A Stable Algorithm for Multi-dimensional Padé Systems and the Inversion of Generalized Sylvester Matrices

Stan Cabay*, Anthony R. Jones† and George Labahn‡

Abstract. For $k+1$ power series $a_0(z), \ldots, a_k(z)$, we present a new iterative, look-ahead algorithm for numerically computing Padé-Hermite systems and simultaneous Padé systems along a diagonal of the associated Padé tables. The algorithm computes the systems at all those points along the diagonal at which the associated striped Sylvester and mosaic Sylvester matrices are well-conditioned. It is shown that a good estimate for the condition numbers of these Sylvester matrices at a point is easily determined from the Padé-Hermite system and simultaneous Padé system computed at that point. The operation and the stability of the algorithm is controlled by a single parameter τ which serves as a threshold in deciding if the Sylvester matrices at a point are sufficiently well-conditioned. We show that the algorithm is weakly stable, and provide bounds for the error in the computed solutions as a function of τ. Experimental results are given which show that the bounds reflect the actual behavior of the error.

The algorithm requires $\mathcal{O}(\|n\|^2 + s^2\|n\|)$ operations, to compute Padé-Hermite and simultaneous Padé systems of type $n = [n_0, \ldots, n_k]$, where $\|n\| = n_0 + \cdots + n_k$ and s is the largest step-size taken along the diagonal. An additional application of the algorithm is the stable inversion of striped and mosaic Sylvester matrices.

Key Words. Padé-Hermite approximants, simultaneous Padé approximants, striped Hankel inverses, striped Sylvester inverses, mosiac Sylvester inverses, numerical algorithm, numerical stability

AMS(MOS) Subject Classification. 41A21, 65F05, 65G05

1. Introduction. Let $A^t(z) = [a_0(z), \ldots, a_k(z)]$, $k \geq 1$, be a vector of formal power series over the real numbers[1] with $a_0(0) \neq 0$ and let $n = [n_0, \ldots, n_k]$ be a vector of integers with $n_\beta \geq -1, 0 \leq \beta \leq k$, and with at least one $n_\beta \geq 0$. A *Padé-Hermite approximant* of type n for $A(z)$ is a nontrivial vector $[q_0(z), \ldots, q_k(z)]$ of polynomials $q_\beta(z)$ over the real numbers having degrees[2] at most $n_\beta, 0 \leq \beta \leq k$, such that

$$a_0(z)q_0(z) + \cdots + a_k(z)q_k(z) = c_{\|n\|+k} z^{\|n\|+k} + c_{\|n\|+k+1} z^{\|n\|+k+1} + \cdots, \tag{1}$$

with $\|n\| = n_0 + \ldots + n_k$.

The *Padé-Hermite approximation problem* was introduced in 1873 by Hermite and has been widely studied by several authors (for a bibliography, see, for example [2, 3, 5, 6, 23]). Note that for $A^t(z) = [-1, a(z)]$, equation (1) becomes

$$a(z)q_1(z) - q_0(z) = O(z^{n_0+n_1+1}).$$

Thus, as a special case we have the classical Padé approximation problem for the power series $a(z)$. The Padé-Hermite approximation problem also includes other

* Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada, T6G 2H1. The research of this author as partially supported by Natural Sciences and Engineering Research Council of Canada grant A8035.

† Bell Northern Research, P.O. Box 3511, Station C, Ottawa, Ontario, Canada, K1Y 4H7

‡ Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, N2L3G1. The research of this author was partially supported by Natural Sciences and Engineering Research Council of Canada grant FS1525C.

[1] The restriction to real numbers is made in order to simplify floating point analysis. All of the results given in this paper also hold with minor modifications for the field of complex numbers.

[2] By convention, a polynomial of degree -1 is the zero polynomial.

classical approximation problems such as the algebraic approximants where $A^t(z) = [1, a(z), a(z)^2, \ldots, a(z)^k]$ (see [26] for the special case $k = 2$) and G^3J approximants where $A^t(z) = [1, a(z), a'(z)]$. Additional examples can be found in [1].

Closely related to Padé-Hermite approximants are *simultaneous Padé approximants.* A simultaneous Padé approximant of type n for $A(z)$ is a nontrivial vector $[q_0^*(z), \ldots, q_k^*(z)]$ of polynomials $q_\beta^*(z)$ over the real numbers having degrees of at most $\|n\| - n_\beta, 0 \le \beta \le k$, such that

$$(2) \qquad q_0^*(z) \cdot a_\beta(z) + q_\beta^*(z) \cdot a_0(z) = c_{\|n\|+k}^{(\beta)} z^{\|n\|+k} + c_{\|n\|+k+1}^{(\beta)} z^{\|n\|+k+1} + \cdots,$$

for $\beta = 1, \cdots, k$. Simultaneous Padé approximants were also defined by Hermite and were used in his famous proof of the transcendence of e. Again, for $A^t(z) = [-1, a(z)]$, the simultaneous Padé approximation problem becomes the classical Padé approximation problem for $a(z)$.

By equating coefficients in (1), the Padé-Hermite approximation problem can be viewed as solving a system of linear equations of size $\|n\| \times \|n\|$. Thus, one can use Gaussian elimination to solve this problem with a complexity of $\mathcal{O}(\|n\|^3)$ operations. However, the coefficient matrix of the corresponding linear system has a type of "structured" form so it is not surprising that there are a number of fast [2, 12] $\mathcal{O}(\|n\|^2)$ and superfast [6, 10] $\mathcal{O}(\|n\| \log^2 \|n\|)$ algorithms for determining Padé-Hermite approximants. All these algorithms have the property that they work for any input vector of power series. In addition, these algorithms all make important use of exact arithmetic; in particular, they all depend on knowing that certain quantities are known to be 0 or not. A similar statement also applies for the fast and superfast computation of simultaneous Padé approximants.

In the special case of Padé approximants, it has long been known that existing fast and superfast Padé algorithms all had problems with numerical stability for certain problems. In this case the first known numerically stable algorithm for fast Padé approximation was presented by Cabay and Meleshko [13]. Alternate algorithms for fast Padé computation that also consider the issue of numerical stability include [11, 14, 16] and [18], and for superfast computation [20]. An insightful look into the connection between stable algorithms for computing Padé approximants and other algorithms in numerical analysis is given by Gutknecht and Gragg [19]. Algorithms dealing with the closely associated problem of stably computing fast rational interpolation include [8].

In this paper, we present a new algorithm for the computation of Padé-Hermite and simultaneous Padé *systems.* These systems are matrix polynomials which contain the desired multi-dimensional Padé approximant along with quantities that can be used to recursively or iteratively compute the next approximant along a well defined diagonal path. The algorithm works for all vectors of power series and is fast in the sense that it computes a system in $\mathcal{O}(\|n\|^2)$ operations in the generic case. In addition, we show that this algorithm is *weakly stable* in the sense that it provides good answers to well-conditioned problems. The algorithm is a look-ahead procedure that computes the systems of type n by computing all the Padé systems at the well-conditioned locations along a diagonal path in the associated Padé tables passing through the point n. In the case of Padé approximation ($k = 1$), the algorithm reduces to the Cabay and Meleshko algorithm.

It is known (cf. [10] or [23]) that in exact arithmetic a Padé-Hermite system exists uniquely if and only if the striped Sylvester coefficient matrix of the corresponding associated linear system is nonsingular. This is also true for simultaneous Padé systems where the coefficient matrix of the associated linear system is now a mosaic Sylvester rather than a striped Sylvester matrix. However, in the case of floating point arithmetic determining that such coefficient matrices are nonsingular is not good enough. Instead one must know, at least in a reasonably computable way that the linear systems are also well-conditioned. Central to the stable operation of our algorithm is the ability to estimate the condition numbers of the associated striped Sylvester and mosaic Sylvester matrices. The estimates follow from some "near" inverse formulae for these matrices that are derived in this paper and which are expressed in terms of both Padé-Hermite and simultaneous Padé systems. This is the reason why our algorithm computes Padé-Hermite and simultaneous Padé systems in tandem; the inverse formulae, and consequently the estimates for the condition numbers, require that both the Padé-Hermite and the simultaneous Padé systems be available. The striped Sylvester and mosaic Sylvester matrices are deemed to be well-conditioned if the computed estimates of the condition numbers are bounded by some specified "stability" tolerance τ.

As a corollary to our results, there is a formula which gives the inverse of a striped Sylvester matrix expressed in terms of the associated Padé-Hermite system only. One attempt to use this formula to develop a stable algorithm for computing Padé-Hermite systems (independent of simultaneous Padé systems) was only partly successful [22]; bounds for the inverse of the associated striped Sylvester matrix (and consequently bounds for its condition number) using the formula were often too pessimistic and impractical.

This paper is organized as follows. Preliminary definitions and basic facts about Padé-Hermite and simultaneous Padé systems are given in the next two sections. §4 gives a near commutativity relationship between these two systems in floating point arithmetic while §5 gives the algorithm for computing these systems. The remainder of the paper is devoted to showing that the algorithm is weakly stable for the computation of either system. To this end, §6 discusses norms for matrix polynomials and power series while §7 and §8 discuss the errors that result from the iterative steps of the algorithm. §9 and §10 provide the necessary material for determining our stability parameter by creating approximate inversion formulae for striped and mosaic Sylvester matrices. §11 completes the proof of stability while §12 provides results of some numerical experiments that reflect the theoretic results of the previous sections. The final section gives some conclusions and a discussion of further areas of research.

2. Padé-Hermite Systems. In this section, we introduce the notion of a Padé-Hermite system for a vector of formal power series. Let

$$A^t(z) = [a_0(z), \ldots, a_k(z)]\,, \tag{3}$$

where

$$a_\alpha(z) = \sum_{\ell=0}^{\infty} a_\alpha^{(\ell)}\, z^\ell, \qquad \alpha = 0, \ldots, k,$$

with $a_\alpha^{(\ell)} \in \mathcal{F}$, the field of real numbers. Assume that $a_0^{(0)} \neq 0$, which means that $a_0^{-1}(z)$ exists. Let $n = [n_0, \ldots, n_k]$ and $\|n\| = n_0 + \cdots + n_k$. Then the $(k+1) \times (k+1)$

matrix of polynomials

$$S(z)=\left[\begin{array}{c|c} z^2p(z) & U^t(z) \\ \hline z^2Q(z) & V(z) \end{array}\right]=\left[\begin{array}{c|ccc} z^2p(z) & u_1(z) & \cdots & u_k(z) \\ \hline z^2q_1(z) & v_{1,1}(z) & \cdots & v_{1,k}(z) \\ \vdots & \vdots & & \vdots \\ z^2q_k(z) & v_{k,1}(z) & \cdots & v_{k,k}(z) \end{array}\right] \tag{4}$$

is a Padé-Hermite system (PHS) [12] of type n for $A(z)$ if the following conditions are satisfied.
I. **(Degree conditions)**: For $1 \le \alpha, \beta \le k$,

$$\begin{aligned} p(z) &= \sum_{\ell=0}^{n_0-1} p^{(\ell)} z^\ell, & u_\beta(z) &= \sum_{\ell=0}^{n_0} u_\beta^{(\ell)} z^\ell, \\ q_\alpha(z) &= \sum_{\ell=0}^{n_\alpha-1} q_\alpha^{(\ell)} z^\ell, & v_{\alpha,\beta}(z) &= \sum_{\ell=0}^{n_\alpha} v_{\alpha,\beta}^{(\ell)} z^\ell. \end{aligned} \tag{5}$$

II. **(Order condition)**:

$$A^t(z)S(z) = z^{\|n\|+1}T^t(z), \tag{6}$$

where $T^t(z) = [r(z), W^t(z)]$ with $W^t(z) = [w_1(z), \ldots, w_k(z)]$ is the residual.
III. **(Nonsingularity condition)**: The constant term of $V(z)$ is a diagonal matrix,

$$V(0) = diag\ [\gamma_1, \ldots, \gamma_k], \tag{7}$$

and

$$\gamma \equiv (a_0^{(0)})^{-1} \prod_{\alpha=0}^{k} \gamma_\alpha \ \neq 0, \tag{8}$$

where $\gamma_0 = r(0)$.

Remark 1: Only the first column of $S(z)$ is a Padé-Hermite approximant as defined in §1; this being of type $[n_0-1, \ldots, n_k-1]$. The remaining columns $S(z)$ do not quite satisfy the order condition (1) and are therefore not Padé-Hermite approximants; these columns serve primarily to facilitate the computation of the first column using the algorithm given later in §5. But there are other uses for these columns of $S(z)$, such as that of expressing the inverse of a striped Sylvester matrix (see the inverse formula (90)).

Remark 2: The nonsingularity condition III is equivalent to the condition that $r(0) \neq 0$ and that $V(0)$ be a nonsingular diagonal matrix.

Remark 3: The PHS is said to be **normalized** [12] if the nonsingularity condition III is replaced by $r(0) = 1$ and $V(0) = I_k$. This can be achieved by multiplying $S(z)$ on the right by Γ^{-1}, where

$$\Gamma = diag\,[\gamma_0, \ldots, \gamma_k]\,. \tag{9}$$

The PHS is said to be **scaled** [22] if each column of $S(z)$ has norm equal to 1 for some norm and if, in addition, $\gamma_\beta > 0$, $0 \le \beta \le k$. Here, also, scaling a PHS is accomplished by multiplying it on the right by an appropriate diagonal matrix.

Remark 4: The nonsingularity condition III, namely $\gamma \neq 0$, refers to the nonsingularity of $S(z)$; that is, $S(z)$ is nonsingular iff $\gamma \neq 0$. Equivalently, the nonsingularity condition refers to the nonsingularity of the associated striped Sylvester matrix $\mathcal{M}_n$ defined in (14) below; in [12] it is shown that a PHS (with $\gamma \neq 0$) exists iff $\mathcal{M}_n$ is nonsingular.

If the order condition (6) is not satisfied exactly, but rather

$$A^t(z)S(z) = z^{\|n\|+1}T^t(z) + \delta T^t(z), \tag{10}$$

where $\delta T^t(z) = \left[z^2\, \delta r(z), \delta W^t(z)\right]$ with $\delta W^t(z) = [\delta w_1(z), \ldots, \delta w_k(z)]$ is a relatively "small" residual error, then $S(z)$ is called a numerical Padé-Hermite system (NPHS). In (10), for $1 \leq \beta \leq k$,

$$\begin{aligned} \delta r(z) &= \sum_{\ell=0}^{\|n\|-2} \delta r^{(\ell)}\, z^{\ell}, \\ \delta w_\beta(z) &= \sum_{\ell=0}^{\|n\|} \delta w_\beta^{(\ell)}\, z^{\ell}. \end{aligned}$$

If $\delta T^t(z) = 0$, then $S(z)$ is an exact (rather than a numerical) Padé-Hermite system. To distinguish it from a NPHS $S(z)$, an exact system is denoted by $S_E(z)$.

The following lemma shows that Remark 4 applies to a NPHS as well; that is, $S(z)$ is nonsingular for sufficiently small $\delta T^t(z)$ if $\gamma \neq 0$.

Lemma 1. *If $S(z)$ is a NPHS of type n for $A(z)$, then*

$$det[S(z)] = z^{\|n\|+1}\gamma + \theta_I(z), \tag{11}$$

where

$$\theta_I(z) = a_0^{-1}(z)\delta T^t(z) \begin{bmatrix} \det[V(z)] \\ -z^2\, V^{adj}(z)Q(z) \end{bmatrix} (mod\ z^{\|n\|+2}).$$

Proof. Let

$$\phi(z) = \det[V(z)]$$

and

$$\Omega(z) = -V^{\mathrm{adj}}(z)Q(z).$$

From (4), the first column of $S^{\mathrm{adj}}(z)$ is $\left[\phi(z),\ z^2\Omega^t(z)\right]^t$ and satisfies

$$S(z) \begin{bmatrix} \phi(z) \\ z^2\, \Omega(z) \end{bmatrix} = \begin{bmatrix} z^2\, p(z) & U^t(z) \\ z^2\, Q(z) & V(z) \end{bmatrix} \begin{bmatrix} \phi(z) \\ z^2\, \Omega(z) \end{bmatrix} = \begin{bmatrix} \det[S(z)] \\ 0 \end{bmatrix}. \tag{12}$$

Multiplying both sides of (10) on the right by $\left[\phi(z),\ z^2\, \Omega^t(z)\right]^t$, it follows from (12) that

$$a_0(z)\det[S(z)] \;=\; A^t(z) \begin{bmatrix} \det[S(z)] \\ 0 \end{bmatrix} \tag{13}$$

$$
\begin{aligned}
&= A^t(z)\, S(z) \begin{bmatrix} \phi(z) \\ z^2\, \Omega(z) \end{bmatrix} \\
&= z^{\|n\|+1} T^t(z) \begin{bmatrix} \phi(z) \\ z^2\, \Omega(z) \end{bmatrix} + \delta T^t(z) \begin{bmatrix} \phi(z) \\ z^2\, \Omega(z) \end{bmatrix}.
\end{aligned}
$$

But, from the degree conditions (5), $\partial(\det[S(z)]) \leq \|n\| + 1$, and so (13) becomes

$$
\begin{aligned}
\det[S(z)] &= z^{\|n\|+1} r(0)\, \phi(0) (a_0^{(0)})^{-1} + a_0^{-1}(z) \delta T^t(z) \begin{bmatrix} \phi(z) \\ z^2\, \Omega(z) \end{bmatrix} \ (mod\ z^{\|n\|+2}) \\
&= z^{\|n\|+1} \gamma + \theta_I(z).
\end{aligned}
$$

∎

Associated with $A(z)$, let $\mathcal{M}_n$ be the striped Sylvester matrix of order $\|n\|$,

$$
(14) \qquad \mathcal{M}_n = \left[\begin{array}{ccc|c|ccc}
a_0^{(0)} & & & & a_k^{(0)} & & \\
 & \ddots & & & & \ddots & \\
\vdots & & a_0^{(0)} & \cdots & \vdots & & a_k^{(0)} \\
 & & \vdots & & & & \vdots \\
a_0^{(\|n\|-1)} & \cdots & a_0^{(\|n\|-n_0)} & & a_k^{(\|n\|-1)} & \cdots & a_k^{(\|n\|-n_k)}
\end{array} \right].
$$

Then $S(z)$ can be obtained by solving two sets of linear equations with $\mathcal{M}_n$ as the coefficient matrix [12]. From (10),

$$
(15) \qquad a_0(z)\, p(z) + \sum_{\alpha=1}^{k} a_\alpha(z)\, q_\alpha(z) = z^{\|n\|-1} r(z) + \delta r(z),
$$

which gives rise to

$$
(16) \qquad \mathcal{M}_n \cdot \mathcal{X} = [0, \ldots, 0, \gamma_0]^t,
$$

where

$$
\mathcal{X} = \left[p^{(0)}, \ldots, p^{(n_0-1)} | q_1^{(0)}, \ldots, q_1^{(n_1-1)} | \cdots | q_k^{(0)}, \ldots, q_k^{(n_k-1)} \right]^t.
$$

The solution $\mathcal{X}$ yields the first column $S_{0,0}(z), S_{1,0}(z), \ldots, S_{k,0}(z)$ of $S(z)$. In (16), we require that $\gamma_0 = r(0) \neq 0$; $\gamma_0 = 1$ for a normalized NPHS. The existence of a solution to (16) is assured if $\mathcal{M}_n$ is nonsingular. The term $\delta r(z)$ in (15) represents the residual error made in solving (16)

Next, to compute $U^t(z)$ and $V(z)$ (i.e., the remaining columns of $S(z)$), again we use (6), namely,

$$
(17) \quad a_0(z)\, u_\beta(z) + \sum_{\alpha=1}^{k} a_\alpha(z)\, v_{\alpha,\beta}(z) = z^{\|n\|+1} w_\beta(z) + \delta w_\beta(z),\ 1 \leq \beta \leq k.
$$

For $\alpha, \beta = 1, \ldots, k$, set

$$
(18) \qquad
\begin{aligned}
u_\beta^{(0)} &= -\frac{a_\beta^{(0)}}{a_0^{(0)}} \gamma_\beta, \\
v_{\alpha,\beta}^{(0)} &= \begin{cases} \gamma_\beta, & \alpha = \beta, \\ 0, & \alpha \neq \beta. \end{cases}
\end{aligned}
$$

This yields the constant terms $U^t(0)$ and $V(0)$ of $U^t(z)$ and $V(z)$, respectively. The remaining components

$$(19)\quad \mathcal{Y} = \left[\begin{array}{ccc|ccc|c|ccc} u_1^{(1)} & \cdots & u_1^{(n_0)} & v_{1,1}^{(1)} & \cdots & v_{1,1}^{(n_1)} & & v_{k,1}^{(1)} & \cdots & v_{k,1}^{(n_k)} \\ \vdots & & \vdots & \vdots & & \vdots & \cdots & \vdots & & \vdots \\ u_k^{(1)} & \cdots & u_k^{(n_0)} & v_{1,k}^{(1)} & \cdots & v_{1,k}^{(n_1)} & & v_{k,k}^{(1)} & \cdots & v_{k,k}^{(n_k)} \end{array}\right]^t$$

can be obtained by solving

$$(20)\qquad \mathcal{M}_n \cdot \mathcal{Y} = -\begin{bmatrix} a_0^{(1)} & \cdots & a_k^{(1)} \\ \vdots & & \vdots \\ a_0^{(\|n\|)} & \cdots & a_k^{(\|n\|)} \end{bmatrix} \left[\frac{U^t(0)}{V(0)}\right].$$

In (20), we require that $\gamma_\beta \neq 0, 1 \leq \beta \leq k$; $\gamma_\beta = 1$ for a normalized NPHS. Again, the existence of a solution to (20) is assured if $\mathcal{M}_n$ is nonsingular. The terms $\delta w_\beta(z)$, $1 \leq \beta \leq k$, in (17) represent the residual errors made in solving (20).

For the special case when $n = [n_0, 0, \ldots, 0]$ the NPHS becomes

$$(21)\qquad S(z) = \left[\begin{array}{c|c} [a_0^{(0)}]^{-1} z^{n_0+1} & U^t(z) \\ \hline 0 & I_k \end{array}\right] \cdot diag[\gamma_0, \cdots, \gamma_k],$$

where $U^t(z) = -[a_0(z)]^{-1} \cdot [a_1(z), \cdots, a_k(z)]\ (mod\ z^{n_0+1})$. For initialization purposes in the algorithm given later in §5, we adopt (21) even in the cases $n_0 = 0$ and $n_0 = -1$, despite the fact that it no longer strictly meets all the requirements of an NPHS.

EXAMPLE 2. For the power series $A(z) = [a_0(z), a_1(z), a_2(z)]^t$, where

$$\begin{aligned} a_0(z) &= 1 - z + 2z^2 - 2z^3 + 3z^4 - 3z^5 + 4z^6 - 4z^7 + 5z^8 - 5z^9 \ldots, \\ a_1(z) &= 2z + 3z^3 + 4z^5 + 5z^7 + 6z^9 \ldots, \\ a_2(z) &= -1 + z + 5z^2 + 3z^3 + 2z^4 - 2z^5 - 6z^6 + z^7 - 8z^8 + 5z^9 \ldots, \end{aligned}$$

the associated striped Sylvester matrix of type n=[2,3,1] is

$$(22)\qquad \mathcal{M}_n = \left[\begin{array}{rr|rrr|r} 1 & 0 & 0 & 0 & 0 & -1 \\ -1 & 1 & 2 & 0 & 0 & 1 \\ 2 & -1 & 0 & 2 & 0 & 5 \\ -2 & 2 & 3 & 0 & 2 & 3 \\ 3 & -2 & 0 & 3 & 0 & 2 \\ -3 & 3 & 4 & 0 & 3 & -2 \end{array}\right]$$

To obtain the normalized Padé-Hermite system $S_E(z)$ of type n=[2,3,1] for $A(z)$, first solve (16) which yields

$$(23)\qquad \mathcal{X} = \frac{1}{37}[-4, 44, -22, 36, -9, -4]^t.$$

Next, the system (20) becomes

$$\mathcal{M}_n \cdot \mathcal{Y} = \begin{bmatrix} -2 & 0 \\ 0 & -7 \\ -3 & -1 \\ 0 & -5 \\ -4 & 5 \\ 0 & 2 \end{bmatrix},$$

which gives

$$(24) \qquad \mathcal{Y} = \frac{1}{37} \begin{bmatrix} -73 & -44 \\ -48 & 3 \\ -13 & -131 \\ -9 & 137 \\ -7 & 123 \\ 1 & -44 \end{bmatrix}.$$

The solutions (23) and (24) then give

$$(25)\ S_E(z) = \frac{1}{37} \left[\begin{array}{c|cc} z^2(-4+44z) & -73z-48z^2 & 37-44z+3z^2 \\ \hline z^2(-22+36z-9z^2) & 37-13z-9z^2-7z^3 & -131z+137z^2+123z^3 \\ z^2(-4) & z & 37-44z \end{array} \right].$$

Note that

$$A^t(z)\, S_E(z) = z^7\, T^t(z),$$

where

$$(26) \quad T^t(z) = \frac{1}{37}\left[37+20z+42z^2+\ldots,\ -5+8z-4z^2+\ldots,\ 516-130z+805z^2+\ldots\right].$$

3. Simultaneous Padé Systems. A Padé-Hermite system gives an approximation to a vector of formal power series using matrix multiplication on the right. In this section we give the definition of a simultaneous Padé system which corresponds to a similar approximation but with matrix multiplication on the left and with degree constraints that can be thought of as being "dual" to the degree constraints of a Padé-Hermite system. As in the previous section, a simultaneous Padé system exists if and only if a particular matrix of Sylvester type is nonsingular, in this case it is a mosaic Sylvester matrix.

Let

$$(27) \qquad A^*(z) = \left[\begin{array}{ccc} a^*_{0,1}(z) & \cdots & a^*_{0,k}(z) \\ \hline a^*_{1,1}(z) & \cdots & a^*_{1,k}(z) \\ \vdots & & \vdots \\ a^*_{k,1}(z) & \cdots & a^*_{k,k}(z) \end{array} \right] = \left[\begin{array}{c} B^{*t}(z) \\ \hline C^*(z) \end{array} \right]$$

be a $(k+1)\times k$ matrix of power series with $\det(C^*(0)) \neq 0$. The $(k+1)\times(k+1)$ matrix of polynomials

$$(28)\ S^*(z) = \left[\begin{array}{c|c} v^*(z) & U^{*t}(z) \\ \hline z^2Q^*(z) & z^2P^*(z) \end{array} \right] = \left[\begin{array}{c|ccc} v^*(z) & u_1^*(z) & \cdots & u_k^*(z) \\ \hline z^2q_1^*(z) & z^2p^*_{1,1}(z) & \cdots & z^2p^*_{1,k}(z) \\ \vdots & \vdots & & \vdots \\ z^2q_k^*(z) & z^2p^*_{k,1}(z) & \cdots & z^2p^*_{k,k}(z) \end{array} \right]$$

is a simultaneous Padé system (SPS) [10, 12] of type n for $A^*(z)$ if the following conditions are satisfied.

I. **(Degree conditions)**: For $1 \le \alpha, \beta \le k$,

$$(29) \qquad v^*(z) = \sum_{\ell=0}^{\|n\|-n_0} v^{*(\ell)} z^\ell, \qquad u^*_\beta(z) = \sum_{\ell=0}^{\|n\|-n_\beta} u_\beta^{*(\ell)} z^\ell,$$

$$q^*_\alpha(z) = \sum_{\ell=0}^{\|n\|-n_0-1} q_\alpha^{*(\ell)} z^\ell, \qquad p^*_{\alpha,\beta}(z) = \sum_{\ell=0}^{\|n\|-n_\beta-1} p^{*(\ell)}_{\alpha,\beta} z^{(\ell)}.$$

II. **(Order condition)**:

$$S^*(z)A^*(z) = z^{\|n\|+1}T^*(z), \tag{30}$$

where $T^{*t}(z) = [W^*(z)|R^{*t}(z)]$ with $R^*(z)$ a $k \times k$ matrix.
III. **(Nonsingularity condition)**: The constant term of $R^*(z)$ is a diagonal matrix

$$R^*(0) = diag\ [\gamma_1^*, \ldots, \gamma_k^*], \tag{31}$$

and

$$\gamma^* \equiv (a_0^{(0)})^{-1} \prod_{\alpha=0}^{k} \gamma_\alpha^* \neq 0, \tag{32}$$

where $\gamma_0^* = v^*(0)$.

<u>Remark 5:</u> The SPS is said to be **normalized** [10] if the nonsingularity condition III is replaced by $v^*(0) = 1$ and $R^*(0) = I_k$. This can be achieved by multiplying $S^*(z)$ on the left by Γ^{*-1}, where

$$\Gamma^* = diag\,[\gamma_0^*, \ldots, \gamma_k^*]\,. \tag{33}$$

The SPS is said to be **scaled** when each row of $S^*(z)$ has norm equal to 1 for some norm and if, in addition, $\gamma_\alpha^* > 0$, $0 \leq \alpha \leq k$. Here, also, scaling a SPS is accomplished by multiplying it on the left by an appropriate diagonal matrix.

<u>Remark 6:</u> The nonsingularity condition III, namely $\gamma^* \neq 0$, refers to the nonsingularity of $S^*(z)$; that is, $S^*(z)$ is nonsingular iff $\gamma^* \neq 0$. Equivalently, the nonsingularity condition refers to the nonsingularity of the associated mosaic Sylvester matrix $\mathcal{M}_n^*$ defined in (35); in [10] it is shown that a SPS exists iff $\mathcal{M}_n^*$ is nonsingular.

As for the Padé-Hermite system, if the order condition (30) is not satisfied exactly, but rather

$$S^*(z)A^*(z) = z^{\|n\|+1}T^*(z) + \delta T^*(z), \tag{34}$$

where $\delta T^{*t}(z) = \left[\delta W^*(z)|z^2\,\delta R^{*t}(z)\right]$ (with $\delta R^*(z)$ a $k \times k$ matrix) is a relatively "small" residual error, then $S^*(z)$ is called a numerical simultaneous Padé system (NSPS). In (34), for $1 \leq \alpha, \beta \leq k$,

$$\begin{aligned} \delta w_\beta^*(z) &= \sum_{\ell=0}^{\|n\|} \delta w_\beta^{*(\ell)}\, z^\ell, \\ \delta r_{\alpha,\beta}^*(z) &= \sum_{\ell=0}^{\|n\|-2} \delta r_{\alpha,\beta}^{*(\ell)}\, z^{(\ell)}. \end{aligned}$$

As with the NPHS $S(z)$, a NSPS for which $\delta T^*(z) = 0$ is denoted by $S_E^*(z)$.

Associated with $A^*(z)$, let $\mathcal{M}_n^*$ be the mosaic Sylvester matrix of order $k\|n\|$,

$$\mathcal{M}_n^* = \begin{bmatrix} \mathcal{S}_{0,1}^* & \cdots & \mathcal{S}_{0,k}^* \\ \vdots & & \vdots \\ \mathcal{S}_{k,1}^* & \cdots & \mathcal{S}_{k,k}^* \end{bmatrix}, \tag{35}$$

where, for $0 \le \alpha \le k$ and $1 \le \beta \le k$,

$$\mathcal{S}^*_{\alpha,\beta} = \begin{bmatrix} a^{*(0)}_{\alpha,\beta} & & \cdots & a^{*(\|n\|-1)}_{\alpha,\beta} \\ & \ddots & & \vdots \\ & & a^{*(0)}_{\alpha,\beta} \quad \cdots & a^{*(n_\alpha)}_{\alpha,\beta} \end{bmatrix}.$$

Also define the order $k(\|n\|+1)$ matrix

$$(36) \qquad \mathcal{N}^*_n = \left[\begin{array}{c|ccc|c|ccc} & a^{*(1)}_{1,1} & \cdots & a^{*(\|n\|)}_{1,1} & & a^{*(1)}_{1,k} & \cdots & a^{*(\|n\|)}_{1,k} \\ C^*(0) & \vdots & & \vdots & \cdots & \vdots & & \vdots \\ & a^{*(1)}_{k,1} & \cdots & a^{*(\|n\|)}_{k,1} & & a^{*(1)}_{k,k} & \cdots & a^{*(\|n\|)}_{k,k} \\ \hline \mathbf{0} & \multicolumn{7}{c}{\mathcal{M}^*_n} \end{array}\right].$$

Then, as for the NPHS, $S^*(z)$ can be obtained by solving two sets of linear equations with $\mathcal{M}^*_n$ and $\mathcal{N}^*_n$ as the coefficient matrices (also see [12]).

To obtain $S^*_{0,1}(z), \ldots, S^*_{0,k}(z)$ of $S^*(z)$, we use

$$(37)\ v^*(z)\ a^*_{0,\beta}(z) + \sum_{\alpha=1}^{k} u^*_\alpha(z) a^*_{\alpha,\beta}(z) \ = z^{\|n\|+1} w^*_\beta(z) + \delta w^*_\beta(z), \quad 1 \le \beta \le k,$$

which is the first row of (34). Matching coefficients of $1, z, \ldots, z^{\|n\|}$ in (37) gives

$$(38) \quad \mathcal{X}^{*t} \cdot \mathcal{N}^*_n = \\ -v^{*(0)} \left[\ B^{*t}(0) \ \middle|\ a^{*(1)}_{0,1}, \quad \ldots, \quad a^{*(\|n\|)}_{0,1} \ \middle|\ \cdots \ \middle|\ a^{*(1)}_{0,k}, \quad \ldots, \quad a^{*(\|n\|)}_{0,k} \ \right],$$

where

$$\mathcal{X}^{*t} = [u^{*(0)}_1, \ldots, u^{*(0)}_k | v^{*(1)}, \ldots, v^{*(\|n\|-n_0)} | u^{*(1)}_1, \ldots, u^{*(\|n\|-n_1)}_1 | \cdots \\ \cdots | u^{*(1)}_k, \ldots, u^{*(\|n\|-n_k)}_k].$$

With $v^{*(0)} = \gamma^*_0 \ne 0$ specified ($\gamma^*_0 = 1$ for a normalized NSPS), a unique solution to (38) is assured if $\mathcal{M}^*_n$ is nonsingular, since by assumption $\det[C^*(0)] \ne 0$. The terms $\delta w^*_\beta(z)$ in (37) represent the residual errors made in solving (38).

Next, to compute $P^*(z)$ and $Q^*(z)$ (i.e., the remaining rows of $S^*(z)$), again we use (34), namely,

$$(39)\ q^*_\alpha(z)\ a^*_{0,\beta}(z) + \sum_{\rho=1}^{k} p^*_{\alpha,\rho}(z)\ a^*_{\rho,\beta}(z) = z^{\|n\|-1} r^*_{\alpha,\beta}(z) + \delta r^*_{\alpha,\beta}(z), \quad 1 \le \alpha, \beta \le k.$$

Let

$$\mathcal{Y}^{*t}_\alpha = \left[q^{*(0)}_\alpha, \ldots, q^{*(\|n\|-n_0-1)}_\alpha | p^{*(0)}_{\alpha,1}, \ldots, p^{*(\|n\|-n_1-1)}_{\alpha,1} | \cdots | p^{*(0)}_{\alpha,k}, \ldots, p^{*(\|n\|-n_k-1)}_{\alpha,k}\right].$$

Then, (39) and the requirement that $R^*(0) = diag[\gamma^*_1, \ldots, \gamma^*_k]$ yields

$$(40) \qquad \mathcal{Y}^{*t}_\alpha \cdot \mathcal{M}^*_n = \gamma^*_\alpha E^t_{\alpha\|n\|}, \quad 1 \le \alpha \le k,$$

where $E^t_{\alpha\|n\|}$ is the unit row vector of length $k\|n\|$ with a single 1 in position $\alpha\|n\|$. With $diag[\gamma_1^*, \ldots, \gamma_k^*]$ specified ($\gamma_\alpha^* = 1$ for a normalized NSPS), a solution of (40) exists uniquely if $\mathcal{M}_n^*$ is nonsingular. The solution $\mathcal{Y}_\alpha^*$ provides the αth row of $S^*(z)$; namely, $S^*_{\alpha,0}(z) = z^2 \cdot q^*_\alpha(z)$ and $S^*_{\alpha,\beta}(z) = z^2 \cdot p^*_{\alpha,\beta}(z)$, $1 \le \beta \le k$. The terms $\delta r^*_{\alpha,\beta}(z)$ in (39) represent the residual errors made in solving (40).

In the remainder of the paper, without loss of generality, we make the simplifying assumption that

$$A^*(z) = \left[\begin{array}{ccc} -a_1(z) & \cdots & -a_k(z) \\ \hline a_0(z) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & a_0(z) \end{array} \right]. \tag{41}$$

In this case, there is an important commutativity relationship between Padé-Hermite systems and simultaneous Padé systems, given later in §4. But, in our presentation, the residual $T^*(z)$ continues to take the more general form (27) rather than (41); because, for the computation of the NSPS for $T^*(z)$, which is required by the algorithm given in §5, the conversion of $T^*(z)$ from the form (27) to the form (41) by means of multiplication on the right by $R^{*-1}(z)$ introduces undesirable instabilities.

For the special case when $n = [n_0, 0, \ldots, 0]$, with $A^*(z)$ defined by (41), the NSPS becomes

$$S^*(z) \;=\; diag[\gamma_0^*, \cdots, \gamma_k^*] \left[\begin{array}{c|c} 1 & U^{*t}(z) \\ \hline 0 & [a_0^{(0)}]^{-1} z^{n_0+1} I_k \end{array} \right], \tag{42}$$

where $U^{*t}(z) = [a_0(z)]^{-1} \cdot [a_1(z), \cdots, a_k(z)]\ (mod\ z^{n_0+1})$. For initialization purposes in the algorithm given in §5, we adopt (42) even in the case when $n_0 = 0$ and $n_0 = -1$, despite the fact that it no longer strictly meet all the requirements of a NSPS.

With $A^*(z)$ defined by (41), it is easy to see that $\mathcal{M}_n$ is nonsingular if and only if $\mathcal{M}_n^*$ is. Indeed, we will later provide a relationship between the condition numbers of $\mathcal{M}_n$ and $\mathcal{M}_n^*$.

Example 3. Continuing with Example 2, the associated mosaic Sylvester matrix of type $n = [2, 3, 1]$ is

$$\mathcal{M}_n^* = \left[\begin{array}{rrrrrr|rrrrrr}
0 & -2 & 0 & -3 & 0 & -4 & 1 & -1 & -5 & -3 & -2 & 2 \\
 & 0 & -2 & 0 & -3 & 0 & & 1 & -1 & -5 & -3 & -2 \\
 & & 0 & -2 & 0 & -3 & & & 1 & -1 & -5 & -3 \\
 & & & 0 & -2 & 0 & & & & 1 & -1 & -5 \\
\hline
1 & -1 & 2 & -2 & 3 & -3 & & & & & & \\
 & 1 & -1 & 2 & -2 & 3 & & & & & & \\
 & & 1 & -1 & 2 & -2 & & & & & & \\
\hline
 & & & & & & 1 & -1 & 2 & -2 & 3 & -3 \\
 & & & & & & & 1 & -1 & 2 & -2 & 3 \\
 & & & & & & & & 1 & -1 & 2 & -2 \\
 & & & & & & & & & 1 & -1 & 2 \\
 & & & & & & & & & & 1 & -1
\end{array} \right], \tag{43}$$

and so

$$\mathcal{N}_n^* = \left[\begin{array}{rr|rrrrrr|rrrrrr}
1 & 0 & -1 & 2 & -2 & 3 & -3 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -2 & 3 & -3 & 4 \\
\hline
\mathbf{0} & & & & & & & \mathcal{M}_n^* & & & & & &
\end{array} \right].$$

The solution of (38),

$$\mathcal{X}^{*t} \cdot \mathcal{N}_n^* = \left[\begin{array}{rr|rrrrrr|rrrrrr} 0 & 1 & -2 & 0 & -3 & 0 & -4 & 0 & -1 & -5 & -3 & -2 & 2 & 6 \end{array} \right],$$

is

$$\mathcal{X}^{*t} = \frac{1}{37}\left[\begin{array}{cc|cccc|ccc|ccccc} 0 & 37 & -57 & 10 & 0 & 5 & 74 & -40 & -57 & 57 & 249 & -103 & -428 & -159 \end{array}\right]$$

and the solution of (40),

$$\mathcal{Y}^{*t}\cdot\mathcal{M}_n^* = \left[\begin{array}{cccccc|cccccc} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right],$$

is

$$\mathcal{Y}^{*t} = \frac{1}{37}\left[\begin{array}{cccc|ccc|ccccc} 22 & -48 & 37 & -24 & 0 & 44 & -52 & -22 & 48 & 117 & -136 & -147 \\ 4 & -2 & 0 & -1 & 0 & 8 & 4 & -4 & 2 & 28 & 19 & -20 \end{array}\right].$$

From $\mathcal{X}^{*t}$ and $\mathcal{Y}^{*t}$, it follows that the normalized simultaneous Padé system of type (2,3,1) is

$$S^*(z) = \frac{1}{37}\left[\begin{array}{l|l|l} 37-57z+10z^2+5z^4 & 74z-40z^2-57z^3 & -37+57z+249z^2-103z^3-428z^4-159z^5 \\ \hline z^2(22-48z+37z^2-24z^3) & z^2(44z-52z^2) & z^2(-22+48z+117z^2-136z^3-147z^4) \\ z^2(4-2z-z^3) & z^2(8z+4z^2) & z^2(-4+2z+28z^2+19z^3-20z^4) \end{array}\right]. \tag{44}$$

Note that

$$S^*(z)A^*(z) = z^7T^*(z),$$

where

$$T^*(z) = \frac{1}{37}\left\{\begin{bmatrix} 5 & -516 \\ 37 & 0 \\ 0 & 37 \end{bmatrix} + \begin{bmatrix} 0 & 329 \\ -24 & 131 \\ -1 & 7 \end{bmatrix} z + \begin{bmatrix} 10 & -772 \\ 74 & -373 \\ 0 & 23 \end{bmatrix} z^2 + \cdots\right\}. \tag{45}$$

4. Duality. Theorem 4 below gives a relationship between Padé-Hermite and simultaneous Padé systems which is crucial to the results of the subsequent sections. It generalizes earlier results of Mahler and their extensions to block matrices ([13, 23, 24, 25])

THEOREM 4. *If $S(z)$ is a NPHS of type n for $A(z)$ and $S^*(z)$ is a NSPS of type n for $A^*(z)$, then*

$$S^*(z)\cdot S(z) = z^{\|n\|+1}(a_0^{(0)})^{-1}\Gamma^*\Gamma + \theta_{II}(z), \tag{46}$$

where

$$\theta_{II}(z) \;=\; a_0^{-1}(z)\left\{\begin{bmatrix} v^*(z) \\ z^2Q^*(z) \end{bmatrix}\delta T^t(z) + \delta T^*(z)\left[\begin{array}{c|c} z^2Q(z) & V(z) \end{array}\right]\right\} \pmod{z^{D+1}}$$

with

$$D = \left[\begin{array}{c|ccc} \|n\|+1 & \|n\| & \cdots & \|n\| \\ \hline \|n\|+2 & \|n\|+1 & \cdots & \|n\|+1 \\ \vdots & \vdots & & \vdots \\ \|n\|+2 & \|n\|+1 & \cdots & \|n\|+1 \end{array}\right]$$

and with the modulo operation applied componentwise.

Proof. The theorem (in the case that $\delta T(z) = 0$ and $\delta T^*(z) = 0$) follows from [23]. The arguments used in the following proof, however, are considerably simpler. Let

$$B^t(z) = [a_1(z), \cdots, a_k(z)].$$

Then, using (10) and (34),

$$\begin{aligned}
(47)\ \ & a_0(z)\, S^*(z) \cdot S(z) \\
&= a_0(z) \left\{ \left[\begin{array}{c} v^*(z) \\ z^2 Q^*(z) \end{array} \right] \left[\ z^2 p(z) \mid U^t(z) \ \right] + \left[\begin{array}{c} U^{*t}(z) \\ z^2 P^*(z) \end{array} \right] \left[\ z^2 Q(z) \mid V(z) \ \right] \right\} \\
&= \left[\begin{array}{c} v^*(z) \\ z^2 Q^*(z) \end{array} \right] \left\{ a_0(z) \ \left[\ z^2 p(z) \mid U^t(z) \ \right] + B^t(z) \left[\ z^2 Q(z) \mid V(z) \ \right] \right\} \\
&\qquad + \left\{ a_0(z) \left[\begin{array}{c} U^{*t}(z) \\ z^2 P^*(z) \end{array} \right] - \left[\begin{array}{c} v^*(z) \\ z^2 Q^*(z) \end{array} \right] B^t(z) \right\} \left[\ z^2 Q(z) \mid V(z) \ \right] \\
&= \left[\begin{array}{c} v^*(z) \\ z^2 Q^*(z) \end{array} \right] A^t(z) S(z) + S^*(z) A^*(z) \left[\ z^2 Q(z) \mid V(z) \ \right] \\
&= z^{\|n\|+1} \left\{ \left[\begin{array}{c} v^*(z) \\ z^2 Q^*(z) \end{array} \right] \left[\ r(z) \mid W^t(z) \ \right] + \left[\begin{array}{c} W^{*t}(z) \\ R^*(z) \end{array} \right] \left[\ z^2 Q(z) \mid V(z) \ \right] \right\} \\
&\qquad + \left[\begin{array}{c} v^*(z) \\ z^2 Q^*(z) \end{array} \right] \delta T^t(z) + \delta T^*(z) \left[\ z^2 Q(z) \mid V(z) \ \right].
\end{aligned}$$

But, from (5) and (29), the degrees of $S^*(z)S(z)$ are bounded componentwise by D. It then follows from (47) that

$$\begin{aligned}
S^*(z)S(z) &= z^{\|n\|+1}(a_0^{(0)})^{-1} \left[\begin{array}{c|c} v^*(0)r(0) & 0 \\ \hline 0 & R^*(0)V(0) \end{array} \right] + \theta_{II}(z) \\
&= z^{\|n\|+1}(a_0^{(0)})^{-1} \Gamma^* \Gamma + \theta_{II}(z),
\end{aligned}$$

which is (46). ∎

Corollary 5. *If $S(z)$ is a* **normalized** *NPHS of type n for $A(z)$ and $S^*(z)$ is a* **normalized** *NSPS of type n for $A^*(z)$, then*

$$(48)\qquad S(z) \cdot S^*(z) = z^{\|n\|+1}(a_0^{(0)})^{-1} I_{k+1} + \theta_{III}(z),$$

where

$$\theta_{III}(z) = S(z)\ \Gamma^{*-1} \theta_{II}(z) \Gamma^{-1}\ S^{-1}(z).$$

Proof. Multiplying both sides of (46) on the left by Γ^{*-1} and on the right by Γ^{-1}, we obtain

$$(49)\qquad S^*(z) \cdot S(z) = z^{\|n\|+1}(a_0^{(0)})^{-1} I_{k+1} + \Gamma^{*-1} \theta_{II}(z) \Gamma^{-1}.$$

The result now follows by multiplying both sides of (49) on the left by $S(z)$ and on the right by $S^{-1}(z)$.

∎

Note that $S(z)$ and $S^*(z)$ in (48) are now normalized, but $\theta_{II}(z)$ continues to be associated with systems which are not.

Corollary 6. *The residuals $T(z)$ for a* **normalized** *NPHS of type n for $A(z)$ and $T^*(z)$ for a* **normalized** *NSPS of type n for $A^*(z)$ satisfy*

$$T^t(z)\, S^*(z) = (a_0^{(0)})^{-1}\, A^t(z) + \theta^t_{IV}(z), \tag{50}$$

where

$$\theta^t_{IV}(z) = \left\{ A^t(z)\theta_{III}(z) - \delta T^t(z) S^*(z)) \right\} / z^{\|n\|+1}.$$

Proof. From (10) and (48), it follows that

$$\begin{aligned} \left\{ z^{\|n\|+1}\, T^t(z) + \delta T^t(z) \right\} S^*(z) &= A^t(z)\, S(z)\, S^*(z) \\ &= A^t(z) \left\{ z^{\|n\|+1}\, (a_0^{(0)})^{-1} + \theta_{III}(z) \right\} \end{aligned}$$

and so (50) is true. ∎

5. The Algorithm. To compute a NPHS of type n for $A(z)$ and a NSPS of type n for $A^*(z)$, the systems (16), (20), (38) and (40) can be solved using a method such as Gaussian elimination. This method, while not restricting the input power series, does not take advantage of the inherent structure of the coefficient matrices $\mathcal{M}_n$ and $\mathcal{M}^*_n$. Alternatively, a variety of recurrence relations which do take advantage of this structure have been described in the literature ([2],[5], [10],[12]). These recurrence relations usually lead to much more efficient algorithms for algebraically computing Padé-Hermite systems and simultaneous Padé systems. The recurrence relations given in [10] and [12] appear to be the most easily adaptable to numerical computation and it is the detailed study of the numerical behavior of these recurrences that we devote the remainder of this paper. We begin by briefly describing these recurrences in the algebraic case.

Let $e_0 = [1, 0, \ldots, 0]$ be a $1 \times k+1$ vector, set

$$M = \min \left\{ n_0,\ \max_{1 \le \beta \le k} \{n_\beta\} \right\} + 1,$$

and define integer vectors $n^{(i)} = (n_0^{(i)}, \ldots, n_k^{(i)})$ for $0 \le i \le M$ by $n^{(0)} = -e_0$ and, for $i > 0$,

$$n_\beta^{(i)} = \max\{0,\, n_\beta - M + i\}, \qquad \beta = 0, \ldots, k.$$

Then the sequence $\{n^{(i)}\}_{i=0,1,\ldots}$ lies on a piecewise linear path with $n_\beta^{(i+1)} \ge n_\beta^{(i)}$ for each i, β and[3] $n^{(M)} = n$. The sequence $\{n^{(i)}\}$ contains a subsequence $\{m^{(\sigma)}\}$ called the **sequence of nonsingular points** for $A(z)$ and $A^*(z)$. This sequence is defined by $m^{(\sigma)} = n^{(i_\sigma)}$, where

$$i_\sigma = \begin{cases} 0, & \sigma = 0, \\ \min\{i > i_{\sigma-1} : \det(\mathcal{M}_{n^{(i)}}) \ne 0\}, & \sigma \ge 1, \end{cases}$$

[3] We assume here with loss of generality that $n_\beta \ge 0, 0 \le \beta \le k$, because if $n_\beta = -1$ for some β, we can simply remove n_β from n and $a_\beta(z)$ from $A^t(z)$ and decrease k by 1.

where $\det(\mathcal{M}_{n^{(i)}})$ is the determinant[4] of $\mathcal{M}_{n^{(i)}}$. Corresponding to the sequence of nonsingular points $\{m^{(\sigma)}\}$ is the sequence $\left\{S_E^{(\sigma)}(z)\right\}$ of Padé-Hermite systems with residuals $\left\{T_E^{(\sigma)^t}(z)\right\}$ and the sequence $\left\{S_E^{*(\sigma)}(z)\right\}$ of Padé-Hermite systems with residuals $\left\{T_E^{*(\sigma)}(z)\right\}$. We have that

$$A^t(z) \cdot S_E^{(\sigma)}(z) \;=\; z^{\|m^{(\sigma)}\|+1}\, T_E^{(\sigma)t}(z)$$

and

$$S_E^{*(\sigma)}(z) \cdot A^*(z) \;=\; z^{\|m^{(\sigma)}\|+1}\, T_E^{*(\sigma)}(z).$$

The following theorem provides a relation of the $(\sigma+1)$th exact systems in terms of the σth exact systems.

THEOREM 7. *For $\sigma \geq 1$ and $i > i_\sigma$, let $\nu = n^{(i)} - m^{(\sigma)} - e_0$. Then, the following statements are equivalent.*

1. *$n^{(i)}$ is a nonsingular point for $A(z)$ and $A^*(z)$.*
2. *ν is a nonsingular point for $T_E^{(\sigma)}(z)$.*
3. *ν is a nonsingular point for $T_E^{*(\sigma)}(z)$.*

Furthermore, we have the recurrence relations

$$S_E^{(\sigma+1)}(z) \;=\; S_E^{(\sigma)}(z) \,\cdot\, \widehat{S}_E(z), \quad T_E^{(\sigma+1)}(z) \;=\; \widehat{T}_E(z), \tag{51}$$

and

$$S_E^{*(\sigma+1)}(z) \;=\; \widehat{S}_E^*(z) \,\cdot\, S_E^{*(\sigma)}(z), \quad T_E^{*(\sigma+1)}(z) \;=\; \widehat{T}_E^*(z), \tag{52}$$

where $\widehat{S}_E(z)$ is the Padé-Hermite system of type $(m^{(\sigma+1)} - m^{(\sigma)} - e_0)$ for $T_E^{(\sigma)}(z)$ with residual $\widehat{T}_E(z)$ and $\widehat{S}_E^*(z)$ is the simultaneous Padé system of type $(m^{(\sigma+1)} - m^{(\sigma)} - e_0)$ for $T_E^{*(\sigma)}(z)$ with residual $\widehat{T}_E^*(z)$.

Proof. The proof for the NPHS is given in [12] and for the NSPS in [10]. ∎

Theorem 7 reduces the problem of determining a Padé-Hermite system and a simultaneous Padé system of types $m^{(\sigma+1)}$ to two smaller problems: determine systems of type $m^{(\sigma)}$ for the original power series and then determine systems of type $\nu = m^{(\sigma+1)} - m^{(\sigma)} - e_0$ for the residual power series. For the residual power series, the system $\widehat{S}_E(z)$ is obtained by solving the linear equations (16) and (20), where in the following the associated matrix is now denoted by $\widehat{\mathcal{M}}_\nu$ rather than by $\mathcal{M}_\nu$; and, the system $\widehat{S}_E^*(z)$ is obtained by solving the linear equations (38) and (40), where in the following the associated matrix is now denoted by $\widehat{\mathcal{M}}_\nu^*$ rather than by $\mathcal{M}_\nu^*$. The overhead cost of each step of this iterative scheme is the cost of determining the residual power series and the cost of combining the solutions, i.e., the cost of computing $S_E^{(\sigma+1)}(z)$ and $S_E^{*(\sigma+1)}(z)$ in (51) and (52). This overhead cost summed over all the steps, in general, is an order of magnitude less than the cost of solving the linear systems (16), (20), (38) and (40) directly.

EXAMPLE 8. Continuing with Example 2, we can compute the Padé-Hermite system of type $[3,4,2]$ by utilizing (25) and the recurrence relation (51). In order to do this, we compute the Padé-Hermite system of type $\nu = [3,4,2]-[2,3,1]-[1,0,0] =$

[4] By convention, the determinant of a null matrix is defined to be equal to 1.

$[0,1,1]$ for the residual $T_E(z)$ in (26). The striped Sylvester matrix associated with $T_E(z)$ is

$$\widehat{\mathcal{M}}_\nu = \frac{1}{37}\left[\begin{array}{r|r} -5 & 516 \\ 8 & -130 \end{array}\right]. \tag{53}$$

Using (53), equations (16) and (20) are solved to obtain the Padé-Hermite system

$$\widehat{S}_E(z) = \begin{bmatrix} 0 & \frac{105}{148} & -\frac{3096}{4255} \\ \frac{2064}{1739}z^2 & 1-\frac{24}{37}z & -\frac{10432}{60865}z \\ \frac{4025}{3478}z^2 & -\frac{805}{296}z & 1+\frac{2175}{3478}z \end{bmatrix}. \tag{54}$$

of type ν for $T_E(z)$. By multiplying $S_E(z)$ in (25) on the right by $\widehat{S}_E(z)$, we obtain the new Padé-Hermite system of type $[3,4,2]$,

$$S_E(z) = \frac{1}{94}\begin{bmatrix} 5z^2-1024z^3-669z^4 & -188z+94z^3 & 94-53z+3278z^2+549z^3 \\ 516z^2-199z^3-107z^4-81z^5 & 94-94z & -1954z+1489z^2-351z^3+821z^4 \\ 5z^2+8z^3 & 0 & 94-53z+28z^2 \end{bmatrix}.$$

Similarly, continuing with Example 3, we can compute the simultaneous Padé system of type $[3,4,2]$ by utilizing (44) and the recurrence relation (52). In order to do this we compute the simultaneous Padé system of type $\nu=[0,1,1]$ for the residual $T^*_E(z)$ in (45). The mosaic Sylvester matrix associated with $T^*_E(z)$ is

$$\widehat{\mathcal{M}}^*_\nu = \frac{1}{37}\begin{bmatrix} 5 & 0 & -516 & 329 \\ 0 & 5 & 0 & -516 \\ 37 & -24 & 0 & 131 \\ 0 & -1 & 37 & 7 \end{bmatrix}. \tag{55}$$

Using (55), equations (38) and (40) are solved to obtain the simultaneous Padé system

$$\widehat{S}^*_E(z) = \frac{1}{3478}\begin{bmatrix} 3478-81z-3032z^2 & -47+1017z & 48504-39568z \\ z^2(-19092-15130z) & 2580z^2 & -266256z^2 \\ z^2(-185-396z) & 25z^2 & -2580z^2 \end{bmatrix} \tag{56}$$

of type ν for $T^*_E(z)$. By multiplying $S^*_E(z)$ in (25) on the left by $\widehat{S}^*_E(z)$, we obtain the new simultaneous Padé system of type $(3,4,2)$,

$$S^*_E(z) = \frac{1}{94}\left[\begin{matrix} 94-147z+81z^2-28z^3 & 188z-106z^2-38z^3+53z^4-28z^5 \\ z^2(-516+386z-246z^2+188z^3+94z^5) & z^2(-1032z-260z^2-236z^3-246z^4) \\ z^2(-5-3z+8z^2) & z^2(-10z-16z^2+5z^3+8z^4) \end{matrix}\right.$$
$$\left.\begin{matrix} -94+147z+577z^2-249z^3-703z^4-153z^5-351z^6+821z^7 \\ z^2(516-386z-3366z^2-1614z^3+1882z^4+2996z^5+5370z^6) \\ z^2(5+3z-43z^2-61z^3+37z^4+107z^5+81z^6) \end{matrix}\right].$$

Numerically, the recurrences (51) and (52) perform badly if $\mathcal{M}_{m^{(\sigma)}}$ and $\mathcal{M}^*_{m^{(\sigma)}}$ are ill-conditioned at any point $m^{(\sigma)}$. Rather than moving from nonsingular point to nonsingular point along the diagonal, what we would like to do is move from a well-conditioned point to the next well-conditioned point. This is the motivation for the algorithm VECTOR_PADE given below, where the points $m^{(\sigma)}$, $\sigma=0,1,\ldots$, correspond to stable points rather than to nonsingular points and we step over unstable blocks.

A quantitative measure of the stability of a point $m^{(\sigma)}$ is provided by the stability parameter

$$\kappa^{(\sigma)} = \sum_{\beta=0}^{k}(\gamma_\beta^{(\sigma)}\gamma_\beta^{*(\sigma)})^{-1}. \tag{57}$$

We will show later in §9 and §10 that $\kappa^{(\sigma)}$ serves as a rough estimate for the condition numbers $\|\mathcal{M}_{m^{(\sigma)}}\|_1 \cdot \|\mathcal{M}_{m^{(\sigma)}}^{-1}\|_1$ of $\mathcal{M}_{m^{(\sigma)}}$ (cf. (95)) and $\|\mathcal{M}^*_{m^{(\sigma)}}\|_\infty \cdot \|\mathcal{M}^{*-1}_{m^{(\sigma)}}\|_\infty$ of $\mathcal{M}^*_{m^{(\sigma)}}$ (cf. (101)). For the estimate (57), it is assumed that $S^{(\sigma)}(z)$ and $S^{*(\sigma)}(z)$ are both scaled and that $\|a_\beta(z)\| \le 1, \quad 0 \le \beta \le k$. The norms used for the various scaling are defined in §6. In (57), it is also assumed that the residual errors $\delta T^{(\sigma)}(z)$ and $\delta T^{*(\sigma)}(z)$ in the order equations

$$A^t(z) \cdot S^{(\sigma)}(z) = z^{\|m^{(\sigma)}\|+1}\, T^{(\sigma)t}(z) + \delta T^{(\sigma)t}(z) \tag{58}$$

and

$$S^{*(\sigma)}(z) \cdot A^*(z) = z^{\|m^{(\sigma)}\|+1}\, T^{*(\sigma)}(z) + \delta T^{*(\sigma)}(z), \tag{59}$$

at the point $m^{(\sigma)}$ are relatively insignificant. We say that $m^{(\sigma)}$ is a **stable point** (or, a well-conditioned point) if for some preassigned tolerance τ, $\kappa^{(\sigma)} \le \tau$. In the algorithm below, the user supplies the tolerance value τ.

VECTOR_PADE($A(z)$, n, k, τ)
$\sigma \leftarrow 0$; $m^{(0)} \leftarrow -e_0$; $S^{(0)} \leftarrow I_{k+1}$; $S^{*(0)} \leftarrow I_{k+1}$;
$M \leftarrow \min\{n_0, \max_{1\le\beta\le k}\{n_\beta\}\} + 1$
$i \leftarrow 0$; *stable* $\leftarrow$ true
While (($i < M$) and *stable*) do
  $\nu \leftarrow n - m^{(\sigma)} - e_0$
  $s \leftarrow 0$; *stable* $\leftarrow$ false
  While ($s < M - i$) and (not *stable*) do
    $s \leftarrow s + 1$
    $\nu_\beta^{(s)} \leftarrow \max\{0, \nu_\beta + i - M + s\}$, $\beta = 0, \ldots, k$
    Compute the residuals $T^{(\sigma)}(z)$ and $T^{*(\sigma)}(z)$ in (58) and (59)
    Construct the matrices $\mathcal{M}_{\nu^{(s)}}$ for $T^{(\sigma)}(z)$ and $\mathcal{M}^*_{\nu^{(s)}}$ for $T^{*(\sigma)}(z)$
    If $\mathcal{M}_{\nu^{(s)}}$ is numerically nonsingular then
      $m^{(\sigma+1)} \leftarrow m^{(\sigma)} + \nu^{(s)} + e_0$
      Obtain $\widehat{S}(z)$ by solving (16) and (20) by Gaussian elimination
      $S^{(\sigma+1)}(z) \leftarrow S^{(\sigma)}(z)\, \widehat{S}(z)$
      Scale $S^{(\sigma+1)}(z)$ and compute $\Gamma^{(\sigma+1)}$
      Obtain $\widehat{S}^*(z)$ by solving (38) and (40) by Gaussian elimination
      $S^{*(\sigma+1)}(z) \leftarrow \widehat{S}^*(z)\, S^{*(\sigma)}(z)$
      Scale $S^{*(\sigma+1)}(z)$ and compute $\Gamma^{*(\sigma+1)}$
      Using (57), compute $\kappa^{(\sigma+1)}$
      stable $\leftarrow \kappa^{(\sigma+1)} \le \tau$
    end If
  end While
  If *stable* then $\sigma \leftarrow \sigma + 1$; $i \leftarrow i + s$
end While
If stable then return $(S^{(\sigma)}(z), S^{*(\sigma)}(z), \kappa^{(\sigma)})$ else return $(S^{(\sigma+1)}(z), S^{*(\sigma+1)}(z), \kappa^{(\sigma+1)})$

6. Norms and Floating Point Errors. In this section, some norms are defined for matrix power series and matrix polynomials. Proofs regarding some of the properties of these norms are straightforward and can be found in [22]. Also given are some results on floating-point errors that are used in later sections.

Let

$$a(z) = \sum_{\ell=0}^{\infty} a^{(\ell)} z^{\ell} \in \mathcal{F}\,[[z]]\,,$$

where $\mathcal{F}[[z]]$ is the domain of power series with coefficients from $\mathcal{F}$. Then a norm for $\mathcal{F}\,[[z]]$ is given by

$$\|a(z)\| = \sup_{0\leq\ell<\infty} \{|\ a^{(\ell)}\ |\} \tag{60}$$

for $a(z) \in \mathcal{F}\,[[z]]$. For some integer ∂, let

$$s(z) = \sum_{\ell=0}^{\partial} s^{(\ell)} z^{\ell}\ \in \mathcal{F}\,[z]\,,$$

where $\mathcal{F}[z]$ is the domain of polynomials with coefficients over $\mathcal{F}$. Then a norm of $s(z)$ is

$$\|s(z)\| = \sum_{\ell=0}^{\partial} |\ s^{(\ell)}\ |\ . \tag{61}$$

It is easy to show that

$$\|a(z)\cdot s(z)\|\ \leq\ \|a(z)\|\cdot\|s(z)\|, \tag{62}$$

and so the norm (61) for $\mathcal{F}[z]$ is compatible with the norm (60) for $\mathcal{F}\,[[z]]$. In addition, for fixed $s(z)$, the bound is reached for $a(z) = 1$. Therefore,

$$\|s(z)\| = \sup_{a(z)\neq 0} \frac{\|a(z)\, s(z)\|}{\|a(z)\|}.$$

Thus, (61) is the operator norm for $\mathcal{F}[z]$ induced by the norm (60) for $\mathcal{F}\,[[z]]$. Finally, for $s(z), t(z) \in \mathcal{F}\,[z]$, it can be shown that

$$\|s(z)+t(z)\|\ \leq\ \|s(z)\|+\|t(z)\| \tag{63}$$

and

$$\|s(z)\cdot t(z)\|\ \leq\ \|s(z)\|\cdot\|t(z)\|. \tag{64}$$

Next, let $A^t(z) = [a_0(z), \ldots, a_k(z)] \in \mathcal{F}_{k+1}\,[[z]]$ be a $1 \times k+1$ vector of power series with

$$a_\alpha(z)\ =\ \sum_{\ell=0}^{\infty} a_\alpha^{(\ell)} z^{\ell}, \quad \alpha = 0, \ldots, k.$$

A norm for $A^t(z)$ is given by

$$\|A^t(z)\| = \max_{0\leq\beta\leq k} \{\|a_\beta(z)\|\}\,. \tag{65}$$

Now, let[5] $S(z) \in \mathcal{F}_{(k+1)\times(k+1)}[z]$. Then $S(z)$ defines a mapping of $A^t(z) \in \mathcal{F}_{(k+1)}[[z]]$ to $A^t(z)S(z) \in \mathcal{F}_{(k+1)}[[z]]$. We use the norm

$$\|S(z)\| = \max_{0\le\beta\le k}\left\{\sum_{\alpha=0}^{k}\|S_{\alpha,\beta}(z)\|\right\} \tag{66}$$

for $\mathcal{F}_{(k+1)\times(k+1)}[z]$. Then,

$$\|S(z)\| = \sup_{A^t(z)\neq 0}\left\{\frac{\|A^t(z)\cdot S(z)\|}{\|A^t(z)\|}\right\},$$

so that (66) is the operator norm induced by the norm (65). Consequently, the compatibility condition

$$\|A^t(z)\cdot S(z)\| \le \|A^t(z)\|\cdot\|S(z)\| \tag{67}$$

is satisfied.

Finally, let $A^*(z) \in \mathcal{F}_{(k+1)\times k}[[z]]$ with

$$A^*(z) = \begin{bmatrix} a^*_{0,1}(z) & \cdots & a^*_{0,k}(z) \\ \vdots & & \vdots \\ a^*_{k,1}(z) & \cdots & a^*_{k,k}(z) \end{bmatrix}, \tag{68}$$

where $a^*_{\alpha,\beta}(z) \in \mathcal{F}[[z]]$. A norm for $A^*(z)$ is given by

$$\|A^*(z)\| = \max_{1\le\beta\le k}\left\{\sum_{\alpha=0}^{k}\|a^*_{\alpha,\beta}(z)\|\right\}. \tag{69}$$

Then, for[6] $S^*(z) \in \mathcal{F}_{(k+1)\times(k+1)}[z]$, we have that

$$\|S^*(z)\cdot A^*(z)\| \le \|S^*(z)\|\cdot\|A^*(z)\|. \tag{70}$$

In addition, for $S(z), S^*(z) \in \mathcal{F}_{(k+1)\times(k+1)}[z]$, it can be shown that

$$\|S(z)\cdot S^*(z)\| \le \|S(z)\|\cdot\|S^*(z)\|. \tag{71}$$

We now give some standard results from the field of floating point error analysis. Let μ denote the unit floating point error and assume that the degrees of all polynomials and the orders of all matrices are bounded by some N, where $N\mu \le 0.01$ (this restriction comes from Forsythe and Moler [15]). Indeed, as an assumption for all the lemmas and theorems below, we require that $(\|n\| + k + 1)\mu \le 0.01$. After Wilkinson [28], we denote a floating point operation by $fl[\cdot]$. In the following results, it is assumed that the operands consist of floating point numbers.

LEMMA 9. *If $\partial\mu \le 0.01$, then*

$$fl[\sum_{k=1}^{\partial} u_k v_k] = \sum_{k=1}^{\partial} u_k v_k(1+\delta_k),$$

[5] We are interested primarily in the case that $S(z)$ is a Padé-Hermite system.
[6] We are interested primarily in the case that $S^*(z)$ is a simultaneous Padé system.

where $|\delta_k| \leq 1.01\partial\mu$.

LEMMA 10. *If* $S(z)$ *is a NPHS of type* n *for* $A(z)$*, then*

$$fl[A^t(z) \cdot S(z)] = A^t(z) \cdot S(z) + \Psi^t(z),$$

where

$$||\Psi^t(z)|| \leq 1.01\mu(||n|| + k + 1)||A^t(z)|| \cdot ||S(z)||.$$

Proof. Using Lemma 9, for $0 \leq \beta \leq k$,

$$\begin{aligned} fl[\sum_{\alpha=0}^{k} a_\alpha(z) S_{\alpha,\beta}(z)] &= \sum_{\ell=0}^{\infty} z^\ell fl[\sum_{\alpha=0}^{k}\sum_{j=0}^{n_\alpha} a_\alpha^{(\ell-j)} S_{\alpha,\beta}^{(j)}] \\ &= \sum_{\ell=0}^{\infty} z^\ell \sum_{\alpha=0}^{k}\sum_{j=0}^{n_\alpha} a_\alpha^{(\ell-j)} S_{\alpha,\beta}^{(j)} (1 + \delta_{\alpha,\beta,j,\ell}), \end{aligned}$$

where $|\delta_{\alpha,\beta,j,\ell}| \leq 1.01(n_\alpha + k + 1)\mu$. So,

$$\Psi_\beta(z) = \sum_{\ell=0}^{\infty} z^\ell \sum_{\alpha=0}^{k}\sum_{j=0}^{n_\alpha} a_\alpha^{(\ell-j)} S_{\alpha,\beta}^{(j)} \delta_{\alpha,\beta,j,\ell},$$

and

$$\begin{aligned} ||\Psi^t(z)|| &= \max_{0\leq\beta\leq k}\{||\Psi_\beta(z)||\} \\ &\leq \max_{0\leq\beta\leq k}\left\{ \sup_{0\leq\ell<\infty} \left[\sum_{\alpha=0}^{k}\sum_{j=0}^{n_\alpha} |a_\alpha^{(\ell-j)}| \cdot |S_{\alpha,\beta}^{(j)}| \cdot |\delta_{\alpha,\beta,j,\ell}| \right] \right\} \\ &\leq 1.01\mu \max_{0\leq\beta\leq k}\left\{ \sum_{\alpha=0}^{k} (n_\alpha + k + 1)||a_\alpha(z)|| \sum_{j=0}^{n_\alpha} |S_{\alpha,\beta}^{(j)}| \right\} \\ &\leq 1.01\mu \max_{0\leq\alpha\leq k}\{n_\alpha + k + 1\}||A^t(z)|| \max_{0\leq\beta\leq k}\left\{ \sum_{\alpha=0}^{k} ||S_{\alpha,\beta}(z)|| \right\} \\ &\leq 1.01\mu(||n|| + k + 1)||A^t(z)|| \cdot ||S(z)||. \end{aligned}$$

∎

LEMMA 11. *If* $S^*(z)$ *be a NSPS of type* n *for* $A^*(z)$*, then*

$$fl[S^*(z) \cdot A^*(z)] = S^*(z) \cdot A^*(z) + \Psi^*(z),$$

where

$$||\Psi^*(z)|| \leq 1.01\mu(||n|| + 1)||S^*(z)|| \cdot ||A^*(z)||.$$

Proof. Using Lemma 9, for $0 \le \alpha \le k$ and $1 \le \beta \le k$,

$$
\begin{aligned}
fl[-S^*_{\alpha,0}(z)a_\beta(z) + S^*_{\alpha,\beta}(z)a_0(z)] &= \sum_{\ell=0}^{\infty} z^\ell fl[-\sum_{j=0}^{\|n\|-n_0} S^{*(j)}_{\alpha,0} a^{(\ell-j)}_\beta + \sum_{j=0}^{\|n\|-n_\beta} S^{*(j)}_{\alpha,\beta} a^{(\ell-j)}_0] \\
&= \sum_{\ell=0}^{\infty} z^\ell \left\{ - \sum_{j=0}^{\|n\|-n_0} S^{*(j)}_{\alpha,0} a^{(\ell-j)}_\beta (1+\delta_{\alpha,0,j,\ell}) \right. \\
&\qquad \left. + \sum_{j=0}^{\|n\|-n_\beta} S^{*(j)}_{\alpha,\beta} a^{(\ell-j)}_0 (1+\delta_{\alpha,\beta,j,\ell}) \right\},
\end{aligned}
$$

where, for all α, β, j and ℓ, $|\delta_{\alpha,\beta,j,\ell}| \le 1.01(\|n\| - n_\beta + 1)\mu$. So,

$$
\Psi^*_{\alpha,\beta}(z) = \sum_{\ell=0}^{\infty} z^\ell \{ - \sum_{j=0}^{\|n\|-n_0} -S^{*(j)}_{\alpha,0} a^{(\ell-j)}_\beta \delta_{\alpha,0,j,\ell} + \sum_{j=0}^{\|n\|-n_\beta} S^{*(j)}_{\alpha,\beta} a^{(\ell-j)}_0 \delta_{\alpha,\beta,j,\ell} \}
$$

and

$$
\begin{aligned}
\|\Psi^*(z)\| &= \max_{1\le\beta\le k} \{ \sum_{\alpha=0}^{k} \|\Psi^*_{\alpha,\beta}(z)\| \} \\
&\le \max_{1\le\beta\le k} \left\{ \sum_{\alpha=0}^{k} \sup_{0\le\ell<\infty} \left[\sum_{j=0}^{\|n\|-n_0} |S^{*(j)}_{\alpha,0}| \cdot |a^{(\ell-j)}_\beta| \cdot |\delta_{\alpha,0,j,\ell}| \right. \right. \\
&\qquad \left. \left. + \sum_{j=0}^{\|n\|-n_\beta} |S^{*(j)}_{\alpha,\beta}| \cdot |a^{(\ell-j)}_0| \cdot |\delta_{\alpha,\beta,j,\ell}| \right] \right\} \\
&\le \max_{1\le\beta\le k} \left\{ \sum_{\alpha=0}^{k} \left[1.01\mu(\|n\| - n_0 + 1)\|a_\beta(z)\| \sum_{j=0}^{\|n\|-n_0} |S^{*(j)}_{\alpha,0}| \right. \right. \\
&\qquad \left. \left. + \ 1.01\mu(\|n\| - n_\beta + 1)\|a_0(z)\| \sum_{j=0}^{\|n\|-n_\beta} |S^{*(j)}_{\alpha,\beta}| \right] \right\} \\
&\le 1.01\mu(\|n\|+1) \max_{1\le\beta\le k} \left\{ \sum_{\alpha=0}^{k} \|a_\beta(z)\| \cdot \|S^*_{\alpha,0}(z)\| + \|a_0(z)\| \cdot \|S^*_{\alpha,\beta}(z)\| \right\} \\
&\le 1.01\mu(\|n\|+1)\|S^*(z)\| \max_{1\le\beta\le k} \{\|a_\beta(z)\| + \|a_0(z)\|\} \\
&\le 1.01\mu(\|n\|+1)\|S^*(z)\| \cdot \|A^*(z)\|.
\end{aligned}
$$

∎

7. Error Analysis for Padé-Hermite Systems. In this section, we obtain bounds for the error in the order condition for the NPHS computed by the algorithm VECTOR_PADE. We begin by first analysing the floating point errors introduced by one iteration of the algorithm. At the σth iteration, the NPHS $S^{(\sigma)}(z)$ of type $m^{(\sigma)}$ for $A^t(z)$ is available and satisfies

$$
A^t(z) \cdot S^{(\sigma)}(z) = \delta T^{(\sigma)^t}(z) + \mathcal{O}(z^{\|m^{(\sigma)}\|+1}).
$$

The algorithm proceeds to compute $S^{(\sigma+1)}(z)$ of type $m^{(\sigma+1)}$.

An iterative step consists of three parts. In the first part, the first $||\nu^{(\sigma)}||+1$ terms of $T^{(\sigma)}(z)$ are computed; a bound for the floating point errors introduced in this part is given in Lemma 12 below. In the second part, the NPHS $\widehat{S}^{(\sigma)}(z)$ of type $\nu^{(\sigma)}$ for $T^{(\sigma)}(z)$ is computed; an error analysis is given Lemma 13. In the third part, Lemma 14 provides bounds for the floating point errors introduced in computing $S^{(\sigma+1)}(z) = S^{(\sigma)}(z) \cdot \widehat{S}^{(\sigma)}(z)$. At this point in the algorithm, $S^{(\sigma+1)}(z)$ is scaled so that the norm of each column is 1. We assume for the sake of simplicity that this scaling introduces no additional errors. This is reasonable assumption because errors due to scaling are comparatively insignificant[7].

Lemma 12. *The computed residual $T^{(\sigma)}(z)$ satisfies*

$$z^{||m^{(\sigma)}||+1} T^{(\sigma)^t}(z) = A^t(z) \cdot S^{(\sigma)}(z) - \delta T^{(\sigma)^t}(z) + z^{||m^{(\sigma)}||+1} \theta_V^{(\sigma)^t}(z),$$

where

$$||\theta_V^{(\sigma)^t}(z)|| \le 1.01(||m^{(\sigma)}|| + k + 1) \cdot \mu.$$

Proof. The algorithm computes the first $||\nu^{(\sigma)}||+1$ terms of the residual only. That is,

$$\begin{aligned} z^{||m^{(\sigma)}||+1} T^{(\sigma)^t}(z) &= fl[A^t(z) \cdot S^{(\sigma)}(z)] \pmod{z^{||m^{(\sigma+1)}||+1}} \\ &\quad - fl[A^t(z) \cdot S^{(\sigma)}(z)] \pmod{z^{||m^{(\sigma)}||+1}}. \end{aligned}$$

Thus,

$$A^t(z) \cdot S^{(\sigma)}(z) \pmod{z^{||m^{(\sigma+1)}||+1}} = \delta T^{(\sigma)^t}(z) + z^{||m^{(\sigma)}||+1}[T^{(\sigma)^t}(z) - \theta_V^{(\sigma)^t}(z)],$$

where $\theta_V^{(\sigma)^t}(z)$ is the error introduced into the computation of $T^{(\sigma)^t}(z)$ by floating point operations. The result now follows from Lemma 10 since $A^t(z)$ and $S^{(\sigma)}(z)$ are both scaled. ∎

Lemma 13. *If $\widehat{\mathcal{M}}_{\nu^{(\sigma)}}$ is nonsingular and $\widehat{S}^{(\sigma)}(z)$ is obtained by solving (16) and (20), then*

$$T^{(\sigma)^t}(z) \cdot \widehat{S}^{(\sigma)}(z) = \theta_{VI}^{(\sigma)^t}(z) + O(z^{||\nu^{(\sigma)}||+1}),$$

where

$$||\theta_{VI}^{(\sigma)^t}(z)|| \le (16||\nu^{(\sigma)}||^4 \cdot \rho_\sigma \cdot \mu + O(\mu^2)) \cdot ||\widehat{S}^{(\sigma)}(z)||.$$

Proof. First we obtain bounds for the first component of $\theta_{VI}^{(\sigma)^t}(z)$. The first column of $\widehat{S}^{(\sigma)}(z)$ corresponds to the solution $\widehat{\mathcal{X}}$ of (16) obtained by Gaussian elimination. $\widehat{\mathcal{X}}$ is the exact solution of

$$(\widehat{\mathcal{M}}_{\nu^{(\sigma)}} + \mathcal{E}) \cdot \widehat{\mathcal{X}} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

[7] Note also that $\widehat{S}^{(\sigma)}(z)$ can be determined apostiori with appropriate values of $\hat{\gamma}^{(\sigma)}$ so that $S^{(\sigma+1)}(z)$ is already scaled. None of the subsequent error bounds would change, and so in reality this assumption is made without loss of generality.

where[8]

$$||\mathcal{E}||_1 \leq 8||\nu^{(\sigma)}||^3 \cdot \rho_\sigma \cdot ||\widehat{\mathcal{M}}_{\nu^{(\sigma)}}||_1 \cdot \mu + O(\mu^2)$$

and ρ_σ is the growth factor associated with the LU-decomposition of $\widehat{\mathcal{M}}^t_{\nu^{(\sigma)}}$ ([17](page 67)).

But, from Lemma 10,

$$||T^{(\sigma)^t}(z)|| \leq 1 + 1.01 \cdot (||m^{(\sigma)}|| + k + 1) \cdot \mu,$$

since $A(z)$ and $S^{(\sigma)}(z)$ are both scaled. So,

$$||\widehat{\mathcal{M}}_{\nu^{(\sigma)}}||_1 \leq ||\nu^{(\sigma)}|| \cdot ||T^{(\sigma)^t}(z)|| \leq ||\nu^{(\sigma)}|| \cdot \left\{1 + 1.01(||m^{(\sigma)}|| + k + 1) \cdot \mu\right\}.$$

Thus,

$$\widehat{\mathcal{M}}_{\nu^{(\sigma)}} \cdot \widehat{\mathcal{X}} - \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \;=\; -\mathcal{E} \cdot \widehat{\mathcal{X}},$$

where

$$\begin{aligned} ||\mathcal{E} \cdot \widehat{\mathcal{X}}||_1 &\leq \left\{8||\nu^{(\sigma)}||^4 \cdot \left[1 + 1.01(||m^{(\sigma)}|| + k + 1) \cdot \mu\right] \cdot \rho_\sigma \cdot \mu \;+\; O(\mu^2)\right\} \cdot ||\widehat{\mathcal{X}}||_1 \\ &\leq \;\left\{16||\nu^{(\sigma)}||^4 \cdot \rho_\sigma \cdot \mu \;+\; O(\mu^2)\right\} \cdot ||\widehat{\mathcal{X}}||_1. \end{aligned}$$

Here, we have used $1.01(||m^{(\sigma)}|| + k + 1) \cdot \mu \leq 1.01(||n|| + k + 1) \cdot \mu \leq 1$. A similar analysis can be done for solving (20) to obtain $\widehat{\mathcal{Y}}$. But $\widehat{\mathcal{X}}$ yields the first column of $\widehat{S}^{(\sigma)}(z)$ with residual error $\mathcal{E} \cdot \widehat{\mathcal{X}}$ and $\widehat{\mathcal{Y}}$ yields the remaining columns of $\widehat{S}^{(\sigma)}(z)$ with a corresponding residual error. Thus,

$$T^{(\sigma)^t}(z) \cdot \widehat{S}^{(\sigma)}(z) = \theta_{VI}^{(\sigma)^t}(z) + O(z^{||\nu^{(\sigma)}||+1}),$$

where

$$||\theta_{VI}^{(\sigma)^t}(z)|| \leq \left\{16||\nu^{(\sigma)}||^4 \cdot \rho_\sigma \cdot \mu + O(\mu^2)\right\} \cdot ||\widehat{S}^{(\sigma)}(z)||.$$

∎

LEMMA 14. *If* $S^{(\sigma+1)}(z) = fl(S^{(\sigma)}(z) \cdot \widehat{S}^{(\sigma)}(z))$, *then*

$$S^{(\sigma+1)}(z) \;=\; S^{(\sigma)}(z) \cdot \widehat{S}^{(\sigma)}(z) \;+\; \theta_{VII}^{(\sigma)}(z),$$

where

$$||\theta_{VII}^{(\sigma)}(z)|| \;\leq\; 1.01(||\nu^{(\sigma)}|| + k + 1) \cdot ||S^{(\sigma)}(z)|| \cdot ||\widehat{S}^{(\sigma)}(z)||\mu.$$

[8] Gaussian elimination is applied to $\widehat{\mathcal{M}}^t_{\nu^{(\sigma)}}$ so that the error bounds given in [17] hold for $||\cdot||_1$ rather than $||\cdot||_\infty$.

Proof. For $1 \leq \alpha, \beta \leq k$, the (α, β)-component of $S^{(\sigma+1)}(z)$ is

$$
\begin{aligned}
& fl\left[z^2 q_\alpha(z) \cdot \widehat{u}_\beta(z) + \sum_{\rho=1}^{k} v_{\alpha,\rho}(z) \cdot \widehat{v}_{\rho,\beta}(z)\right] \\
&= fl\left[z^2 \sum_{\ell=0}^{m_\alpha^{(\sigma)}+\nu_0^{(\sigma)}-1} z^\ell \sum_{j=0}^{\nu_0^{(\sigma)}} q_\alpha^{(\ell-j)} \widehat{u}_\beta^{(j)} \;+\; \sum_{\rho=1}^{k} \sum_{\ell=0}^{m_\alpha^{(\sigma)}+\nu_\rho^{(\sigma)}} z^\ell \sum_{j=0}^{\nu_\rho^{(\sigma)}} v_{\alpha,\rho}^{(\ell-j)} \widehat{v}_{\rho,\beta}^{(j)}\right] \\
&= \sum_{\ell=0}^{m_\alpha^{(\sigma)}+\nu_0^{(\sigma)}-1} z^{\ell+2} \sum_{j=0}^{\nu_0^{(\sigma)}} q_\alpha^{(\ell-j)} \widehat{u}_\beta^{(j)} \cdot (1 + \delta_{\alpha,\beta,j,\ell,0}). \\
&\qquad + \sum_{\ell=0}^{m_\alpha^{(\sigma)}+\nu_\rho^{(\sigma)}} z^\ell \sum_{\rho=1}^{k} \sum_{j=0}^{\nu_\rho^{(\sigma)}} v_{\alpha,\rho}^{(\ell-j)} \widehat{v}_{\rho,\beta}^{(j)} \cdot (1 + \delta_{\alpha,\beta,j,\ell,\rho}),
\end{aligned}
$$

where $|\delta_{\alpha,\beta,j,\ell,\rho}| \leq 1.01 \cdot (\nu_\rho^{(\sigma)} + k + 1) \cdot \mu$. Here, we have used Lemma 9 with the assumption that $(\|\nu^{(\sigma)}\| + k + 1)\mu \leq 0.01$. So,

$$
\begin{aligned}
(\theta_{VII}^{(\sigma)}(z))_{\alpha,\beta} &= z^2 \sum_{\ell=0}^{m_\alpha^{(\sigma)}+\nu_0^{(\sigma)}-1} z^\ell \sum_{j=0}^{\nu_0^{(\sigma)}} q_\alpha^{(\ell-j)} \cdot \widehat{u}_\beta^{(j)} \cdot \delta_{\alpha,\beta,j,\ell,0} \\
&\quad + \sum_{\rho=1}^{k} \sum_{\ell=0}^{m_\alpha^{(\sigma)}+\nu_\rho^{(\sigma)}} z^\ell \sum_{j=0}^{\nu_\rho^{(\sigma)}} v_{\alpha,\rho}^{(\ell-j)} \cdot \widehat{v}_{\rho,\beta}^{(j)} \cdot \delta_{\alpha,\beta,j,\ell,\rho}.
\end{aligned}
$$

Thus, from (64)

$$
\left\| \left(\theta_{VII}^{(\sigma)}(z)\right)_{\alpha,\beta} \right\| \leq 1.01 \cdot (\|\nu^{(\sigma)}\| + k + 1) \cdot \{\|q_\alpha(z)\| \cdot \|\widehat{u}_\beta(z)\| + \sum_{\rho=1}^{k} \|v_{\alpha,\rho}(z)\| \cdot \|\widehat{v}_{\rho,\beta}(z)\|\}\mu.
$$

An equivalent result holds for $\alpha = \beta = 0$. The lemma now follows using (71). ∎

The use of the results of the three lemmas above enables us to express the residual error $\delta T^{(\sigma+1)^t}(z)$ in the order condition at the $(\sigma+1)$th iteration in terms of the residual error $\delta T^{(\sigma)^t}(z)$ at the σth iteration plus the floating point errors introduced "locally" by the σth iteration.

LEMMA 15.

$$
\delta T^{(\sigma+1)^t}(z) = \delta T^{(\sigma)^t}(z) \cdot \widehat{S}^{(\sigma)}(z) + \mathcal{L}^{(\sigma)^t}(z), \tag{72}
$$

where

$$
\begin{aligned}
\mathcal{L}^{(\sigma)^t}(z) &= \Big\{A^t(z) \cdot \theta_{VII}^{(\sigma)}(z) \\
&\quad + z^{\|m^{(\sigma)}\|+1} \left[\theta_{VI}^{(\sigma)^t}(z) - \theta_V^{(\sigma)^t}(z) \cdot \widehat{S}^{(\sigma)}(z)\right]\Big\} \quad (mod \;\; z^{\|m^{(\sigma+1)}\|+1}).
\end{aligned}
$$

Proof. The result is an immediate consequence of Lemmas 12, 13 and 14. ∎

Thus, the residual error $\delta T^{(\sigma+1)^t}(z)$ is composed of the local error $\mathcal{L}^{(\sigma)^t}(z)$ introduced by the σth iteration plus the residual error $\delta T^{(\sigma)^t}(z)$ from the previous iteration propagated by $\widehat{S}^{(\sigma)}(z)$. Applying (72) recursively, we obtain the following.

THEOREM 16. *The residual error satisfies*

$$\delta T^{(\sigma+1)^t}(z) = \sum_{j=0}^{\sigma} \mathcal{L}^{(j)^t}(z) \cdot \mathcal{G}_j^{(\sigma)}(z), \tag{73}$$

where

$$\mathcal{G}_j^{(\sigma)}(z) = \begin{cases} \widehat{S}^{(j+1)}(z) \cdot \widehat{S}^{(j+2)}(z) \cdots \widehat{S}^{(\sigma)}(z), & 0 \le j < \sigma, \\ I_{k+1}, & j = \sigma. \end{cases} \tag{74}$$

Proof. The result follows by induction from Lemma 15. ∎

From (73), we see that the residual error $\delta T^{(\sigma+1)^t}(z)$ is composed of the local errors $\mathcal{L}^{(j)^t}(z)$ propagated by $\mathcal{G}_j^{(\sigma)}$. Lemmas 12, 13 and 14 provide bounds for $\mathcal{L}^{(j)^t}(z)$. To obtain a bound for $\delta T^{(\sigma+1)^t}(z)$, it remains to determine bounds for the propagation matrices $\mathcal{G}_j^{(\sigma)}$. The concern is that the $\widehat{S}^{(j)}(z)$ making up $\mathcal{G}_j^{(\sigma)}$ will cause $\mathcal{G}_j^{(\sigma)}$ to grow exponentially with σ. The next Lemma and Theorem show that this is not case; a bound is obtained for $\mathcal{G}_j^{(\sigma)}$ which is independent of σ. Hence, the local error $\mathcal{L}^{(j)^t}(z)$ introduced at iteration j and propagated to iteration $\sigma+1$ by $\mathcal{G}_j^{(\sigma)}$ does not grow with σ. Thus, in this sense, the error grows additively; that is, $\delta T^{(\sigma+1)^t}(z)$ is bounded by the sum of the bounds of the local errors at each iteration j.

LEMMA 17. *If μ is so small and $\delta T^{(\sigma)^t}(z)$ and $\delta T^{*(\sigma)}(z)$ are not too large so that*

$$\begin{aligned} &\kappa^{(\sigma)} \cdot |a_0^{(0)}| \cdot \Big\{ ||a_0^{-1}(z)|| \Big[(k+1)||\delta T^{(\sigma)^t}(z)|| + ||\delta T^{*(\sigma)}(z)|| \Big] \\ + \quad & 1.01(k+1)(||\nu^{(\sigma)}|| + k+1) \cdot \mu \Big\} \le \frac{1}{2}, \end{aligned}$$

then

$$||\widehat{S}^{(\sigma)}(z)|| \le 2\kappa^{(\sigma)} \cdot (k+1) \cdot |a_0^{(0)}|.$$

Proof. From (57),

$$\begin{aligned} ||(\Gamma^{*(\sigma)} \cdot \Gamma^{(\sigma)})^{-1} \cdot S^{*(\sigma)}(z) \cdot S^{(\sigma+1)}(z)|| \quad &\le \quad ||(\Gamma^{*(\sigma)} \cdot \Gamma^{(\sigma)})^{-1}|| \cdot ||S^{*(\sigma)}(z)|| \cdot ||S^{(\sigma+1)}(z)|| \\ &\le \quad \kappa^{(\sigma)} \cdot (k+1). \end{aligned}$$

But, using Theorem 4 and Lemma 14

$$\begin{aligned} &||(\Gamma^{*(\sigma)} \cdot \Gamma^{(\sigma)})^{-1} \cdot S^{*(\sigma)}(z) \cdot S^{(\sigma+1)}(z)|| \\ &\quad = \quad ||(\Gamma^{*(\sigma)} \cdot \Gamma^{(\sigma)})^{-1} \cdot S^{*(\sigma)}(z) \cdot \Big\{ S^{(\sigma)}(z) \cdot \widehat{S}^{(\sigma)}(z) + \theta_{VII}^{(\sigma)}(z) \Big\} || \\ &\quad = \quad ||(\Gamma^{*(\sigma)} \cdot \Gamma^{(\sigma)})^{-1} \cdot \Big\{ (z^{||m^{(\sigma)}||+1} \cdot (a_0^{(0)})^{-1} \cdot \Gamma^{*(\sigma)} \cdot \Gamma^{(\sigma)} + \theta_{II}^{(\sigma)}(z) \Big\} \cdot \widehat{S}^{(\sigma)}(z) \\ &\qquad + \ (\Gamma^{*(\sigma)} \cdot \Gamma^{(\sigma)})^{-1} \cdot S^{*(\sigma)}(z) \cdot \theta_{VII}^{(\sigma)}(z)|| \\ &\quad \ge \quad |a_0^{(0)}|^{-1} \cdot ||\widehat{S}^{(\sigma)}(z)|| \\ &\qquad - \ ||(\Gamma^{*(\sigma)} \cdot \Gamma^{(\sigma)})^{-1}|| \cdot ||a_0^{-1}(z)|| \cdot \Big\{ (k+1)||\delta T^{(\sigma)^t}(z)|| \ + \ ||\delta T^{*(\sigma)}(z)|| \Big\} \cdot ||\widehat{S}^{(\sigma)}(z)|| \\ &\qquad - \ ||(\Gamma^{*(\sigma)} \cdot \Gamma^{(\sigma)})^{-1}|| \cdot \Big\{ 1.01 \cdot (||\nu^{(\sigma)}|| + k+1) \Big\} \cdot ||S^{(\sigma)}(z)|| \cdot ||\widehat{S}^{(\sigma)}(z)|| \cdot ||S^{*(\sigma)}(z)|| \cdot \mu \end{aligned}$$

$$
\begin{aligned}
&\geq \ ||\widehat{S}^{(\sigma)}(z)|| \cdot \Big\{ |a_0^{(0)}|^{-1} - \kappa^{(\sigma)} \cdot ||a_0^{-1}(z)|| \cdot \Big[(k+1)||\delta T^{(\sigma)^t}(z)|| \ + \ ||\delta T^{*(\sigma)}(z)|| \Big] \\
&\quad - \ 1.01\, \kappa^{(\sigma)} \cdot (||\nu^{(\sigma)}|| + k + 1) \cdot (k+1) \cdot \mu \Big\} \\
&\geq \ |a_0^{(0)}|^{-1} \cdot ||\widehat{S}^{(\sigma)}(z)||/2.
\end{aligned}
$$

The result now follows. ∎

THEOREM 18. *If μ is so small and $\delta T^{(\sigma)^t}(z)$ and $\delta T^{*(\sigma)}(z)$ are not too large so that*

$$
\begin{aligned}
&\kappa^{(j)} \cdot |a_0^{(0)}| \cdot \Big\{ ||a_0^{-1}(z)|| \Big[(k+1)||\delta T^{(j)^t}(z)|| + ||\delta T^{*(j)}(z)|| \Big] \\
&\quad + \quad 1.01(k+1)(||\nu^{(j)}|| + k + 1) \cdot \mu \Big\} \leq \frac{1}{2}, \ j \leq \sigma,
\end{aligned}
$$

then

$$
||\mathcal{G}_{j-1}^{(\sigma)}(z)|| \leq 2\kappa^{(j)} \cdot (k+1) \cdot |a_0^{(0)}| + O(\mu), \quad j \leq \sigma.
$$

Proof. From (74) and from Lemma 14

$$
S^{(\sigma+1)}(z) = S^{(j)}(z) \cdot \mathcal{G}_{j-1}^{(\sigma)}(z) + \sum_{\ell=j}^{\sigma} \theta_{VII}^{(\ell)}(z) \cdot \mathcal{G}_{\ell}^{(\sigma)}(z).
$$

We proceed by induction. Assume the theorem is true for $\mathcal{G}_{\sigma-1}^{(\sigma)}(z)$, $\mathcal{G}_{\sigma-2}^{(\sigma)}(z)$, $\cdots$, $\mathcal{G}_{j}^{(\sigma)}(z)$ (the initial case, $j = \sigma - 1$, is proved in Lemma 17 since $\mathcal{G}_{\sigma-1}^{(\sigma)}(z) = \widehat{S}^{(\sigma)}(z)$). From (57),

$$
||(\Gamma^{*(j)} \cdot \Gamma^{(j)})^{-1} S^{*(j)}(z) \cdot S^{(\sigma+1)}(z)|| \leq \kappa^{(j)}(k+1),
$$

But, using Lemma 14, Theorem 4 and the inductive hypothesis,

$$
\begin{aligned}
&||(\Gamma^{*(j)} \cdot \Gamma^{(j)})^{-1} \cdot S^{*(j)}(z) \cdot S^{(\sigma+1)}(z)|| \\
&\quad \geq \ ||(\Gamma^{*(j)} \cdot \Gamma^{(j)})^{-1} \cdot S^{*(j)}(z) \cdot S^{(j)}(z) \cdot \mathcal{G}_{j-1}^{(\sigma)}(z) \\
&\qquad + \sum_{\ell=j}^{\sigma} (\Gamma^{*(j)} \cdot \Gamma^{(j)})^{-1} \cdot S^{*(j)}(z) \cdot \theta_{VII}^{(\ell)}(z) \cdot \mathcal{G}_{\ell}^{(\sigma)}(z)|| \\
&\quad \geq \ ||(\Gamma^{*(j)} \cdot \Gamma^{(j)})^{-1} \cdot \Big\{ z^{||m^{(j)}||+1} (a_0^{(0)})^{-1} \Gamma^{*(j)} \cdot \Gamma^{(j)} + \theta_{II}^{(j)}(z) \Big\} \cdot \mathcal{G}_{j-1}^{(\sigma)}(z)|| \\
&\qquad -\kappa^{(j)} \sum_{\ell=j}^{\sigma} \{k+1\} \cdot \Big\{ 2.02\kappa^{(\ell)} \cdot (k+1) \cdot (||\nu^{(\ell)}|| + k + 1) \cdot |a_0^{(0)}| \cdot \mu \Big\} \cdot \\
&\qquad\qquad \Big\{ 2\kappa^{(\ell+1)} \cdot (k+1) \cdot |a_0^{(0)}| + O(\mu) \Big\} \\
&\quad \geq \ ||\mathcal{G}_{j-1}^{(\sigma)}(z)|| \Big\{ |a_0^{(0)}|^{-1} - \kappa^{(j)} \Big[||a_0^{-1}(z)||((k+1)||\delta T^{(j)^t}(z)|| + ||\delta T^{*(j)}(z)||) \Big] \Big\} - O(\mu) \\
&\quad \geq \ |a_0^{(0)}|^{-1} ||\mathcal{G}_{j-1}^{(\sigma)}(z)||/2 - O(\mu).
\end{aligned}
$$

∎

In the above theorem, we have taken the liberty of replacing a summation involving terms linear in μ with an $O(\mu)$ expression. We could have left the summation in explicitly, but, as we shall see, this summation becomes quadratic in μ when it is used to obtain a bound on $\delta T^{(\sigma)^t}(z)$.

To simplify the analysis, we now split the local error $\mathcal{L}^{(\sigma)^t}(z)$ into three parts and analyze the propagation of each part separately in each of the next three lemmas below. Let

$$(75)\ \mathcal{L}_1^{(\sigma)^t}(z) = \begin{cases} 0, & \sigma = 0, \\ -z^{||m^{(\sigma)}||+1}\ \theta_V^{(\sigma)^t}(z)\ \widehat{S}^{(\sigma)}(z)\ (\bmod z^{||m^{(\sigma+1)}||+1}), & \sigma \geq 1, \end{cases}$$

$$(76)\ \mathcal{L}_2^{(\sigma)^t}(z) = z^{||m^{(\sigma)}||+1}\theta_{VI}^{(\sigma)^t}(z)\ (\bmod z^{||m^{(\sigma+1)}||+1}), \qquad \sigma \geq 0,$$

$$(77)\ \mathcal{L}_3^{(\sigma)^t}(z) = \begin{cases} 0, & \sigma = 0, \\ A^t(z)\theta_{VII}^{(\sigma)}(z)\ (\bmod z^{||m^{(\sigma+1)}||+1}), & \sigma \geq 1, \end{cases}$$

and define

$$(78) \qquad \mathcal{E}_i^{(\sigma+1)}(z) = \sum_{j=0}^{\sigma} \mathcal{L}_i^{(j)^t}(z) \cdot \mathcal{G}_j^{(\sigma)}(z), \qquad i = 1, 2, 3.$$

Then, according to Lemma 15 and Theorem 16,

$$\delta T^{(\sigma+1)^t}(z) = \sum_{i=1}^{3} \mathcal{E}_i^{(\sigma+1)^t}(z)\ .$$

LEMMA 19.

$$\begin{aligned} ||\mathcal{E}_1^{(\sigma+1)^t}(z)|| \leq\ & 4\kappa^{(\sigma)} \cdot (k+1) \cdot (||m^{(\sigma)}|| + k + 1) \cdot |a_0^{(0)}| \cdot \mu \\ & + 8(k+1)^2 \cdot |a_0^{(0)}|^2 \cdot \mu \sum_{j=0}^{\sigma-1} \kappa^{(j)} \cdot \kappa^{(j+1)} \cdot (||m^{(j)}|| + k + 1) \\ & + O(\mu^2). \end{aligned}$$

Proof. From (75) and (78), from Lemmas 12 and 17 and from Theorem 18,

$$\begin{aligned} ||\mathcal{E}_1^{(\sigma+1)^t}(z)|| &= ||\sum_{j=0}^{\sigma} \mathcal{L}_1^{(j)^t}(z) \cdot \mathcal{G}_j^{(\sigma)}(z)|| \\ &\leq ||\theta_V^{(\sigma)^t}(z)|| \cdot ||\widehat{S}^{(\sigma)}(z)|| + \sum_{j=0}^{\sigma-1} ||\theta_V^{(j)^t}(z)|| \cdot ||\widehat{S}^{(j)}(z)|| \cdot ||\mathcal{G}_j^{(\sigma)}(z)|| \\ &\leq \left\{1.01(||m^{(\sigma)}|| + k + 1) \cdot \mu\right\} \left\{2\kappa^{(\sigma)}(k+1)|a_0^{(0)}|\right\} \\ &\quad + \sum_{j=0}^{\sigma-1} \left\{1.01(||m^{(j)}|| + k + 1)\mu\right\} \cdot \left\{2\kappa^{(j)}(k+1)|a_0^{(0)}|\right\} \\ &\qquad \cdot \left\{2\kappa^{(j+1)}(k+1)|a_0^{(0)}| + O(\mu)\right\}, \end{aligned}$$

and so the result follows. ∎

LEMMA 20.

$$\begin{aligned}||\mathcal{E}_2^{(\sigma+1)^t}(z)|| \quad \leq \quad & 32\cdot\kappa^{(\sigma)}\cdot(k+1)||\nu^{(\sigma)}||^4\cdot\rho_\sigma\cdot|a_0^{(0)}|\cdot\mu \\ & +64(k+1)^2\cdot|a_0^{(0)}|^2\cdot\mu\sum_{j=0}^{\sigma-1}\kappa^{(j)}\cdot\kappa^{(j+1)}\cdot\rho_j\cdot||\nu^{(j)}||^4 \\ & +O(\mu^2).\end{aligned}$$

Proof. From (76) and (78), from Lemmas 13 and 17 and from Theorem 18,

$$\begin{aligned}||\mathcal{E}_2^{(\sigma+1)^t}(z)|| \quad = \quad & ||\sum_{j=0}^{\sigma}\mathcal{L}_2^{(j)^t}(z)\mathcal{G}_j^{(\sigma)}(z)|| \\ \leq \quad & ||\theta_{VI}^{(\sigma)^t}(z)||+\sum_{j=0}^{\sigma-1}||\theta_{VI}^{(j)^t}(z)||\cdot||\mathcal{G}_j^{(\sigma)}(z)|| \\ \leq \quad & \left\{16||\nu^{(\sigma)}||^4\cdot\rho_\sigma\cdot\mu+O(\mu^2)\right\}\cdot||\widehat{S}^{(\sigma)}(z)|| \\ & +\sum_{j=0}^{\sigma-1}\left\{16||\nu^{(j)}||^4\cdot\rho_j\cdot\mu+O(\mu^2)\right\}\cdot||\widehat{S}^{(j)}(z)||\cdot||\mathcal{G}_j^{(\sigma)}(z)|| \\ \leq \quad & \left\{16||\nu^{(\sigma)}||^4\cdot\rho_\sigma\cdot\mu+O(\mu^2)\right\}\cdot\left\{2\kappa^{(\sigma)}(k+1)|a_0^{(0)}|+\right\} \\ & +\sum_{j=0}^{\sigma-1}\left\{16||\nu^{(j)}||^4\cdot\rho_j\cdot\mu+O(\mu^2)\right\}\cdot\left\{2\kappa^{(j)}(k+1)|a_0^{(0)}|+\right\} \\ & \qquad\cdot\left\{2\kappa^{(j+1)}(k+1)|a_0^{(0)}|+O(\mu)\right\}.\end{aligned}$$

The result now follows. ∎

LEMMA 21.

$$\begin{aligned}||\mathcal{E}_3^{(\sigma+1)^t}(z)|| \quad \leq \quad & 4\kappa^{(\sigma)}\cdot(k+1)\cdot(||\nu^{(\sigma)}||+k+1)\cdot|a_0^{(0)}|\cdot\mu \\ & +8(k+1)^2\cdot|a_0^{(0)}|^2\cdot\mu\sum_{j=0}^{\sigma-1}\kappa^{(j)}\kappa^{(j+1)}(||\nu^{(j)}||+k+1) \\ & +O(\mu^2).\end{aligned}$$

Proof. From (77) and (78), from Lemmas 14 and 17 and from Theorem 18,

$$\begin{aligned}||\mathcal{E}_3^{(\sigma+1)^t}(z)|| \quad = \quad & ||\sum_{j=0}^{\sigma}\mathcal{L}_3^{(j)^t}(z)\cdot\mathcal{G}_j^{(\sigma)}(z)|| \\ \leq \quad & ||A^t(z)\cdot\theta_{VII}^{(\sigma)}(z)||+\sum_{j=0}^{\sigma-1}||A^t(z)\cdot\theta_{VII}^{(j)}(z)||\cdot\mathcal{G}_j^{(\sigma)}(z)|| \\ \leq \quad & 1.01(||\nu^{(\sigma)}||+k+1)\cdot||\widehat{S}^{(\sigma)}(z)||\cdot\mu \\ & +\sum_{j=0}^{\sigma-1}\left\{1.01(||\nu^{(j)}||+k+1)\cdot\mu\right\}\cdot||\widehat{S}^{(j)}(z)||\cdot||\mathcal{G}_j^{(\sigma)}(z)||\end{aligned}$$

$$
\begin{aligned}
\leq\ & \left\{1.01\cdot(||\nu^{(\sigma)}||+k+1)\cdot\mu\right\}\cdot\left\{2\kappa^{(\sigma)}(k+1)|a_0^{(0)}|\right\}\\
&+\sum_{j=0}^{\sigma-1}\left\{1.01(||\nu^{(j)}||+k+1)\cdot\mu\right\}\cdot\left\{2\kappa^{(j)}(k+1)|a_0^{(0)}|\right\}\\
&\qquad\qquad\cdot\left\{2\kappa^{(j+1)}(k+1)|a_0^{(0)}|+O(\mu)\right\}.
\end{aligned}
$$

The result now follows. ∎

In the above three lemmas, the bounds obtained involve the products $\kappa^{(j)}\kappa^{(j+1)}$. These result from inequalities involving the expression $||\widehat{S}^{(j)}(z)||\cdot||\mathcal{G}_j^{(\sigma)}(z)||$. However, it is seen that $\widehat{S}^{(j)}(z)\cdot\mathcal{G}_j^{(\sigma)}(z)=\mathcal{G}_{j-1}^{(\sigma)}(z)$, so it is felt that the inequalities are crude and the bounds should just involve a single $\kappa^{(j)}$. Experimental results [9] support this conjecture.

Finally, we can give the bound on the residual error.

THEOREM 22. *If μ is so small and $\delta T^{(j)^t}(z)$ and $\delta T^{*(j)}(z)$ are not too large so that*

$$(||n||+k+1)\mu\leq 0.01$$

and

$$
\begin{aligned}
&\kappa^{(j)}\cdot|a_0^{(0)}|\cdot\left\{||a_0^{-1}(z)||\left[(k+1)||\delta T^{(j)^t}(z)||+||\delta T^{*(j)}(z)||\right]\right.\\
&\left.+\quad 1.01(k+1)(||\nu^{(j)}||+k+1)\cdot\mu\right\}\leq\frac{1}{2},\ j\leq\sigma,
\end{aligned}
$$

then

$$||\delta T^{(\sigma+1)^t}(z)||\leq F_\sigma+2(k+1)\cdot|a_0^{(0)}|\sum_{j=0}^{\sigma-1}\kappa^{(j+1)}F_j, \tag{79}$$

where

$$
\begin{aligned}
F_j\ =\ & 4\kappa^{(j)}(k+1)\cdot|a_0^{(0)}|\cdot\mu\\
&\cdot\left\{(||m^{(j)}||+k+1)+4\rho_j||\nu^{(j)}||^4+(||\nu^{(j)}||+k+1)\right\}.
\end{aligned}
\tag{80}
$$

Proof. Sum the error bounds given in Lemmas 19, 20 and 21 ∎

Theorem 22 assures us that if $||\delta T^{(\sigma)^t}(z)||$ is small and $\kappa^{(\sigma)}$ is not too large, then $||\delta T^{(\sigma+1)^t}(z)||$ will also be small. Thus, $||\delta T^{(\sigma)^t}(z)||$ will remain small for all σ as long as, at every iteration j, a step $\nu^{(j)}$ is chosen (stepping over unstable blocks) so that $\kappa^{(j)}$ is not too large. The same observation is made about $\delta T^{*(\sigma)}(z)$ in §8. Consequently, the assumptions of Theorem 22 are satisfied in practice

8. Error Analysis for Simultaneous Padé Systems. In this section, we obtain bounds for the error in the order condition for the NSPS computed by the algorithm VECTOR_PADE. The approach used in obtaining these bounds follows step by step the approach used in §7. As before, we begin by first analyzing the floating point errors introduced by one iteration of the algorithm. At the σth iteration, the NSPS $S^{*(\sigma)}(z)$ of type $m^{(\sigma)}$ for $A^*(z)$ is available and satisfies

$$S^{*(\sigma)}(z)A^*(z)=\delta T^{*(\sigma)}(z)+\mathcal{O}(z^{||m^{(\sigma)}||+1}).$$

The algorithm proceeds to compute $S^{*(\sigma+1)}(z)$ of type $m^{(\sigma+1)}$.

An iterative step consists of three parts. In the first part, the first $||\nu^{(\sigma)}||+1$ of $T^{*(\sigma)}(z)$ are computed; a bound for the floating point errors introduced by these computations is given in Lemma 23 below. In the second part, the NSPS $\widehat{S}^{*(\sigma)}(z)$ of type $\nu^{(\sigma)}$ for $T^{*(\sigma)}(z)$ is computed; an error analysis is given Lemma 24. In the third part, Lemma 14 provides bounds for the floating point errors introduced in computing $S^{*(\sigma+1)}(z)=\widehat{S}^{*(\sigma)}(z)\cdot S^{*(\sigma)}(z)$. At this point in the algorithm, $S^{*(\sigma+1)}(z)$ is scaled so that the norm of each row is 1. As before, we assume for the sake of simplicity that this scaling introduces no additional errors.

Lemma 23. *The computed residual $T^{*(\sigma)}(z)$ satisfies*

$$z^{||m^{(\sigma)}||+1}T^{*(\sigma)}(z)=S^{*(\sigma)}(z)\cdot A^{*}(z)-\delta T^{*(\sigma)}(z)+z^{||m^{(\sigma)}||+1}\theta_V^{*(\sigma)}(z),$$

where

$$||\theta_V^{*(\sigma)}(z)||\leq 2.02(k+1)(||m^{(\sigma)}||+1)\cdot\mu.$$

Proof. The algorithm computes the first $||\nu^{(\sigma)}||+1$ terms of the residual only. That is,

$$\begin{aligned} z^{||m^{(\sigma)}||+1}T^{*(\sigma)} &= fl[S^{*(\sigma)}(z)]\cdot A^{*}(z)] \mod z^{||m^{(\sigma+1)}||+1}\\ &\qquad -fl[S^{*(\sigma)}(z)\cdot A^{*}(z)] \bmod z^{||m^{(\sigma)}||+1}. \end{aligned}$$

Thus,

$$S^{*(\sigma)}(z)\cdot A^{*}(z) \bmod z^{||m^{(\sigma+1)}||+1}=\delta T^{*(\sigma)}(z)+z^{||m^{(\sigma)}||+1}[T^{*(\sigma)}(z)-\theta_V^{*(\sigma)}(z)],$$

where $\theta_V^{*(\sigma)}(z)$ is the error introduced into the computation of $T^{*(\sigma)}(z)$ by floating point operations. The result now follows from Lemma 10 since $A^{*}(z)$ and $S^{*(\sigma)}(z)$ are both scaled, and therefore $||A^{*}(z)||\leq 2$ and $||S^{*(\sigma)}(z)||\leq k+1$. ∎

Lemma 24. *If $\widehat{\mathcal{M}}^{*}_{\nu(\sigma)}$ is nonsingular and $\widehat{S}^{*(\sigma)}(z)$ is obtained by solving (38) and (40), then*

$$\widehat{S}^{*(\sigma)}(z)\cdot T^{*(\sigma)}(z)=\theta_{VI}^{*(\sigma)}(z)+O(z^{||\nu^{(\sigma)}||+1}),$$

where

$$||\theta_{VI}^{*(\sigma)}(z)||\leq\left\{32(k+1)^5||\nu^{(\sigma)}||^4\cdot\rho_\sigma^{*}\cdot\mu+O(\mu^2)\right\}\cdot||\widehat{S}^{*(\sigma)}(z)||.$$

Proof. First we obtain bounds for rows $1,\ldots,k$ of $\theta_{VI}^{*(\sigma)}(z)$. Row α of $\widehat{S}^{*(\sigma)}(z)$ corresponds to the solution $\widehat{\mathcal{Y}}^{*}_\alpha$ of (40) obtained by Gaussian elimination. $\widehat{\mathcal{Y}}^{*}_\alpha$ is the exact solution of

$$\widehat{\mathcal{Y}}^{*t}_\alpha\cdot(\widehat{\mathcal{M}}^{*}_{\nu(\sigma)}+\mathcal{E}^{*})=E^{t}_{\alpha||\nu(\sigma)||},$$

where

$$||\mathcal{E}^{*}||_\infty\leq 8(k^3||\nu^{(\sigma)}||)^3\cdot\rho_\sigma^{*}\cdot||\widehat{\mathcal{M}}^{*}_{\nu(\sigma)}||_\infty\cdot\mu+O(\mu^2)$$

and ρ_σ^{*} is the growth factor associated with the LU-decomposition of $\widehat{\mathcal{M}}^{*}_{\nu(\sigma)}$.

But, from Lemma 11

$$\begin{aligned} ||T^{*(\sigma)}(z)|| &\leq 2(k+1)[1+1.01\cdot(||m^{(\sigma)}||+1)\cdot\mu] \\ &\leq 4(k+1)\cdot\mu, \end{aligned}$$

since $\|A^*(z)\| \leq 2$ and $\|S^{*(\sigma)}(z)\| \leq k+1$ because of scaling. Here, we have used $1.01(\|m^{(\sigma)}\|+1)\mu \leq 1.01(\|n\|+k+1)\mu \leq 1$. Thus,

$$\widehat{\mathcal{Y}}_\alpha^{*t}\cdot\widehat{\mathcal{M}}^*_{\nu^{(\sigma)}} - E^t_{\alpha\|\nu^{(\sigma)}\|} = -\widehat{\mathcal{Y}}_\alpha^{*t}\cdot\mathcal{E}^*,$$

where

$$\begin{aligned} ||\mathcal{E}^{*t}\cdot\widehat{\mathcal{Y}}^*_\alpha||_1 &\leq ||\widehat{\mathcal{Y}}^{*t}_\alpha\cdot\mathcal{E}^*||_\infty \\ &\leq ||\widehat{\mathcal{Y}}^*_\alpha||_1\cdot||\mathcal{E}^*||_\infty \\ &\leq ||\widehat{\mathcal{Y}}^*_\alpha||_1\cdot\left\{8k^3||\nu^{(\sigma)}||^4\cdot\rho^*_\sigma\cdot||T^{*(\sigma)}(z)||\cdot\mu+O(\mu^2)\right\} \\ &\leq ||\widehat{\mathcal{Y}}^*_\alpha||_1\cdot\left\{32(k+1)^4||\nu^{(\sigma)}||^4\cdot\rho^*_\sigma\cdot\mu+O(\mu^2)\right\}, \end{aligned}$$

since $||\widehat{\mathcal{M}}^*_{\nu^{(\sigma)}}||_\infty \leq (k+1)||\nu^{(\sigma)}||\cdot||T^{*(\sigma)}(z)||$. A similar analysis can be done for solving (38) to obtain $\widehat{\mathcal{X}}^*$. But, $\widehat{\mathcal{Y}}^*_\alpha$ yields row α, $1\leq\alpha\leq k$, of $\widehat{S}^{*(\sigma)}(z)$ with residual error $\mathcal{E}^{*t}\cdot\widehat{\mathcal{Y}}^*_\alpha$ (i.e., $\mathcal{E}^{*t}\cdot\widehat{\mathcal{Y}}^*_\alpha$ gives row α of $\theta^{*(\sigma)}_{VI}(z)$) and $\widehat{\mathcal{X}}^*$ yields the first row of $\widehat{S}^{*(\sigma)}(z)$ with a corresponding residual error. The result now follows. ∎

LEMMA 25. *If* $S^{*(\sigma+1)}(z) = fl\left\{\widehat{S}^{*(\sigma)}(z)\cdot S^{*(\sigma)}(z)\right\}$, *then*

$$S^{*(\sigma+1)}(z) \;=\; \widehat{S}^{*(\sigma)}(z)\cdot S^{*(\sigma)}(z) \;+\; \theta^{*(\sigma)}_{VII}(z),$$

where

$$||\theta^{*(\sigma)}_{VII}(z)|| \;\leq\; 1.01(||\nu^{(\sigma)}||+k+1)\cdot||\widehat{S}^{*(\sigma)}(z)||\cdot||S^{*(\sigma)}(z)||\mu.$$

Proof. The $(0,0)$-component of $S^{*(\sigma+1)}(z)$ is

$$\begin{aligned} S^{*(\sigma+1)}_{0,0}(z) &= fl\left\{\widehat{v}^*(z)\cdot v^*(z) + z^2\sum_{\rho=1}^{k}\widehat{u}^*_\rho(z)\cdot q^*_\rho(z)\right\} \\ &= \sum_{\ell=0}^{\|m^{(\sigma+1)}\|-m_0^{(\sigma+1)}} z^\ell\left\{\sum_{j=0}^{\|\nu^{(\sigma)}\|-\nu_0^{(\sigma)}}\widehat{v}^{*(j)}v^{*(\ell-j)}\cdot(1+\delta^*_{0,0,j,\ell,0})\right. \\ &\quad\left.+\sum_{\rho=1}^{k}\sum_{j=0}^{\|\nu^{(\sigma)}\|-\nu_\rho^{(\sigma)}}\widehat{u}^{*(j)}_\rho q^{*(\ell-j-2)}_\rho\cdot(1+\delta^*_{0,0,j,\ell,\rho})\right\}, \end{aligned}$$

where $|\delta^*_{0,0,j,\ell,\rho}| \;\leq\; 1.01\cdot(\|\nu^{(\sigma)}\|-\nu_\rho^{(\sigma)}+k)\cdot\mu$. Thus,

$$\begin{aligned} (\theta^{*(\sigma)}_{VII}(z))_{0,0} &= \sum_{\ell=0}^{\|m^{(\sigma+1)}\|-m_0^{(\sigma+1)}} z^\ell \sum_{j=0}^{\|\nu^{(\sigma)}\|-\nu_0^{(\sigma)}}\widehat{v}^{*(j)}v^{*(\ell-j)}\cdot\delta^*_{0,0,j,\ell,0} \\ &\quad+z^2\sum_{\rho=1}^{k}\sum_{\ell=0}^{\|m^{(\sigma+1)}\|-m_0^{(\sigma+1)}} z^\ell\sum_{j=0}^{\|\nu^{(\sigma)}\|-\nu_\rho^{(\sigma)}}\widehat{u}^{*(j)}_\rho q^{*(\ell-j)}_\rho\cdot\delta^*_{0,0,j,\ell,\rho}. \end{aligned}$$

Thus, from (64)

$$|| \left(\theta_{VII}^{*(\sigma)}(z)\right)_{0,0} || \;\leq\; 1.01 \cdot (||\nu^{(\sigma)}|| + k) \cdot \left\{ ||\widehat{v}^*(z)|| \cdot ||v^*(z)|| + \sum_{\rho=1}^{k} ||\widehat{u}_\rho^*(z)|| \cdot ||q_\rho^*(z)|| \right\} \mu.$$

An equivalent result holds for $S_{\alpha,\beta}^{*(\sigma+1)}(z)$ for α and β other than $\alpha = \beta = 0$. The lemma now follows using (71). ∎

The use of the results of the three lemmas above enables us to express the residual error $\delta T^{*(\sigma+1)}(z)$ in the order condition at the $(\sigma+1)$th iteration in terms of the residual error $\delta T^{*(\sigma)}(z)$ at the σth iteration plus the floating point errors introduced "locally" by the σth iteration.

LEMMA 26.

$$\delta T^{*(\sigma+1)}(z) \;=\; \widehat{S}^{*(\sigma)}(z) \cdot \delta T^{*(\sigma)}(z) + \mathcal{L}^{*(\sigma)}(z),$$

where

$$\mathcal{L}^{*(\sigma)}(z) \;=\; \left\{ \theta_{VII}^{*(\sigma)}(z) A^*(z) \cdot + z^{||m^{(\sigma)}||+1} \left[\theta_{VI}^{*(\sigma)}(z) \;-\; \widehat{S}^{*(\sigma)}(z) \theta_{V}^{*(\sigma)}(z) \cdot \right] \right\} \; mod \; z^{||m^{i+1}||+1}.$$

Proof. The result is an immediate consequence of Lemmas 23, 24 and 25. ∎

THEOREM 27. *The residual error satisfies*

$$\delta T^{*(\sigma+1)}(z) = \sum_{j=0}^{\sigma} \mathcal{G}_j^{*(\sigma)}(z) \cdot \mathcal{L}^{*(\sigma)}(z),$$

where

$$\mathcal{G}_j^{*(\sigma)}(z) = \begin{cases} \widehat{S}^{*(\sigma)}(z) \cdot \widehat{S}^{*(\sigma+1)}(z) \cdots \widehat{S}^{*(j+1)}(z), & 0 \leq j < \sigma, \\ I_{k+1}, & j = \sigma. \end{cases} \tag{81}$$

Proof. The result follows by induction from Lemma 26. ∎

We see from (81) that the residual error $\delta T^{*(\sigma+1)}(z)$ is composed of the local errors $\mathcal{L}^{*(j)^t}(z)$ propagated by $\mathcal{G}_j^{*(\sigma)}$. The next lemma and theorem give bounds independent of σ for the propagation matrices $\mathcal{G}_j^{*(\sigma-1)}$. Consequently, as for the NPHS, the residual error grows additively with σ.

LEMMA 28. *If μ is so small and $\delta T^{(\sigma)^t}(z)$ and $\delta T^{*(\sigma)}(z)$ are not too large so that*

$$\begin{aligned} \kappa^{(\sigma)} \cdot |a_0^{(0)}| \cdot \Big\{ ||a_0^{-1}(z)|| & \left[(k+1) ||\delta T^{(\sigma)^t}(z)|| + ||\delta T^{*(\sigma)}(z)|| \right] \\ + \quad & 1.01(k+1)(||\nu^{(\sigma)}|| + k + 1) \cdot \mu \Big\} \leq \frac{1}{2}, \end{aligned}$$

then

$$||\widehat{S}^{*(\sigma)}(z)|| \leq 2\kappa^{(\sigma)} \cdot (k+1) \cdot |a_0^{(0)}|.$$

Proof. From (57),

$$\begin{aligned} ||S^{*(\sigma+1)}(z) \cdot S^{(\sigma)}(z) \cdot (\Gamma^{*(\sigma)} \cdot \Gamma^{(\sigma)})^{-1}|| \;&\leq\; ||S^{*(\sigma+1)}(z)|| \cdot ||S^{(\sigma)}(z)|| \cdot ||(\Gamma^{*(\sigma)} \cdot \Gamma^{(\sigma)})^{-1}|| \\ &\leq\; \kappa^{(\sigma)} \cdot (k+1). \end{aligned}$$

But, using Theorem 4 and Lemma 25

$$
\begin{aligned}
&||S^{*(\sigma+1)}(z)\cdot S^{(\sigma)}(z)\cdot(\Gamma^{*(\sigma)}\cdot\Gamma^{(\sigma)})^{-1}|| \\
&\quad = \ ||\left\{\widehat{S}^{*(\sigma)}(z)\cdot S^{*(\sigma)}(z)+\theta^{*(\sigma)}_{VII}(z)\right\}\cdot S^{(\sigma)}(z)(\Gamma^{*(\sigma)}\cdot\Gamma^{(\sigma)})^{-1}|| \\
&\quad = \ ||\widehat{S}^{*(\sigma)}(z)\cdot\left\{z^{||m^{(\sigma)}||+1}\cdot(a_0^{(0)})^{-1}\cdot\Gamma^{*(\sigma)}\cdot\Gamma^{(\sigma)}+\theta^{(\sigma)}_{II}(z)\right\}(\Gamma^{*(\sigma)}\cdot\Gamma^{(\sigma)})^{-1} \\
&\qquad + \ \theta^{*(\sigma)}_{VII}(z)\cdot S^{(\sigma)}(z)(\Gamma^{*(\sigma)}\cdot\Gamma^{(\sigma)})^{-1}|| \\
&\quad \geq \ |a_0^{(0)}|^{-1}\cdot||\widehat{S}^{*(\sigma)}(z)|| \\
&\qquad -||(\Gamma^{*(\sigma)}\cdot\Gamma^{(\sigma)})^{-1}||\cdot||a_0^{-1}(z)||\cdot\left\{(k+1)||\delta T^{(\sigma)^t}(z)|| \ + \ ||\delta T^{*(\sigma)}(z)||\right\}\cdot||\widehat{S}^{*(\sigma)}(z)|| \\
&\qquad -||(\Gamma^{*(\sigma)}\cdot\Gamma^{(\sigma)})^{-1}||\cdot\left\{1.01\cdot(||\nu^{(\sigma)}||+k+1)\right\}\cdot||S^{*(\sigma)}(z)||\cdot||\widehat{S}^{*(\sigma)}(z)||\cdot||S^{(\sigma)}(z)||\cdot\mu \\
&\quad \geq \ ||\widehat{S}^{*(\sigma)}(z)||\cdot\left\{|a_0^{(0)}|^{-1}-\kappa^{(\sigma)}\cdot||a_0^{-1}(z)||\cdot\left[(k+1)||\delta T^{(\sigma)^t}(z)|| \ + \ ||\delta T^{*(\sigma)}(z)||\right]\right. \\
&\qquad \left. - \ 1.01\,\kappa^{(\sigma)}\cdot(||\nu^{(\sigma)}||+k+1)\cdot(k+1)\cdot\mu\right\} \\
&\quad \geq \ |a_0^{(0)}|^{-1}\cdot||\widehat{S}^{*(\sigma)}(z)||/2.
\end{aligned}
$$

The result now follows. ∎

THEOREM 29. *If μ is so small and $\delta T^{(\sigma)^t}(z)$ and $\delta T^{*(\sigma)}(z)$ are not too large so that*

$$
\begin{aligned}
&\kappa^{(j)}\cdot|a_0^{(0)}|\cdot\left\{||a_0^{-1}(z)||\left[(k+1)||\delta T^{(j)^t}(z)||+||\delta T^{*(j)}(z)||\right]\right. \\
&\quad + \quad \left. 1.01(k+1)(||\nu^{(j)}||+k+1)\cdot\mu\right\}\leq\frac{1}{2}, \ j\leq\sigma,
\end{aligned}
$$

then

$$||\mathcal{G}^{*(\sigma)}_{j-1}(z)||\leq 2\kappa^{(j)}\cdot(k+1)\cdot|a_0^{(0)}|+O(\mu), \ \ j\leq\sigma.$$

Proof. From (81) and Lemma 25

$$S^{*(\sigma+1)}(z)=\mathcal{G}^{*(\sigma)}_{j-1}(z)\cdot S^{*(j)}(z)+\sum_{\ell=j}^{\sigma}\mathcal{G}^{*(\sigma)}_{\ell}(z)\cdot\theta^{*(\ell)}_{VII}(z).$$

We proceed by induction. Assume the theorem is true for $\mathcal{G}^{*(\sigma)}_{\sigma-1}(z)$, $\mathcal{G}^{*(\sigma)}_{\sigma-2}(z)$, $\cdots$, $\mathcal{G}^{*(\sigma)}_{j}(z)$ (the initial case, $j=\sigma-1$, is proved in Lemma 28 since $\mathcal{G}^{*(\sigma)}_{\sigma-1}(z)=\widehat{S}^{*(\sigma)}(z)$). From (57),

$$||S^{*(\sigma+1)}(z)\cdot S^{(j)}(z)(\Gamma^{*(j)}\cdot\Gamma^{(j)})^{-1}||\leq\kappa^{(j)}(k+1)$$

But, using Theorem 4, Lemma 25, and the inductive hypothesis,

$$
\begin{aligned}
&||S^{*(\sigma+1)}(z)\cdot S^{(j)}(z)(\Gamma^{*(j)}\cdot\Gamma^{(j)})^{-1}|| \\
&\quad \geq \ ||\mathcal{G}^{*(\sigma)}_{j-1}(z)\cdot S^{*(j)}(z)\cdot S^{(j)}(z)\cdot(\Gamma^{*(j)}\cdot\Gamma^{(j)})^{-1} \\
&\qquad +\sum_{\ell=j}^{\sigma}\mathcal{G}^{*(\sigma)}_{\ell}(z)\cdot\theta^{*(\ell)}_{VII}(z)\cdot S^{(j)}(z)\cdot(\Gamma^{*(j)}\cdot\Gamma^{(j)})^{-1}||
\end{aligned}
$$

$$
\begin{aligned}
&\geq \quad ||\mathcal{G}_{j-1}^{*(\sigma)}(z) \cdot \left\{ z^{||m^{(j)}||+1} (a_0^{(0)})^{-1} \Gamma^{*(j)} \cdot \Gamma^{(j)} + \theta_{II}^{(j)}(z) \right\} (\Gamma^{*(j)} \cdot \Gamma^{(j)})^{-1} || \\
&\quad -\kappa^{(j)} \sum_{\ell=j}^{\sigma} \{k+1\} \cdot \left\{ 2.02\kappa^{(\ell)} \cdot (k+1) \cdot (||\nu^{(\ell)}|| + k + 1) \cdot |a_0^{(0)}| \cdot \mu \right\} \cdot \\
&\qquad\qquad \left\{ 2\kappa^{(\ell+1)} \cdot (k+1) \cdot |a_0^{(0)}| + O(\mu) \right\} \\
&\geq \quad ||\mathcal{G}_{j-1}^{*(\sigma)}(z)|| \left\{ |a_0^{(0)}|^{-1} - \kappa^{(j)} \left[||a_0^{-1}(z)|| ((k+1) ||\delta T^{(j)^t}(z)|| + ||\delta T^{*(j)}(z)||) \right] \right\} - O(\mu) \\
&\geq \quad |a_0^{(0)}|^{-1} ||\mathcal{G}_{j-1}^{*(\sigma)}(z)||/2 - O(\mu).
\end{aligned}
$$

∎

To simplify the analysis, we split the local error $\mathcal{L}^{(\sigma)^t}(z)$ into three parts and analyze the propagation of each part separately. Let

$$
\begin{aligned}
&(82)\ \mathcal{L}_1^{*(\sigma)}(z) = \begin{cases} 0, & \sigma = 0, \\ -z^{||m^{(\sigma)}||+1}\ \widehat{S}^{*(\sigma)}(z) \cdot \theta_V^{*(\sigma)}(z) \ (\bmod z^{||m^{(\sigma+1)}||+1}), & \sigma \geq 1, \end{cases} \\
&(83)\ \mathcal{L}_2^{*(\sigma)}(z) = z^{||m^{(\sigma)}||+1} \theta_{VI}^{*(\sigma)}(z) (\bmod z^{||m^{(\sigma+1)}||+1}), \qquad \sigma \geq 0, \\
&(84)\ \mathcal{L}_3^{*(\sigma)}(z) = \begin{cases} 0, & \sigma = 0, \\ \theta_{VII}^{*(\sigma)}(z) \cdot A^*(z) \ (\bmod z^{||m^{(\sigma+1)}||+1}), & \sigma \geq 1, \end{cases}
\end{aligned}
$$

and define

$$
\mathcal{E}_i^{*(\sigma+1)}(z) = \sum_{j=0}^{\sigma} \mathcal{G}_j^{*(\sigma)}(z) \cdot \mathcal{L}_i^{*(j)^t}(z), \qquad i = 1, 2, 3. \tag{85}
$$

Then, according to Lemma 26 and Theorem 27,

$$
\delta T^{*(\sigma+1)}(z) = \sum_{i=1}^{3} \mathcal{E}_i^{*(\sigma+1)}(z).
$$

Lemma 30.

$$
\begin{aligned}
||\mathcal{E}_1^{*(\sigma+1)}(z)|| \leq\ & 8\kappa^{(\sigma)} \cdot (k+1)^2 \cdot (||m^{(\sigma)}|| + 1) \cdot |a_0^{(0)}| \cdot \mu \\
& +16(k+1)^3 \cdot |a_0^{(0)}|^2 \cdot \mu \sum_{j=0}^{\sigma-1} \kappa^{(j)} \cdot \kappa^{(j+1)} \cdot (||m^{(j)}|| + 1) \\
& +O(\mu^2).
\end{aligned}
$$

Proof. From (82) and (85), from Lemmas 23 and 28 and from Theorem 29,

$$
\begin{aligned}
||\mathcal{E}_1^{*(\sigma+1)}(z)|| &= ||\sum_{j=0}^{\sigma} \mathcal{G}_j^{*(\sigma)}(z) \cdot \mathcal{L}_1^{*(j)}(z)|| \\
&\leq ||\widehat{S}^{*(\sigma)}(z)|| \cdot ||\theta_V^{*(\sigma)}(z)|| + \sum_{j=0}^{\sigma-1} ||\mathcal{G}_j^{*(\sigma)}(z)|| \cdot ||\widehat{S}^{*(j)}(z)|| \cdot ||\theta_V^{*(j)}(z)|| \\
&\leq \left\{ 2.02(k+1)(||m^{(\sigma)}|| + 1) \cdot \mu \right\} \left\{ 2\kappa^{(\sigma)}(k+1) \cdot |a_0^{(0)}| \right\}
\end{aligned}
$$

$$+\sum_{j=0}^{\sigma-1}\left\{2.02(k+1)(||m^{(j)}||+1)\mu\right\}\cdot\left\{2\kappa^{(j)}(k+1)\cdot|a_0^{(0)}|\right\}$$
$$\cdot\left\{2\kappa^{(j+1)}(k+1)\cdot|a_0^{(0)}|+O(\mu)\right\},$$

and so the result follows. ∎

LEMMA 31.

$$\begin{aligned}||\mathcal{E}_2^{*((\sigma+1)}(z)|| \quad \leq \quad & 64\cdot\kappa^{(\sigma)}\cdot(k+1)^6||\nu^{(\sigma)}||^4\cdot\rho_\sigma^*\cdot|a_0^{(0)}|\cdot\mu \\ & +128(k+1)^7\cdot|a_0^{(0)}|^2\cdot\mu\sum_{j=0}^{\sigma-1}\kappa^{(j)}\cdot\kappa^{(j+1)}\cdot\rho_j^*\cdot||\nu^{(j)}||^4 \\ & +O(\mu^2).\end{aligned}$$

Proof. From (83) and (85), from Lemmas 24 and 28 and from Theorem 29,

$$\begin{aligned}||\mathcal{E}_2^{*(\sigma+1)}(z)|| \quad = \quad & ||\sum_{j=0}^{\sigma}\mathcal{G}_j^{*(\sigma)}(z)\mathcal{L}_2^{*(j)}(z)|| \\ \leq \quad & ||\theta_{VI}^{*(\sigma)}(z)||+\sum_{j=0}^{\sigma-1}||\mathcal{G}_j^{*(\sigma)}(z)||\cdot||\theta_{VI}^{*(j)}(z)|| \\ \leq \quad & \left\{32(k+1)^5||\nu^{(\sigma)}||^4\cdot\rho_\sigma^*\cdot\mu+O(\mu^2)\right\}\cdot||\widehat{S}^{*(\sigma)}(z)|| \\ & +\sum_{j=0}^{\sigma-1}\left\{32(k+1)^5||\nu^{(j)}||^4\cdot\rho_j^*\cdot\mu+O(\mu^2)\right\}\cdot||\widehat{S}^{*(j)}(z)||\cdot||\mathcal{G}_j^{*(\sigma)}(z)|| \\ \leq \quad & \left\{32(k+1)^5||\nu^{(\sigma)}||^4\cdot\rho_\sigma^*\cdot\mu+O(\mu^2)\right\}\cdot\left\{2\kappa^{(\sigma)}(k+1)|a_0^{(0)}|\right\} \\ & +\sum_{j=0}^{\sigma-1}\left\{32(k+1)^5||\nu^{(j)}||^4\cdot\rho_j^*\cdot\mu+O(\mu^2)\right\}\cdot\left\{2\kappa^{(j)}(k+1)|a_0^{(0)}|\right\} \\ & \cdot\left\{2\kappa^{(j+1)}(k+1)|a_0^{(0)}|+O(\mu)\right\}.\end{aligned}$$

The result now follows. ∎

LEMMA 32.

$$\begin{aligned}||\mathcal{E}_3^{*(\sigma+1)}(z)|| \quad \leq \quad & 8\kappa^{(\sigma)}\cdot(k+1)^2\cdot(||\nu^{(\sigma)}||+k+1)\cdot|a_0^{(0)}|\cdot\mu \\ & +16(k+1)^3\cdot|a_0^{(0)}|^2\cdot\mu\sum_{j=0}^{\sigma-1}\kappa^{(j)}\kappa^{(j+1)}(||\nu^{(j)}||+k+1) \\ & +O(\mu^2).\end{aligned}$$

Proof. From (84) and (85), from Lemmas 25 and 28 and from Theorem 29,

$$||\mathcal{E}_3^{*(\sigma+1)}(z)|| \quad = \quad ||\sum_{j=0}^{\sigma}\mathcal{G}_j^{*(\sigma)}(z)\cdot\mathcal{L}_3^{*(j)}(z)||$$

$$
\begin{aligned}
&\leq \quad ||\theta^{*(\sigma)}_{VII}(z)\cdot A^*(z)|| + \sum_{j=0}^{\sigma-1}\mathcal{G}_j^{*(\sigma)}(z)\cdot||\theta^{*(j)}_{VII}(z)\cdot A^*(z)|| \\
&\leq \quad 2\left\{1.01(||\nu^{(\sigma)}||+k+1)\cdot||\widehat{S}^{*(\sigma)}(z)||\cdot||S^{*(\sigma)}(z)||\cdot\mu\right\} \\
&\qquad + \sum_{j=0}^{\sigma-1} 2\left\{1.01(||\nu^{(j)}||+k+1)\cdot||\widehat{S}^{*(j)}(z)||\cdot||S^{*(j)}(z)||\cdot\mu\right\}||\mathcal{G}_j^{*(\sigma)}(z)|| \\
&\leq \quad \left\{2.02\cdot(||\nu^{(\sigma)}||+k+1)\cdot\mu\right\}\left\{2\kappa^{(\sigma)}(k+1)|a_0^{(0)}|\right\}\{k+1\} \\
&\qquad + \sum_{j=0}^{\sigma-1}\left\{2.02(||\nu^{(j)}||+k+1)\cdot\mu\right\}\left\{2\kappa^{(j)}(k+1)|a_0^{(0)}|\right\}\{k+1\} \\
&\qquad\qquad \cdot\left\{2\kappa^{(j+1)}(k+1)|a_0^{(0)}|+O(\mu)\right\}.
\end{aligned}
$$

The result now follows. ∎

Using the above three lemmas, we can finally give the bound on the residual error.

THEOREM 33. *If μ is so small and $\delta T^{(j)^t}(z)$ and $\delta T^{*(j)}(z)$ are not too large so that*

$$(\|n\|+k+1)\mu \leq 0.01$$

and

$$
\begin{aligned}
&\kappa^{(j)}\cdot|a_0^{(0)}|\cdot\left\{||a_0^{-1}(z)||\left[(k+1)||\delta T^{(j)^t}(z)||+||\delta T^{*(j)}(z)||\right]\right. \\
&\quad + \quad \left.1.01(k+1)(||\nu^{(j)}||+k+1)\cdot\mu\right\} \leq \frac{1}{2},\ j\leq\sigma,
\end{aligned}
$$

then

$$
||\delta T^{*(\sigma+1)}(z)|| \leq F_\sigma^* + 2(k+1)\cdot|a_0^{(0)}|\sum_{j=0}^{\sigma-1}\kappa^{(j+1)}F_j^*, \tag{86}
$$

where

$$
\begin{aligned}
F_j^* &= 8\kappa^{(j)}(k+1)^2\cdot|a_0^{(0)}|\cdot\mu \\
&\left\{(||m^{(j)}||+1)+8(k+1)^4\rho_j^*||\nu^{(j)}||^4+(||\nu^{(j)}||+k+1)\right\}.
\end{aligned} \tag{87}
$$

Proof. Sum the error bounds given in Lemmas 30, 31 and 32 ∎

9. The Inverse of a Striped Sylvester Matrix. In this section, a formula is given for the inverse of $\mathcal{M}_n$ expressed in terms of both $S(z)$ and $S^*(z)$. This enables estimating the condition number of $\mathcal{M}_n$ without explicitly computing $\mathcal{M}_n^{-1}$.

Associated with the NPHS $S(z)$, define the order $\|n\|$ matrices

$$\mathcal{P} = \left[\begin{array}{ccc|ccc|c|ccc} p^{(0)} & \cdots & p^{(n_0-1)} & q_1^{(0)} & \cdots & q_1^{(n_1-1)} & & q_k^{(0)} & \cdots & q_k^{(n_k-1)} \\ \vdots & \iddots & 0 & \vdots & \iddots & 0 & & \vdots & \iddots & 0 \\ p^{(n_0-1)} & \iddots & & q_1^{(n_1-1)} & \iddots & & \cdots & q_k^{(n_k-1)} & \iddots & \\ 0 & & \vdots & 0 & & \vdots & & 0 & & \vdots \\ \vdots & & & \vdots & & & & \vdots & & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & & 0 & \cdots & 0 \end{array}\right]$$

(88)

and, for $\beta = 1, 2, \ldots, k$,

$$\mathcal{U}_\beta = \left[\begin{array}{ccc|ccc|c|ccc} u_\beta^{(1)} & \cdots & u_\beta^{(n_0)} & v_{1,\beta}^{(1)} & \cdots & v_{1,\beta}^{(n_1)} & & v_{k,\beta}^{(1)} & \cdots & v_{k,\beta}^{(n_k)} \\ \vdots & \iddots & 0 & \vdots & \iddots & 0 & & \vdots & \iddots & 0 \\ u_\beta^{(n_0)} & \iddots & & v_{1,\beta}^{(n_1)} & \iddots & & \cdots & v_{k,\beta}^{(n_k)} & \iddots & \\ 0 & & \vdots & 0 & & \vdots & & 0 & & \vdots \\ \vdots & & & \vdots & & & & \vdots & & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & & 0 & \cdots & 0 \end{array}\right]. \tag{89}$$

Finally, for any power series $a(z) = \sum_{\ell=0}^{\infty} a^{(\ell)} z^\ell$, and any integer function $f(i,j)$, $i, j = 1, 2, \ldots$, let $\left[a^{(f(i,j))}\right]$ denote a matrix of order $\|n\|$ whose element in position (i,j) is $a^{(f(i,j))}$.

The main result of this section is Theorem 34 below which gives the inverse of $\mathcal{M}_n$ in terms of the NPHS $S(z)$ and the NSPS $S^*(x)$ of types n for $A(z)$.

THEOREM 34. *In terms of the* **normalized** *NPHS $S(z)$ and the* **normalized** *NSPS $S^*(x)$ of types n for $A(z)$, the inverse of $\mathcal{M}_n$ satisfies*

$$\mathcal{M}_n^{-1} \left\{ \left[a_0^{(i-j)}\right] + \theta_{VIII} \right\} = a_0^{(0)} \left\{ \mathcal{P}^t \left[v^{*(\|n\|-i-j+1)}\right] + \sum_{\beta=1}^{k} \mathcal{U}_\beta^t \left[q_\beta^{*(\|n\|-i-j)}\right] \right\}, \tag{90}$$

where

$$\begin{aligned} \theta_{VIII} &= a_0^{(0)} \left\{ \left[(\theta_{IV})_0^{(i-j)}\right] - \sum_{\alpha=0}^{k} \left[a_\alpha^{(\|n\|+i-j)}\right] \left[(\theta_{III})_{\alpha,0}^{(i-j+1)}\right] \right. \\ &\quad \left. + \left[\delta r^{(i+j-2)}\right] \left[v^{*(\|n\|-i-j+1)}\right] + \sum_{\beta=1}^{k} \left[\delta w_\beta^{(i+j-1)}\right] \left[q_\beta^{*(\|n\|-i-j)}\right] \right\}. \end{aligned}$$

Proof. The coefficient of z^{i+j-2}, for $i, j = 1, 2, \ldots, \|n\|$, in the first component of (10), namely,

$$a_0(z)\; p(z) + \sum_{\alpha=1}^{k} a_\alpha(z)\; q_\alpha(z) = z^{\|n\|-1} r(z) + \delta r(z),$$

is

$$\sum_{\ell=0}^{n_0} a_0^{(i+j-\ell-2)} p^{(\ell)} + \sum_{\alpha=1}^{k} \sum_{\ell=0}^{n_\alpha-1} a_\alpha^{(i+j-\ell-2)} q_\alpha^{(\ell)} = r^{(-\|n\|+i+j-1)} + \delta r^{(i+j-2)}.$$

This is the $(i,j)th$ component of

$$\begin{aligned} (91) \quad \left[r^{(-\|n\|+i+j-1)}\right] + \left[\delta r^{(i+j-2)}\right] &= \left[a_0^{(\|n\|+i-j)}\right] \left[p^{(-\|n\|+i+j-2)}\right] \\ &+ \sum_{\alpha=1}^{k} \left[a_\alpha^{(\|n\|+i-j)}\right] \left[q_\alpha^{(-\|n\|+i+j-2)}\right] + \mathcal{M}_n \, \mathcal{P}^t. \end{aligned}$$

Similarly, the coefficient of z^{i+j-1}, for $i,j = 1,2,\ldots,\|n\|$, in the $(\beta+1)st$ component, $\beta = 1,\ldots,k$, of (10), namely,

$$a_0(z) \; u_\beta(z) + \sum_{\alpha=1}^{k} a_\alpha(z) \; v_{\alpha,\beta}(z) = z^{\|n\|+1} w_\beta(z) + \delta w_\beta(z),$$

is

$$\sum_{\ell=0}^{n_0} a_0^{(i+j-\ell-1)} u_\beta^{(\ell)} + \sum_{\alpha=1}^{k} \sum_{\ell=0}^{n_\alpha} a_\alpha^{(i+j-\ell-1)} v_{\alpha,\beta}^{(\ell)} = w_\beta^{(-\|n\|+i+j-2)} + \delta w_\beta^{(i+j-1)}.$$

This is the $(i,j)th$ component of

$$\begin{aligned} (92) \quad \left[w_\beta^{(-\|n\|+i+j-2)}\right] + \left[\delta w_\beta^{(i+j-1)}\right] &= \left[a_0^{(\|n\|+i-j)}\right] \left[u_\beta^{(-\|n\|+i+j-1)}\right] \\ &+ \sum_{\alpha=1}^{k} \left[a_\alpha^{(\|n\|+i-j)}\right] \left[v_{\alpha,\beta}^{(-\|n\|+i+j-1)}\right] + \mathcal{M}_n \, \mathcal{U}_\beta^t. \end{aligned}$$

Next, the coefficient of z^{i-j-1} for $i,j = 1,\ldots,\|n\|$, in the first row and first column of (48) for a normalized NPHS and a normalized NSPS, namely,

$$p(z) \; v^*(z) + \sum_{\beta=1}^{k} u_\beta(z) q_\beta^*(z) = z^{\|n\|-1} (a_0^{(0)})^{-1} + z^{-2} (\theta_{III})_{0,0}(z),$$

is

$$\sum_{\ell=0}^{n_0-1} v^{*(i-j-\ell-1)} \; p^{(l)} + \sum_{\beta=1}^{k} \sum_{\ell=0}^{n_0} q_\beta^{*(i-j-\ell-1)} u_\beta^{(\ell)} = (\theta_{III})_{0,0}^{(i+j-1)}.$$

This is the $(i,j)th$ component of

$$\begin{aligned} (93) \quad \left[p^{(-\|n\|+i+j-2)}\right] \left[v^{*(\|n\|-i-j+1)}\right] &+ \sum_{\beta=1}^{k} \left[u_\beta^{(-\|n\|+i+j-1)}\right] \left[q_\beta^{*(\|n\|-i-j)}\right] \\ &= \left[(\theta_{III})_{0,0}^{(i-j+1)}\right]. \end{aligned}$$

The coefficient of z^{i-j-1} in the first column and the $(\alpha+1)st$ row, $\alpha = 1,\ldots,k$, of (48), namely,

$$q_\alpha(z) \; v^*(z) + \sum_{\beta=1}^{k} v_{\alpha,\beta}(z) q_\beta^*(z) = z^{-2} (\theta_{III})_{\alpha,0}(z)$$

is

$$\sum_{\ell=0}^{n_\alpha} v^{*(i-j-\ell-1)}\, q_\alpha^{(l)} + \sum_{\beta=1}^{k}\sum_{\ell=0}^{n_\alpha} q_\beta^{*(i-j-\ell-1)} v_{\alpha,\beta}^{(\ell)} = (\theta_{III})_{\alpha,0}^{(i-j+1)}.$$

This is the $(i,j)th$ component of

$$\text{(94)}\quad \left[q_\alpha^{(-\|n\|+i+j-2)}\right]\left[v^{*(\|n\|-i-j+1)}\right] + \sum_{\beta=1}^{k}\left[v_{\alpha,\beta}^{(-\|n\|+i+j-1)}\right]\left[q_\beta^{*(\|n\|-i-j)}\right] = \left[(\theta_{III})_{\alpha,0}^{(i-j+1)}\right].$$

Also, the coefficient of z^{i-j}, for $i,j=1,\ldots,\|n\|$ in the first component of (50) for a normalized NPHS and NSPS, namely,

$$r(z)v^*(z) \; + \; z^2\sum_{\beta=1}^{k} w_\beta(z)q_\beta^*(z) = (a_0^{(0)})^{-1}\; a_0(z) + (\theta_{IV})_0(z).$$

is the $(i,j)th$ component of

$$\begin{aligned}(a_0^{(0)})^{-1}\;\left[a_0^{(i-j)}\right] + \left[(\theta_{IV})_0^{(i-j)}\right] \;&=\; \left[r^{(-\|n\|+i+j-1)}\right]\left[v^{*(\|n\|-i-j+1)}\right]\\ &\quad+\; \sum_{\beta=1}^{k}\left[w_\beta^{(-\|n\|+i+j-2)}\right]\left[q_\beta^{*(\|n\|-i-j)}\right].\end{aligned}$$

We are finally ready to prove the theorem. From (91), (92), (93), (94) and (95),

$$\begin{aligned}
&\mathcal{M}_n\left\{\mathcal{P}^t\left[v^{*(\|n\|-i-j+1)}\right] + \sum_{\beta=1}^{k}\mathcal{U}_\beta^t\left[q_\beta^{*(\|n\|-i-j)}\right]\right\}\\
&= \;\Bigg\{\left[r^{(-\|n\|+i+j-1)}\right] + \left[\delta r^{(i+j-2)}\right] - \left[a_0^{(\|n\|+i-j)}\right]\left[p^{(-\|n\|+i+j-2)}\right]\\
&\qquad -\sum_{\alpha=1}^{k}\left[a_\alpha^{(\|n\|+i-j)}\right]\left[q_\alpha^{(-\|n\|+i+j-2)}\right]\Bigg\}\left[v^{*(\|n\|-i-j+1)}\right]\\
&\qquad +\sum_{\beta=1}^{k}\Bigg\{\left[w_\beta^{(-\|n\|+i+j-2)}\right] + \left[\delta w_\beta^{(i+j-1)}\right] - \left[a_0^{(\|n\|+i-j)}\right]\left[u_\beta^{(-\|n\|+i+j-1)}\right]\\
&\qquad -\sum_{\alpha=1}^{k}\left[a_\alpha^{(\|n\|+i-j)}\right]\left[v_{\alpha,\beta}^{(-\|n\|+i+j-1)}\right]\Bigg\}\left[q_\beta^{*(\|n\|-i-j)}\right]\\
&= \;\left[r^{(-\|n\|+i+j-1)}\right]\left[v^{*(\|n\|-i-j+1)}\right] + \sum_{\beta=1}^{k}\left[w_\beta^{(-\|n\|+i+j-2)}\right]\left[q_\beta^{*(\|n\|-i-j)}\right]\\
&+ \;\left[\delta r^{(i+j-2)}\right]\left[v^{*(\|n\|-i-j+1)}\right] + \sum_{\beta=1}^{k}\left[\delta w_\beta^{(i+j-1)}\right]\left[q_\beta^{*(\|n\|-i-j)}\right]\\
&- \;\sum_{\alpha=0}^{k}\left[a_\alpha^{(\|n\|-i-j)}\right]\left[(\theta_{III})_{\alpha,0}^{(i-j+1)}\right]
\end{aligned}$$

$$= \quad (a_0^{(0)})^{-1} \left[a_0^{(i-j)}\right] + \theta_{VIII}.$$

The result (90) now follows. ∎

Corollary 35 below drops the requirement in Theorem 34 that $S(z)$ and $S^*(z)$ be normalized. In particular, the results of the corollary apply when $S(z)$ and $S^*(z)$ are scaled.

COROLLARY 35. *In terms of the NPHS $S(z)$ (unnormalized) of type n for $A(z)$ and the NSPS $S^*(z)$ (unnormalized) of type n for $A^*(z)$, the inverse of $\mathcal{M}_n$ is given by*

$$\mathcal{M}_n^{-1}\left\{\left[a_0^{(i-j)}\right] + \theta_{IX}\right\} = a_0^{(0)}\left\{(\gamma_0\gamma_0^*)^{-1}\mathcal{P}^t\left[v^{*(\|n\|-i-j+1)}\right] + \sum_{\beta=1}^{k}(\gamma_\beta\gamma_\beta^*)^{-1}\mathcal{U}_\beta^t\left[q_\beta^{*(\|n\|-i-j)}\right]\right\}, \tag{95}$$

where

$$\begin{aligned}\theta_{IX} \quad = \quad & a_0^{(0)}\left\{\left[(\theta_{IV})_0^{(i-j)}\right] - \sum_{\alpha=0}^{k}\left[a_\alpha^{(\|n\|+i-j)}\right]\left[(\theta_{III})_{\alpha,0}^{(i-j+1)}\right]\right.\\ & +(\gamma_0\gamma_0^*)^{-1}\left[\delta r^{(i+j-2)}\right]\left[v^{*(\|n\|-i-j+1)}\right]\\ & \left.+\sum_{\beta=1}^{k}(\gamma_\beta\gamma_\beta^*)^{-1}\left[\delta w_\beta^{(i+j-1)}\right]\left[q_\beta^{*(\|n\|-i-j)}\right]\right\}.\end{aligned}$$

Proof. The normalized NPHS is obtained from an unnormalized one by multiplying it on the right by the diagonal matrix $diag[\gamma_0^{-1},\ldots,\gamma_k^{-1}]$. Similarly, the normalized NSPS is obtained from an unnormalized one by multiplying it on the left by the diagonal matrix $diag[\gamma_0^{*-1},\ldots,\gamma_k^{*-1}]$. The result now follows directly from (90). Note that in the definition of θ_{IX}, we continue to associate $\theta_{III}(x)$ and $\theta_{IV}(z)$ with a normalized NSPS. ∎

EXAMPLE 36. Continuing with Examples 2 and 3, according to Theorem 34, the NPHS (25) and the NSPS (44) give the inverse of the striped Sylvester matrix $\mathcal{M}_n$ in (22) as

$$\mathcal{M}_n^{-1}\begin{bmatrix}1&0&0&0&0&0\\-1&1&0&0&0&0\\2&-1&1&0&0&0\\-2&2&-1&1&0&0\\3&-2&2&-1&1&0\\-3&3&-2&2&-1&1\end{bmatrix}$$

$$= \frac{1}{(37)^2}\left\{\begin{bmatrix}-4&44&0&0&0&0\\44&0&0&0&0&0\\-22&36&-9&0&0&0\\36&-9&0&0&0&0\\-9&0&0&0&0&0\\-4&0&0&0&0&0\end{bmatrix}\begin{bmatrix}0&5&0&10&-57&37\\5&0&10&-57&37&0\\0&10&-57&37&0&0\\10&-57&37&0&0&0\\-57&37&0&0&0&0\\37&0&0&0&0&0\end{bmatrix}\right.$$

$$
+ \begin{bmatrix} -73 & -48 & 0 & 0 & 0 & 0 \\ -48 & 0 & 0 & 0 & 0 & 0 \\ -13 & -9 & -7 & 0 & 0 & 0 \\ -9 & -7 & 0 & 0 & 0 & 0 \\ -7 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -24 & 37 & -48 & 22 & 0 \\ -24 & 37 & -48 & 22 & 0 & 0 \\ 37 & -48 & 22 & 0 & 0 & 0 \\ -48 & 22 & 0 & 0 & 0 & 0 \\ 22 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}
$$

$$
\left. + \begin{bmatrix} -44 & 3 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 0 \\ -131 & 137 & 123 & 0 & 0 & 0 \\ 137 & 123 & 0 & 0 & 0 & 0 \\ 123 & 0 & 0 & 0 & 0 & 0 \\ -44 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 & -2 & 4 & 0 \\ -1 & 0 & -2 & 4 & 0 & 0 \\ 0 & -2 & 4 & 0 & 0 & 0 \\ -2 & 4 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \right\}
$$

$$
= \frac{1}{37} \begin{bmatrix} 37 & 0 & 1 & 0 & 2 & -4 \\ 0 & 37 & -48 & 74 & -96 & 44 \\ 0 & 0 & 24 & -37 & 48 & -22 \\ 0 & 0 & -9 & 37 & -55 & 36 \\ 0 & 0 & -7 & 0 & 23 & -9 \\ 0 & 0 & 1 & 0 & 2 & -4 \end{bmatrix}.
$$

10. The Inverse of a Mosaic Sylvester Matrix. In this section, a formula is given for the inverse of $\mathcal{M}_n^*$ expressed in terms of both $S(z)$ and $S^*(z)$. This enables estimating the condition number of $\mathcal{M}_n^*$ without explicitly computing $\mathcal{M}_n^{*-1}$.

Associated with the NPHS $S(z)$ and the NSPS $S^*(z)$, for $\beta = 1, 2, \ldots, k$, define the $\|n\| \times k\|n\|$ matrices

$$
\mathcal{V}_\beta = \left[\begin{array}{ccc|c|ccc} v_{1,\beta}^{(\|n\|-1)} & \cdots & v_{1,\beta}^{(0)} & & v_{k,\beta}^{(\|n\|-1)} & \cdots & v_{k,\beta}^{(0)} \\ \vdots & \iddots & & \cdots & \vdots & \iddots & \\ v_{1,\beta}^{(0)} & & & & v_{k,\beta}^{(0)} & & \end{array} \right],
$$

$$
\mathcal{Q} = \left[\begin{array}{ccc|c|ccc} q_1^{(\|n\|-2)} & \cdots & q_1^{(0)} \quad 0 & & q_k^{(\|n\|-2)} & \cdots & q_k^{(0)} \quad 0 \\ \vdots & \iddots & & \cdots & \vdots & \iddots & \\ q_1^{(0)} & \iddots & & & q_k^{0)} & \iddots & \\ 0 & & & & 0 & & \end{array} \right],
$$

$$
\mathcal{V}^* = \left[\begin{array}{ccc|ccc|c|ccc} v^{*(1)} & \cdots & v^{*(\eta_0)} & u_1^{*(1)} & \cdots & u_1^{*(\eta_1)} & & u_k^{*(1)} & \cdots & u_k^{*(\eta_k)} \\ \vdots & \iddots & 0 & \vdots & \iddots & 0 & & \vdots & \iddots & 0 \\ v^{*(\eta_0)} & \iddots & & u_1^{*(\eta_1)} & \iddots & & \cdots & u_k^{*(\eta_k)} & \iddots & \\ 0 & & \vdots & 0 & & \vdots & & 0 & & \vdots \\ \vdots & & & \vdots & & & & \vdots & & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & & 0 & \cdots & 0 \end{array} \right]
$$

and

$$\mathcal{Q}^*_\beta = \left[\begin{array}{ccc|ccc|c|ccc} q_\beta^{*(0)} & \cdots & q_\beta^{*(\eta_0-1)} & p_{\beta,1}^{*(0)} & \cdots & p_{\beta,1}^{*(\eta_1-1)} & & p_{\beta,k}^{*(0)} & \cdots & p_{\beta,k}^{*(\eta_k-1)} \\ \vdots & \iddots & 0 & \vdots & \iddots & 0 & & \vdots & \iddots & 0 \\ q_\beta^{*(\eta_0-1)} & \iddots & & p_{\beta,1}^{*(\eta_1-1)} & \iddots & & \cdots & p_{\beta,k}^{*(\eta_k-1)} & \iddots & \\ 0 & & \vdots & 0 & & \vdots & & 0 & & \vdots \\ \vdots & & & \vdots & & & & \vdots & & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & & 0 & \cdots & 0 \end{array}\right].$$

where $\eta_\beta = \|n\| - n_\beta$. For $\beta = 1, 2, \ldots, k$, also define the $\|n\| \times k\|n\|$ residual error matrices

$$\delta W_\beta = [\delta \bar{W}_\beta, \mathbf{0}_{n_1}, \ldots, \mathbf{0}_{n_k}]$$

and

$$\delta R = [\delta \bar{R}, \mathbf{0}_{n_1}, \ldots, \mathbf{0}_{n_k}],$$

where

$$\delta \bar{W}_\beta = \begin{bmatrix} \delta w_\beta^{(\|n\|-1)} & \cdots & \delta w_\beta^{(n_0)} \\ & & \vdots \\ \vdots & & \delta w_\beta^{(0)} \\ & \iddots & \\ \delta w_\beta^{(0)} & & \end{bmatrix}, \quad \delta \bar{R} = \begin{bmatrix} \delta r^{(\|n\|-2)} & \cdots & \delta r^{(n_0-1)} \\ & & \vdots \\ \vdots & & \delta r^{(0)} \\ & \iddots & 0 \\ \delta r^{(0)} & \iddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix},$$

and $\mathbf{0}_{n_\beta}$ is a $\|n\| \times \|n\| - n_\beta$ matrix of zeroes. Also, let

$$\theta = \begin{bmatrix} \theta_{0,0} & \cdots & \theta_{0,k} \\ \vdots & & \vdots \\ \theta_{k,0} & \cdots & \theta_{k,k} \end{bmatrix},$$

where each $\theta_{\alpha,\beta}$ is an $(\|n\| - n_\alpha) \times (\|n\| - n_\beta)$ matrix given by

$$\theta_{\alpha,\beta} = \begin{bmatrix} (\theta_{III})_{\alpha,\beta}^{(\|n\|+1)} & \cdots & (\theta_{III})_{\alpha,\beta}^{(2\|n\|-n_\beta)} \\ \vdots & & \vdots \\ (\theta_{III})_{\alpha,\beta}^{(n_\alpha+2)} & \cdots & (\theta_{III})_{\alpha,\beta}^{(\|n\|+n_\alpha-n_\beta+1)} \end{bmatrix}$$

with $\theta_{III}(z)$ the error appearing in equation (48). Finally, let $\left[a_0^{(i-j)}\right]$ denote an order $\|n\|$, lower triangular, matrix as in §9.

The main result of this section is Theorem 37 below which gives the inverse of $\mathcal{M}_n^*$ in terms of the NPHS $S(z)$ and the NSPS $S^*(x)$ of types n for $A(z)$.

THEOREM 37. *In terms of the* **normalized** *NPHS $S(z)$ and the* **normalized** *NSPS $S^*(x)$ of types n for $A(z)$, the inverse of $\mathcal{M}_n^*$ satisfies*

$$(96)\quad \mathcal{M}_n^{*-1}\left\{(a_0^{(0)})^{-1} I_{k\|n\|} + \theta^*_{VIII}\right\} = \mathcal{Q}^t \left[a_0^{(i-j)}\right]^{-1} \mathcal{V}^* + \sum_{\beta=1}^{k} \mathcal{V}_\beta^t \left[a_0^{(i-j)}\right]^{-1} \mathcal{Q}^*_\beta,$$

where

$$\theta^*_{VIII} \quad = \quad \theta - \delta R^t \left[a_0^{(i-j)}\right]^{-1} \mathcal{V}^* - \sum_{\beta=1}^{k} \delta W^t_\beta \left[a_0^{(i-j)}\right]^{-1} \mathcal{Q}^*_\beta \tag{97}$$

Proof. Let

$$\bar{\mathcal{Q}} = \left[\begin{array}{ccc|ccc|c|ccc} p^{(\|n\|-2)} & \cdots & p^{(n_0-1)} & q_1^{(\|n\|-2)} & \cdots & q_1^{(n_1-1)} & & q_k^{(\|n\|-2)} & \cdots & q_k^{(n_k-1)} \\ & & \vdots & & & \vdots & & & & \vdots \\ \vdots & & p^{(0)} & \vdots & & q_1^{(0)} & \cdots & \vdots & & q_k^{(0)} \\ & \iddots & 0 & & \iddots & 0 & & & \iddots & 0 \\ p^{(0)} & \iddots & & q_1^{(0)} & \iddots & & & q_k^{(0)} & \iddots & \\ 0 & & & 0 & & & & 0 & & \end{array}\right].$$

Then, the order condition (10) for an NPHS implies that

$$\mathcal{M}^*_n \cdot \mathcal{Q}^t = \bar{\mathcal{Q}}^t \cdot \left[a_0^{(i-j)}\right] - \delta R^t. \tag{98}$$

To see this, note the $(i,j)th$ component, $1 \le i \le \|n\| - n_0$, $1 \le j \le \|n\|$, of (98) is the coefficient of $z^{\|n\|-i-j}$ in

$$a_0(z)\; p(z) + \sum_{\alpha=1}^{k} a_\alpha(z)\; q_\alpha(z) = z^{\|n\|-1} r(z) + \delta r(z).$$

The remaining components of (98) are obvious identities.

Similarly, for $1 \le \beta \le k$, let

$$\bar{\mathcal{V}}_\beta = \left[\begin{array}{ccc|ccc|c|ccc} u_\beta^{(\|n\|-1)} & \cdots & u_\beta^{(n_0)} & v_{1,\beta}^{(\|n\|-1)} & \cdots & v_{1,\beta}^{(n_1)} & & v_{k,\beta}^{(\|n\|-1)} & \cdots & v_{k,\beta}^{(n_k)} \\ & & \vdots & & & \vdots & & & & \vdots \\ \vdots & & u_\beta^{(0)} & \vdots & & v_{1,\beta}^{(0)} & \cdots & \vdots & & v_{k,\beta}^{(0)} \\ & \iddots & & & \iddots & & & & \iddots & \\ u_\beta^{(0)} & & & v_{1,\beta}^{(0)} & & & & v_{k,\beta}^{(0)} & & \end{array}\right].$$

Then, the coefficient of $z^{\|n\|-i-j+1}$, $1 \le i \le \|n\| - n_0$, $1 \le j \le \|n\|$ in the order condition (10) for an NPHS, namely,

$$a_0(z)\; u_\beta(z) + \sum_{\alpha=1}^{k} a_\alpha(z)\; v_{\alpha,\beta}(z) = z^{\|n\|+1} w_\beta(z) + \delta w_\beta(z),$$

gives the $(i,j)th$ component of

$$\mathcal{M}^*_n \cdot \mathcal{V}^t_\beta = \bar{\mathcal{V}}^t_\beta \cdot \left[a_0^{(i-j)}\right] - \delta W^t. \tag{99}$$

The remaining components of (99) are easy to verify.

Next, observe that the duality theorem 4 and its corollary 5 imply that

$$
\bar{\mathcal{Q}}^t \cdot \mathcal{V}^* + \sum_{\beta=1}^{k} \bar{\mathcal{V}}_\beta^t \cdot \mathcal{Q}_\beta^* = (a_0^{(0)})^{-1} I_{k\|n\|} + \theta. \tag{100}
$$

Combining (98), (99) and (100), we obtain the result (96). ∎

Corollary 38 below drops the requirement in Theorem 37 that $S(z)$ and $S^*(z)$ be normalized. In particular, the results of the corollary apply when $S(z)$ and $S^*(z)$ are scaled.

COROLLARY 38. *In terms of the NPHS $S(z)$ (unnormalized) of type n for $A(z)$ and the NSPS $S^*(z)$ (unnormalized) of type n for $A^*(z)$, the inverse of $\mathcal{M}_n^*$ is given by*

$$
\begin{aligned}
(101) \qquad & \mathcal{M}_n^{*-1}\left\{(a_0^{(0)})^{-1} I_{k\|n\|} + \theta_{IX}^*\right\} \\
&= (\gamma_0\gamma_0^*)^{-1}\mathcal{Q}^t \left[a_0^{(i-j)}\right]^{-1} \mathcal{V}^* + \sum_{\beta=1}^{k} (\gamma_\beta\gamma_\beta^*)^{-1}\mathcal{V}_\beta^t \left[a_0^{(i-j)}\right]^{-1} \mathcal{Q}_\beta^*,
\end{aligned}
$$

where

$$
(102)\theta_{IX}^* \;=\; \theta - (\gamma_0\gamma_0^*)^{-1}\delta R^t \left[a_0^{(i-j)}\right]^{-1} \mathcal{V}^* - \sum_{\beta=1}^{k} (\gamma_\beta\gamma_\beta^*)^{-1}\delta W_\beta^t \left[a_0^{(i-j)}\right]^{-1} \mathcal{Q}_\beta^*
$$

Proof. The normalized NPHS is obtained from an unnormalized one by multiplying it on the right by the diagonal matrix $diag[\gamma_0^{-1}, \ldots, \gamma_k^{-1}]$. Similarly, the normalized NSPS is obtained from an unnormalized one by multiplying it on the left by the diagonal matrix $diag[\gamma_0^{*-1}, \ldots, \gamma_k^{*-1}]$. The result now follows directly from (96). ∎

EXAMPLE 39. Continuing with Examples 2 and 3, according to Theorem 37, the NPHS (25) and the NSPS (44) give the inverse of the mosaic Sylvester matrix. The relevant matrices in the inverse formula (96) are

$$
\mathcal{M}_n^* = \left[\begin{array}{cccccc|cccccc}
0 & -2 & 0 & -3 & 0 & -4 & 1 & -1 & -5 & -3 & -2 & 2 \\
0 & 0 & -2 & 0 & -3 & 0 & 0 & 1 & -1 & -5 & -3 & -2 \\
0 & 0 & 0 & -2 & 0 & -3 & 0 & 0 & 1 & -1 & -5 & -3 \\
0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 1 & -1 & -5 \\
\hline
1 & -1 & 2 & -2 & 3 & -3 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 2 & -2 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 2 & -2 & 3 & -3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 2 & -2 & 3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 2 & -2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1
\end{array}\right],
$$

$$
[a_0^{(i-j)}]^{-1} = \begin{bmatrix}
1 & 1 & -1 & -1 & 0 & 0 \\
0 & 1 & 1 & -1 & -1 & 0 \\
0 & 0 & 1 & 1 & -1 & -1 \\
0 & 0 & 0 & 1 & 1 & -1 \\
0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix},
$$

$$\mathcal{Q} = \frac{1}{37}\left[\begin{array}{rrrrrr|rrrrrr}
0 & 0 & -9 & 36 & -22 & 0 & 0 & 0 & 0 & 0 & -4 & 0 \\
0 & -9 & 36 & -22 & 0 & 0 & 0 & 0 & 0 & -4 & 0 & 0 \\
-9 & 36 & -22 & 0 & 0 & 0 & 0 & 0 & -4 & 0 & 0 & 0 \\
36 & -22 & 0 & 0 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 0 \\
-22 & 0 & 0 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right],$$

$$\mathcal{V}^* = \frac{1}{37}\left[\begin{array}{rrrr|rrr|rrrrr}
-57 & 10 & 0 & 5 & 74 & -40 & -57 & 57 & 249 & -103 & -428 & -159 \\
10 & 0 & 5 & 0 & -40 & -57 & 0 & 249 & -103 & -428 & -159 & 0 \\
0 & 5 & 0 & 0 & -57 & 0 & 0 & -103 & -428 & -159 & 0 & 0 \\
5 & 0 & 0 & 0 & 0 & 0 & 0 & -428 & -159 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -159 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right],$$

$$\mathcal{V}_1 = \frac{1}{37}\left[\begin{array}{rrrrrr|rrrrrr}
0 & 0 & -7 & -9 & -13 & 37 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & -7 & -9 & -13 & 37 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
-7 & -9 & -13 & 37 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
-9 & -13 & 37 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
-13 & 37 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
37 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right],$$

$$\mathcal{Q}_1^* = \frac{1}{37}\left[\begin{array}{rrrr|rrr|rrrrr}
22 & -48 & 37 & -24 & 0 & 44 & -52 & -22 & 48 & 117 & -136 & -147 \\
-48 & 37 & -24 & 0 & 44 & -52 & 0 & 48 & 117 & -136 & -147 & 0 \\
37 & -24 & 0 & 0 & -52 & 0 & 0 & 117 & -136 & -147 & 0 & 0 \\
-24 & 0 & 0 & 0 & 0 & 0 & 0 & -136 & -147 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -147 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right],$$

$$\mathcal{V}_2 = \frac{1}{37}\left[\begin{array}{rrrrrr|rrrrrr}
0 & 0 & 123 & 137 & -131 & 0 & 0 & 0 & 0 & 0 & -44 & 37 \\
0 & 123 & 137 & -131 & 0 & 0 & 0 & 0 & 0 & -44 & 37 & 0 \\
123 & 137 & -131 & 0 & 0 & 0 & 0 & 0 & -44 & 37 & 0 & 0 \\
137 & -131 & 0 & 0 & 0 & 0 & 0 & -44 & 37 & 0 & 0 & 0 \\
-131 & 0 & 0 & 0 & 0 & 0 & -44 & 37 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 37 & 0 & 0 & 0 & 0 & 0
\end{array}\right]$$

and

$$\mathcal{Q}_2^* = \frac{1}{37}\left[\begin{array}{rrrr|rrr|rrrrr}
4 & -2 & 0 & -1 & 0 & 8 & 4 & -4 & 2 & 28 & 19 & -20 \\
-2 & 0 & -1 & 0 & 8 & 4 & 0 & 2 & 28 & 19 & -20 & 0 \\
0 & -1 & 0 & 0 & 4 & 0 & 0 & 28 & 19 & -20 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 19 & -20 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -20 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right].$$

Using (96), we now obtain

$$\mathcal{M}_n^{*-1} = \frac{1}{37^2}\left[\begin{array}{rrrrrrr}
-333 & 851 & 0 & -259 & 1369 & 703 & -333 \\
999 & -1184 & 1369 & -592 & 0 & 3367 & 999 \\
851 & -1110 & 0 & 814 & 0 & 1702 & 851 \\
-1813 & 2960 & -2738 & 1480 & 0 & -3626 & 2294 \\
-518 & 259 & 0 & -555 & 0 & -1036 & -518 \\
814 & -1776 & 1369 & -888 & 0 & 1628 & -1924 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-148 & 74 & 0 & 37 & 0 & -296 & -148 \\
-148 & 74 & 0 & 37 & 0 & -296 & -148 \\
148 & -74 & 0 & -37 & 0 & 296 & 148 \\
148 & -74 & 0 & -37 & 0 & 296 & 148
\end{array}\right.$$

$$\left.\begin{matrix}
333 & -851 & -2331 & 3552 & 7141 \\
-999 & 1184 & 5624 & 296 & -888 \\
-851 & 1110 & 5957 & -1776 & -9731 \\
1813 & -2960 & -9953 & 4736 & 6327 \\
518 & -259 & -3626 & -1776 & 2590 \\
-814 & 1776 & 4329 & -5032 & -5439 \\
1369 & 1369 & -1369 & -1369 & 0 \\
0 & 1369 & 1369 & -1369 & -1369 \\
148 & -74 & 333 & 666 & -629 \\
148 & -74 & -1036 & 666 & 2109 \\
-148 & 74 & 1036 & 703 & 629 \\
-148 & 74 & 1036 & 703 & -740
\end{matrix}\right]$$

11. Stability. In this section, bounds for the errors $\delta S(z) = S(z) - S_E(z)$ and $\delta S^*(z) = S^*(z) - S_E^*(z)$ are obtained. Since $S(z)$ and $S^*(z)$ are scaled, these same bounds serve also as bounds for the relative errors in $S(z)$ and $S^*(z)$. To make the comparisons meaningful in the above, we insist that $S_E(z)$ and $S_E^*(z)$ are such that

$$\begin{aligned} V_E(0) &= V(0) = diag[\gamma_1, \ldots, \gamma_k], \\ r_E(0) &= r(0) = \gamma_0, \end{aligned}$$

and

$$\begin{aligned} v_E^*(0) &= v^*(0) = \gamma_0^*, \\ R_E^*(0) &= R^*(0) = diag[\gamma_1^*, \ldots, \gamma_k^*]. \end{aligned}$$

We begin by first finding bounds for $\delta S(z)$. From (6) and (10)

$$A^t(z) \cdot \delta S(z) = \delta T^t(z) + \mathcal{O}(z^{\|n\|+1}).$$

So, the constant terms[9] $\delta u_\beta^{(0)}$ and $\delta v_{\alpha,\beta}^{(0)}$ for $0 \le \alpha, \beta \le k$ of $S(z)$ are zero. It then follows that the remaining components of $\delta S(z)$ satisfy

$$\mathcal{M}_n \cdot \delta\mathcal{X} = [\delta r^{(0)}, \ldots, \delta r^{(\|n\|-1)}]^t, \tag{103}$$

where

$$\delta\mathcal{X} = \left[\delta p^{(0)}, \ldots, \delta p^{(n_0-1)} | \delta q_1^{(0)}, \ldots, \delta q_1^{(n_1-1)} | \cdots | \delta q_k^{(0)}, \ldots, \delta q_k^{(n_k-1)}\right]^t,$$

and

$$\mathcal{M}_n \cdot \delta\mathcal{Y} = \begin{bmatrix} \delta w_1^{(1)} & \cdots & \delta w_k^{(1)} \\ \vdots & & \vdots \\ \delta w_1^{(\|n\|)} & \cdots & \delta w_k^{(\|n\|)} \end{bmatrix}, \tag{104}$$

where

$$\delta\mathcal{Y} = \left[\begin{array}{ccc|ccc|c|ccc} \delta u_1^{(1)} & \cdots & \delta u_1^{(n_0)} & \delta v_{1,1}^{(1)} & \cdots & \delta v_{1,1}^{(n_1)} & & \delta v_{k,1}^{(1)} & \cdots & \delta v_{k,1}^{(n_k)} \\ \vdots & & \vdots & \vdots & & \vdots & \cdots & \vdots & & \vdots \\ \delta u_k^{(1)} & \cdots & \delta u_k^{(n_0)} & \delta v_{1,k}^{(1)} & \cdots & \delta v_{1,k}^{(n_1)} & & \delta v_{k,k}^{(1)} & \cdots & \delta v_{k,k}^{(n_k)} \end{array}\right]^t.$$

[9] In actual fact, the computations in (18) may yield errors resulting in nonzero values of $\delta u_\beta^{(0)}$ for $1 \le \beta \le k$. But, these errors, each resulting from two floating point operations, are comparatively small and are ignored in order to simplify the analysis.

From (103) and (104), it follows that

$$
\begin{aligned}
(105) \qquad \|\delta S(z)\| &\leq \max\{\|\delta\mathcal{X}\|_1, \|\delta\mathcal{Y}\|_1\} \\
&\leq \|n\|\cdot\|\mathcal{M}_n^{-1}\|_1 \cdot \max\{\|\delta r(z)\|, \|\delta W^t(z)\|\} \\
&\leq \|n\|\cdot\|\mathcal{M}_n^{-1}\|_1 \cdot \|\delta T^t(z)\|.
\end{aligned}
$$

Thus, to obtain a bound for $\delta S(z)$, we need only to obtain bounds for $\mathcal{M}_n^{-1}$ and $\delta T^t(z)$. This is done formally in Theorem 40 below. In the theorem, $\delta T^t(z)$ is the residual error corresponding to the NPHS computed by the algorithm of §5 in $\sigma + 1$ steps. So, $n = m^{(\sigma+1)}$ and a bound for $\|\delta T^t(z)\|$ is given by Theorem 22 in which $\delta T^{(\sigma+1)^t}(z) = \delta T^t(z)$. At the point $m^{(\sigma+1)}$, we drop the superscript $\sigma+1$ so that $S(z) = S^{(\sigma+1)}(z)$, $\kappa = \kappa^{(\sigma+1)}$, and so on. A bound for $\mathcal{M}_n^{-1}$ is then obtained directly from Corollary 35 without changes to notation. The point $m^{(\sigma)}$ is the last stable point (i.e., $\kappa^{(\sigma)} \leq \tau$) prior to the point n along the diagonal passing through n. The point n itself need not be stable.

THEOREM 40. *If μ is so small and $\delta T^t(z)$ and $\delta T^*(z)$ are not too large so that the conditions of Theorem 22 are satisfied and*

$$(106) \qquad 2k!\cdot\|n\|\cdot\|a_0^{-1}(z)\|\cdot\|\delta T^t(z)\| \leq \gamma$$

and

$$
\begin{aligned}
(107)\kappa\|n\|^2\cdot\|a_0^{-1}(z)\|\cdot\Big\{&(k+2)\|\delta T^t(z)\| \\
+ &\frac{2(k+1)!\cdot(\|n\|+1)\|a_0^{-1}(z)\|}{\gamma}\left[(k+1)\|\delta T^t(z)\| + \|\delta T^*(z)\|\right]\Big\} \leq 1/2,
\end{aligned}
$$

then

$$(108)\ \|\delta S(z)\| \leq 2\kappa\|n\|^2\cdot|a_0^{(0)}|\cdot\|a_0^{-1}(z)\|\left\{F_\sigma + 2(k+1)\cdot|a_0^{(0)}|\sum_{j=0}^{\sigma-1}\kappa^{(j+1)}F_j\right\},$$

where F_j is defined in (80).

Proof. We begin by finding a bound for θ_{IX} appearing in the inverse formula (95) for $\mathcal{M}_n$. A bound for θ_{IX} depends on bounds for $\theta_I(z)$, $\theta_{II}(z)$, $\theta_{III}(z)$ and $\theta_{IV}(z)$. Using (106) and Hadamard's inequality, a bound for $\theta_I(z)$ in (11) for a scaled NSPS is given by

$$\|\theta_I(z)\| \leq k!\cdot\|n\|\cdot\|a_0^{-1}(z)\|\cdot\|\delta T^t(z)\| \leq \gamma/2.$$

So,

$$|det(S(z))| \geq \gamma/2.$$

Next, $\theta_{II}(z)$ in (46) for a scaled NSPS is bounded by

$$\|\theta_{II}(z)\| \leq \|a_0^{-1}(z)\|\cdot\left\{(k+1)\|\delta T^t(z)\| + \|\delta T^*(z)\|\right\}.$$

Consequently, $\theta_{III}(z)$ in (48) for a scaled NSPS (note the change from a normalized NSPS) is bounded by

$$
\begin{aligned}
(109) \qquad \|\theta_{III}(z)\| &= \|S(z)\ (\Gamma^*\Gamma)^{-1}\theta_{II}(z)\ S^{-1}(z)\| \\
&= \|S(z)\ (\Gamma^*\Gamma)^{-1}\theta_{II}(z)\ S^{adj}(z)/det(S(z)\| \\
&\leq 2\kappa(k+1)!\|\theta_{II}(z)\|/\gamma \\
&\leq \frac{2\kappa(k+1)!\|a_0^{-1}(z)\|}{\gamma}\left\{(k+1)\|\delta T^t(z)\| + \|\delta T^*(z)\|\right\}.
\end{aligned}
$$

In addition, a bound for $\theta_{IV}(z)$ in (50) for a scaled NSPS (here, also, note the change from a normalized NSPS) is given by

$$\begin{aligned}
\|\theta^t_{IV}(z)\| &= \| \left\{ A^t(z)\theta_{III}(z) - \delta T^t(z)(\Gamma^*\Gamma)^{-1}S^*(z)) \right\} / z^{\|n\|+1} \| \\
&\leq \kappa(k+1)\|\delta T^t(z)\| + \frac{2\kappa(k+1)! \cdot \|n\| \cdot \|a_0^{-1}(z)\|}{\gamma} \left\{ (k+1)\|\delta T^t(z)\| + \|\delta T^*(z)\| \right\}.
\end{aligned}$$

We are almost ready to give a bound for θ_{IX} appearing in the inverse formula (95). But, first observe that

$$\| \left[(\theta_{IV})_0^{(i-j)} \right] \|_1 \quad \leq \quad \|n\| \cdot \|\theta^t_{IV}(z)\|$$

and that

$$\begin{aligned}
\| \sum_{\alpha=0}^{k} \left[a_\alpha^{(\|n\|-i-j)} \right] \left[(\theta_{III})_{\alpha,0}^{(i+j-1)} \right] \|_1 &\leq \sum_{\alpha=0}^{k} \|n\| \cdot \|(\theta_{III})_{\alpha,0}(z)\| \\
&\leq \|n\| \cdot \|(\theta_{III})(z)\|.
\end{aligned}$$

Thus,

$$\begin{aligned}
\|\theta_{IX}\|_1 &= \|a_0^{(0)} \left\{ \left[(\theta_{IV})_0^{(i-j)} \right] - \sum_{\alpha=0}^{k} \left[a_\alpha^{(\|n\|+i-j)} \right] \left[(\theta_{III})_{\alpha,0}^{(i-j+1)} \right] \right. \\
&\quad + (\gamma_0\gamma_0^*)^{-1} \left[\delta r^{(i+j-2)} \right] \left[v^{*(\|n\|-i-j+1)} \right] \\
&\quad \left. + \sum_{\beta=1}^{k} (\gamma_\beta\gamma_\beta^*)^{-1} \left[\delta w_\beta^{(i+j-1)} \right] \left[q_\beta^{*(\|n\|-i-j)} \right] \right\} \|_1 \\
&\leq \|n\| \cdot \|\theta^t_{IV}(z)\| + \|n\| \cdot \|\theta_{III}(z)\| \\
&\quad + (\gamma_0\gamma_0^*)^{-1}\|n\| \cdot \|\delta T^t(z)\| + \sum_{\beta=1}^{k} (\gamma_\beta\gamma_\beta^*)^{-1} \|n\| \cdot \|\delta T^t(z)\| \\
&\leq \|n\| \left\{ \|\theta^t_{IV}(z)\| + \|\theta_{III}(z)\| + \kappa\|\delta T^t(z)\| \right\} \\
&\leq \kappa\|n\| \left\{ (k+2)\|\delta T^t(z)\| \right. \\
&\qquad \left. + \frac{2(k+1)!(\|n\|+1)\|a_0^{-1}(z)\|}{\gamma} \left[(k+1)\|\delta T^t(z)\| + \|\delta T^*(z)\| \right] \right\}.
\end{aligned}$$

It then follows from (107) that

$$\text{(110)} \qquad \| \left[a_0^{(i-j)} \right]^{-1} \theta_{IX} \|_1 \leq \|n\| \cdot \|a_0^{-1}(z)\| \cdot \|\theta_{IX}\|_1 \leq 1/2,$$

and so $I_{\|n\|} + \left[a_0^{(i-j)} \right]^{-1} \theta_{IX}$ is invertible (Stewart [27, page 187]). Consequently,

$$\begin{aligned}
\text{(111)} \quad \| \left\{ \left[a_0^{(i-j)} \right] + \theta_{IX} \right\}^{-1} \|_1 &\leq \| \left\{ I_{\|n\|} + \left[a_0^{(i-j)} \right]^{-1} \theta_{IX} \right\}^{-1} \left[a_0^{(i-j)} \right]^{-1} \|_1 \\
&\leq \frac{\| \left[a_0^{(i-j)} \right]^{-1} \|_1}{1 - \| \left[a_0^{(i-j)} \right]^{-1} \theta_{IX} \|_1} \\
&\leq 2 \| \left[a_0^{(i-j)} \right]^{-1} \|_1.
\end{aligned}$$

Therefore, a bound for $\mathcal{M}_n^{-1}$ in (95) is given

$$
\begin{aligned}
(112)\ \|\mathcal{M}_n^{-1}\|_1 \quad &\leq \quad \|\left\{\left[a_0^{(i-j)}\right] + \theta_{IX}\right\}^{-1}\|_1 \cdot \|a_0^{(0)}\ \left\{(\gamma_0\gamma_0^*)^{-1}\mathcal{P}_n^t\left[v^{*(\|n\|-i-j+1)}\right]\right. \\
&\quad\left. + \sum_{\beta=1}^{k}(\gamma_\beta\gamma_\beta^*)^{-1}\mathcal{U}_{n,\beta}^t\left[q_\beta^{*(\|n\|-i-j)}\right]\right\}\|_1 \\
&\leq \quad 2|a_0^{(0)}|\cdot\|\left[a_0^{(i-j)}\right]^{-1}\|_1\sum_{\beta=0}^{k}(\gamma_\beta\gamma_\beta^*)^{-1} \\
&\leq \quad 2\kappa\|n\|\cdot|a_0^{(0)}|\cdot\|a_0^{-1}(z)\|.
\end{aligned}
$$

The result (108) now follows from (105) using (79) and (112). ∎

Next, we find bounds for $\delta S^*(z)$. From (30) and (34)

$$S^*(z)A^*(z) = \delta T^*(z) + \mathcal{O}(z^{\|n\|+1}).$$

As for the NSPS, for the sake of simplicity, here again we ignore that the constant term errors, $\delta w_\beta^{*(0)}$, for $1 \leq \beta \leq k$. This is done with no great loss of generality, since these are the comparatively small errors made in computing $\delta u_\beta^{*(0)}(z)$ from

$$u_\beta^{*(0)}a_0^{(0)} + v^{*(0)}a_\beta^{(0)} = 0$$

with $v^{*(0)} = \gamma_0^*$. It then follows, in a fashion similar to solving (38) and (40) that the remaining components of $\delta S^*(z)$ satisfy

$$\text{(113)}\qquad \delta\mathcal{X}^{*t}\cdot\mathcal{M}_n^* = [\delta w_1^{*(1)},\ldots,\delta w_1^{*(\|n\|)}|\cdots|\delta w_k^{*(1)},\ldots,\delta w_k^{*(\|n\|)}],$$

where

$$\delta\mathcal{X}^{*t} = \left[\delta v^{*(1)},\ldots,\delta v^{*(\|n\|-n_0)}|\delta u_1^{*(1)},\ldots,\delta u_1^{*(\|n\|-n_1)}|\cdots|\delta u_k^{*(1)},\ldots,\delta u_k^{*(\|n\|-n_k)}\right],$$

and, for $1 \leq \alpha \leq k$,

$$\text{(114)}\qquad \delta\mathcal{Y}_\alpha^{*t}\cdot\mathcal{M}_n^* = [\delta r_{\alpha,1}^{*(0)},\ldots,\delta r_{\alpha,1}^{*(\|n\|-1)}|\cdots|\delta r_{\alpha,k}^{*(0)},\ldots,\delta r_{\alpha,k}^{*(\|n\|-1)}],$$

where

$$\delta\mathcal{Y}_\alpha^{*t} = \left[\delta q_\alpha^{*(0)},\ldots,\delta q_\alpha^{*(\|n\|-n_0-1)}|\delta p_{\alpha,1}^{*(0)},\ldots,\delta p_{\alpha,1}^{*(\|n\|-n_1-1)}|\cdots|\delta p_{\alpha,k}^{*(0)},\ldots,\delta p_{\alpha,k}^{*(\|n\|-n_k-1)}\right].$$

From (113) and (114), we get

$$
\text{(115)}\qquad
\begin{aligned}
\|\delta S^*(z)\| \quad &\leq \quad (k+1)\max\{\|\delta\mathcal{X}^*\|_1, \|\delta\mathcal{Y}^*\|_1\} \\
&\leq \quad (k+1)^2\|\mathcal{M}_n^{*-1}\|_\infty\cdot\|\delta T^*(z)\|.
\end{aligned}
$$

Thus, to obtain a bound for $\delta S^*(z)$, we need only to obtain bounds for $\mathcal{M}_n^{*-1}$ and $\delta T^*(z)$. This is done formally in Theorem 41 below. In the theorem, $\delta T^*(z)$ is the residual error corresponding to the NSPS computed by the algorithm of §5 in $\sigma+1$ steps. So, as for the NSPS, $n = m^{(\sigma+1)}$ and a bound for $\|\delta T^*(z)\|$ is given by Theorem 33 in which $\delta T^{*(\sigma+1)}(z) = \delta T^*(z)$. At the point $m^{(\sigma+1)}$, we drop the superscript $\sigma+1$ so that $S^*(z) = S^{*(\sigma+1)}(z)$, $\kappa = \kappa^{(\sigma+1)}$, and so on. A bound for

$\mathcal{M}_n^{*-1}$ is then obtained directly from Corollary 38 without changes to notation. The point $m^{(\sigma)}$ is the last stable point (i.e., $\kappa^{(\sigma)} \le \tau$) prior to the point n along the diagonal passing through n. The point n itself need not be stable.

THEOREM 41. *If μ is so small and $\delta T^t(z)$ and $\delta T^*(z)$ are not too large so that the conditions of Theorem 33 are satisfied and*

$$2k! \cdot \|n\| \cdot \|a_0^{-1}(z)\| \cdot \|\delta T^t(z)\| \le \gamma \tag{116}$$

and

$$\begin{aligned} &\kappa(k+1)|a_0^{(0)}| \cdot \|a_0^{-1}(z)\| \cdot \big\{(k+1)\|n\|^2 \cdot \|\delta T^t(z)\| \\ &\quad + \frac{2(k+1)!}{\gamma}\left[(k+1)\|\delta T^t(z)\| + \|\delta T^*(z)\|\right]\Big\} \le 1/2, \end{aligned} \tag{117}$$

then

$$\begin{aligned} \|\delta S^*(z)\| \quad &\le \quad 2\kappa(k+1)^2\|n\| \cdot |a_0^{(0)}| \cdot \|a_0^{-1}(z)\| \\ &\quad \cdot \left\{ F_\sigma^* + 2(k+1) \cdot |a_0^{(0)}| \sum_{j=0}^{\sigma-1} \kappa^{(j+1)} F_j^* \right\}, \end{aligned} \tag{118}$$

where F_j^* is defined in (87).

Proof. From (109),

$$\begin{aligned} \|\theta\|_\infty \quad &\le \quad (k+1)\|\theta_{III}(z)\| \\ &\le \quad \frac{2\kappa(k+1)(k+1)!\|a_0^{-1}(z)\|}{\gamma}\left\{(k+1)\|\delta T^t(z)\| + \|\delta T^*(z)\|\right\}. \end{aligned}$$

Note that the assumption (116) is used in (109) to derive the bound for $\theta_{III}(z)$. Thus, θ_{IX}^* appearing in the inverse formula (101) is bounded by

$$\begin{aligned} \|\theta_{IX}^*\|_\infty \quad &= \quad \|\theta - (\gamma_0\gamma_0^*)^{-1}\delta R^t \left[a_0^{(i-j)}\right]^{-1} \mathcal{V}^* - \sum_{\beta=1}^{k} (\gamma_\beta\gamma_\beta^*)^{-1}\delta W_\beta^t \left[a_0^{(i-j)}\right]^{-1} \mathcal{Q}_\beta^* \|_\infty \\ &\le \quad \frac{2\kappa(k+1)(k+1)!\|a_0^{-1}(z)\|}{\gamma}\left\{(k+1)\|\delta T^t(z)\| + \|\delta T^*(z)\|\right\} \\ &\quad + (\gamma_0\gamma_0^*)^{-1}(k+1)\|n\|^2 \cdot \|\delta T^t(z)\| \cdot \|a_0^{-1}(z)\| \cdot \|S^*(z)\| \\ &\quad + \sum_{\beta=1}^{k} (\gamma_\beta\gamma_\beta^*)^{-1}(k+1)\|n\|^2 \cdot \|\delta T^t(z)\| \cdot \|a_0^{-1}(z)\| \cdot \|S^*(z)\| \\ &\le \quad \kappa(k+1) \cdot \|a_0^{-1}(z)\| \\ &\quad \cdot \left\{(k+1)\|n\|^2\|\delta T^t(z)\| + \frac{2(k+1)!}{\gamma}\left[(k+1)\|\delta T^t(z)\| + \|\delta T^*(z)\|\right]\right\}. \end{aligned}$$

Therefore, using the assumption (117),

$$\|(a_0^{(0)})^{-1} I_{k\|n\|} + \theta_{IX}^*{}^{-1}\|_\infty \le 2|a_0^{(0)}|$$

and so

$$\|\mathcal{M}_n^{*-1}\|_\infty \quad \le \quad \|\left\{(a_0^{(0)})^{-1} I_{k\|n\|} + \theta_{IX}^*\right\}^{-1}\|_\infty \cdot \|(\gamma_0\gamma_0^*)^{-1}\mathcal{Q}^t \left[a_0^{(i-j)}\right]^{-1} \mathcal{V}^* \tag{119}$$

$$
\begin{aligned}
&+\sum_{\beta=1}^{k}(\gamma_\beta\gamma_\beta^*)^{-1}\mathcal{V}_\beta^t\left[a_0^{(i-j)}\right]^{-1}\mathcal{Q}_\beta^*\|_\infty \\
\leq\ & 2\kappa(k+1)^2\|n\|\cdot|a_0^{(0)}|\cdot\|a_0^{-1}(z)\|.
\end{aligned}
$$

The result (118) now follows from (115) using (86) and (119). ∎

THEOREM 42. *The algorithm VECTOR_PADE for computing $S(z)$ and $S^*(z)$ is weakly stable.*

Proof. If the conditions of Theorem 40 hold, then from (108)

$$
\|\delta S(z)\| \leq 2\kappa\|n\|^2\cdot|a_0^{(0)}|\cdot\|a_0^{-1}(z)\|\left\{\bar{F}_\sigma + 2\tau(k+1)\cdot|a_0^{(0)}|\cdot\sum_{j=0}^{\sigma-1}\bar{F}_j\right\},
$$

where

$$
\bar{F}_j = 4\tau(k+1)\cdot|a_0^{(0)}|\cdot\mu\left\{(||m^{(j)}||+k+1)+4\rho_j||\nu^{(j)}||^4+(||\nu^{(j)}||+k+1)\right\}
$$

Likewise, if the conditions of Theorem 41 hold, then from (118)

$$
\|\delta S^*(z)\| \leq 2\kappa(k+1)^4\|n\|\cdot|a_0^{(0)}|\cdot\|a_0^{-1}(z)\|\left\{\bar{F}_\sigma^* + 2\tau(k+1)\cdot|a_0^{(0)}|\sum_{j=0}^{\sigma-1}\bar{F}_j^*\right\},
$$

where

$$
\begin{aligned}
F_j^* &= 8\tau(k+1)^2\cdot|a_0^{(0)}|\cdot\mu \\
&\quad\left\{(||m^{(j)}||+1)+8(k+1)^4\rho_j^*||\nu^{(j)}||^4+(||\nu^{(j)}||+k+1)\right\}.
\end{aligned}
$$

Thus, if the problem is well-conditioned (i.e., if the condition number κ associated with the matrices $\mathcal{M}_n$ and $\mathcal{M}_n^*$ is not too large), then the computed solution $S(z)$ is close to the exact solution $S_E(z)$ and $S^*(z)$ is close to the exact solution $S_E^*(z)$. The algorithm is therefore weakly stable [7]. ∎

12. Experimental Results. Numerical experiments have been performed to compare the analysis of the algorithm with its practice. A summary of the conclusions is presented here; details appear in [9].

The algorithm VECTOR_PADE was implemented using Sun Fortran 1.3.1. All calculations were performed in double precision. The linear systems (16), (20), (38) and (40) arising at intermediate steps of the algorithm were solved using the LINPACK routines SGEFA and SGESL. The results were then compared to the exact answers, obtained via the Maple computer algebra system. Tables 1 and 2 give the results of one small but typical experiment for which $n = (18, 19, 19)$ and $A^t(z) = [a_0(z), a_1(z), a_2(z)]$ with $a_0(z) = 1$ and with coefficients of $a_1(z)$, $a_2(z)$ randomly and uniformly distributed between -1 and 1. The tables give results at all intermediate points along the diagonal through n. In these tables, the errors (represented in scientific notation with two digits of accuracy and the exponent enclosed in parenthesis) in the computed $S^{(j)}(z)$ and $S^{*(j)}$ and in the order conditions are given for two values of the stability parameter τ. The value $\tau = 10^4$ in Table 1 indicates a willingness to accept only those striped Sylvester matrices $\mathcal{M}_{m^{(j)}}$ and mosaic Sylvester matrices $\mathcal{M}_{m^{(j)}}^*$ with condition numbers less than 10^4, approximately (i.e.,

Table 1
Errors at intermediate steps: $\tau = 10^4$

j	$\kappa^{(j)}$	$\|\|\delta T^{(j)^t}(z)\|\|$	$\frac{\|\|\delta S^{(j)}(z)\|\|}{\|\|S_E^{(j)}(z)\|\|}$	$\|\|\delta T^{*(j)}(z)\|\|$	$\frac{\|\|\delta S^{*(j)}(z)\|\|}{\|\|S_E^{*(j)}(z)\|\|}$
1	1.2(2)	5.6(-17)	1.4(-16)	2.8(-17)	7.0(-17)
2	1.8(2)	2.2(-16)	6.5(-16)	2.5(-16)	6.5(-16)
3	1.6(2)	2.8(-16)	9.8(-16)	5.2(-16)	1.8(-15)
4	9.5(2)	1.8(-16)	8.3(-16)	8.6(-16)	2.0(-15)
5	6.6(2)	1.9(-16)	1.7(-15)	8.7(-16)	2.9(-15)
-	4.1(7)	1.7(-16)	1.2(-15)	1.0(-15)	2.1(-15)
6	1.1(3)	3.2(-16)	2.3(-15)	9.7(-16)	4.8(-15)
7	1.5(3)	1.5(-16)	1.2(-15)	9.0(-16)	6.6(-15)
8	9.1(3)	3.0(-16)	1.8(-15)	8.2(-16)	4.4(-15)
9	3.7(3)	3.8(-16)	4.9(-15)	6.8(-16)	1.3(-14)
10	2.9(3)	3.7(-16)	3.3(-15)	1.1(-15)	1.2(-14)
-	3.2(6)	3.1(-16)	5.6(-15)	1.0(-15)	4.6(-14)
11	2.0(3)	1.1(-15)	8.1(-15)	1.5(-15)	1.4(-14)
-	1.6(4)	1.2(-15)	7.4(-15)	1.9(-15)	1.8(-14)
12	2.9(3)	7.9(-16)	9.5(-15)	2.4(-15)	2.2(-14)
-	4.1(4)	8.3(-16)	1.0(-14)	2.6(-15)	2.3(-14)
-	6.3(4)	1.0(-15)	2.4(-14)	2.5(-15)	2.9(-14)
-	1.1(4)	7.6(-16)	1.7(-14)	2.8(-15)	3.3(-14)
-	1.1(5)	6.7(-16)	1.3(-13)	3.2(-15)	1.4(-13)

those for which $\kappa^{(j)} \leq 10^4$). Striped and mosaic Sylvester matrices not satisfying this criterion are assumed to lie in an unstable block and are skipped over. An unstable point is identified by the value "-" in the column labeled "j". In Table 2, the value $\tau = 10^9$ permits a much greater tolerance for ill-conditioning and results in an expected deterioration in the accuracy.

It was observed that the large constants and powers of $||m^{(j)}||$ and $||\nu^{(j)}||$ that occur in the bounds derived above are not manifested in the experiments. Also, $||\delta T^t(z)||$ and $||\delta T^*(z)||$ depends on $\kappa^{(j)}$ and not $\kappa^{(j)}\kappa^{(j+1)}$ and the overall error is proportional to the largest $\kappa^{(j)}$ encountered. As for the case k=1 reported in [13], operational bounds on the errors in the order conditions are

$$||\delta T^t(z)|| \leq C(k+1)\mu \left(\sum_{j=0}^{\sigma} \kappa^{(j)} \rho_j ||m^{(j)}|| \right) + O(\mu^2)$$

and

$$||\delta T^*(z)|| \leq C(k+1)^2\mu \left(\sum_{j=0}^{\sigma} \kappa^{(j)} \rho_j ||m^{(j)}||^2 \right) + O(\mu^2),$$

where C is a moderate constant. In addition, for the errors in the solutions, operational bounds are

$$||\delta S(z)|| \leq C\kappa(k+1)\mu \left(\sum_{j=0}^{\sigma} \kappa^{(j)} \rho_j ||m^{(j)}|| \right) + O(\mu^2)$$

Table 2
Errors at intermediate steps: $\tau = 10^9$

j	$\kappa^{(j)}$	$\|\|\delta T^{(j)^t}(z)\|\|$	$\frac{\|\|\delta S^{(j)}(z)\|\|}{\|\|S_E^{(j)}(z)\|\|}$	$\|\|\delta T^{*(j)}(z)\|\|$	$\frac{\|\|\delta S^{*(j)}(z)\|\|}{\|\|S_E^{*(j)}(z)\|\|}$
1	1.2(2)	5.6(-17)	1.4(-16)	2.8(-17)	7.0(-17)
2	1.8(2)	2.2(-16)	6.5(-16)	2.5(-16)	6.5(-16)
3	1.6(2)	2.8(-16)	9.8(-16)	5.2(-16)	1.8(-15)
4	9.5(2)	1.8(-16)	8.3(-16)	8.6(-16)	2.0(-15)
5	6.6(2)	1.9(-16)	1.7(-15)	8.7(-16)	2.9(-15)
6	4.1(7)	1.7(-16)	1.2(-15)	1.0(-15)	2.1(-15)
7	1.1(3)	3.1(-12)	3.4(-11)	9.2(-12)	1.2(-10)
8	1.5(3)	2.8(-12)	1.9(-11)	1.5(-11)	1.5(-10)
9	9.1(3)	4.5(-12)	3.7(-11)	2.8(-11)	5.6(-10)
10	3.7(3)	4.9(-12)	9.5(-11)	2.9(-11)	3.7(-10)
11	2.9(3)	4.3(-12)	7.3(-11)	4.0(-11)	3.5(-10)
12	3.2(6)	3.8(-12)	1.7(-10)	4.7(-11)	1.9(-9)
13	2.0(3)	1.3(-11)	1.3(-10)	2.9(-11)	2.2(-10)
14	1.6(4)	1.3(-11)	1.4(-10)	2.0(-11)	1.9(-10)
15	2.9(3)	8.5(-12)	1.1(-10)	3.2(-11)	3.5(-10)
16	4.1(4)	6.3(-12)	1.1(-10)	3.4(-11)	3.5(-10)
17	6.3(4)	6.5(-12)	1.5(-10)	3.6(-11)	6.5(-10)
18	1.1(4)	6.8(-12)	1.8(-10)	3.6(-11)	4.4(-10)
19	1.1(5)	9.0(-12)	2.2(-10)	3.3(-11)	8.1(-10)

and

$$||\delta S^*(z)|| \leq C\kappa(k+1)^3\mu\left(\sum_{j=0}^{\sigma}\kappa^{(j)}\rho_j||m^{(j)}||^2\right) + O(\mu^2).$$

13. Conclusions. In this paper we have presented a new fast, weakly stable algorithm for the computation of Padé-Hermite and simultaneous Padé systems. The algorithm requires $\mathcal{O}(||n||^2 + s^2||n||)$ operations to compute a Padé-Hermite system and a simultaneous Padé system of type $n = [n_0, \ldots, n_k]$, where $||n|| = n_0 + \cdots + n_k$ and s is the largest distance from one well-conditioned subproblem to the next. The algorithm can also be used for fast stable inversion of striped or mosaic Sylvester matrices (see ([21] for the case $k = 1$ and $a_0(z) = 1$). The algorithm relies on the ability to specify when a given subproblem is well conditioned. The stability estimates come as a result of "approximate" inversion formulae for striped and mosaic Sylvester matrices derived in this paper. In addition to a complete stability analysis, we have also provided some numerical experiments that verify that the algorithm performs as theoretic results imply.

There is a number of open research problems that result from this work. The algorithm that has been presented is fast rather than superfast as is possible in the case of exact arithmetic [10]. It is possible to modify the algorithm so that it takes steps in a quadratic fashion as done in [10]. However, while this approach will work in the generic case, it is possible to find examples where not all the required subproblems are stable. In these cases the algorithm might not be numerically stable. It would be of interest to find a superfast algorithm that works in all cases and in addition is numerically stable.

In cases where the largest step-size is small the algorithm has complexity $\mathcal{O}(\|n\|^2)$. However, there are cases where the algorithm may require a very large step-size and and then have a higher cost than Gaussian elimination. This will happen if there is a very large unstable block, or if the stability parameter τ is chosen to be too low. It would be of interest to find a fast, stable algorithm that has complexity $\mathcal{O}(\|n\|^2)$ in all cases.

Our algorithm proceeds along a diagonal path in the corresponding Padé tables of our approximants. It would be of interest to find fast, stable algorithms that proceed along alternate paths in the Padé tables. An example of this in the Padé case is found in [18] where the computation proceeds along straight-line paths. In the context of matrix solvers this is the difference between giving a Toeplitz solver instead of a Hankel solver as is done in [13].

The *M-Padé approximation problem* is a generalization of the Padé-Hermite approximation problem which requires that the residual in (1) vanishes at a given set of knots $z_0, z_1, \ldots, z_{N-1}$, counting multiplicities ([3, 4, 5, 25]). The case where all the z_i are equal to 0 is just the Padé-Hermite problem. In this case the coefficient matrix for the associated linear system is the matrix of divided differences. It would be of interest to determine stability parameters for such matrices, with a view to developing fast, stable algorithms for computing this approximation problem. Along these lines, some experiments for the case k=2 are reported in [8].

REFERENCES

[1] G. Baker and P. Graves-Morris, *Padé Approximants, Part II*, Addison-Wesley, Reading, MA, 1981.

[2] M. V. Barel and A. Bultheel, *The computation of non-perfect Padé-Hermite approximants*, Numerical Algorithms, 1 (1991), pp. 285–304.

[3] B. Beckermann, *Zur Interpolation mit polynomialen Linearkombinationen beliebiger Funktionen*, PhD thesis, Institut für Angewandte Mathematik, Universität Hannover, 1988.

[4] B. Beckermann, *The structure of the singular solution table of the m-Padé approximation problem*, Journal of Computational and Applied Mathematics, 32 (1990), pp. 3–15.

[5] ———, *A reliable method for computing M-Padé approximants on arbitrary staircases*, Journal of Computational and Applied Mathematics, 40 (1992), pp. 19–42.

[6] B. Beckermann and G. Labahn, *A uniform approach for the fast computation of matrix-type Padé approximants*, SIAM Journal on Matrix Analysis and Applications, (to appear).

[7] J. R. Bunch, *The weak and strong stability of algorithms in numerical linear algebra*, Linear Algebra and Its Applications, 88/89 (1987), pp. 49–66.

[8] S. Cabay, M. Gutknecht, and R. Meleshko, *Stable rational interpolation?*, Tech. Rep. IPS 93–12, (to appear in Proceedings of MTNS 93), IPS-Zürich, 1993.

[9] S. Cabay, A. Jones, and G. Labahn, *Experiments with a stable algorithm for computing Padé-Hermite and simultaneous Padé approximants.* in prepartion, 1994.

[10] S. Cabay and G. Labahn, *A superfast algorithm for multi-dimensional Padé systems*, Numerical Algorithms, 2 (1992), pp. 201–224.

[11] ———, *Fast,stable inversion of mosaic Hankel matrices*, Proceedings of MTNS 93, (to appear).

[12] S. Cabay, G. Labahn, and B. Beckermann, *On the theory and computation of non-perfect Padé-Hermite approximants*, Journal of Computational and Applied Mathematics, 39 (1992), pp. 295–313.

[13] S. Cabay and R. Meleshko, *A weakly stable algorithm for Padé approximants and inversion of Hankel matrices*, SIAM Journal on Matrix Analysis and Applications, 14 (1993), pp. 735–765.

[14] T. Chan and P. Hansen, *A stable Levinson algorithm for general Toeplitz systems*, SIAM Journal on Matrix Analysis and Applications, 13 (1992), pp. 490–506.

[15] G. E. Forsythe and C. B. Moler, *Computer Solution of Linear Algebraic Systems*, Prentice-Hall, 1967.

[16] R. Freund and H. Zha, *A look-ahead algorithm for the solution of general Hankel systems*,

Numerische Mathematik, 64 (1993), pp. 295–322.
[17] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, Johns Hopkins University Press, 1983.
[18] M. GUTKNECHT, *Stable row recurrences in the Padé table and generically superfast lookahead solvers for non-hermitian toeplitz solvers*, Linear Algebra and Its Applications, 188/189 (1993), pp. 351–421.
[19] M. GUTKNECHT AND W. GRAGG, *Stable look-ahead versions of the Euclidean abd Chebyshev algorithms*. preprint, 1994.
[20] M. H. GUTKNECHT AND M. HOCHBRUCK, *Look-ahead Levinson and Schur algorithms for non-Hermitian Toeplitz systems*, Tech. Rep. IPS 93–11, IPS-Zürich, 1993.
[21] ———, *The stability of inversion formulas for Toeplitz matrices*, Tech. Rep. IPS 93–13, IPS-Zürich, 1993.
[22] A. JONES, *The numerical computation of Padé-Hermite systems*, Master's thesis, Dept. Comp. Sci., Univ. Alberta, 1992.
[23] G. LABAHN, *Inversion components for block Hankel-like matrices*, Linear Algebra and Its Applications, 177 (1992), pp. 7–48.
[24] G. LABAHN, D. K. CHOI, AND S. CABAY, *The inverses of block Hankel and block Toeplitz matrices*, SIAM J. Comput., 19 (1990), pp. 98–123.
[25] K. MAHLER, *Perfect systems*, Compositio Math., 19 (1968), pp. 95–166.
[26] R. SHAFER, *On quadratic approximation*, SIAM J. Numerical Analysis, 11 (1974), pp. 447–460.
[27] G. W. STEWART, *Introduction to Matrix Computations*, Academic Press, 1973.
[28] J. WILKINSON, *Rounding Errors in Algebraic Processes*, Prentice-Hall, 1963.