1973), pp. 841-848.

[11] J.S.R. CHISHOLM, *Generalizations of Padé approximants*, Circuits Systems Signal Process., 1 (1982), pp. 279-287.

[12] C.K. CHUI, P.W. SMITH and L.Y. SU, *A minimization problem related to Padé synthesis of recursive digital filters*, in "*Padé and rational approximation: theory and applications*", eds. E.B. Saff and R.S. Varga, Academic press, London/New York, 1977.

[13] G.E. COLLINS, *Subresultants and reduced polynomial remainder sequences*, J. Assoc. Comput. Mach., 14 (1967), pp. 128-142.

[14] A.K. COMMON and P.R. GRAVES-MORRIS, *Some properties of Chisholm approximants*, J. Inst. Maths. Applics., 13 (1974), pp. 229-232.

[15] A.A.M. CUYT, *A review of multivariate Padé approximation theory*, J. Comp. Appl. Math., 12 and 13 (1985), pp. 221-232.

[16] A.A.M. CUYT, *The ε-algorithm and multivariate Padé-approximants*, Numer. Math., 40 (1982), pp. 39-46.

[17] A.A.M. CUYT, *The QD-algorithm and multivariate Padé-approximants*, Numer. Math., 42 (1983), pp. 259-269.

[18] A.A.M. CUYT, *A comparison of some multivariate Padé-approximants*, SIAM J. Math. Anal., 14 (1983), pp. 195-202.

[19] K.O. GEDDES, *Symbolic computation of Padé approximants*, ACM Trans. Math. Software, 5 (1979), pp. 218-233.

[20] W.B. GRAGG, *The Padé table and its relation to certain algorithms of numerical analysis*, SIAM Review, 14 (1972), p. 1-62.

[21] P.R. GRAVES-MORRIS, R. HUGHES JONES and G.J. MAKINSON, *The calculation of some rational approximants in two variables*, J. Inst. Maths. Applics., 13 (1974), pp. 311-320.

[22] R. HUGHES JONES, *General rational approximants in N-variables*, J. Approx. Theory, 16 (1976), pp. 201-233.

[23] R. HUGHES JONES and G.J. MAKINSON, *The generation of Chisholm rational polynomial approximants to power series in two variables*, J. Inst. Maths. Applics., 13 (1974), pp. 299-310.

[24] J. KARLSSON and H. WALLIN, *Theory of Padé approximants and generalizations*, in "*Padé and rational approximation: theory and applications*", eds. E.B. Saff and R.S. Varga, Academic Press, London/New York, 1977.

[25] D.E. KNUTH, *The art of computer programming: seminumerical algorithms*, Vol 2, Addison-Wesley, Reading, MA, 1969.

[26] D. LEVIN, *General order Padé-type rational approximants defined from double power series*, J. Inst. Maths. Applics., 18 (1976), pp. 1-8.

[27] M. MARCUS, *Finite Dimensional Multilinear Algebra*, Part I, Marcel Dekker, Inc., New York, 1973.

[28] M.I. SOBHY, *Applications of Padé approximants in electrical network problems*, in "*Padé approximants and their applications*", ed. P.R. Graves-Morris, Academic Press, London/New York, 1973.

[29] P.J.S. WATSON, *Algorithms for differentiation and integration*, in "*Padé approximants and their applications*", ed. P.R. Graves-Morris, Academic Press, London/New York, 1973.

[30] M. WAX and T. KAILATH, *Efficient inversion of Toeplitz-block Toeplitz matrix*, IEEE Trans. Acoust. Speech Signal Process., 31 (1983), pp. 1218-1221.