BIT 18 (1978), 475–489

ORDER STARS AND STABILITY THEOREMS

G. WANNER, E. HAIRER and S. P. NØRSETT

Abstract.

This paper clears up to the following three conjectures:

1. The conjecture of Ehle [1] on the A-acceptability of Padé approximations to e^z, which is true;
2. The conjecture of Nørsett [5] on the zeros of the "E-polynomial", which is false;
3. The conjecture of Daniel and Moore [2] on the highest attainable order of certain A-stable multistep methods, which is true, generalizing the well-known Theorem of Dahlquist.

We further give necessary as well as sufficient conditions for A-stable (acceptable) rational approximations, bounds for the highest order of "restricted" Padé approximations and prove the non-existence of A-acceptable restricted Padé approximations of order greater than 6.

The method of proof, just looking at "order stars" and counting their "fingers", is very natural and geometric and never uses very complicated formulas.

1. How we came to order stars.

In the A-stability analysis of many classes of one-step methods, such as implicit Runge–Kutta, Collocation, Rosenbrock type or multiderivative formulas, for the numerical integration of stiff differential equations, there is the question if certain rational approximations to the exponential function $R(z)=P_k(z)/Q_j(z)$ are bounded by 1 on the entire left half plane $\operatorname{Re} z \leqq 0$.

In many cases $R(z)$ is a Padé approximation of order $k+j$ where

$$P_k(z) = 1+\frac{k}{j+k}z+\frac{k(k-1)}{(j+k)(j+k-1)}\frac{z^2}{2!}+\ldots+\frac{k(k-1)\ldots 1}{(j+k)\ldots(j+1)}\frac{z^k}{k!}$$

$$Q_j(z) = 1-\frac{j}{k+j}z+\frac{j(j-1)}{(k+j)(k+j-1)}\frac{z^2}{2!}-+\ldots\pm\frac{j(j-1)\ldots 1}{(k+j)\ldots(k+1)}\frac{z^j}{j!}.$$

Nørsett [5] defined the "E-polynomial"

$$E(y) = |Q_j(iy)|^2-|P_k(iy)|^2$$

in order to study the boundedness $|R(iy)| \leqq 1$ on the imaginary axis. He conjectured that, apart from a multiple zero at the origin, $E(y)$ has only real single roots. We computed the general formula (not easily)

Received April 17, 1978.

$$E(y) = \left[\frac{k!}{(k+j)!}\right]^2 \sum_{r=r_0}^{j} \frac{(-1)^{j-r}}{(j-r)!} \left[\prod_{q=1}^{j-r} (j-q+1)(k+q)(r-k-q)\right] y^{2r}$$

$$r_0 = \left[\frac{k+j+2}{2}\right]$$

for the (k,j) Padé approximation $(j \geq k)$. It turned out that the conjecture of Nørsett is false (first counter example $j=6$, $k=0$) and only true for $k \to \infty$, $k-j = \text{const}$.

Numerical computations showed a close connection between the number of complex pairs of zeros of $E(y)$ and the number of poles of $R(z)$ in the left half plane. See Table 1 below:

Table 1

	Number of complex pairs of zeros of $E(y)$ with $\mathrm{Re}\,(y)>0$													Number of complex pairs of poles of $R(z)$ with $\mathrm{Re}\,(z)<0$											
	$k=0$	1	2	3	4	5	6	7	8	9	10	11		$k=0$	1	2	3	4	5	6	7	8	9	10	11
$j=4$	0	0	0	0	0								$j=4$	0	0	0	0	0							
$j=5$	0	0	0	0	0	0							$j=5$	1	0	0	0	0	0						
$j=6$	1	0	0	0	0	0	0						$j=6$	1	0	0	0	0	0	0					
$j=7$	1	0	0	0	0	0	0	0					$j=7$	1	1	0	0	0	0	0	0				
$j=8$	1	1	0	0	0	0	0	0	0				$j=8$	1	1	1	0	0	0	0	0	0			
$j=9$	1	1	1	0	0	0	0	0	0	0			$j=9$	1	1	1	0	0	0	0	0	0	0		
$j=10$	1	1	1	0	0	0	0	0	0	0	0		$j=10$	2	1	1	1	0	0	0	0	0	0	0	
$j=11$	2	1	1	1	0	0	0	0	0	0	0	0	$j=11$	2	1	1	1	1	0	0	0	0	0	0	0
$j=12$	2	1	1	1	1	0	0	0	0	0	0	0	$j=12$	2	2	1	1	1	1	0	0	0	0	0	0
$j=13$	2	2	1	1	1	1	0	0	0	0	0	0	$j=13$	2	2	1	1	1	1	1	0	0	0	0	0
$j=14$	2	2	1	1	1	1	0	0	0	0	0	0	$j=14$	2	2	2	1	1	1	1	0	0	0	0	0
$j=15$	2	2	2	1	1	1	1	0	0	0	0	0	$j=15$	2	2	2	2	1	1	1	1	0	0	0	0
$j=16$	2	2	2	2	1	1	1	1	0	0	0	0	$j=16$	3	2	2	2	1	1	1	1	1	0	0	0
$j=17$	3	2	2	2	1	1	1	1	1	0	0	0	$j=17$	3	2	2	2	2	1	1	1	1	1	0	0
$j=18$	3	2	2	2	2	1	1	1	1	1	0	0	$j=18$	3	3	2	2	2	2	1	1	1	1	1	0
$j=19$	3	3	2	2	2	2	1	1	1	1	0	0	$j=19$	3	3	3	2	2	2	1	1	1	1	1	0
$j=20$	3	3	3	2	2	2	1	1	1	1	1	0	$j=20$	3	3	3	2	2	2	2	1	1	1	1	1
$j=21$	3	3	3	2	2	2	2	1	1	1	1	1	$j=21$	4	3	3	3	2	2	2	2	1	1	1	1

In order to understand this regularity, we searched for curves that link the zeros of $E(y)$ to the poles of $R(z)$ in some manner. This led us to the definition of the "set A" as given below.

2. Properties of order stars.

First we study rational approximations to the exponential function

$$R(z) = \frac{P_k(z)}{Q_j(z)} \tag{1}$$

where $P_k(z)$ and $Q_j(z)$ are real polynomials of degree k and j respectively. We assume that $Q_j(0) \neq 0$ and that the fraction is reduced, so that P_k and Q_j have no common zeros.

Our aim is to study the *stability region* of R, namely

$$D = \{z \in \mathbf{C} \;;\; |R(z)| \leqq 1\} \,. \tag{2}$$

One says that R is *A-acceptable* (and hence the corresponding method is A-stable), if

$$D \supset \mathbf{C}^- = \{z \in \mathbf{C} \;;\; \operatorname{Re}(z) \leqq 0\} \,.$$

The main tool of this paper is to study instead the region

$$A = \{z \in \mathbf{C} \;;\; |R(z)| > |e^z|\} = \{z \in \mathbf{C} \;;\; |S(z)| > 1\} \tag{3}$$

where

$$S(z) = R(z)/e^z \,. \tag{4}$$

Note that $R(z)$ and $S(z)$ have the same zeros and the same poles.

PROPOSITION 1. *R is A-acceptable if and only if*

(i) *A has no intersection with the imaginary axis, and*

(ii) *R has no poles in $\mathbf{C}^-$.*

PROOF. This follows from the fact that on the imaginary axis, where $|e^z| = 1$, D and A are complementary, and from the maximum principle. ∎

Examples of the set A for Padé approximations are illustrated in the following Figure 1. Looking at these figures, one immediately understands for which reason Padé approximations are A-acceptable exactly if $j-2 \leqq k \leqq j$ (see Theorem 7 below).

The following Propositions 2, 3, and 4 are very elementary but fundamental for the discussion of A.

PROPOSITION 2. *Let the set B_r be defined as $B_r = \{t \in S^1 \,;\, re^{it} \in A\}$. Then there is a number r_0 such that for $r \geqq r_0$ B_r is just an interval in S^1, which for $r \to \infty$ tends to $[\pi/2, 3\pi/2]$. So the border ∂A possesses only two branches that go to infinity.*

PROOF. Since e^x increases quicker and e^{-x} decreases quicker than any rational function, for t fixed,

$$\lim_{r\to\infty} \left|\frac{R(re^{it})}{e^{re^{it}}}\right| = \begin{cases} 0 & \text{if } -\dfrac{\pi}{2} < t < \dfrac{\pi}{2} \\[2mm] \infty & \text{if } \dfrac{\pi}{2} < t < \dfrac{3\pi}{2} \end{cases} .$$

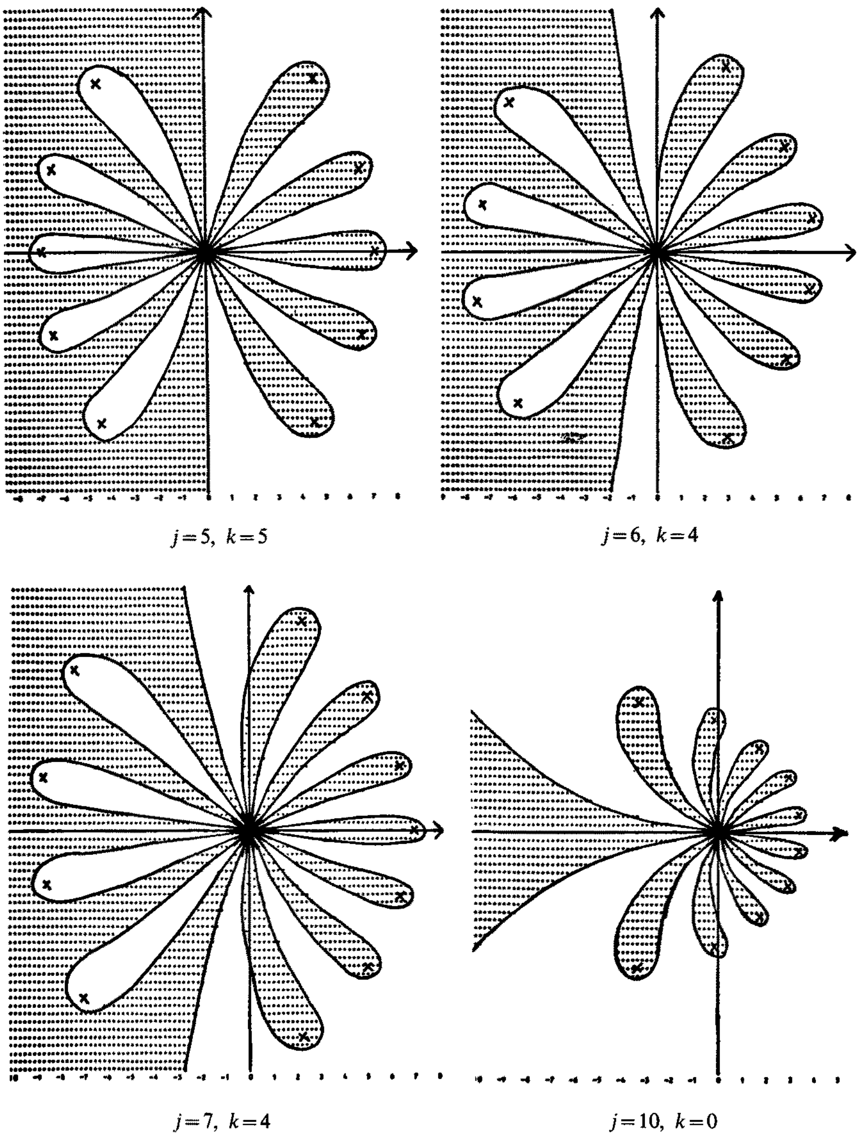


Figure 1. Order stars for Padé approximations.

Thus the boundary ∂A has at least two intersections with the circle $z=re^{it}$, $r>r_0$. In order to show that this circle has at most two intersections, we compute for $|R(re^{it})|=e^{r\cos t}$ the derivative

$$\frac{d}{dt}\left((e^{r\cos t})^2-|R(re^{it})|^2\right) = 2re^{2r\cos t}\left(-\sin t-\operatorname{Re}\left(ie^{it}\frac{R'(re^{it})}{R(re^{it})}\right)\right).$$

Since $|R'/R| \to 0$ for $r \to \infty$, this derivative is <0 for $0<t<\pi$ and >0 for $\pi<t<2\pi$ for large values of r. Hence there can only be two crossing points. ∎

The next Proposition relates the shape of A to the order of approximation.

One says that R is an *approximation of order p*, if there exists a constant $C \neq 0$ so that

$$e^z - R(z) = Cz^{p+1} + O(z^{p+2}) \quad \text{for } z \to 0 . \tag{5}$$

PROPOSITION 3. *R is an approximation of order p if and only if for $z \to 0$ A consists of $p+1$ sectors of width $\pi/(p+1)$, separated by $p+1$ sectors of the complement of A, each of the same width.*

PROOF. By (3) $z=re^{it}$ lies in A iff $|R(re^{it})e^{-r\cos t}|>1$. We insert (5) to obtain for $r \to 0$ the condition

$$|1 - Ce^{-r\cos t} r^{p+1} e^{i(p+1)t}| > 1$$

which leads to

$$C\,\mathrm{Re}\,(e^{i(p+1)t}) = C\cos(p+1)t < 0 .$$

This is satisfied in consecutive intervals of length $\pi/(p+1)$. ∎

For this reason we use the name *order star* for the set A. We further call *fingers* the connected components of each of these sectors. If m sectors join together to one finger, we call it a *finger of multiplicity m*. The analogous sets for the complement CA we call *dual fingers* and *dual fingers of multiplicity m*.

PROPOSITION 4. *Each bounded finger of multiplicity m contains at least m poles of R (counted with their multiplicity); each bounded dual finger of multiplicity m contains at least m zeros of R.*

PROOF. Let $c(t)$, $t_0 \leqq t \leqq t_1$, be a parametrization of the positively oriented boundary of a finger F, $\boldsymbol{a}=(c_1'(t), c_2'(t))$ a tangent vector, $\boldsymbol{n}=(c_2'(t), -c_1'(t))$ an outside normal vector. We write $S(z)=r(x,y)e^{i\varphi(x,y)}$ $(z=x+iy)$ and since the modulus of S increases inside F, we have $\partial(\log r)/\partial\boldsymbol{n}<0$. Now the Cauchy–Riemann differential equations in polar-coordinate-form

$$\frac{\partial(\log r)}{\partial x} = \frac{\partial\varphi}{\partial y}; \quad \frac{\partial(\log r)}{\partial y} = -\frac{\partial\varphi}{\partial x}$$

(see e.g. [13], p. 67) imply $\partial\varphi/\partial\boldsymbol{a}<0$. Thus the argument of S decreases along c.

The difference between the number of zeros and the number of poles inside c is

$$Z - P = \frac{1}{2\pi i}\int_c \frac{S'(z)}{S(z)}\,dz = \text{number of rotations of } \arg(S) \text{ along } c .$$

If F is an m-fold finger, the boundary returns m times to the origin, thus at least m times arg (S) has the same direction, so the number of rotations is at least $-m$ and $P \geqq m$ (see Fig. 2 where $m=3$).

For dual fingers the argumentation is the same starting from $\partial(\log r)/\partial \boldsymbol{n} > 0$. ∎

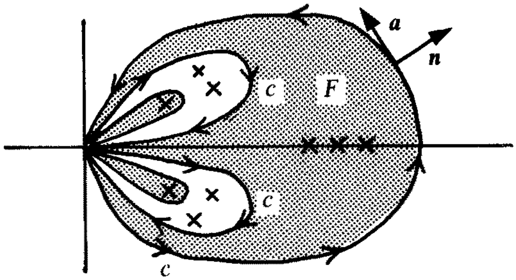


Figure 2.

3. Stability theorems for rational functions.

THEOREM 5. *If $R(z)=P_k(z)/Q_j(z)$ is A-acceptable and an approximation to e^z of order p, then*

$$p \leqq 2j \quad \text{and} \quad p \leqq 2k+2\,.$$

PROOF. By Proposition 3 at least $[(p+1)/2]$ fingers of A start in the left half plane $\mathbf{C}^-$ (see Fig. 3 where $p+1=11$). These fingers cannot cross the imaginary axis (Prop. 1) and cannot be bounded (Prop. 4). So (Prop. 2) they all must collapse and include at least $[(p+1)/2]-1$ bounded dual fingers. So Prop. 4 gives that the total number of zeros of R satisfies $k \geqq [(p+1)/2]-1$ or $p \leqq 2k+2$.

The other inequality, which follows trivially from $p \leqq k+j$ and $k \leqq j$ could be proved similarly (see Theorem 12). ∎

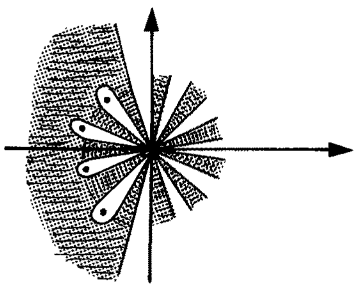


Figure 3.

Let us next give a simple proof of a result similar to that of Crouzeix and Ruamps [11].

THEOREM 6. *Suppose that for* $R(z)=P_k(z)/Q_j(z)$

(i) $p \geqq 2j-2$;

(ii) $\lim_{z\to\infty}|R(z)| \leqq 1$;

(iii) *the coefficients of* Q *have alternating signs.*

Then R *is A-acceptable.*

PROOF. It follows from (ii) that $k \leqq j$. Now the polynomial of degree $2j$ which is even because of symmetry

$$E(y) = |Q_j(iy)|^2 - |P_k(iy)|^2 = (|Q|+|P|)(|Q|-|P|)$$

satisfies $E(y)=O(y^{p+1})$ because of (5). Thus (i) gives us $E(y)=Ky^{2j}$ and (ii) implies $K \geqq 0$, so the order star can nowhere meet the imaginary axis.

At least $[(p+1)/2]$ fingers of A start in the right half plane $\mathbf{C}^+$ and must be bounded (Prop. 2). Hence there must be at least $[(p+1)/2]$ poles of S in $\mathbf{C}^+$. Since from (i) $[(p+1)/2] \geqq j-1$, there can be at most one (and hence real) pole of S in $\mathbf{C}^-$, which is impossible because of (iii). ∎

THEOREM 7. ("*Theorem and Conjecture of Ehle*"). *Any Padé approximation* $R(z) = P_k(z)/Q_j(z)$ *to the exponential function is A-acceptable if and only if*

$$j-2 \leqq k \leqq j\,.$$

PROOF. Since Padé approximations have optimal order $p=k+j$, this is an immediate consequence of Theorems 5 and 6. ∎

4. The attainable order with real and multiple singularities.

For the treatment of large stiff systems it is preferable to use rational approximations for which the denominator can be factorized into real linear factors, since then the evaluation of $y_{n+1}=(Q_j(A))^{-1}P_k(A)y_n$ can be decomposed into a sequence of real linear equations. Another reason for the interest in these types of approximations is the fact that they are related to Rosenbrock type methods as well as semi-implicit or singly-implicit Runge–Kutta methods.

Let us give the following extension of a result of Nørsett and Wolfbrandt [6, 14].

THEOREM 8. *Let* $R(z)=P_k(z)/Q_j(z)$ *be such that* $Q_j(z)$ *has only* m *complex different zeros. If in addition* $Q_j(z)$ *possesses real zeros, then the order* p *satisfies*

$$p \leqq k+m+1\,.$$

If $Q_j(z)$ *has no real zeros at all, then we have*

$$p \leqq k+m\,.$$

PROOF. At most $\lambda = m+2$ of the $p+1$ dual fingers can be infinite (see Fig. 4). If there are no real singularities, $\lambda = m+1$. So at least $p+1-\lambda$ dual fingers are bounded and hence (Prop. 4) the number of zeros k must satisfy $k \geqq p+1-\lambda$. This gives the stated estimates. ∎

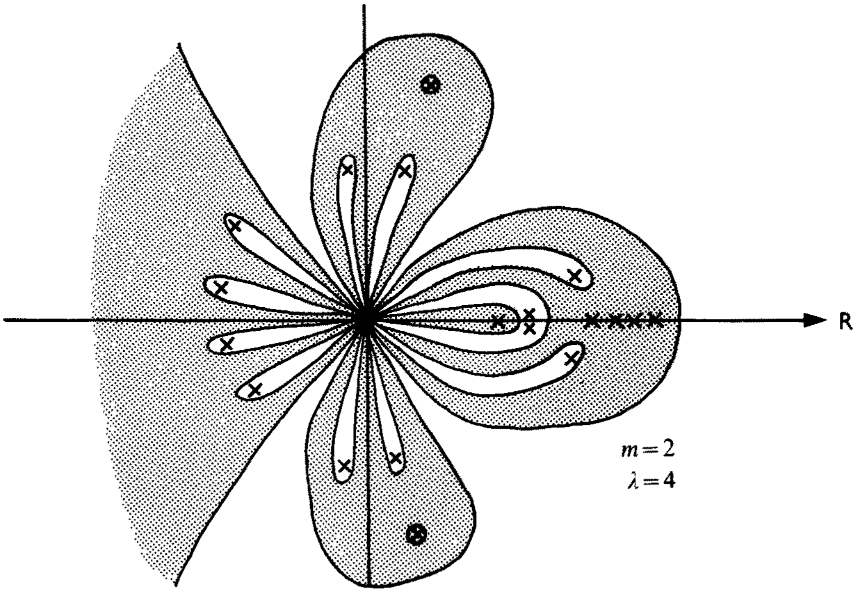


Figure 4.

5. A-acceptability of restricted Padé approximations.

In this section we study the particular class of approximations

$$R(z) = \left(\sum_{m=0}^{k} (-1)^k L_k^{(k-m)}(1/\gamma)(\gamma z)^m\right)/(1-\gamma z)^k \tag{7}$$

which are of order k for all $\gamma \in \mathbf{R}$ (see Nørsett [9] Corollary 2.1; L_k denotes the kth Laguerre polynomial). Here the denominator has just a k-fold zero. This makes them very useful for large sparse matrices.

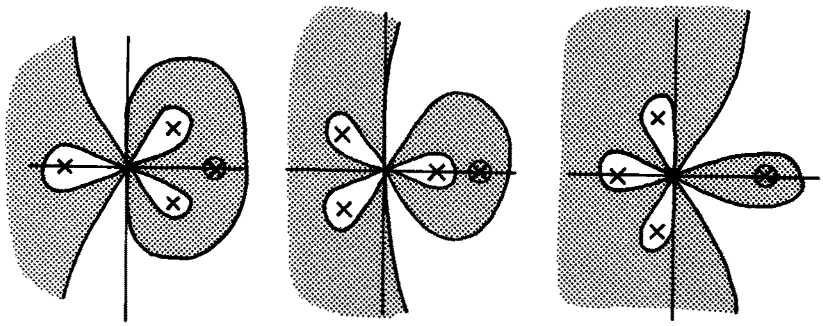
Since the error constant for this approximation is

$$C = (-1)^{k+1} \frac{\gamma^k}{k+1} L'_{k+1}\left(\frac{1}{\gamma}\right), \tag{8}$$

R has maximal order $k+1$ (see Theorem 8), when

$$\gamma = \gamma_v, \; L'_{k+1}\left(\frac{1}{\gamma_v}\right) = 0, \quad \frac{1}{\gamma_1} < \frac{1}{\gamma_2} < \ldots < \frac{1}{\gamma_k}, \qquad v = 1, \ldots, k\,. \tag{9}$$

Figure 5 shows the order stars of these optimal approximations for the case $k=3$.



$\gamma=\gamma_1$ $\gamma=\gamma_2$ $\gamma=\gamma_3$

Figure 5.

PROPOSITION 9. *The order star for the restricted Padé approximation* (7) *with* γ *as in* (9) *contains just one bounded finger of multiplicity* $k-\nu+1$.

PROOF. Since there is only one k-fold singularity, all bounded fingers, say m, must collapse to one m-fold finger. Thus k of the $k+2$ dual fingers must be finite and each contains one zero of $P_k(z)$ (Prop. 4). So the order star is uniquely determined once we know how many zeros of $P_k(z)$ lie to the right of the two infinite branches of ∂A (Compare with Fig. 5 where this number, from left to right, is equal to 2, 1, and 0).

We write $P_k(z)$ by putting $\gamma z=w$ and $1/\gamma=\beta$ as

$$L_k(\beta)w^k+L_k'(\beta)w^{k-1}+L_k''(\beta)w^{k-2}+\ldots = 0 . \tag{10}$$

When β increases say from $1/\gamma_\nu$ to $1/\gamma_{\nu+1}$, all w's depend continuously on β if $L_k(\beta) \neq 0$ and the zeros cannot change their position vis-à-vis ∂A. From properties of orthogonal polynomials there is exactly one $\bar{\beta}$ in this interval where $L_k(\beta)$ has a single zero and thus exactly one solution of (10), say w_1, tends to infinity. To study its behavior near $\bar{\beta}$ we write $L_k(\beta)=(\beta-\bar{\beta})L_k'(\bar{\beta})$ and neglect lower order terms in (10). Because $L_k'(\bar{\beta})\neq 0$ this leads to $w_1 \cong 1/(\bar{\beta}-\beta)$, showing that for increasing β, one zero of $P_k(z)$ tends at the right to $+\infty$ and comes back at the left from $-\infty$. So m has decreased by one and there is a constant M such that $m=M-\nu$ ($\nu=1,\ldots,k$). Finally the overall inequality $1 \leqq m \leqq k$ (Prop. 4) gives $M=k+1$. ∎

Figure 6 illustrates how the order star for $k=3$ changes when γ varies from γ_1 to γ_2.

THEOREM 10. *The restricted Padé approximation* (7) *with optimal order* (9) *can only be A-acceptable if*

$$q \leqq \nu \leqq q+1 \quad \textit{for } k=2q+1$$

$$q = \nu \quad \textit{for } k=2q .$$

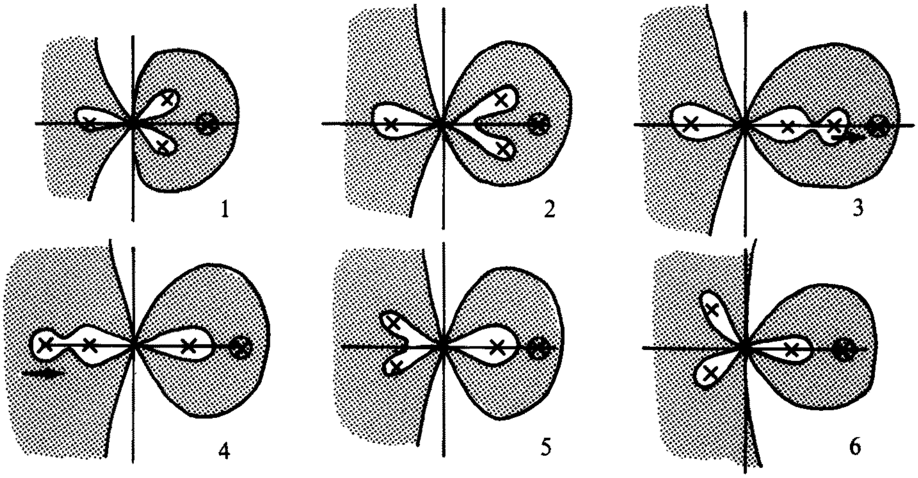


Figure 6.

PROOF. $k+2$ fingers start at the origin. According to proposition 9, $k-\nu+1$ go to the right, $\nu+1$ go to the left. At most $[(k+3)/2]$ fingers can start in $\mathbf{C}^+$ or in $\mathbf{C}^-$. So if not

$$\left[\frac{k+3}{2}\right] \geqq k-\nu+1 \quad \text{and} \quad \left[\frac{k+3}{2}\right] \geqq \nu+1 ,$$

some fingers must cross the imaginary axis and R cannot be A-acceptable. ∎

LEMMA. *The approximation* (7) *and* (9) *can only be A-acceptable if*

$$k \geqq \begin{cases} 1 & \text{if } \nu=1 \\ 5 & \text{if } \nu=2 \\ 9 & \text{if } \nu=3 \\ 6\nu-10 & \text{if } \nu\geqq 4 . \end{cases}$$

PROOF. From (7) we have $\lim_{z\to\infty} |R(z)| = |L_k(1/\gamma_\nu)|$. So it is necessary that $|L_k(1/\gamma_\nu)| \leqq 1$. For small values of k the stated bounds are obtained by numerical computations (see Nørsett [9a], p. A.7). For large values we use Hilb's asymptotic formula (see Szegö [12], page 193)

$$L_k(x) = \frac{e^{x/2}}{(\pi)^{1/2}(xk)^{1/4}} \left(\cos\left(2(xk)^{1/2} - \pi/4\right) + (xk)^{-1/2}O(1)\right)$$

which gives for the νth extremum point of L_{k+1} the approximation

$$x_\nu = (\pi^2(\nu+1/4)^2)/4(k+1) .$$

By inserting this into the above formula for L_{k+1} and using $L_{k+1}(\gamma_\nu^{-1}) = L_k(\gamma_\nu^{-1})$, some modifications show that $|L_k(x_\nu)| > 1$, if not

$$k+1 \geqq \pi^2(\nu+1/4)^2/(4\log(\pi^2(\nu+1/4)/2)) .$$

Computing this expression for different value of ν, one obtains the above stated estimates except for the case $\nu = 1$, where the asymptotic formula is not yet sufficiently close. ∎

THEOREM 11. *The restricted Padé approximation* (7) *with optimal order* (9) *is A-acceptable if and only if*

$$\begin{array}{lcccc} k= & 1 & 2 & 3 & 5 \\ \nu= & 1 & 1 & 1 & 2 \end{array} .$$

PROOF. These are the only possible cases left by the two foregoing results. The A-acceptability of these remaining cases has been proved in [10]. ∎

6. The multistep case.

The stability analysis of multistep methods, multistep-multiderivative formulas, PC-schemes, composite or cyclic multistep methods, multistep Runge–Kutta methods etc. (see e.g. [3, 4, 7, 8]) leads to a characteristic algebraic equation

$$(11) \qquad Q(z,R) := Q_0(z)R^k + Q_1(z)R^{k-1} + \ldots + Q_k(z) = 0$$

for the eigenvalues of the resulting difference equation.

We suppose

$$(12) \quad Q(z,R) \text{ irreducible, } Q_0(0) \neq 0,\ \frac{\partial Q}{\partial R}(0,1) \neq 0,\ \deg Q_r \leqq j \quad (r = 0,1,\ldots,k) .$$

For linear multistep methods all $Q_r(z)$ are linear, hence in this case $j = 1$. For other classes of methods j indicates the number of used derivatives or stages while k is the number of steps involved in the method. For $k = 1$ we obtain a rational R as considered in the foregoing sections.

For k greater than 1 the solution of (11) becomes a multivalued function. We thus introduce the corresponding Riemann surface M on which R becomes singlevalued again. With the exception of some pathological cases M can most easily be written as

$$M = \{(z,w) \in \mathbf{C}^2 \ ; \ Q(z,w) = 0\}$$

with the projections

$$\begin{array}{ccc} M \xrightarrow{R} \mathbf{C} & \qquad & (z,w) \xmapsto{R} w \\ \pi\downarrow & & \pi\downarrow \\ \mathbf{C} & & z \end{array}$$

For each $z \in \mathbf{C}$, the inverse $\pi^{-1}(z)$ of the covering projection consists in general of k points $(z, w_1), \ldots, (z, w_k)$ where $w_1, \ldots, w_k$ are the k solutions of (11). So M is an orientable surface consisting locally of k sheets lying above $\mathbf{C}$ and interacting in a finite number of branch points, i.e. the finite number of points where (11) has multiple solutions. (See e.g. [13], chapter V).

Again we define the *stability domain* as

$$D = \{z \in \mathbf{C} \ ; \ |R(\pi^{-1}(z))| \leqq 1 \text{ for all inverses and } < 1 \text{ at branch points}\}$$

and call the function R *A-acceptable* if D contains $\mathbf{C}^-$.

The *order star* is now a subset of M

$$A = \{z \in M \ ; \ |R(z)| > |e^{\pi(z)}|\} = \{z \in M \ ; \ |S(z)| > 1\}$$

where

$$S(z) = R(z)/e^{\pi(z)} \quad \text{for } z \in M \ .$$

We suppose that the methods considered are *consistent*, so that for $z=0$ $R=1$ is a solution of (11) and by (12) this solution is simple in a neighborhood of the origin. Thus it can be continued in a neighbourhood of 0 to a *principal solution* $R_1(z)$ defined on its *principal sheet.*

The method is of *order p*, if there exists a constant $C \neq 0$ such that for the principal solution

$$e^z - R_1(z) = Cz^{p+1} + O(z^{p+2}) \quad \text{for } z \to 0 \ . \tag{13}$$

This is, because of

$$Q(z, e^z) = Q(z, e^z) - Q(z, R_1(z)) = \frac{\partial Q}{\partial R}(0,1)Cz^{p+1} + O(z^{p+2})$$

and (12) equivalent to the usual definition.

Now the Propositions 1–4 remain true with the following modifications:

PROPOSITION 1. *R is A-acceptable if and only if*

(i) *A has no intersection with $\pi^{-1}(i\mathbf{R})$ and $\bar{A}$ never touches $\pi^{-1}(i\mathbf{R})$ in a branch point;*

(ii) *R has no poles in $\pi^{-1}(\mathbf{C}^-)$, i.e. $Q_0(z)$ has no zeros in $\mathbf{C}^-$.*

The proof is trivial in one direction (the one which is actually used below) and uses the maximum principle for Riemann surfaces for the other direction.

PROPOSITION 2. *The same statement now holds on each sheet. The sheets of M outside all finite branch points are either all separated or, if ∞ is itself a branch point, some of them spiral together in the usual way. The formula $|R'/R| \to 0$ for $z \to \infty$, used in the proof, is best seen from the expansion*

$$R(z) = \sum_{q=q_0}^{\infty} a_q z^{-q/m} \ .$$

PROPOSITION 3. *It remains true, but this time on the principal sheet only.*

We again say *fingers* and *multiple fingers* for the connected components of these sectors in M.

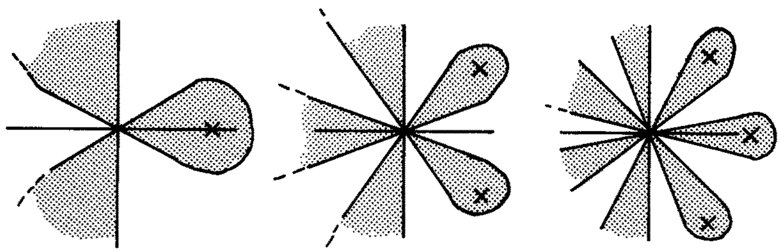
PROPOSITION 4. *It remains the same.*

For the proof one has to take care that the border of F can be composed of *several* closed loops, since M may be no longer simply connected. So the integration of $(1/2\pi i)\int_c S'(z)/S(z)\,dz$ has to be extended over all of these loops and the sum of the integrals is $\leqq -m$ since in total m times one of these loops visits the origin.

THEOREM 12 ("Conjecture of Daniel and Moore"). *If R is A-acceptable and satisfies* (11) *and* (12), *then $p \leqq 2j$ and* sign $(C) = (-1)^j$ *for $p = 2j$.*

PROOF. If R is of order p, at least $[(p+1)/2]$ sectors start in $\mathbf{C}^+$ on the principal sheet (Prop. 3, Fig. 7). Its fingers cannot cross $\pi^{-1}(i\mathbf{R})$ and thus must be bounded (Prop. 1 and 2). The total number of poles available on M is j, the degree of $Q_0(z)$. So by Propositon 4 $[(p+1)/2] \leqq j$ or $p \leqq 2j$.

The second assertion can be seen from Figure 7 and the fact (see the proof of Prop. 3) that the real positive axis (for z small) belongs to A iff $C<0$. ∎



$p=2$ (the Dahlquist case) $\qquad$ $p=4$ $\qquad$ $p=6$

Figure 7.

THEOREM 13 (Second part of the conjecture). *The error constant C of an A-acceptable R of maximal order $p=2j$ satisfies*

$$|C| \geqq |\tilde{C}|$$

where

$$\tilde{C} = (-1)^j \frac{(j!)^2}{(2j)!(2j+1)!}$$

is the error constant of $R_{jj}(z)$, the diagonal Padé approximation of order $2j$.

PROOF. Subtracting (5) from (13) we get

$$R_{jj}(z)-R_1(z) = (C-\tilde{C})z^{2j+1}+O(z^{2j+2}) . \tag{14}$$

So we consider $R_1(z)$ as approximation of order $2j$ to $R_{jj}(z)$ and look at the *relative* order star

$$B = \{z \in M \ ; \ |R_1(z)|>|R_{jj}(\pi(z))|\} = \{z \in M \ ; \ |S(z)|>1\}$$

where

$$S(z) = R_1(z)/R_{jj}(\pi(z)) .$$

Since $|R_{jj}(iy)|=1$ and R_1 is A-acceptable, B cannot cross $\pi^{-1}(i\mathbf{R})$ and since $R_{jj} \neq 0$ on $\mathbf{C}^+$ (see e.g. Fig. 1 or Theorem 7), $S(z)$ has no more poles on $\pi^{-1}(\mathbf{C}^+)$ than $R(z)$.

In spite of the fact that the fingers in $\mathbf{C}^+$ are no longer necessarily bounded, Proposition 4 applies as well, since both R_1 and R_{jj} are A-acceptable and the point ∞ is no longer a singularity.

Suppose now $|C|<|\tilde{C}|$ and, for example, j even. It follows from (14) that for z real, small and positive, $R_{jj}<R_1$, and hence the positive real axis for z small belongs to B. We thus have the situation contrary to that in Figure 7, so that now $j+1$ fingers start in $\mathbf{C}^+$ requiring $j+1$ poles of R_1, a contradiction. ∎

Remarks.

1. Several authors have proved parts of this conjecture. Genin [3] arrives at different results, since his methods are not stable in the usual sense. He has been corrected later by Jeltsch ([4] and other papers).

2. In the multistep case it is no longer easy to derive similar conditions on the number of zeros as in Theorem 5, since here M is in general not simply connected so that two collapsing fingers do not necessarily contain a bounded dual finger. A counter-example is Gear's 2nd order backward difference formula which leads to $(3/2-z)R^2-2R+1/2=0$. Here R has no zero at all.

3. The authors wish to acknowledge a long discussion with Mr. B. Kaup on Riemann surfaces.

REFERENCES

1. B. L. Ehle, *On Padé approximations to the exponential function and A-stable methods for the numerical solution of initial value problems*, Thesis 1969, CSRR2010, Department of Computer Science, University of Waterloo, Ontario, Canada, conjecture 3.1.
Also see a paper of similar title in SIAM J. Math. Anal. 4 (1973), 671–680.
2. J. W. Daniel and R. E. Moore, *Computation and theory in ordinary differential equations*, Freeman and Co., 1970, page 80.
3. Y. Genin, *An algebraic approach to A-stable linear multistep-multiderivative integration formulas*, BIT 14 (1974), 382–406.

4. R. Jeltsch, *Note on A-stability of multistep-multiderivative methods*, BIT 16 (1976), 74–78.
5. S. P. Nørsett, *C-polynomials for rational approximations to the exponential function*, Numer. Math. 25 (1975), 39–56.
6. S. P. Nørsett and A. Wolfbrandt, *Attainable order of rational approximations to the exponential function with only real poles*, BIT 17 (1977), 200–208.
7. W. B. Rubin, *A-stability and composite multistep methods*, Ph. D. Thesis, EE Dept., Syracuse University, New York 1973.
8. H. J. Stetter, *Analysis of discretization methods in ordinary differential equations*, Springer-Verlag 1973.
9. S. P. Nørsett, *Restricted Padé approximations to the exponential function*, to appear in SIAM J. Numer. Anal. 1978.
9a. *id.*, similar title, Math. and Computation, No. 474, Trondheim.
10. S. P. Nørsett, *One step methods of Hermite type for numerical integration of stiff systems*, BIT 14 (1974), 63–77.
11. M. Crouzeix and F. Ruamps, *On rational approximations to the exponential*, R.A.I.R.O. Analyse Numérique 11, n° 3 (1977), 241–243.
12. G. Szegö, *Orthogonal polynomials*, AMS 1939.
13. Behnke-Sommer, *Theorie der analytischen Funktionen*, Springer 1962.
14. A. Wolfbrandt, *A note on a recent result of rational approximations to the exponential function*, BIT 17 (1977), 367–368.

DÉPARTEMENT DE MATHÉMATIQUES
CASE POSTALE 124
2-4 RUE DU LIÈVRE
CH-1211 GENÈVE 24
SUISSE