Experiments with a Weakly Stable Algorithm for Computing Padé-Hermite and Simultaneous Padé Approximants

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

January 8, 1997

Abstract

In a recent paper [7], the authors develop a fast, iterative, look-ahead algorithm for numerically computing Padé-Hermite systems and simultaneous Padé systems along a diagonal of the associated Padé tables. Included in [7] is a detailed error analysis showing that the algorithm is weakly stable. In this paper, we describe a Fortran implementation, VECTOR_PADE, of this algorithm together with a number of numerical experiments. These experiments show that the theoretical error bounds obtained in [7] reflect the general behavior of the actual error, but that in practice these bounds are large over-estimates.

Categories and Subject Descriptors: G.1 [Numerical Analysis]: G.1.2 Approximation Theory - Rational approximation; G.1.3 Numerical Linear Algebra - Error analysis, linear systems, matrix inversion

General Terms: Algorithms, experimentation

Additional Key Words and Phrases: Padé-Hermite approximants, simultaneous Padé approximants, Sylvester matrix, Toeplitz matrix, Hankel matrix, numerical stability

1 Introduction

Let $A^t(z) = [a_0(z), \ldots, a_k(z)]$, $k \geq 1$, with

$$a_\beta(z) = \sum_{\ell=0}^{\infty} a_\beta^{(\ell)} z^\ell,$$

*Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada, T6G 2H1. The research of this author was 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.

be a vector of formal power series over the real numbers with $a_0^{(0)} \neq 0$ and let $n = [n_0, \ \ldots \ , n_k]$ be a vector[1] of integers with $n_\beta \geq 0, 0 \leq \beta \leq k$. 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 at most $n_\beta, 0 \leq \beta \leq k$, such that

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

with $\|n\| = n_0 + \ldots + n_k$. 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 $\|n\| - n_\beta, 0 \leq \beta \leq k$, such that

$$q_0^*(z) \cdot a_\beta(z) + q_\beta^*(z) \cdot a_0(z) = O(z^{\|n\|+1}), \tag{2}$$

for $\beta = 1, \cdots, k$. For the special case $A^t(z) = [-1, a(z)]$, the Padé-Hermite and the simultaneous Padé approximation problems each become the classical Padé approximation problem for $a(z)$. Padé-Hermite approximation also includes other classical approximation problems such as algebraic approximants with $A^t(z) = [1, a(z), a^2(z), \ldots, a^k(z)]$ (e.g. [13] for the special case $k = 2$) and G^3J approximants with $A^t(z) = [1, a(z), a'(z)]$. Simultaneous Padé approximants were first used by Hermite in 1873 in his famous proof of the transcendence of e. Additional examples, along with historical motivations and applications of these approximants, can be found in numerous references (cf., [1, 7, 10, 14]).

By equating coefficients in (1) and (2), the problem of computing a Padé-Hermite or simultaneous Padé approximant of type n becomes that of solving some systems of linear equations of order $\mathcal{O}(\|n\|)$. As such, 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 systems has a structured form so it is not surprising that there are a number of fast $\mathcal{O}(\|n\|^2)$ methods (cf., [14, 7, 10]) and superfast $\mathcal{O}(\|n\| \log^2 \|n\|)$ (cf., [3, 9]) methods for determining these approximants. However, unlike the Gaussian elimination method, which is weakly stable (in the class of problems involving these structured matrices) [5], the fast and superfast methods, with one exception, can encounter problems with numerical instabilities. The exception, for general k, is the algorithm VECTOR_PADE described in [7], which is proven there to be weakly stable.

The primary focus of this paper is the numerical experimentation with a Fortran implementation of the VECTOR_PADE algorithm. This algorithm is a look-ahead procedure that iteratively computes the approximants at all the "well-conditioned points" along a piece-wise diagonal path passing through the point n. In the case of the classical Padé approximation problem ($k = 1$), the algorithm reduces to the Cabay-Meleshko algorithm [11].

The paper is organized as follows: §2 and §3 give a summary of those results obtained in [7] that are relevant to this discussion, including the basic iterative step in the VECTOR_PADE algorithm. The results of the numerical experiments, which confirm the weak stability of the algorithm, are discussed in §4. This section also includes a discussion of some practical error bounds that are seen to be much better than those obtained theoretically in [7]. Some concluding remarks are made in §5. The paper has two appendices. Appendix A gives a detailed example of how VECTOR_PADE advances the computation of approximants

[1]Zero subscripting is used here and in the remainder of this presentation.

along the diagonal path from one well-conditioned point to the next. Appendix B gives an upper bound for a stability parameter that is central to the operation of VECTOR_PADE and to the error bounds.

2 Multi-dimensional Padé Systems

In this section, we review some basic results from [7], which gives the VECTOR_PADE algorithm and a detailed error analysis. The analysis provides some insight into the behavior of the algorithm and is used in particular to show that it is weakly stable.

The VECTOR_PADE algorithm actually computes numerical Padé-Hermite and simultaneous Padé *systems* (denoted by NPHS and NSPS, respectively) rather than just the respective approximants. For $k+1$ power series, these systems are square matrix polynomials of order $k+1$ which can be partitioned into their respective Padé approximants along with weakened versions of these approximants (cf., [10]). The NPHS and NSPS are defined in a way similar to their corresponding approximants in that they are required to satisfy an order condition, a set of degree bounds and a non-singularity condition.

An NPHS of type n is represented by

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

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

$$p(z) = \sum_{\ell=0}^{n_0-1} p^{(\ell)} z^\ell, \qquad u_\beta(z) = \sum_{\ell=0}^{n_0} u_\beta^{(\ell)} z^\ell,$$
$$q_\alpha(z) = \sum_{\ell=0}^{n_\alpha-1} q_\alpha^{(\ell)} z^\ell, \qquad v_{\alpha,\beta}(z) = \sum_{\ell=0}^{n_\alpha} v_{\alpha,\beta}^{(\ell)} z^\ell.$$

The order condition for a NPHS $S(z)$ of type n, in the case of floating point arithmetic, is given by

$$A^t(z) \cdot S(z) = z^{\|n\|+1} T^t(z) + \delta T^t(z) \tag{3}$$

with $T^t(z)$ the residual vector of power series and $\delta T^t(z)$ a vector of small error terms ($\delta T^t(z) = 0$ using exact arithmetic). The non-singularity condition requires that, for $1 \le \alpha, \beta \le k$,

$$v_{\alpha,\beta}^{(0)} = \begin{cases} \gamma_\beta \neq 0, & \alpha = \beta \\ 0, & \alpha \neq \beta \end{cases}$$

and, in addition, that $[T^t(0)]_0 = \gamma_0 \neq 0$.

Only the first column of $S(z)$ gives 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 VECTOR_PADE described briefly in §3. But there are other uses for these columns of $S(z)$, such as that of expressing the inverse of a striped Sylvester matrix [6, 8]. Note that a Padé-Hermite approximant of type $[n_0 - 1, \ldots, n_k - 1]$ satisfying (1) requires terms of $A^t(z)$ up to and including $z^{\|n\|-2}$ only, whereas the NSPS of type n satisfying (3) requires terms up to and including $z^{\|n\|}$. So, to compute a Padé-Hermite approximant of type $[n_0 - 1, \ldots, n_k - 1]$ using VECTOR_PADE, the user may first have to manufacture artificial and arbitrary values for the coefficients of $z^{\|n\|-1}$ and $z^{\|n\|}$ in $A^t(z)$.

An NSPS of type n is represented by[2]

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

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

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

The order condition for a NSPS $S^*(z)$ of type n has a similar form except that multiplication is on the right rather than the left; namely,

$$S^*(z) \cdot A^*(z) = z^{\|n\|+1} T^*(z) + \delta T^*(z) \tag{4}$$

with $A^*(z)$ given by

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

The non-singularity condition requires that, for $1 \le \alpha, \beta \le k$,

$$[T^*(0)]_{\alpha,\beta} = \begin{cases} \gamma_\beta^* \ne 0, & \alpha = \beta \\ 0, & \alpha \ne \beta \end{cases}$$

and, in addition, that $v^{*(0)} = \gamma_0^* \ne 0$.

Only the first row of $S^*(z)$ is a simultaneous Padé approximant as defined in §1, this being of type n. The remaining rows $S^*(z)$ do not quite satisfy the order condition (2) and

[2]For notational convenience, in [8], the z^2 terms in the last k rows of $S^*(z)$ are replaced by z and, similarly, the z^2 terms in the first column of $S(z)$ are replaced by z.

are therefore not simultaneous Padé approximants; these rows serve primarily to facilitate the computation of the first row using VECTOR_PADE.

One of the main reasons for working with Padé systems rather than approximants is their relation to the linear systems of equations associated with (1) and (2). A Padé-Hermite system of type n with the non-singularity conditions above exists if and only if the coefficient matrix, a generalized striped Sylvester matrix

$$\mathcal{M}_n = \left[\begin{array}{ccc|c|ccc} a_0^{(0)} & & & & a_k^{(0)} & & \\ & \ddots & & & & \ddots & \\ \vdots & & a_0^{(0)} & \cdots & \vdots & & a_k^{(0)} \\ & & \vdots & & & & \vdots \\ a_0^{(\|n\|-1)} & \cdots & a_0^{(\|n\|-n_0)} & & a_k^{(\|n\|-1)} & \cdots & a_k^{(\|n\|-n_k)} \end{array}\right], \tag{5}$$

of the linear system generated by equation (1) is nonsingular. Similarly, a simultaneous Padé system of type n with the non-singularity conditions above exists if and only if the coefficient matrix of the linear system generated by equation (2) is nonsingular. The coefficient matrix is now a certain mosaic Sylvester matrix

$$\mathcal{M}_n^* = \begin{bmatrix} \mathcal{S}_{0,1}^* & \cdots & \mathcal{S}_{0,k}^* \\ \vdots & & \vdots \\ \mathcal{S}_{k,1}^* & \cdots & \mathcal{S}_{k,k}^* \end{bmatrix}, \tag{6}$$

where, for $1 \le \beta \le k$,

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

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

$$\mathcal{S}_{\alpha,\alpha}^* = \begin{bmatrix} a_0^{(0)} & \cdots & a_0^{(\|n\|-1)} \\ & \ddots & \vdots \\ & a_0^{(0)} & \cdots \quad a_0^{(n_\alpha)} \end{bmatrix}$$

and the remaining $\mathcal{S}_{\alpha,\beta}^* = 0$. In addition, the inverses of these matrices are completely determined by the components of the two Padé systems [8].

In the case of exact arithmetic, the above observations lead to an efficient, iterative algorithm [10] for computing all the nonsingular Padé systems along a particular path of the corresponding Padé tables. The success of the algorithm depends on the ability to recognize nonsingular systems and is provided by the non-singularity condition above. In the case of floating-point arithmetic, it is necessary instead to be able to recognize situations where the associated Sylvester matrices are unstable. The central result of [8] gives the quantity (called the stability parameter)

$$\kappa = \sum_{\beta=0}^{k} \frac{1}{|\gamma_\beta \cdot \gamma_\beta^*|} \tag{7}$$

as the estimate of the condition number of these matrices.

3 The Algorithm

Let τ be a fixed number, the stability tolerance set by a user. For the point n, let

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

and define integer vectors $n^{(i)} = (n_0^{(i)}, \ldots, n_k^{(i)})$ for $0 \le i \le N$ by $n^{(0)} = e_0 = [1, 0, \ldots, 0]$ and, for $i > 0$,

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

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 $n^{(N)} = n$. A subsequence $\{m^{(\sigma)}\}_{i=0,\ldots}$ of $\{n^{(i)}\}$ is called a **sequence of stable points** for $A(z)$ and $A^*(z)$ if at each point $m^{(\sigma)} = n^{(i_\sigma)}$, with $i_0 = 0$ and $i_{\sigma+1} > i_\sigma$, the stability criterion $\kappa^{(\sigma)} < \tau$ is satisfied ($\kappa^{(\sigma)}$ is the stability parameter (7) computed from the NPHS and NSPS of type $m^{(\sigma)}$).

A single iteration of the algorithm is given as follows. Assume that $S(z)$ and $S^*(z)$ are the NPHS and NSPS, respectively, of type $m^{(\sigma)}$ (when $\sigma = 0$, $S(z) = S^*(z) = I$, the identity matrix) and that $m^{(\sigma)}$ is i_σ units from the start (i.e., $m^{(\sigma)} = n^{(i_\sigma)}$). Let $T(z)$ and $T^*(z)$ be the residuals of $S(z)$ and $S^*(z)$, respectively, and initialize $j = 1$.

- Step 1: Let $\nu^{(j)} = n^{(i_\sigma+j)} - m^{(\sigma)} - e_0$ and compute these residuals $T(z)$ and $T^*(z)$ up to order $\|\nu^{(j)}\|$. Use the Gaussian elimination method to triangulate the Sylvester matrices $\widehat{\mathcal{M}}_\nu$ and $\widehat{\mathcal{M}}_\nu^*$ associated with $T(z)$ and $T^*(z)$, respectively, to determine the NPHS $\widehat{S}(z)$ and SPS $\widehat{S}^*(z)$ of type $\nu^{(j)}$ for the residuals.

- Step 2: Compute the products $S(z) \cdot \widehat{S}(z)$ and $\widehat{S}^*(z) \cdot S^*(z)$. These are, respectively, the NPHS of type $n^{(i_\sigma+j)}$ for $A(z)$ and the NSPS of type $n^{(i_\sigma+j)}$ for $A^*(z)$. Scale the products so that each column of the NPHS and each row of the NSPS has norm 1 (the norms used for this scaling are given in the next section).

- Step 3: Compute the stability parameter from the scaled NPHS and NSPS of type $n^{(i_\sigma+j)}$. If the stability parameter $\kappa^{(\sigma+1)}$ computed from the scaled NPHS and NSPS of type $n^{(i_\sigma+j)}$ is less than the stability tolerance τ, then the iterative step is complete and we set $m^{(\sigma+1)} = n^{(i_\sigma+j)}$. Otherwise increment j by 1 and go to Step 1.

The algorithm fails when the associated Sylvester matrices at the last point, n, are numerically singular.[3] Nevertheless, in this case, VECTOR_PADE still computes a Padé-Hermite approximant (the first column of $S(z)$) of type $(n_0 - 1, \ldots, n_k - 1)$ satisfying (1) except for a "small" residual error, and a simultaneous Padé approximant (the first row of $S^*(z)$) of type n satisfying (2) except for a "small" residual error. When the associated Sylvester matrices at the last point n are not numerically singular, the algorithm succeeds in computing a NPHS and a NSPS even in the case when the stability criterion at n is not satisfied. In this latter case, there may be a large error in the solutions, but the residual errors are known to remain "small".

[3]By numerical singularity, we mean that a zero pivot element is encountered during the triangular decomposition process of Gaussian elimination with partial pivoting.

4 Theoretical versus Experimental Results

Numerous numerical experiments have been performed to compare the analysis of the algorithm [7] with its practice. These experiments were performed on a "SPARCCompiler Fortran 4.0 for Solaris 2.x and 1.x" implementation of the algorithm VECTOR_PADE. All calculations were performed in double precision with unit error $\mu = 2^{-56}$. The linear systems arising at intermediate steps of the algorithm were solved using the LINPACK routines SGEFA and SGESL. The results were then compared with more accurate answers, obtained via the Maple computer algebra system using 50 decimal digits of precision.

Tables 1, 2 and 3 summarize the results of experiments with one particular vector of power series. These results are typical of numerous other experiments which were performed but which are not reported here. The coefficients of the power series used in the experiments were randomly and uniformly generated with values between -1 and 1 and then modified so as to introduce some pronounced instabilities. To introduce an instability at $m^{(\sigma+1)}$, the coefficients of $a_\beta(z)$, $1 \leq \beta \leq k$, were modified to make almost dependent the columns of the coefficient matrix $\widehat{\mathcal{M}}_{\nu^{(\sigma)}}$ corresponding to the residual $T(z)$ at the point $m^{(j)}$. The power series were then scaled.

The tables give results at all intermediate points along the diagonal passing through a specified point n. These results include the errors in the solutions $\delta S^{(\sigma)}(z) = S^{(\sigma)}(z) - S_E^{(\sigma)}(z)$ and $\delta S^{*(\sigma)}(z) = S^{*(\sigma)}(z) - S_E^{*(\sigma)}(z)$ (where the subscript E denotes the "exact" solution, obtained using MAPLE) and the corresponding residuals errors $\delta T^{(\sigma)^t}(z)$ and $\delta T^{*(\sigma)}(z)$, respectively, at the point $m^{(\sigma)}$. In Tables 1 and 2, for purposes of comparison, also given are the errors $\delta S_G^{(\sigma)}(z) = S_G^{(\sigma)}(z) - S_E^{(\sigma)}(z)$ and $\delta S_G^{*(\sigma)}(z) = S_G^{*(\sigma)}(z) - S_E^{*(\sigma)}(z)$ in the solutions obtained directly by the Gaussian elimination method using the LINPACK routines SGEFA and SGESL. Since the input power series as well as $S_E^{(\sigma)}(z)$ and $S_E^{*(\sigma)}(z)$ are scaled, these also give the relative errors. The floating-point entries in the tables are represented in scientific notation with two digits of accuracy and with the exponent enclosed in parenthesis.

The caption in each table specifies the stability threshold τ used for the experiment. This threshold indicates a willingness to accept only those striped Sylvester matrices $\mathcal{M}_{m^{(\sigma)}}$ and mosaic Sylvester matrices $\mathcal{M}^*_{m^{(\sigma)}}$ with condition numbers less than τ (i.e., those for which $\kappa^{(\sigma)} \leq \tau$). 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 labelled "σ".

Tables 1 and 2 give the results of identical experiments but with different stability thresholds τ. The value $\tau = 10^9$ in Table 2, as opposed to $\tau = 10^5$ in Table 1, permits a much greater tolerance for ill-conditioning and results in an expected deterioration in the accuracy. A comparison of the results of these two tables illustrates the efficacy of the look-a-head strategy of the algorithm in stepping over instabilities.

Table 3 gives the results of the same experiment as in Table 2 with the exception that $a_0^{(0)}$ has been changed so that $\|a_0(z)(mod\ z^{\|n\|+1}) = 2.7$ in Table 2 becomes 8.2×10^{14} in Table 3. These two tables illustrate the deterioration of the accuracy of results obtained by VECTOR_PADE as the size of the inverse of $a_0(z)$ increases.

In order to compare the results presented in these tables with theoretical ones, we now briefly summarize the errors bounds derived in [7]. Denote the last stable point prior

σ	$\kappa^{(\sigma)}$	$\kappa\left(\mathcal{M}_{m(\sigma)}\right)$	$\kappa(\mathcal{M}^*_{m(\sigma)})$	$\|\delta T^{(\sigma)^t}(z)\|$	$\|\delta S^{(\sigma)}(z)\|$	$\|\delta S_G^{(\sigma)}(z)\|$	$\|\delta T^{*(\sigma)}(z)\|$	$\|\delta S^{*(\sigma)}(z)\|$	$\|\delta S_G^{*(\sigma)}(z)\|$
1	3.2	-	-	0.0	9.8(-17)	9.8(-17)	6.9(-18)	7.6(-17)	7.6(-17)
2	3.9(3)	2.8(2)	3.8(2)	1.5(-17)	7.1(-17)	8.7(-17)	1.7(-17)	4.7(-16)	2.7(-16)
3	3.7(3)	4.6(2)	5.9(2)	3.6(-17)	6.6(-16)	3.5(-16)	2.5(-17)	2.9(-15)	4.8(-16)
4	7.7(3)	6.3(2)	8.2(2)	1.0(-16)	5.7(-15)	2.6(-16)	3.6(-17)	2.7(-15)	5.8(-16)
-	6.4(14)	1.3(8)	1.3(8)	1.1(-16)	1.0(-14)	5.2(-16)	4.5(-17)	3.6(-10)	6.8(-11)
5	1.1(4)	1.1(3)	1.2(3)	9.3(-17)	1.5(-14)	1.7(-15)	5.7(-17)	8.4(-15)	1.5(-15)
-	3.8(5)	1.0(4)	8.2(3)	9.2(-17)	1.3(-14)	6.1(-16)	4.1(-16)	2.0(-14)	2.8(-15)
6	1.1(4)	1.1(3)	1.1(3)	1.1(-16)	8.5(-15)	6.9(-16)	4.2(-16)	2.2(-14)	1.4(-15)
-	1.3(14)	9.6(7)	1.3(8)	1.1(-16)	2.1(-14)	6.2(-16)	2.1(-16)	8.2(-10)	3.2(-11)
7	3.9(4)	1.6(3)	2.0(3)	1.2(-16)	7.7(-15)	1.2(-15)	4.2(-16)	3.5(-14)	2.8(-15)
-	3.8(8)	3.6(7)	3.4(7)	9.4(-17)	3.2(-11)	2.1(-11)	4.3(-16)	5.1(-10)	5.0(-11)
-	1.9(9)	1.2(8)	1.3(8)	8.9(-17)	1.7(-10)	1.9(-10)	4.1(-16)	7.1(-10)	3.9(-10)
-	1.1(15)	9.8(8)	1.5(9)	9.0(-17)	2.7(-10)	1.4(-10)	4.0(-16)	2.8(-9)	8.3(-10)
-	1.3(9)	9.1(7)	1.2(8)	9.2(-17)	1.9(-10)	6.2(-11)	4.5(-16)	1.0(-9)	5.9(-11)
-	2.1(5)	9.7(3)	8.0(3)	3.3(-16)	6.2(-14)	1.5(-15)	4.4(-16)	3.5(-14)	5.2(-15)
8	3.0(4)	2.4(3)	2.7(3)	3.2(-16)	6.9(-14)	2.8(-15)	4.2(-16)	4.9(-14)	3.7(-15)
-	1.4(13)	7.9(7)	8.4(7)	3.2(-16)	2.3(-13)	5.8(-15)	5.1(-16)	6.9(-10)	1.1(-10)
9	6.4(4)	6.5(3)	7.6(3)	5.4(-16)	7.6(-13)	9.5(-15)	6.0(-16)	2.1(-13)	1.6(-14)
-	2.3(5)	1.4(4)	1.2(4)	5.5(-16)	5.3(-13)	9.7(-15)	2.3(-15)	4.6(-13)	7.1(-15)
-	1.9(9)	4.4(7)	9.2(7)	6.0(-16)	6.5(-10)	1.5(-11)	4.3(-15)	1.4(-8)	1.2(-10)
-	1.8(14)	1.2(8)	2.8(8)	6.2(-16)	2.9(-9)	1.2(-10)	2.9(-15)	7.8(-9)	9.8(-11)
10	6.1(4)	4.9(3)	4.4(3)	5.5(-16)	2.1(-13)	3.8(-15)	2.3(-15)	8.1(-13)	2.3(-14)
-	1.6(13)	3.2(7)	4.2(7)	2.2(-16)	1.9(-13)	3.3(-15)	1.4(-15)	6.7(-10)	1.9(-11)
11	8.9(4)	2.7(3)	3.3(3)	7.1(-16)	9.7(-14)	3.5(-15)	2.5(-15)	4.0(-13)	7.5(-15)
-	3.9(12)	2.6(7)	2.5(7)	2.7(-16)	1.3(-13)	4.2(-15)	3.8(-15)	1.7(-9)	8.1(-11)
12	7.7(4)	3.7(3)	3.9(3)	1.2(-15)	2.8(-13)	3.7(-15)	3.3(-15)	9.6(-13)	8.1(-15)
13	4.3(4)	2.1(3)	2.8(3)	2.0(-15)	1.0(-13)	1.6(-15)	4.2(-15)	1.4(-12)	1.0(-14)
14	9.3(4)	2.4(3)	5.2(3)	2.1(-15)	1.8(-13)	3.9(-15)	5.3(-15)	6.6(-12)	7.6(-15)
15	9.0(4)	3.6(3)	6.0(3)	2.5(-15)	1.5(-13)	4.4(-15)	8.1(-15)	1.8(-11)	1.6(-14)
16	8.7(4)	3.6(3)	7.9(3)	3.0(-15)	2.6(-13)	1.7(-15)	7.6(-15)	7.4(-12)	2.4(-14)
17	1.1(4)	8.0(2)	2.0(3)	4.7(-15)	4.5(-13)	1.5(-15)	6.5(-15)	1.5(-12)	2.8(-15)
18	2.9(4)	1.4(3)	3.2(3)	5.6(-15)	5.6(-13)	2.0(-15)	1.1(-14)	2.1(-12)	5.8(-15)
19	2.4(4)	2.4(3)	4.5(3)	5.5(-15)	4.0(-13)	4.0(-15)	1.8(-14)	7.6(-12)	1.1(-14)
20	6.0(4)	1.0(4)	1.2(4)	6.1(-15)	3.5(-13)	3.0(-15)	1.8(-14)	6.8(-12)	6.1(-15)
21	1.2(4)	3.7(3)	3.2(3)	5.8(-15)	3.0(-13)	2.3(-15)	1.7(-14)	4.1(-12)	5.8(-15)
22	6.5(3)	1.8(3)	1.3(3)	6.7(-15)	6.7(-13)	2.0(-15)	1.6(-14)	9.5(-13)	2.5(-15)
-	6.9(5)	3.8(4)	1.7(4)	4.9(-15)	6.4(-12)	3.4(-15)	2.5(-14)	1.1(-12)	2.5(-15)
23	3.2(4)	9.9(3)	3.4(3)	8.7(-15)	7.8(-12)	6.0(-15)	3.2(-14)	8.3(-12)	1.1(-14)

Table 1: $k = 2$; $n = (37, 37, 37)$; $\|a_0^{-1}(z)(mod\, z^{\|n\|+1})\| = 2.7(1)$; $\tau = 10^5$

σ	$\kappa^{(\sigma)}$	$\kappa\left(\mathcal{M}_{m(\sigma)}\right)$	$\kappa(\mathcal{M}^*_{m(\sigma)})$	$\|\delta T^{(\sigma)^t}(z)\|$	$\|\delta S^{(\sigma)}(z)\|$	$\|\delta S_G^{(\sigma)}(z)\|$	$\|\delta T^{*(\sigma)}(z)\|$	$\|\delta S^{*(\sigma)}(z)\|$	$\|\delta S_G^{*(\sigma)}(z)\|$
1	3.2(0)	-	-	0.0	9.8(-17)	9.8(-17)	6.9(-18)	7.6(-17)	7.6(-17)
2	3.9(3)	2.8(2)	3.8(2)	1.5(-17)	7.1(-17)	8.7(-17)	1.7(-17)	4.7(-16)	2.7(-16)
3	3.7(3)	4.6(2)	5.9(2)	3.6(-17)	6.6(-15)	3.5(-16)	2.5(-17)	2.9(-15)	4.8(-16)
4	7.7(3)	6.3(2)	8.2(2)	1.0(-16)	5.7(-15)	2.6(-16)	3.6(-17)	2.7(-15)	5.8(-16)
-	6.4(14)	1.3(8)	1.3(8)	1.1(-16)	1.0(-14)	5.2(-16)	4.5(-17)	3.6(-10)	6.8(-11)
5	1.1(4)	1.1(3)	1.2(3)	9.3(-17)	1.5(-14)	1.7(-15)	5.7(-17)	8.4(-15)	1.5(-15)
6	3.8(5)	1.0(4)	8.2(3)	9.2(-17)	1.3(-14)	6.1(-16)	4.1(-16)	2.0(-14)	2.8(-15)
7	1.1(4)	1.1(3)	1.1(3)	2.2(-16)	1.1(-14)	6.9(-16)	1.6(-15)	1.1(-13)	1.4(-15)
-	1.3(14)	9.6(7)	1.3(8)	1.1(-16)	1.1(-14)	6.2(-16)	6.7(-15)	7.7(-9)	3.2(-11)
8	3.9(4)	1.6(3)	2.0(3)	2.5(-16)	1.1(-14)	1.2(-15)	4.8(-15)	2.3(-13)	2.8(-15)
9	3.8(8)	3.6(7)	3.4(7)	1.7(-16)	1.6(-10)	2.1(-11)	6.0(-15)	4.1(-9)	5.0(-11)
-	1.9(9)	1.2(8)	1.3(8)	1.6(-16)	2.9(-10)	1.9(-10)	8.9(-15)	1.6(-8)	3.9(-10)
-	1.1(15)	9.8(8)	1.5(9)	1.1(-16)	1.0(-9)	1.4(-10)	8.2(-15)	4.1(-8)	8.3(-10)
-	1.3(9)	9.1(7)	1.2(8)	1.3(-16)	1.6(-10)	6.2(-11)	6.9(-15)	1.3(-8)	5.9(-11)
10	2.1(5)	9.7(3)	8.0(3)	1.3(-12)	1.9(-10)	1.5(-15)	2.2(-13)	2.1(-10)	5.2(-15)
11	3.0(4)	2.4(3)	2.7(3)	1.9(-11)	2.3(-9)	2.8(-15)	8.3(-13)	2.8(-10)	3.7(-15)
-	1.4(13)	7.9(7)	8.4(7)	7.2(-12)	1.1(-9)	5.8(-15)	1.6(-12)	1.4(-6)	1.1(-10)
12	6.4(4)	6.5(3)	7.6(3)	1.7(-11)	1.3(-9)	9.5(-15)	3.8(-12)	1.0(-9)	1.6(-14)
13	2.3(5)	1.4(4)	1.2(4)	3.4(-11)	1.1(-9)	9.7(-15)	2.1(-11)	3.7(-9)	7.1(-15)
-	1.9(9)	4.4(7)	9.2(7)	4.2(-11)	1.8(-6)	1.5(-11)	4.6(-11)	7.7(-6)	1.2(-10)
-	1.8(14)	1.2(8)	2.8(8)	3.8(-11)	7.9(-6)	1.2(-10)	3.9(-11)	3.7(-5)	9.8(-11)
14	6.1(4)	4.9(3)	4.4(3)	3.4(-11)	1.5(-9)	3.8(-15)	3.0(-11)	1.2(-8)	2.3(-14)
-	1.6(13)	3.2(7)	4.2(7)	5.2(-12)	9.7(-10)	3.3(-15)	2.0(-11)	1.1(-5)	1.9(-11)
15	8.9(4)	2.7(3)	3.3(3)	3.8(-11)	2.0(-9)	3.5(-15)	2.5(-11)	4.6(-9)	7.5(-15)
-	3.9(12)	2.6(7)	2.5(7)	1.0(-11)	1.4(-9)	4.2(-15)	3.8(-11)	2.9(-5)	8.1(-11)
16	7.7(4)	3.7(3)	3.9(3)	8.8(-11)	9.9(-9)	3.7(-15)	3.3(-11)	4.9(-9)	8.1(-15)
17	4.3(4)	2.1(3)	2.8(3)	1.4(-10)	5.6(-9)	1.6(-15)	6.8(-11)	5.6(-9)	1.0(-14)
18	9.3(4)	2.4(3)	5.2(3)	1.6(-10)	8.0(-9)	3.9(-15)	5.8(-11)	2.0(-8)	7.6(-15)
19	9.0(4)	3.6(3)	6.0(3)	1.8(-10)	8.4(-9)	4.4(-15)	8.3(-11)	7.0(-8)	1.6(-14)
20	8.7(4)	3.6(3)	7.9(3)	1.8(-10)	7.2(-9)	1.7(-15)	7.9(-11)	3.4(-8)	2.4(-14)
21	1.1(4)	8.0(2)	2.0(3)	2.1(-10)	7.2(-9)	1.5(-15)	6.3(-11)	7.7(-9)	2.8(-15)
22	2.9(4)	1.4(3)	3.2(3)	2.3(-10)	1.0(-8)	2.0(-15)	4.5(-11)	4.0(-9)	5.8(-15)
23	2.4(4)	2.4(3)	4.5(3)	2.5(-10)	9.0(-9)	4.0(-15)	1.1(-10)	3.2(-8)	1.1(-14)
24	6.0(4)	1.0(4)	1.2(4)	2.6(-10)	1.5(-8)	3.0(-15)	1.2(-10)	2.6(-8)	6.1(-15)
25	1.2(4)	3.7(3)	3.2(3)	2.4(-10)	1.0(-8)	2.3(-15)	1.1(-10)	1.6(-8)	5.8(-15)
26	6.5(3)	1.8(3)	1.3(3)	3.9(-10)	4.3(-8)	2.0(-15)	9.4(-11)	3.7(-9)	2.5(-15)
27	6.9(5)	3.8(4)	1.7(4)	2.9(-10)	2.1(-7)	3.4(-15)	4.2(-11)	1.1(-9)	2.5(-15)
28	3.2(4)	9.9(3)	3.4(3)	5.2(-10)	3.0(-7)	6.0(-15)	3.5(-11)	1.4(-9)	1.1(-14)

Table 2: $k = 2$; $n = (37, 37, 37)$; $\|a_0^{-1}(z)(mod\, z^{\|n\|+1})\| = 2.7(1)$; $\tau = 10^9$

σ	$\kappa^{(\sigma)}$	$\kappa\left(\mathcal{M}_{m(\sigma)}\right)$	$\kappa(\mathcal{M}^*_{m(\sigma)})$	$\|\delta T^{(\sigma)^t}(z)\|$	$\|\delta S^{(\sigma)}(z)\|$	$\|\delta T^{*(\sigma)}(z)\|$	$\|\delta S^{*(\sigma)}(z)\|$
1	1.7(3)	-	-	1.7(-18)	1.2(-17)	8.7(-19)	6.4(-17)
2	1.4(4)	2.6(1)	1.3(2)	2.4(-18)	2.7(-16)	2.2(-18)	4.8(-16)
3	5.7(4)	7.3(1)	1.5(3)	9.7(-18)	6.0(-16)	4.8(-18)	7.0(-15)
4	1.7(7)	1.7(3)	8.2(4)	1.6(-17)	1.2(-14)	4.1(-18)	1.5(-14)
5	6.6(5)	3.4(2)	2.4(4)	3.8(-16)	3.0(-14)	2.2(-16)	1.1(-13)
6	9.3(4)	1.7(2)	1.2(4)	2.3(-15)	4.7(-13)	2.5(-15)	2.3(-13)
7	1.2(5)	2.1(2)	2.8(4)	2.7(-15)	6.8(-13)	5.3(-15)	3.9(-12)
8	1.2(5)	2.1(2)	8.8(4)	3.4(-15)	7.6(-13)	5.1(-15)	5.0(-12)
9	4.8(5)	4.5(2)	2.8(5)	4.1(-15)	8.0(-13)	6.4(-15)	8.9(-12)
10	3.3(6)	1.8(3)	2.5(6)	3.7(-15)	1.9(-12)	3.9(-15)	4.2(-11)
11	3.9(6)	2.2(3)	9.1(6)	3.1(-15)	2.9(-12)	4.1(-15)	6.9(-11)
12	8.4(6)	2.5(3)	5.1(7)	3.3(-15)	4.6(-12)	4.1(-15)	7.2(-11)
13	2.4(6)	2.6(3)	6.3(7)	3.5(-15)	1.7(-12)	4.1(-15)	8.9(-10)
14	9.8(5)	1.6(3)	6.3(7)	2.6(-15)	2.6(-12)	4.2(-15)	3.2(-10)
15	3.2(6)	2.7(3)	3.9(8)	6.6(-15)	2.4(-12)	4.1(-15)	2.3(-10)
16	2.1(6)	4.3(3)	9.3(8)	4.7(-15)	6.8(-12)	3.3(-15)	1.6(-9)
17	4.4(7)	1.6(4)	2.7(9)	4.6(-15)	3.9(-12)	4.4(-15)	3.3(-10)
18	3.6(6)	5.3(3)	8.4(8)	4.5(-15)	1.7(-12)	4.7(-15)	4.8(-10)
19	1.7(6)	4.3(3)	4.7(9)	5.8(-15)	3.6(-12)	7.9(-15)	7.6(-9)
20	1.4(7)	8.5(3)	1.3(10)	4.1(-15)	2.6(-12)	8.8(-15)	1.2(-8)
21	3.8(6)	8.3(3)	2.4(10)	5.8(-15)	2.5(-12)	9.8(-15)	7.2(-8)
22	7.5(5)	2.7(3)	5.5(10)	4.1(-15)	3.2(-12)	1.0(-14)	1.5(-7)
23	5.5(6)	4.3(3)	5.9(11)	2.9(-15)	3.6(-12)	1.1(-14)	1.0(-6)
24	1.3(7)	1.2(4)	1.1(12)	7.1(-15)	1.3(-11)	9.0(-15)	1.0(-6)
25	5.0(8)	8.1(4)	5.9(12)	9.3(-15)	7.7(-12)	9.7(-15)	3.4(-7)
26	2.1(6)	1.2(4)	2.3(12)	5.0(-15)	3.6(-12)	1.0(-14)	1.5(-6)
27	2.5(6)	1.0(4)	2.3(12)	4.9(-15)	3.2(-12)	9.3(-15)	6.5(-7)
-	1.3(9)	1.2(5)	2.0(13)	3.0(-15)	1.9(-12)	8.7(-15)	3.6(-7)
28	6.3(6)	9.2(3)	4.6(12)	8.3(-15)	1.6(-11)	8.6(-15)	2.3(-6)
29	1.0(6)	6.6(3)	4.5(12)	8.5(-15)	4.8(-12)	1.1(-14)	2.0(-6)
30	1.5(6)	9.1(3)	1.1(13)	6.4(-15)	5.5(-12)	1.2(-14)	8.5(-6)
31	8.4(5)	1.0(4)	2.1(13)	5.3(-15)	3.5(-12)	1.1(-14)	6.7(-6)
32	2.0(6)	1.4(4)	7.3(13)	2.4(-15)	1.6(-12)	2.0(-14)	1.2(-5)
33	4.5(6)	7.2(4)	2.2(14)	2.8(-15)	3.9(-12)	1.6(-14)	1.9(-5)
34	6.8(6)	6.6(4)	7.8(14)	3.4(-15)	4.5(-12)	2.1(-14)	2.6(-5)
35	1.0(6)	1.7(4)	5.7(14)	3.4(-15)	2.3(-12)	2.8(-14)	6.9(-5)
36	3.4(5)	1.2(4)	1.1(15)	5.7(-15)	2.3(-12)	4.2(-14)	3.3(-4)
37	4.1(5)	1.2(4)	9.6(15)	6.3(-15)	3.3(-12)	5.6(-14)	1.8(-3)

Table 3: $k = 2$; $n = (37, 37, 37)$; $\|a_0^{-1}(z)(mod\, z^{\|n\|+1})\| = 8.2(14)$; $\tau = 10^9$

to the point n along the diagonal passing through n by $m^{(\sigma_f)}$ (i.e., $\kappa^{(\sigma_f)} \leq \tau$), and let $\nu^{(\sigma_f)} = n - m^{(\sigma_f)} - e_0$. Then the bounds derived in [7] at the point n, for a numerically nonsingular point n, are

$$\|\delta T^t(z)\| \leq F_{\sigma_f} + 2(k+1) \cdot |a_0^{(0)}| \sum_{\sigma=0}^{\sigma_f - 1} \kappa^{(\sigma+1)} F_\sigma, \tag{8}$$

$$\|\delta T^*(z)\| \leq F^*_{\sigma_f} + 2(k+1) \cdot |a_0^{(0)}| \sum_{\sigma=0}^{\sigma_f - 1} \kappa^{(\sigma+1)} F^*_\sigma, \tag{9}$$

$$\begin{aligned} \|\delta S(z)\| &\leq 2\kappa \cdot |a_0^{(0)}| \cdot \|a_0^{-1}(z)(mod\ z^{\|n\|+1})\| \\ &\quad \cdot \left\{ F_{\sigma_f} + 2(k+1) \cdot |a_0^{(0)}| \sum_{\sigma=0}^{\sigma_f - 1} \kappa^{(\sigma+1)} F_\sigma \right\}, \end{aligned} \tag{10}$$

$$\begin{aligned} \|\delta S^*(z)\| &\leq 2\kappa(k+1)^2 \cdot |a_0^{(0)}| \cdot \|a_0^{-1}(z)(mod\ z^{\|n\|+1})\| \\ &\quad \cdot \left\{ F^*_{\sigma_f} + 2(k+1) \cdot |a_0^{(0)}| \sum_{\sigma=0}^{\sigma_f - 1} \kappa^{(\sigma+1)} F^*_\sigma \right\}, \end{aligned} \tag{11}$$

where

$$\begin{aligned} F_\sigma &= 4\kappa^{(\sigma)}(k+1) \cdot |a_0^{(0)}| \cdot \mu \\ &\quad \cdot \left\{ (||m^{(\sigma)}|| + k + 1) + 4\rho_\sigma ||\nu^{(\sigma)}||^3 + (||\nu^{(\sigma)}|| + k + 1) \right\}, \\ F^*_\sigma &= 8\kappa^{(\sigma)}(k+1)^2 \cdot |a_0^{(0)}| \cdot \mu \\ &\quad \cdot \left\{ (||m^{(\sigma)}|| + 1) + 4(k+1)^5 \rho^*_\sigma ||\nu^{(\sigma)}||^3 + (||\nu^{(\sigma)}|| + k + 1) \right\}. \end{aligned}$$

The constants $\rho_\sigma, \rho^*_\sigma$ are growth factors (in practice of size $\mathcal{O}(10)$) associated with the Gaussian elimination method when solving the subsystems at the point $m^{(\sigma)}$.

In the above, the norms used are as follows. For the polynomial $s(z) = \sum_{\ell=0}^{\partial} s^{(\ell)} z^\ell$, define $\|s(z)\| = \sum_{\ell=0}^{\partial} |s^{(\ell)}|$; and, for the power series $a(z) = \sum_{\ell=0}^{\infty} a^{(\ell)} z^\ell$, define $\|a(z)\| = \sum_{\ell=0}^{\infty} |a^{(\ell)}|$ (this norm exists in practice since the power series are truncated). For vectors and matrices over these power series and polynomial domains, the 1-norm is used. So, $\|A^t(z)\| = \max_{0 \leq \beta \leq k} \{\|a_\beta(z)\|\}$ and $\|S(z)\| = \max_{0 \leq \beta \leq k} \left\{ \sum_{\alpha=0}^{k} \|S_{\alpha,\beta}(z)\| \right\}$. It is assumed in the analysis that $A^t(z)$ is scaled; that is, $\|a_\beta(z)\| = 1,\ 0 \leq \beta \leq k$.

Observation 1: The experimental results imply that the large powers of $\|m^{(\sigma)}\|$ and $\|\nu^{(\sigma)}\|$ that occur in the bounds above are not manifested in the experiments. Also, $\|\delta T^t(z)\|$ and $\|\delta T^*(z)\|$ appear to depend on $\kappa^{(\sigma)}$ and not on $\kappa^{(\sigma)}\kappa^{(\sigma+1)}$. In addition, the overall error is proportional to the largest $\kappa^{(\sigma)}$ encountered. Thus, the bounds are crude, but they do appear to reflect the behavior of the error. As Wilkinson points out [15, page 567], "The main object of such an analysis is to expose the potential instabilities, if any, of an algorithm so that hopefully from the insight thus obtained one might be led to improved algorithms. Usually the bound itself is weaker than it might have been because of the necessity of restricting the mass of detail to a reasonable level and because of the limitations imposed by expressing the errors in terms of matrix norms."

Operational bounds on the errors in the order conditions (as for the case k=1 reported in [11]) appear to be

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

and

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

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

$$\|\delta S(z)\| \leq C\kappa(k+1)\mu \cdot \| \left(\sum_{\sigma=0}^{\sigma_f} \kappa^{(\sigma)} \rho_\sigma \|m^{(\sigma)}\| \right)$$

and

$$\|\delta S^*(z)\| \leq C\kappa(k+1)^3\mu \cdot |a_0^{(0)}| \cdot \|a_0^{-1}(z)(mod\, z^{\|n\|+1})\| \left(\sum_{\sigma=0}^{\sigma_f} \kappa^{(\sigma)} \rho^*_\sigma \|m^{(\sigma)}\|^2 \right).$$

Note that, as predicted by the bound (11), the accuracy of $S^*(z)$ deteriorates dramatically with an increase in $\|a_0^{-1}(z)\,(mod\, z^{\|n\|+1})\|$ (see Tables 2 and 3) The accuracy of $S(z)$, on the other hand, has remained unaffected in these and many other experiments contrary to what is suggested by the bound (10).

Observation 2: For $\delta T^t(z)$ and $\delta T^*(z)$ sufficiently small, bounds for $\mathcal{M}_n^{-1}$ and $\mathcal{M}_n^{*-1}$ are derived in [8] to be

$$\|\mathcal{M}_n^{-1}\|_1,\ \|\mathcal{M}_n^{*-1}\|_\infty \leq 2\kappa\ \cdot |a_0^{(0)}| \cdot \|a_0^{-1}(z)\,(mod\, z^{\|n\|+1})\|.$$

This gives lower bounds for κ in terms of $\kappa_1(\mathcal{M}_n)$, or $\kappa_\infty(\mathcal{M}_n^*)$ (note that $\|\mathcal{M}_n\|_1 = 1$ and $\|\mathcal{M}_n^*\|_\infty = k$). On the other hand, Appendix B gives the upper bound

$$\kappa \leq \frac{6(k+1)}{|a_0^{(0)}|} \cdot \frac{\kappa_1(\mathcal{M}_n)\,\kappa_\infty(\mathcal{M}_n^*)}{[1-\epsilon\cdot\kappa_1(\mathcal{M}_n)]\,[1-\epsilon\cdot\kappa_\infty(\mathcal{M}_n^*)]},$$

where ϵ is an upper bound for $\|\delta T^t(z)\|_1$ and $\|\delta T^*(z)\|_\infty$. In the experiments, the term $|a_0^{(0)}| \cdot \|a_0^{-1}(z)\,(mod\, z^{\|n\|+1})\|$ does not seem to appear in the upper bound for $\|\mathcal{M}_n^{-1}\|_1$. In addition, κ is usually closer to the lower bounds than to the upper bound; however, at "unstable" points, κ appears to drift towards its upper bound.

Observation 3: The most suitable choice for the stability tolerance τ is difficult to determine, a priori. Too small a choice can cause the algorithm to take large steps (to step over large unstable blocks) and thereby increasing computational costs (up to as much as $O(\|n\|^4)$ operations if τ is set so small that all points are deemed unstable). Too large a choice can result in needless loss of accuracy when a smaller choice of τ would give better

accuracy with about the same computational costs. As an aid in selecting τ, we recommend a trial run through just a few steps of the algorithm in order to derive some insight into the behavior of subsequent $\kappa^{(i)}$'s. The algorithm VECTOR_PADE provides a list of such $\kappa^{(i)}$'s. Alternatively, a dynamic strategy for choosing the step size may be possible.

Observation 4: The residual errors appear to increase only upon exit from an unstable block (see, for example, Table 2: $\sigma = 10$) Indeed, the residual errors do not appear to increase as long $\kappa^{(\sigma)}$ continues to grow. This observation suggests a behavior that is not captured by the residual error bounds (8) and (9). Its truth would be useful in some applications, such as in accelerating convergence of vector sequence [12].

5 Conclusions

The experiments have both verified that the VECTOR_PADE algorithm is indeed weakly stable and illustrated the pessimism of the error bounds derived in [7]. They have also illustrated that the bounds do give an indication of what mostly contributes to the growth of errors in the system. We see that the error is primarily determined by the intermediate system that is most ill conditioned. By skipping over such systems using the look-ahead strategy, the error is kept small and accurate results are obtained. The numerical results also confirm that large errors in the NSPS may arise when the power series $a_0(z)$ has a large inverse.

We remark that the computation of a NPHS and NSPS of type n provide all the components necessary to invert the coefficient matrices of the associated linear systems [8]. When the initial power series $[a_0(z)]^{-1}$ does not have large terms, then the formulae given in [8] enable the computation of these inverses in a fast, numerically stable manner. This is the case, for example, when one is interested simply in inverting striped and mosaic Hankel matrices (since in this case this corresponds to the case where $a_0(z) = 1$). However, when $[a_0(z)]^{-1}$ is large, then it is inadvisable to use these formulae, because in this case VECTOR_PADE may not accurately compute $S^*(z)$ upon which the formulae depend.

It has been noted that the cost complexity of VECTOR_PADE becomes $\mathcal{O}(\|n\|^4)$ in the exceptional case that length of unstable block is $\mathcal{O}(\|n\|)$ (e.g., if the stability tolerance τ is set too low). By modifying the algorithm to use QR factorization with bordering (as in [2] for the case $k = 1$) rather than the Gaussian elimination method to solve the intermediate systems, we expect to be able to maintain stability while guaranteeing a worst case complexity of $\mathcal{O}(\|n\|^3)$. This needs confirmation. Recently, another algorithm [4] for the QR factorization of a Toeplitz matrix (which is the generalized Sylvester matrix with $a_0(z) = 1$ and $k = 1$) with a worst case complexity of $\mathcal{O}(\|n\|^2)$ has been shown to be weakly stable. Perhaps this algorithm, too, can be generalized to arbitrary k.

References

[1] G. Baker and P. Graves-Morris. *Padé Approximants*, volume 14 of *Encyclopedia of Mathematics*. Addison-Wesley, 1981.

[2] B. Beckermann. The stable computation of formal orthogonal polynomials. Preprint, 1995.

[3] B. Beckermann and G. Labahn. A uniform approach for the fast computation of matrix-type Padé approximants. *SIAM Journal on Matrix Analysis and Applications*, pages 804–823, 1994.

[4] A. W. Bojanczyk, R. P. Brent, and F. D. De Hoog. A weakly stable algorithm for general Toeplitz systems. *Numerical Algorithms*, to appear.

[5] James R. Bunch. The weak and strong stability of algorithms in numerical linear algebra. *Linear Algebra and Its Applications*, 88/89:49–66, 1987.

[6] S. Cabay, A. Jones, and G. Labahn. A stable algorithm for multi-dimensional Padé systems and the inversion of generalized Sylvester matrices. Technical Report TR 94-07, Dept. Comp. Sci., Univ. Alberta, 1994.

[7] S. Cabay, A. Jones, and G. Labahn. Computation of numerical Padé-Hermite and simultaneous Padé systems II: A weakly-stable algorithm. *SIAM Journal on Matrix Analysis and Applications*, to appear.

[8] S. Cabay, A. Jones, and G. Labahn. Computation of numerical Padé-Hermite and simultaneous Padé systems I: Near inversion of generalized Sylvester matrices. *SIAM Journal on Matrix Analysis and Applications*, to appear.

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

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

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

[12] P.R. Graves-Morris. A review of Padé methods for the acceleration of convergence of a sequence of vectors. Technical Report NA 92-31, Dept. of Mathematics, Univ. of Bradford, 1992.

[13] R.E. Shafer. On quadratic approximation. *SIAM J. Numerical Analysis*, 11:447–460, 1974.

[14] M. Van Barel and A. Bultheel. The computation of non-perfect Padé-Hermite approximants. *Numerical Algorithms*, 1:285–304, 1991.

[15] James H. Wilkinson. Modern error analysis. *SIAM Review*, 13:548–568, 1971.

Appendix A
An Example

Consider 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 Padé-Hermite system $S(z)$ of type n=[2,3,1] for $A(z)$ is

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

Only the first column of $S(z)$ yields a Padé-Hermite approximant, this being of type $[1, 2, 0]$. Note that

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

where

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

If we wish to compute the NPHS of type $[3, 4, 2]$, we can advance the solution $S(z)$ above using the steps outlined in §3. To do this, we first compute the NPHS of type $\nu = [3, 4, 2] - [2, 3, 1] - [1, 0, 0] = [0, 1, 1]$ for the residual $T(z)$ in (12). This is given by

$$\widehat{S}(z) = \left[\begin{array}{l|ll} 0 & 5 & -48504 \\ \hline z^2(19092) & 37 - 24z & -59984z \\ z^2(185) & -z & 3478 + 2175z \end{array}\right].$$

By multiplying $S(z)$ in (12) on the right by $\widehat{S}(z)$, we obtain the NPHS of type $[3, 4, 2]$,

$$(37)^2 \left[\begin{array}{l|ll} z^2(5 - 1024z - 669z^2) & -2z + z^3 & 94 - 53z + 3278z^2 + 549z^3 \\ \hline z^2(516 - 199z - 107z^2 - 81z^3) & 1 - z & -1954z + 1489z^2 - 351z^3 + 821z^4 \\ z^2(5 + 8z) & 0 & 94 - 53z + 28z^2 \end{array}\right],$$

the first column of $S(z)$ yields a Padé-Hermite approximant of type $[2, 3, 1]$.

The simultaneous Padé system of type [2,3,1] for

$$A^*(z) = \left[\begin{array}{rr} -a_1(z) & -a_2(z) \\ a_0(z) & 0 \\ 0 & a_0(z) \end{array}\right]$$

is

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

Only the first row of $S^*(z)$ is a simultaneous Padé approximant, this being of type $[2, 3, 1]$. Note that

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

where

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

To compute the NSPS of type $[3, 4, 2]$, we proceed as before by first computing an NSPS of type $\nu = [0, 1, 1]$ for the residual $T^*(z)$. This is given by

$$\widehat{S}^*(z) = \left[\begin{array}{c|cc} 3478 - 81z - 3032z^2 & -470 + 1017z & 48504 - 39568z \\ \hline z^2(-9546 - 7565z) & z^2(2580) & z^2(-266256) \\ z^2(-185 - 396z) & z^2(25) & z^2(-2580) \end{array} \right].$$

By multiplying $S^*(z)$ on the left by $\widehat{S}^*(z)$, we obtain the NSPS of type $[3, 4, 2]$, namely

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

where the first row of $S^*(z)$ is a simultaneous Padé approximant of type $[3, 4, 2]$.

Note that each of the Padé-Hermite systems computed above is unique except for the scaling of columns (VECTOR_PADE scales the columns to have a 1-norm length of 1), and each of the simultaneous Padé systems is unique except for the scaling of rows (VECTOR_PADE scales the rows to have a 1-norm length of 1).

Appendix B

An upper bound for the stability parameter κ

THEOREM: Let $\epsilon = \max\{\|\delta T^t(z)\|_1, \|\delta T^*(z)\|_\infty\}$. If $\epsilon \cdot \kappa_1(\mathcal{M}_n)$, $\epsilon \cdot \kappa_\infty(\mathcal{M}_n^*) < 1$, then[4]

$$\kappa \leq \frac{6(k+1)}{|a_0^{(0)}|} \cdot \frac{\kappa_1(\mathcal{M}_n)\,\kappa_\infty(\mathcal{M}_n^*)}{[1-\epsilon\cdot\kappa_1(\mathcal{M}_n)]\,[1-\epsilon\cdot\kappa_\infty(\mathcal{M}_n^*)]}. \tag{12}$$

PROOF: We begin by obtaining lower bounds for γ_β and $\gamma_\beta^*, 0 \leq \beta \leq k$. First, we obtain a lower bound for γ_0. ¿From (3) (see also [7, eqn. (13)]),

$$\mathcal{M}_n \cdot \mathcal{X} = [0, \ldots, 0, \gamma_0]^t + [\delta r^{(0)}, \ldots, \delta r^{(\|n\|-1)}]^t,$$

where

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

So,

$$\|\mathcal{X}\|_1 \;\leq\; \|\mathcal{M}_n^{-1}\|_1 \cdot (|\gamma_0| + \|\delta r(z)\|) \leq \|\mathcal{M}_n^{-1}\|_1 \cdot (|\gamma_0| + \epsilon).$$

However, because $A^t(z)$ is scaled, then $\|\mathcal{M}_n\|_1 = 1$; and because $S(z)$ is scaled, then $\|\mathcal{X}\|_1 = 1$. Thus,

$$1 \;\leq\; \kappa_1(\mathcal{M}_n) \cdot (|\gamma_0| + \epsilon);$$

or,

$$\frac{1}{\gamma_0} \;\leq\; \frac{\kappa_1(\mathcal{M}_n)}{1-\epsilon\cdot\kappa_1(\mathcal{M}_n)}. \tag{13}$$

Next, to obtain a lower bound for the remaining γ_β, $1 \leq \beta \leq k$, we observe from (3) (see also [7, eqn. (17)]) that

$$\begin{aligned}\mathcal{M}_n \cdot \mathcal{Y}_\beta \;&=\; -\gamma_\beta \cdot [a_\beta^{(1)}, \cdots, a_\beta^{(\|n\|)}]^t + \gamma_\beta \cdot \frac{a_\beta^{(0)}}{a_0^{(0)}} \cdot [a_0^{(1)}, \cdots, a_0^{(\|n\|)}]^t \\ &\quad + [\delta w_\beta^{(1)}, \ldots, \delta w_\beta^{(\|n\|)}]^t,\end{aligned}$$

where

$$\mathcal{Y}_\beta \;=\; \left[u_\beta^{(1)}, \ldots, u_\beta^{(n_0)} | v_{1,\beta}^{(1)}, \ldots, v_{1,\beta}^{(n_1)} | \cdots | v_{k,\beta}^{(1)}, \ldots, v_{k,\beta}^{(n_k)}\right]^t.$$

[4] If the power series are ordered so that $|a_\beta^{(0)}| \leq |a_0^{(0)}|, 1 \leq \beta \leq k$, then the term $|a_0^{(0)}|$ need not appear in upper bound (12).

So,

$$\begin{aligned}\|\mathcal{Y}_\beta\|_1 &\leq \|\mathcal{M}_n^{-1}\|_1 \cdot \left\{ |\gamma_\beta| \left(\frac{|a_\beta^{(0)}|}{|a_0^{(0)}|} \|a_0(z)\| + \|a_\beta(z)\| \right) + \|\delta w_\beta(z)\| \right\} \\ &\leq \|\mathcal{M}_n^{-1}\|_1 \cdot \left\{ |\gamma_\beta| \left(\frac{|a_\beta^{(0)}|}{|a_0^{(0)}|} + 1 \right) + \epsilon \right\}.\end{aligned}$$

In addition, because $S(z)$ is scaled and normalized, we get

$$\|\mathcal{Y}_\beta\| = 1 - |u_\beta^{(0)}| - |\gamma_\beta| = 1 - |\gamma_\beta| \frac{|a_\beta^{(0)}| + |a_0^{(0)}|}{|a_0^{(0)}|}.$$

Therefore,

$$\begin{aligned}1 &\leq |\gamma_\beta| \frac{|a_\beta^{(0)}| + |a_0^{(0)}|}{|a_0^{(0)}|} + \|\mathcal{M}_n^{-1}\|_1 \cdot \left\{ |\gamma_\beta| \frac{|a_\beta^{(0)}| + |a_0^{(0)}|}{|a_0^{(0)}|} + \epsilon \right\} \\ &\leq \kappa_1(\mathcal{M}_n) \cdot \left\{ 2|\gamma_\beta| \frac{|a_\beta^{(0)}| + |a_0^{(0)}|}{|a_0^{(0)}|} + \epsilon \right\},\end{aligned}$$

from which it follows that

$$\frac{1}{\gamma_\beta} \leq 2 \frac{|a_\beta^{(0)}| + |a_0^{(0)}|}{|a_0^{(0)}|} \cdot \frac{\kappa_1(\mathcal{M}_n)}{1 - \epsilon \cdot \kappa_1(\mathcal{M}_n)}. \tag{14}$$

Next, to obtain a lower bound for γ_0^*, from (4) (see also [7, eqn. (30)]), we have

$$\begin{aligned}\mathcal{X}^{*t} \cdot \mathcal{M}_n^* &= v^{*(0)} \left[a_1^{*(1)}, \ldots, a_1^{*(\|n\|)}, \cdots, a_k^{*(1)}, \ldots, a_k^{*(\|n\|)} \right] \\ &- \left[u_1^{*(0)} [a_0^{*(1)}, \ldots, a_0^{*(\|n\|)}], \cdots, u_k^{*(0)} [a_0^{*(1)}, \ldots, a_0^{*(\|n\|)}] \right] \\ &+ \left[\delta w_1^{*(1)}, \ldots, \delta w_1^{*(\|n\|)}, \cdots, \delta w_k^{*(1)}, \ldots, \delta w_k^{*(\|n\|)} \right],\end{aligned}$$

where

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

Consequently,

$$\begin{aligned}\|\mathcal{X}^{*t}\|_\infty &\leq \|\mathcal{M}_n^{*-1}\|_\infty \{ |\gamma_0^*| (\|a_1(z)\| + \cdots + \|a_k(z)\|) \\ &\quad + \frac{|\gamma_0^*|}{|a_0^{(0)}|} \cdot \|a_0(z)\| \cdot \left(|a_1^{(0)}| + \cdots + |a_k^{(0)}| \right) + (\|\delta w_1^*(z)\| + \cdots + \|\delta w_k^*(z)\|) \Big\} \\ &\leq \|\mathcal{M}_n^{*-1}\|_\infty \left\{ |\gamma_0^*| \cdot \|\mathcal{M}_n^*\|_\infty + \frac{|\gamma_0^*|}{|a_0^{(0)}|} \cdot \|\mathcal{M}_n^*\|_\infty \cdot \left(|a_1^{(0)}| + \cdots + |a_k^{(0)}| \right) + \epsilon \right\} \\ &\leq \kappa_\infty(\mathcal{M}_n^*) \left\{ \frac{|\gamma_0^*|}{|a_0^{(0)}|} \cdot \left(|a_0^{(0)}| + \cdots + |a_k^{(0)}| \right) + \epsilon \right\}.\end{aligned}$$

But, since $S^*(z)$ is scaled and normalized, we get

$$\begin{aligned}\|\mathcal{X}^{*t}\|_\infty &= 1 - |\gamma_0^*| - u_1^{*(0)}| - \cdots - |u_k^{*(0)} \\ &= 1 - \frac{|\gamma_0^*|}{|a_0^{(0)}|} \cdot \left(|a_0^{(0)}| + \cdots + |a_k^{(0)}|\right),\end{aligned}$$

and so

$$1 \le \kappa_\infty\left(\mathcal{M}_n^*\right)\left\{2 \cdot \frac{|\gamma_0^*|}{|a_0^{(0)}|} \cdot \left(|a_0^{(0)}| + \cdots + |a_k^{(0)}|\right) + \epsilon\right\}.$$

It now follows that

$$\frac{1}{|\gamma_0^*|} \le \frac{2\left(|a_0^{(0)}| + \cdots + |a_k^{(0)}|\right)}{|a_0^{(0)}|} \cdot \frac{\kappa_\infty\left(\mathcal{M}_n^*\right)}{1 - \epsilon \cdot \kappa_\infty\left(\mathcal{M}_n^*\right)}. \tag{15}$$

Finally, to obtain lower bounds for the remaining γ_α^*, $1 \le \alpha \le k$, from (4) (see also [7, eqn. (32)]), we have

$$\mathcal{Y}_\alpha^{*t} \cdot \mathcal{M}_n^* = \gamma_\alpha^* E_{\alpha\|n\|}^t + \left[\delta r_{\alpha,1}^{*(0)}, \ldots, \delta r_{\alpha,1}^{*(\|n\|-1)}, \cdots, \delta r_{\alpha,k}^{*(0)}, \ldots, \delta r_{\alpha,k}^{*(\|n\|-1)}\right], \quad 1 \le \alpha \le k,$$

where

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

and $E_{\alpha\|n\|}^t$ is the unit row vector of length $k\|n\|$ with a single 1 in position $\alpha\|n\|$. Because $S^*(z)$ is scaled,

$$1 = \mathcal{Y}_\alpha^{*t}\|_\infty \le \|\mathcal{M}_n^{*-1}\|_\infty \cdot (|\gamma_\alpha^*| + 1)$$

and so

$$\frac{1}{|\gamma_\alpha^*|} \le \frac{\kappa_\infty\left(\mathcal{M}_n^*\right)}{1 - \epsilon \cdot \kappa_\infty\left(\mathcal{M}_n^*\right)}. \tag{16}$$

The bound (12) for $\kappa = \sum_{\beta=0}^k 1/|\gamma_\beta\gamma_\beta^*|$ now follows from (13), (14), (15) and (16), since $|a_\beta^{(0)}| \le 1, 1 \le \beta \le k$.