Acta Numerica (1995), *pp.* 53–103

A Taste of Padé Approximation

C. Brezinski
Laboratoire d'Analyse Numérique et d'Optimisation
UFR IEEA
Université des Sciences et Technologies de Lille
59655–Villeneuve d'Ascq cedex, France
E-mail: brezinsk@omega.univ-lille1.fr

J. Van Iseghem
UFR de Mathématiques Pures et Appliquées
Université des Sciences et Technologies de Lille
59655–Villeneuve d'Ascq cedex, France
E-mail: jvaniseg@ano.univ-lille1.fr

The aim of this paper is to provide an introduction to Padé approximation and related topics. The emphasis is put on questions relevant to numerical analysis and applications.

CONTENTS

Let f be a formal power-series. A Padé approximant of f is a rational function whose numerator and denominator are chosen so that its power series expansion (which is obtained by dividing the numerator by the denominator) agrees with f as far as possible, that is, at least up to the term whose degree equals the sum of the degrees of the numerator and the denominator of the rational function.

Such approximants have a long history and they have played an important rôle in the solution of many problems such as the transcendence of the numbers e and π and have given birth to some fundamental ideas in mathematics such as the spectral theory of operators. They are also closely connected to continued fractions; see Brezinski (1990) and Lorentzen and Waadeland (1992). Thirty years ago, Padé approximants were rediscovered by physicists and they proved to be a very efficient tool not only for improving existing methods but also for extracting important information from power series and thus leading to new possibilities which were not open before.

Let us give some examples and begin with a purely mathematical one. We shall consider the series

$$f(z) = z - z^2/2 + z^3/3 - z^4/4 + \cdots,$$

which is known to converge to $\ln(1 + z)$ for $|z| \leq 1, z \neq -1$. Thus the simplest process for obtaining an approximate value of $\ln(1 + z)$ is to sum the series up to a certain term. Let us call $f_k(z)$ the partial sum of f up to the term of degree k inclusive and let $[p/q]_f(z)$ be the Padé approximant of f whose numerator and denominator have the respective degrees p and q at most. As we shall see below, the computation of this Padé approximant requires the knowledge of the coefficients of f up to that of the power $p+q$.

For $z = 1$ we have $\ln 2 = 0.6931471805599453\ldots$. For $z = 2$, the series diverges and we have $\ln 3 = 1.098612288668110\ldots$. The Padé approximants give the results presented in the following table.

k	$f_{2k}(1)$	$[k/k]_f(1)$
1	0.830	0.7
2	0.783	0.6933
3	0.759	0.693152
4	0.745	0.69314733
5	0.736	0.6931471849
6	0.730	0.69314718068
7	0.725	0.693147180563
8	0.721	0.69314718056000
9	0.718	0.6931471805599485
10	0.716	0.6931471805599454

k	$f_{2k}(2)$	$[k/k]_f(2)$
1	$0.260 \cdot 10^1$	1.14
2	$0.506 \cdot 10^1$	1.101
3	$0.126 \cdot 10^2$	1.0988
4	$0.375 \cdot 10^2$	1.098625
5	$0.121 \cdot 10^3$	1.0986132
6	$0.410 \cdot 10^3$	1.09861235
7	$0.142 \cdot 10^4$	1.098612293
8	$0.504 \cdot 10^4$	1.0986122890
9	$0.181 \cdot 10^5$	1.098612288692
10	$0.655 \cdot 10^5$	1.0986122886698

Figures 1 and 2 below display respectively the partial sums $[3/0], [5/0]$ and $[7/0]$ of the series for $\arctan x$ for real values of x and its Padé approximants $[2/1], [3/2]$ and $[4/3]$. Each of them is easily recognizable. They clearly show that the domain of convergence of the series has been increased.

For other examples from physics, the interested reader is referred to the very complete book of Baker and Graves-Morris (1981), to the work of A. P. Magnus (1988) and to Guttmann (1989). Applications to numerical analysis, through the use of continued fractions, are described by Jones and Thron (1988). See Brezinski (1991a) for a bibliography.

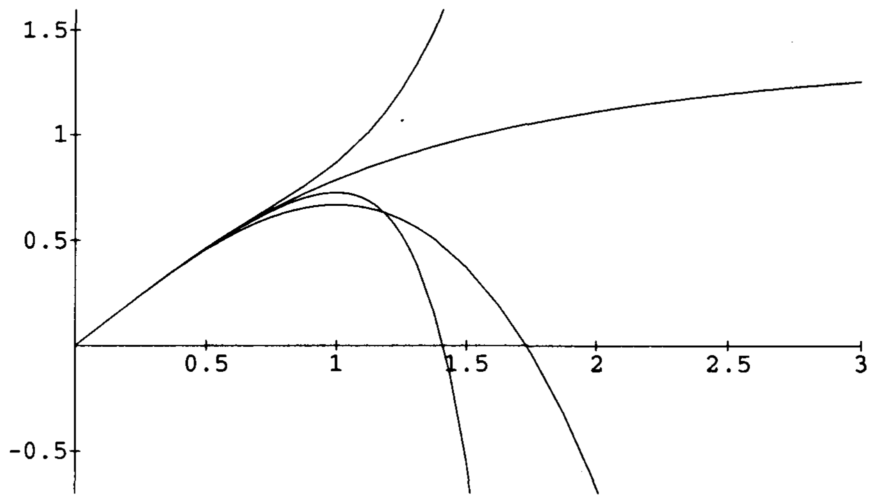


Fig. 1. Partial sums of $\arctan x$

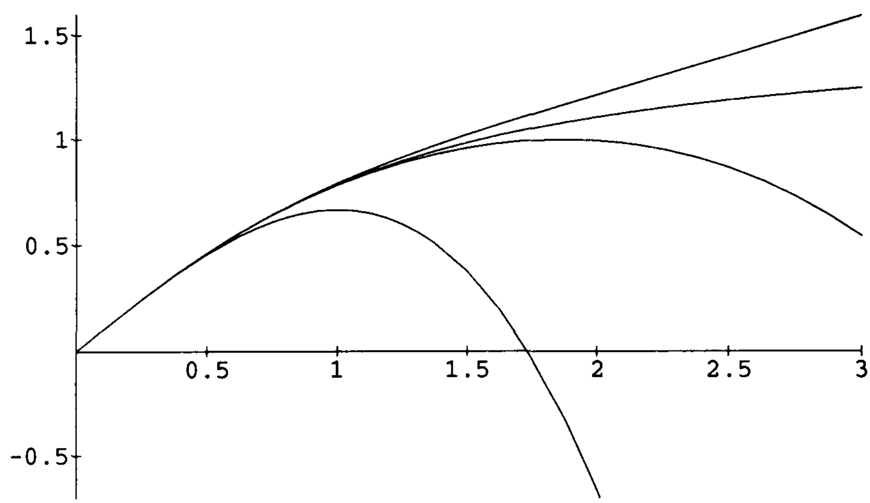


Fig. 2. Padé approximants of $\arctan x$

1. Algebraic theory

Let us first begin with some definitions.

1.1. Definitions

Let us now give the exact definition of Padé approximants. We shall give two approaches to the subject: a direct one which is sufficient to understand it and a more complicated one which leads to a better grasp of its numerous relations with the theory of formal orthogonal polynomials and will serve as a basic tool for developing recurrence relations for the computation of Padé approximants and for other purposes that will be discussed later.

Let f be a formal power series with complex coefficients

$$f(z) = c_0 + c_1 z + c_2 z^2 + c_3 z^3 + \cdots .$$

Definition 1.1 The Padé approximant $[p/q]_f(z)$ is a rational function $N(z)/D(z)$ such that degree $N \leq p$, degree $D \leq q$ and

$$N(z) - f(z)D(z) = \mathcal{O}\left(z^{p+q+1}\right), \qquad z \to 0.$$

Let us write

$$N(z) = a_0 + a_1 z + \cdots + a_p z^p,$$

$$D(z) = b_0 + b_1 z + \cdots + b_q z^q.$$

Then the conditions of the definition lead to

$$\begin{array}{rcl} a_0 & = & c_0 b_0, \\ a_1 & = & c_1 b_0 + c_0 b_1, \\ \vdots & & \vdots \\ a_p & = & c_p b_0 + c_{p-1} b_1 + \cdots + c_{p-q} b_q, \\ 0 & = & c_{p+1} b_0 + c_p b_1 + \cdots + c_{p-q+1} b_q, \\ \vdots & & \vdots \\ 0 & = & c_{p+q} b_0 + c_{p+q-1} b_1 + \cdots + c_0 b_q \end{array}$$

with the convention that $c_i = 0$ for $i < 0$.

The last q equations contain $q+1$ unknowns $b_0, \ldots, b_q$ and, thus, this system has a non-trivial solution. With knowledge of the b_i's, the first $p+1$ equations directly give the a_i's.

Two solutions of the problem lead to the same rational function since $N_1(z) - f(z)D_1(z) = \mathcal{O}(z^{p+q+1})$ and $N_2(z) - f(z)D_2(z) = \mathcal{O}(z^{p+q+1})$ implies $N_1(z)D_2(z) - D_1(z)N_2(z) = \mathcal{O}(z^{p+q+1})$. But the degree of $N_1D_2 - D_1N_2$ is at most $p+q$ and thus $N_1(z)D_2(z)$ is identical to $D_1(z)N_2(z)$. N and D can have a common factor. In particular if z^k is a factor of D, it is also a factor of N, as can be seen from the previous system, and thus $N(z)/D(z)$ cannot have a pole at the origin. Dividing by the highest power k in z contained in D gives a solution with $D(0) \neq 0$ and degree $N \leq p-k$, degree $D \leq q-k, N(z) - f(z)D(z) = \mathcal{O}\left(z^{p+q+1-k}\right)$, and we have

Theorem 1.1 Let $R(z) = N(z)/D(z)$ be an irreducible rational function with degree $N = p - k$, degree $D = q - k, k \geq 0$, and

$$N(z) - f(z)D(z) = \mathcal{O}\left(z^{p+q+1-k}\right), \qquad z \to 0.$$

Then, for $i, j = 0, \ldots, k$,

$$[p - k + i/q - k + j]_f(z) \equiv R(z)$$

and no other Padé approximant is identical to R if k is a maximum.

The result follows from the definition of Padé approximants and their uniqueness. This identity between Padé approximants can hold for all i and j (see Property 1.5) and, in that case, f is a rational function with a numerator of degree $p - k$ and a denominator of degree $q - k$ or there may exist a maximal value of k for which it holds and, in that case, no other Padé approximant is identical to R.

Usually the Padé approximants are arranged in a double-entry table known as the Padé table

$$\begin{array}{cccc} [0/0] & [0/1] & [0/2] & \cdots \\ [1/0] & [1/1] & [1/2] & \cdots \\ [2/0] & [2/1] & [2/2] & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

The theorem given above was proved by Henri Padé in 1892 (see Padé, 1984). It says that identical Padé approximants can only occur in square blocks of the table, a property known as the block structure of the Padé table. If the Padé table does not contain blocks, it is said to be normal; otherwise it is called non-normal. This block structure corresponds in fact to the block structure of the table of formal orthogonal polynomials (see Subsection 1.4) which itself mimics the block structure of the table of Hankel determinants (see Subsection 1.3). On these questions see Gragg (1972), de Bruin and Van Rossum (1975), Gilewicz (1978) and Draux (1983).

Other algebraic properties of Padé approximants will be given in Subsection 1.2.

Let us now come to the second approach to the subject.

Let c be the linear functional on the space of complex polynomials defined by

$$c(x^i) = c_i.$$

The functional c can be extended to the space of formal power series, thus leading to formal (that is, term-by-term) identities.

Our second approach is based on the following obvious formal identity which is given as a theorem since it is fundamental.

Theorem 1.2

$$f(z) = c\left(\frac{1}{1-xz}\right).$$

The problem of approximating $f(z)$ is classical in numerical analysis. For example, if an approximation of

$$I = \int_a^b g(x)w(x)\,\mathrm{d}x$$

is wanted, one can replace g by an interpolation polynomial and integrate it. This procedure leads to a so-called interpolatory quadrature formula which is exact on the space of polynomials of degree at most $k-1$ if k interpolation points are used. If these interpolation points are the zeros of the polynomial of degree k belonging to the family of orthogonal polynomials on $[a,b]$ with respect to w, the quadrature formula, called a Gaussian quadrature formula, becomes exact on the space of polynomials of degree at most $2k-1$ (instead of $k-1$). Thus, in order to obtain an approximation of $f(z)$, let us replace $1/(1-xz)$ by its (Hermite) interpolation polynomial and then apply the functional c (which is analogous to integration). We have

Theorem 1.3 Let $v_k(x) = (x-x_1)^{k_1}\cdots(x-x_n)^{k_n}$, where $x_1,\ldots,x_n$ are distinct points in the complex plane and $k = k_1+\cdots+k_n$. The polynomial

$$R_k(x) = \frac{1}{1-xz}\left(1-\frac{v_k(x)}{v_k(z^{-1})}\right)$$

is the Hermite interpolation polynomial of degree $k-1$ of $(1-xz)^{-1}$, that is, the polynomial such that

$$R_k^{(j)}(x_i) = \frac{d^j}{dx^j}(1-xz)^{-1}\Big|_{x=x_i}$$

for $i=1,\ldots,n$ and $j=0,\ldots,k_i-1$.

The proof of this result was given by Brezinski (1983a). Let us now apply the functional c to R_k in order to obtain an approximation of $f(z)$. We have

$$c(R_k(x)) = \frac{1}{v_k(z^{-1})}c\left(\frac{v_k(z^{-1})-v_k(x)}{1-xz}\right).$$

Setting

$$w_k(z) = c\left(\frac{v_k(z)-v_k(x)}{z-x}\right),$$

where c acts on x and z is a parameter, it is easy to see that w_k is a polynomial of degree $k-1$ in z and that

$$c(R_k(x)) = \tilde{w}_k(z)/\tilde{v}_k(z),$$

where $\tilde{w}_k(z) = z^{k-1}w_k(z^{-1})$ and $\tilde{v}_k(z) = z^k v_k(z^{-1})$. Thus $c(R_k(x))$ is a rational function whose numerator has degree $k-1$ at most and whose denominator has degree k at most. Moreover

$$\begin{aligned} c(R_k(x)) &= c\left(\frac{1}{1-xz}\right) - \frac{z^k}{\tilde{v}_k(z)} c\left(\frac{v_k(x)}{1-xz}\right) \\ &= f(z) + \mathcal{O}\left(z^k\right). \end{aligned}$$

This property is quite similar to the property of interpolatory quadrature formulae to be exact on the space of polynomials of degree at most $k-1$. Thus $c(R_k(x))$ appears as a generalization of such formulae for the function $(1-xz)^{-1}$. Such rational functions, whose poles (the zeros of $\tilde{v}_k$) are arbitrarily chosen, are called Padé-type approximants of f. They will be denoted by $(p/q)_f(z)$. They generalize the Padé approximants. They have interesting properties and will be studied in Subsection 3.1.

From the above formula for the error, we have

$$c(R_k(x)) = f(z) - \frac{z^k}{\tilde{v}_k(z)} c\left(v_k(x)\left(1 + xz + \cdots + x^{k-1}z^{k-1} + \frac{x^k z^k}{1-xz}\right)\right).$$

The polynomial v_k, called the generating polynomial of the Padé-type approximant $(k-1/k)$, can be arbitrarily chosen and we have k degrees of freedom (its k zeros or k among its $k+1$ coefficients since the numerator and the denominator of a rational function are uniquely defined apart from a multiplying factor). Thus, let us take v_k such that

$$c(x^i v_k(x)) = 0 \quad \text{for } i = 0, \ldots, k-1.$$

In that case we shall have

$$c(R_k(x)) = f(z) + \mathcal{O}\left(z^{2k}\right).$$

But $2k = (k-1)+k+1$ which shows that $c(R_k(x))$ matches the original series f up to the degree of the numerator plus the degree of the denominator. Thus $c(R_k(x))$ is the Padé approximant $[k-1/k]$ of f. It can be understood as a generalization of Gaussian quadrature formulae for the function $(1-xz)^{-1}$ since it is exact on the space of polynomials of degree at most $2k-1$.

The relations $c(x^i v_k(x)) = 0$ for $i = 0, \ldots, k-1$ show that v_k is the polynomial of degree k belonging to the family of (formal) orthogonal polynomials with respect to the functional c. In that case v_k will be denoted by P_k. Thus formal orthogonal polynomials appear in a very natural way in the theory of Padé approximants. They form the basis of their algebraic study and lead to recurrence relationships for their computation. These questions will be studied in Subsection 1.5.

Moreover, by construction, we have the following error formula

$$f(z) - [k-1/k]_f(z) = \frac{z^{2k}}{\tilde{P}_k(z)} c\left(\frac{x^k P_k(x)}{1-xz}\right).$$

But $(P_k(x) - P_k(z^{-1}))/(1-xz)$ is a polynomial of degree $k-1$ in x and, due to the orthogonality relations of P_k,

$$c\left(\frac{P_k(x) - P_k(z^{-1})}{1-xz} P_k(x)\right) = 0 = c\left(\frac{P_k^2(x)}{1-xz}\right) - P_k(z^{-1}) c\left(\frac{P_k(x)}{1-xz}\right)$$

and thus

$$f(z) - [k-1/k]_f(z) = \frac{z^{2k}}{\tilde{P}_k^2(z)} c\left(\frac{P_k^2(x)}{1-xz}\right).$$

These expressions are useful for estimating the error in Padé approximation. It is easy to see that

$$f(z) - [k-1/k]_f(z) = \frac{z^{2k}}{\tilde{P}_k(z)} \sum_{i=0}^{\infty} d_{i+k} z^i$$

with $d_i = c(x^i P_k(x)) = b_0 c_i + b_1 c_{i+1} + \cdots + b_k c_{i+k}$ and $P_k(x) = b_0 + \cdots + b_k x^k$.

Obviously, by the orthogonality property of P_k, $d_i = 0$ for $i = 0, \ldots, k-1$.

1.2. Algebraic properties

In this subsection we shall give some algebraic properties of the Padé approximants. The first one is a determinantal formula which was obtained by Jacobi in 1846 using a determinantal formula due to Cauchy for interpolating rational functions.

We set

$$\begin{aligned} f_k(z) &= \sum_{i=0}^{k} c_i z^i \quad \text{for } k \geq 0, \\ &= 0 \quad \text{for } k < 0. \end{aligned}$$

Then, the following property holds:

Property 1.1

$$[p/q]_f(z) = \frac{\begin{vmatrix} z^q f_{p-q}(z) & z^{q-1} f_{p-q+1}(z) & \cdots & f_p(z) \\ c_{p-q+1} & c_{p-q+2} & \cdots & c_{p+1} \\ \vdots & \vdots & & \vdots \\ c_p & c_{p+1} & \cdots & c_{p+q} \end{vmatrix}}{\begin{vmatrix} z^q & z^{q-1} & \cdots & 1 \\ c_{p-q+1} & c_{p-q+2} & \cdots & c_{p+1} \\ \vdots & \vdots & & \vdots \\ c_p & c_{p+1} & \cdots & c_{p+q} \end{vmatrix}}.$$

Let us now assume that $f(0) = c_0 \neq 0$ and let g be the reciprocal series of f formally defined by

$$f(z)g(z) = 1.$$

Setting $g(z) = d_0 + d_1 z + d_2 z^2 + \cdots$ we have

$$\begin{aligned} &c_0 d_0 = 1, \\ &c_0 d_i + c_1 d_{i-1} + \cdots + c_i d_0 = 0, \qquad i \geq 1. \end{aligned}$$

Then we have

Property 1.2
$\forall p, q, [p/q]_f(z)[q/p]_g(z) = 1.$

This property is very useful since it relates the two halves of the Padé table.

The other algebraic properties deal with transformations of the variable and of the series. They have been gathered in the two following properties:

Property 1.3

1. Let $g(z) = f(az), a \neq 0$. Then $[p/q]_g(z) = [p/q]_f(az)$.
2. Let $g(z) = f(z^k), k > 0$. Then, $\forall i, j$ such that $i + j \leq k - 1, [pk + i/qk + j]_g(z) = [p/q]_f(z^k)$.
3. Let $T(z) = Az^k/R(z), A \neq 0$, with R a polynomial of the degree $k > 0$ such that $R(0) \neq 0$. Let $g(z) = f(T(z))$. Then, $\forall i, j$ such that $i + j \leq k - 1, [pk + i/qk + j]_g(z) = [p/q]_f(T(z))$.

Property 1.4

1. Let $g(z) = z^k f(z)$. Then $[p + k/q]_g(z) = z^k [p/q]_f(z)$.
2. If $c_0 = \cdots = c_{k-1} = 0$ and $c_k \neq 0$ and if we set $g(z) = z^{-k} f(z)$ then $[p/q]_g(z) = z^{-k}[p + k/q]_f(z)$.
3. Let R be a polynomial of degree k. If $p \geq q + k$ then $[p/q]_{f+R}(z) = [p/q]_f(z) + R(z)$.

4. Let $g(z) = (A + Bf(z))/(C + Df(z))$ with $C + Dc_0 \neq 0$. Then

$$[p/p]_g(z) = \frac{A + B[p/p]_f(z)}{C + D[p/p]_f(z)}.$$

5. Let $g(z) = af(z), a \neq 0$. Then $[p/q]_g(z) = a[p/q]_f(z)$.

An important property is that of consistency.

Property 1.5 Let f be the power-series expansion of a rational function with a numerator of degree p and a denominator of degree q. Then $\forall i, j \geq 0, [p + i/q + j]_f(z) \equiv f(z)$.

A useful formula is the so-called Nuttall compact formula, obtained by Nuttall (1967). A generalization of it is

Property 1.6 Let $\{q_n\}$ be an arbitrary family of polynomials such that $\forall n, q_n$ has the exact degree n. Let V be the $k \times k$ matrix with elements $v_{ij} = c((1 - xz)q_{i-1}(x)q_{j-1}(x))$ for $i, j = 1, \ldots, k$, let v' be the vector with components $v'_i = c(q_{i-1}(x)(1 - v_k(x)/v_k(z^{-1})))$ for $i = 1, \ldots, k$ and let u be the vector with components $u_i = c(q_{i-1}(x))$ for $i = 1, \ldots, k$. Then

$$(k - 1/k)_f(z) = (u, V^{-1}v'),$$

where v_k is the generating polynomial of $(k - 1/k)$. If $v_k \equiv P_k$ then

$$[k - 1/k]_f(z) = (u, V^{-1}v').$$

If $q_n(x) = x^n$, then $v_{ij} = c_{i+j-2} - zc_{i+j-1}, u_i = c_{i-1}$ and the formula for $[k - 1/k]$ exactly reduces to Nuttall's. Since $(k - 1/k)$ only depends on $c_0, \ldots, c_{k-1}$ then, in the preceding formula, $c_k, \ldots, c_{2k-1}$ can be arbitrarily chosen. In particular they can be set to zero. If $q_n(x) = P_n(x)$, the preceding extension of Nuttall's formula is closely related to the matrix interpretation of Padé approximants; see Gragg (1972).

1.3. Formal orthogonal polynomials

As seen in Subsection 1.1, Padé approximants are based on formal orthogonal polynomials. Thus we shall now digress to treat this subject. The other approximants of the table are related to other families of orthogonal polynomials, called adjacent families of orthogonal polynomials, which will be studied in Subsection 1.4.

Let c be the linear functional on the space of complex polynomials defined by its moments c_i

$$c(x^i) = c_i, \qquad i \geq 0.$$

Let $\{P_k\}$ be a family of polynomials. $\{P_k\}$ is said to be the family of formal orthogonal polynomials with respect to c if, $\forall k \geq 0$,

1. P_k has the exact degree k;
2. $c(x^i P_k(x)) = 0$ for $i = 0, \ldots, k-1$.

Conditions (2) are the so-called orthogonality relations. They are equivalent to the condition $c(p(x)P_k(x)) = 0$ for any polynomial p of degree $k-1$ at most or to $c(P_n(x)P_k(x)) = 0, \forall n \neq k$. The usual orthogonal polynomials (that is, those orthogonal with respect to $c(\cdot) = \int_a^b (\cdot)\, d\alpha(x)$ with α bounded and nondecreassing in $[a,b]$) are known to satisfy a bunch of interesting properties such as a three-term recurrence relation, the Christoffel–Darboux identity, properties of their zeros, etc. Most of these properties still hold for formal orthogonal polynomials. However, in that case, the first question is that of existence. Let us write P_k as

$$P_k(x) = a_0 + a_1 x + \cdots + a_k x^k.$$

Then the orthogonality relations are equivalent to the system

$$a_0 c_i + a_1 c_{i+1} + \cdots + a_k c_{i+k} = 0, \qquad i = 0, \ldots, k-1.$$

Since P_k must have the exact degree k, a_k must be different from zero or, in other words, the Hankel determinant

$$H_k^{(0)} = \begin{vmatrix} c_0 & \cdots & c_{k-1} \\ c_1 & \cdots & c_k \\ \vdots & & \vdots \\ c_{k-1} & \cdots & c_{2k-2} \end{vmatrix}$$

must not vanish. Thus, in the sequel, we shall assume that $\forall k > 0, H_k^{(0)} \neq 0$. In that case, we shall say that the functional c is definite, a property clearly related to the normality of the Padé table. The case where the functional c is non-definite has been extensively studied by Draux (1983).

In the definite case, the polynomial P_k is uniquely determined apart from an arbitrary non-zero constant. Moreover we have the following determinantal formula

$$P_k(x) = D_k \begin{vmatrix} c_0 & c_1 & \cdots & c_k \\ c_1 & c_2 & \cdots & c_{k+1} \\ \vdots & & & \vdots \\ c_{k-1} & c_k & \cdots & c_{2k-1} \\ 1 & x & \cdots & x^k \end{vmatrix}$$

with $D_k \neq 0$ and $P_0(x) = D_0$. Let us set $P_k(x) = t_k x^k + s_k x^{k-1} + \cdots$. For a family of formal orthogonal polynomials we have

Theorem 1.4 $\forall k \geq 0$,

$$P_{k+1}(x) = (A_{k+1}x + B_{k+1})P_k(x) - C_{k+1}P_{k-1}(x)$$

with $P_{-1}(x) = 0, P_0(x) = t_0$,

$$A_{k+1} = t_{k+1}/t_k, \quad B_{k+1} = -\frac{\alpha_k t_{k+1}}{h_k t_k},$$

$$C_{k+1} = \frac{t_{k-1}t_{k+1}}{t_k^2}\frac{h_k}{h_{k-1}},$$

$$\alpha_k = c(xP_k^2(x)), h_k = c(P_k^2(x)).$$

Since P_k is determined apart from a multiplying factor, the t_k's in the preceding recurrence relation can be arbitrarily chosen and thus this relation can be used for computing recursively the P_k's. In particular the choice $t_k = 1$ leads to monic orthogonal polynomials.

The reciprocal of this theorem was first proved by Favard (1935) for the usual orthogonal polynomials. It was extended by Shohat (1938) (see also Van Rossum (1953)) to the formal case.

Theorem 1.5 Let $\{P_k\}$ be a family of polynomials such that the relation of Theorem 1.4 holds with $t_0 \neq 0$ and $\forall k, A_k C_k \neq 0$. Then $\{P_k\}$ is a family of formal orthogonal polynomials with respect to a linear functional c whose moments c_i can be computed.

Let us now define the associated polynomials Q_k by

$$Q_k(z) = c\left(\frac{P_k(x) - P_k(z)}{x - z}\right),$$

where c acts on x and where z is a parameter. It is easy to see that Q_k is a polynomial of degree $k-1$ in z, that $Q_0(z) = 0$ and that, for $k > 1$,

$$Q_k(z) = D_k \begin{vmatrix} c_0 & c_1 & c_2 & \cdots & c_k \\ \vdots & \vdots & \vdots & & \vdots \\ c_{k-1} & c_k & c_{k+1} & \cdots & c_{2k-1} \\ 0 & c_0 & c_0 z + c_1 & \cdots & (c_0 z^{k-1} + c_1 z^{k-2} + \cdots + c_{k-1}) \end{vmatrix}.$$

Theorem 1.6 The family $\{Q_k\}$ satisfies the three-term recurrence relation of Theorem 1.4 with $Q_{-1}(x) = -1, Q_0(x) = 0$ and $C_1 = A_1 c(P_0(x))$. Moreover, $\forall k \geq 0$

$$P_k(x)Q_{k+1}(x) - Q_k(x)P_{k+1}(x) = A_{k+1}h_k.$$

Some other relations satisfied by the P_k's and the Q_k's are given in the definite case by Brezinski (1980, chapter 2).

It follows from Theorem 1.6 that

$$[k/k+1]_f(z) = [k-1/k]_f(z) + \frac{A_{k+1}h_k}{\tilde{P}_k(z)\tilde{P}_{k+1}(z)} z^{2k},$$

a relation known as the Euler–Minding identity and which follows directly from the theory of continued fractions.

All these relations have been extended to the non-definite case by Draux (1983).

The zeros of the classical orthogonal polynomials are known to possess some properties. Not all of them extend to the formal case. In particular the zeros of formal orthogonal polynomials need not be simple or real. However, we have

Theorem 1.7 If c is definite, then $\forall k \geq 0$

1. P_k and P_{k+1} have no common zero,
2. Q_k and Q_{k+1} have no common zero,
3. P_k and Q_k have no common zero.

To end this subsection let us mention that a matrix formalism of orthogonality can be given via tridiagonal matrices. Orthogonal polynomials are known to play an important rôle in numerical analysis. In particular they are closely connected with projection methods used in the theory of linear operators, for example, with the method of moments, Lanczos's method and the conjugate gradient algorithms (see Section 4). All these connections were reviewed by Brezinski (1980, section 2.7, 1994); see also the works of Gutknecht (1990, 1992).

The notion of orthogonality studied in this subsection is a particular case of the more general notion of biorthogonality between a family of elements of a vector space and a family of elements of its dual. The notion of biorthogonality was extensively studied in Brezinski (1991b).

1.4. Adjacent families of orthogonal polynomials

Let us now define the linear functionals $c^{(n)}$ by

$$c^{(n)}(x^i) = c_{n+i}.$$

With the same convention as above, namely, that $c_i = 0$ for $i < 0$, these linear functionals $c^{(n)}$ can be defined even for negative values of the upper index n.

Let us denote by $\{P_k^{(n)}\}$ the family of formal orthogonal polynomials with respect to $c^{(n)}$. The family $\{P_k\}$ studied in Subsection 1.3 corresponds to $n = 0$. Such families are called adjacent families of orthogonal polynomials. They satisfy the same properties as above after replacing c by $c^{(n)}$ or, in other words, the sequence $c_0, c_1, \ldots$ by the sequence $c_n, c_{n+1}, \ldots$. In particular

$\{P_k^{(n)}\}$ exists only if $\forall k, n$

$$H_k^{(n)} = \begin{vmatrix} c_n & \cdots & c_{n+k-1} \\ \vdots & & \vdots \\ c_{n+k-1} & \cdots & c_{n+2k-2} \end{vmatrix} \neq 0.$$

In that case we shall say that the linear functional c is completely definite. For the non–completely definite case, we again refer the interested reader to Draux (1983).

The polynomials $P_k^{(n)}$ are usually placed in a double-entry table similar to the Padé table

$$\begin{array}{ccccc} P_{-1}^{(0)} & P_0^{(-1)} & P_1^{(-2)} & P_2^{(-3)} & \cdots \\ P_{-1}^{(1)} & P_0^{(0)} & P_1^{(-1)} & P_2^{(-2)} & \cdots \\ P_{-1}^{(2)} & P_0^{(1)} & P_1^{(0)} & P_2^{(-1)} & \cdots \\ P_{-1}^{(3)} & P_0^{(2)} & P_1^{(1)} & P_2^{(0)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

Many relationships exist between adjacent polynomials of this table. First of all, each family of orthogonal polynomials satisfies a three-term recurrence relation similar to that of Theorem 1.4. Assuming all the polynomials to be monic we shall write this relation as

$$P_{-1}^{(n)}(x) = 0, \qquad P_0^{(n)}(x) = 1,$$

$$P_{k+1}^{(n)}(x) = (x - q_{k+1}^{(n)} - e_k^{(n)})P_k^{(n)}(x) - q_k^{(n)} e_k^{(n)} P_{k-1}^{(n)}(x). \qquad \begin{array}{ccc} \bullet & & \\ & \bullet & \\ & & * \end{array}$$

A $\bullet$ indicates the position of a polynomial that is known in the table, while a $*$ indicates the position of the polynomial that is computed by the relation.

We also have

$$P_k^{(n+1)}(x) = P_k^{(n)}(x) - e_k^{(n)} P_{k-1}^{(n+1)}(x), \qquad \begin{array}{cc} \bullet & \bullet \\ & * \end{array}$$

$$P_{k+1}^{(n)}(x) = xP_k^{(n+1)}(x) - q_{k+1}^{(n)} P_k^{(n)}(x). \qquad \begin{array}{cc} \bullet & \\ \bullet & * \end{array}$$

Using alternatively these two relations allows us to compute recursively the two adjacent families $\{P_k^{(n)}\}$ and $\{P_k^{(n+1)}\}$.

It can be proved (see, for example, Brezinski (1980, section 2.8) where all these relations and the following ones are given) that the numbers $e_k^{(n)}$ and $q_k^{(n)}$ are related by

$$e_0^{(n)} = 0, \qquad q_1^{(n)} = c_{n+1}/c_n,$$

$$
\begin{aligned}
q_{k+1}^{(n)} + e_{k+1}^{(n)} &= q_{k+1}^{(n+1)} + e_k^{(n+1)}, \\
e_k^{(n)} q_{k+1}^{(n)} &= e_k^{(n+1)} q_k^{(n+1)}
\end{aligned}
$$

which is the so-called Qd algorithm. This algorithm can be used for their recursive computation. It is due to Rutishauser (1954) (see also Henrici (1974)) and was the basis for the development of the LR-algorithm for the computation of the eigenvalues of a matrix. Setting $x = 0$ in the preceding relations, it is easy to see that

$$
\begin{aligned}
q_k^{(n)} &= H_k^{(n+1)} H_{k-1}^{(n)} / H_k^{(n)} H_{k-1}^{(n+1)}, \\
e_k^{(n)} &= H_{k-1}^{(n+1)} H_{k+1}^{(n)} / H_k^{(n)} H_k^{(n+1)}.
\end{aligned}
$$

From these determinantal expressions and from the three preceding relations we can obtain the following ones

$$
\begin{aligned}
H_k^{(n+2)} H_k^{(n)} P_k^{(n)}(x) &= [H_k^{(n+1)}]^2 P_k^{(n+1)}(x) + H_{k+1}^{(n)} H_{k-1}^{(n+2)} P_{k-1}^{(n+2)}(x), \\
H_k^{(n+1)} H_{k+1}^{(n-1)} P_{k+1}^{(n-1)}(x) &= x H_{k+1}^{(n-1)} H_k^{(n+1)} P_k^{(n+1)}(x) - H_{k+1}^{(n)} H_k^{(n)} P_k^{(n)}(x),
\end{aligned}
$$

$$
\begin{aligned}
x H_k^{(n)} H_{k+1}^{(n-1)} H_k^{(n+1)} P_k^{(n+1)}(x) &= [x H_k^{(n+1)} H_{k+1}^{(n-1)} + H_{k+1}^{(n)} H_k^{(n)}] H_k^{(n)} P_k^{(n)}(x) \\
&\quad - H_{k+1}^{(n)} H_k^{(n+1)} H_k^{(n-1)} P_k^{(n-1)}(x), \\
H_k^{(n)} H_k^{(n+1)} H_{k+1}^{(n-1)} P_{k+1}^{(n-1)}(x) &= [x H_{k+1}^{(n-1)} H_k^{(n+1)} - H_k^{(n)} H_{k+1}^{(n)}] H_k^{(n)} P_k^{(n)}(x) \\
&\quad - x H_{k+1}^{(n-1)} H_{k+1}^{(n)} H_{k-1}^{(n+1)} P_{k-1}^{(n+1)}(x).
\end{aligned}
$$

Combining these relations leads to many other ones. However, the preceding eight relations are sufficient to follow any path in the table of the adjacent families of orthogonal polynomials. Of course similar relations hold among the associated polynomials

$$
Q_k^{(n)}(z) = c^{(n)} \left(\frac{P_k^{(n)}(x) - P_k^{(n)}(z)}{x - z} \right).
$$

1.5. Recursive computation of Padé approximants

Let us first relate all the approximants of the Padé table to the adjacent families of orthogonal polynomials defined in the Subsection 1.4. Thus the recurrence relations given there will provide recursive methods for computing any sequence of Padé approximants.

In Subsection 1.1, we saw that

$$
[k-1/k]_f(z) = \tilde{Q}_k^{(0)}(z) / \tilde{P}_k^{(0)}(z).
$$

Making use of the convention that a sum with a negative upper index is equal to zero, it is easy to see that

Theorem 1.8 $\forall k \geq 0, \forall n \geq -k$

$$[n+k/k]_f(z) = \sum_{i=0}^{n} c_i z^i + z^{n+1} \tilde{Q}_k^{(n+1)}(z) / \tilde{P}_k^{(n+1)}(z)$$

with $\tilde{P}_k^{(n+1)}(z) = z^k P_k^{(n+1)}(z^{-1})$ and $\tilde{Q}_k^{(n+1)}(z) = z^{k-1} Q_k^{(n+1)}(z^{-1})$.

Let us set

$$[n+k/k]_f(z) = \tilde{N}_k^{(n+1)}(z) / \tilde{P}_k^{(n+1)}(z)$$

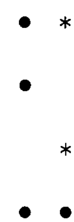
and

$$\begin{aligned} \tilde{P}_k^{(n)}(z) &= \sum_{i=0}^{k} b_i^{(k,n)} z^i, \\ \tilde{N}_k^{(n)}(z) &= \sum_{i=0}^{n+k-1} a_i^{(k,n)} z^i. \end{aligned}$$

The relations of the preceding subsection give

$$\frac{\tilde{N}_{k+1}^{(n)}(z)}{\tilde{P}_{k+1}^{(n)}(z)} = \frac{a_{n+k}^{(k,n+1)} \tilde{N}_k^{(n+2)}(z) - z a_{n+k+1}^{(k,n+2)} \tilde{N}_k^{(n+1)}(z)}{a_{n+k}^{(k,n+1)} \tilde{P}_k^{(n+2)}(z) - z a_{n+k+1}^{(k,n+2)} \tilde{P}_k^{(n+1)}(z)},$$

$$\frac{\tilde{N}_k^{(n)}(z)}{\tilde{P}_k^{(n)}(z)} = \frac{a_{n+k}^{(k-1,n+2)} \tilde{N}_k^{(n+1)}(z) - a_{n+k}^{(k,n+1)} \tilde{N}_{k-1}^{(n+2)}(z)}{a_{n+k}^{(k-1,n+2)} \tilde{P}_k^{(n+1)}(z) - a_{n+k}^{(k,n+1)} \tilde{P}_{k-1}^{(n+2)}(z)}.$$



These two relations are identical with a method due to Longman (1971) for computing recursively approximants located on an ascending staircase of the Padé table. They also cover an algorithm due to Baker (1970).

We have

$$\frac{\tilde{N}_k^{(n+1)}(z)}{\tilde{P}_k^{(n+1)}(z)} = \frac{\tilde{N}_k^{(n)}(z) - e_k^{(n)} z \tilde{N}_{k-1}^{(n+1)}(z)}{\tilde{P}_k^{(n)}(z) - e_k^{(n)} z \tilde{P}_{k-1}^{(n+1)}(z)},$$

$$\frac{\tilde{N}_{k+1}^{(n)}(z)}{\tilde{P}_{k+1}^{(n)}(z)} = \frac{\tilde{N}_k^{(n+1)}(z) - q_{k+1}^{(n)} z \tilde{N}_k^{(n)}(z)}{\tilde{P}_k^{(n+1)}(z) - q_{k+1}^{(n)} z \tilde{P}_k^{(n)}(z)}$$

with $e_k^{(n)} = h_k^{(n)} / h_{k-1}^{(n+1)}, q_{k+1}^{(n)} = h_k^{(n+1)} / h_k^{(n)}$ and $h_k^{(n)} = c^{(n)}(P_k^{(n)^2}(x)) = \sum_{i=0}^{k} c_{n+k+i} b_{k-i}^{(k,n)}$. These two relations are identical to a method due to Watson (1973) to compute recursively approximants located on a descending staircase of the Padé table.

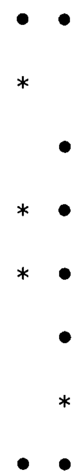
We also have

$$\frac{\tilde{N}_k^{(n+2)}(z)}{\tilde{P}_k^{(n+2)}(z)} = \frac{b_k^{(k,n+1)}\tilde{N}_{k+1}^{(n)}(z) - zb_{k+1}^{(k+1,n)}\tilde{N}_k^{(n+1)}(z)}{b_k^{(k,n+1)}\tilde{P}_{k+1}^{(n)}(z) - zb_{k+1}^{(k+1,n)}\tilde{P}_k^{(n+1)}(z)},$$

$$\frac{\tilde{N}_{k-1}^{(n+2)}(z)}{\tilde{P}_{k-1}^{(n+2)}(z)} = \frac{b_k^{(k,n+1)}\tilde{N}_k^{(n)}(z) - b_k^{(k,n)}\tilde{N}_k^{(n+1)}(z)}{b_k^{(k,n+1)}\tilde{P}_k^{(n)}(z) - b_k^{(k,n)}\tilde{P}_k^{(n+1)}(z)},$$

$$\frac{\tilde{N}_{k-1}^{(n+1)}(z)}{\tilde{P}_{k-1}^{(n+1)}(z)} = \frac{\tilde{N}_k^{(n)}(z) - \tilde{N}_k^{(n+1)}(z)}{\tilde{P}_k^{(n)}(z) - \tilde{P}_k^{(n+1)}(z)},$$

$$\frac{\tilde{N}_k^{(n)}(z)}{\tilde{P}_k^{(n)}(z)} = \frac{\tilde{N}_k^{(n+1)}(z) - \tilde{N}_{k+1}^{(n)}(z)}{\tilde{P}_k^{(n+1)}(z) - \tilde{P}_{k+1}^{(n)}(z)}.$$



Combining these relations together allows one to obtain all the other possible ones. Eight of them were used in a conversational program to compute recursively any sequence of Padé approximants in the normal case; see Brezinski (1980, appendix).

All these recurrence relations were extended by Draux (1983) to the non-normal case. A universal conversational program for computing any sequence of Padé approximants in the non-normal case was given by Draux and Van Ingelandt (1986). In order to avoid numerical instability and also for the detection of the block structure of the Padé table, it was necessary to program these relations in exact arithmetic that is in rational arithmetic coded on several words. All these programs are written in FORTRAN.

Setting for simplicity

$$[n+k-1/k] = N,$$

$$[n+k/k-1] = W, \qquad [n+k/k] = C, \qquad [n+k/k+1] = E,$$

$$[n+k+1/k] = S,$$

we obtain, after elimination among the preceding identities, the so-called cross rule of Wynn (1966)

$$(N-C)^{-1} + (S-C)^{-1} = (W-C)^{-1} + (E-C)^{-1}.$$

When a square block of size m occurs in the Padé table, this cross rule was extended by Cordellier (1979) who proved that

$$(N_i - C)^{-1} + (S_i - C)^{-1} = (W_i - C)^{-1} + (E_i - C)^{-1}$$

for $i = 0, \ldots, m$, where the N_i's and the W_i's are numbered from the upper left corner of the block and the S_i's and the E_i's from its lower right one. The initial values are

$$[-1/q]_f(z) = 0, \qquad [p/-1]_f(z) = \infty,$$

$$[p/0]_f(z) = \sum_{i=0}^{p} c_i z^i, \qquad [0/q]_f(z) = ([q/0]_g(z))^{-1} = \left(\sum_{i=0}^{q} d_i z^i\right)^{-1},$$

where the d_i's are the coefficients of the reciprocal series g of f.

Thus, the theory of formal orthogonal polynomials provides a basis for rediscovering known recursive methods for the computation of sequences of Padé approximants which were found more or less heuristically by their authors. This theory also gives us the possibility of computing any sequence of approximants by new recursive algorithms. It has been possible to extend the theory to the non-normal case, thus leading for the first time to all the possible recurrence relationships among the entries of a non-normal Padé table and to write the only existing complete subroutine.

1.6. The ε-algorithm

We shall now deal with a subject which, a priori, has nothing to do with Padé approximation but which is, in fact, closely related to it: convergence acceleration.

Let (S_n) be a sequence converging to S. If the convergence is slow, one can try to accelerate it. For that purpose, we shall transform the sequence (S_n) into another sequence (T_n) such that, if possible, (T_n) converges to S faster than (S_n), that is,

$$\lim_{n\to\infty} (T_n - S)/(S_n - S) = 0.$$

One of the most popular sequence transformations for that purpose is certainly Aitken's Δ^2 process which corresponds to

$$T_n = S_n - (\Delta S_n)^2/\Delta^2 S_n \quad \text{for } n = 0, 1, \ldots.$$

If the sequence (S_n) is such that $\exists a \neq 1, \lim_{n\to\infty}(S_{n+1} - S)/(S_n - S) = a$ then (T_n) obtained by Aitken's process converges to S faster than (S_n).

In 1955, Shanks (1955) gave a generalization of Aitken's process. He considered the various transformations $e_k : (S_n) \to (e_k(S_n))$ where

$$e_k(S_n) = \frac{\begin{vmatrix} S_n & \cdots & S_{n+k} \\ \vdots & & \vdots \\ S_{n+k} & \cdots & S_{n+2k} \end{vmatrix}}{\begin{vmatrix} \Delta^2 S_n & \cdots & \Delta^2 S_{n+k-1} \\ \vdots & & \vdots \\ \Delta^2 S_{n+k-1} & \cdots & \Delta^2 S_{n+2k-2} \end{vmatrix}}.$$

A recursive algorithm to compute the $e_k(S_n)$'s without computing the determinants involved in their definition was found one year later by Wynn (1956). It is the ε-algorithm whose rules are

$$\varepsilon_{-1}^{(n)} = 0, \qquad \varepsilon_0^{(n)} = S_n, \quad n = 0, 1, \ldots,$$

$$\varepsilon_{k+1}^{(n)} = \varepsilon_{k-1}^{(n+1)} + \left[\varepsilon_k^{(n+1)} - \varepsilon_k^{(n)}\right]^{-1}, \qquad n, k = 0, 1, \ldots.$$

It is related to Shanks's transformation by

$$\varepsilon_{2k}^{(n)} = e_k(S_n).$$

The $\varepsilon_{2k+1}^{(n)}$'s are only intermediate quantities. The ε-algorithm is a quite powerful acceleration process which has been widely studied. For its theory one can consult, for example, Brezinski (1977). Subroutines and applications can be found in Brezinski and Redivo-Zaglia (1991).

The ε-algorithm is related to Padé approximants in the following way: if it is applied to the partial sums of the series f, that is, if

$$S_n = \sum_{i=0}^{n} c_i z^i, \qquad n, = 0, 1, \ldots,$$

then

$$\varepsilon_{2k}^{(n)} = [n + k/k]_f(z).$$

Thus the ε-algorithm can be used to compute recursively the lower half of the Padé table. The upper half of the Padé table can be computed by applying the ε-algorithm to the partial sums of the reciprocal series g of f as stated in Property 1.2. Let us mention that the elimination of the ε's with an odd lower index leads to Wynn's cross rule mentioned in the preceding subsection.

2. Convergence

2.1. Introduction

More complete results about convergence can be found in Brezinski and Van Iseghem (1994) or Baker and Graves-Morris (1981).

The problem of convergence of Padé approximants, which means the convergence of a sequence of Padé approximants when at least one of the degrees tends to infinity, is a difficult problem which can be studied from different points of view. The first one is to study all the abilities of convergence to one function and the first theorem (due to Padé) is an example of such a study for the exponential function. The history of numbers such as e or π tells us, through the link with continued fractions, that it is also possible to do so for functions such as $\tan x$, $\arctan x$ and some others. At the other end, it is possible to hope for convergence results for a whole class of functions. The most useful and well known examples are the meromorphic functions with a fixed number of poles in a disc (convergence of the columns) and the Stieltjes functions, for which the classical uniform convergence on compact subsets of $\mathbb{C}$ can be proved for the diagonal and paradiagonal sequences.

In each case, the problems are different if the sequences considered are in a column, a diagonal or a paradiagonal; close to these cases are, for example, sectorial sequences where m/n has lower and upper bounds as n goes to infinity. Other cases could also be considered.

We will first quote two results which show the most optimistic result that can be expected and a counterexample which limits our ambitions.

Theorem 2.1 For any sequence $(m_i, n_i), i \geq 1$, where $m_i + n_i$ tends to infinity, the poles of the Padé approximants of e^z tend to infinity and

$$\lim_{i \to \infty} [m_i/n_i](z) = \mathrm{e}^z$$

uniformly on any compact set of $\mathbb{C}$.

The following result is due to Wallin (1974).

Theorem 2.2 There exists an entire function f such that the sequence of diagonal Padé approximants $([n/n]_f)$ is unbounded at every point of the complex plane except zero, and so no convergence result can be expected in any open set of the plane.

As we shall see below, the location of the poles of the approximants is of primary importance for studying convergence: for meromorphic functions, they are supposed to be known, and so Montessus de Ballore's theorem is obtained. For Stieltjes functions, the link with orthogonal polynomials is extensively used; they are, in that case, defined by a positive-definite functional, and so properties about the zeros are known.

As a consequence, functions with branch points, for example, are outside of our study, and convergence will be obtained on extremal subsets of $\mathbb{C}$ that localize the set of zeros and poles of the Padé approximants as a barrier to convergence.

Let us now give a very simple example showing the difficulties related with the convergence of Padé approximants (that is, the convergence of a sequence of approximants). We consider the series given by Bender and Orszag (1978)

$$f(z) = \frac{10 + z}{1 - z^2} = \sum_{i=0}^{\infty} c_i z^i$$

with $c_{2i} = 10$ and $c_{2i+1} = 1$. It converges for $|z| < 1$. We have

$$[k/1]_f(z) = \sum_{i=0}^{k-1} c_i z^i + \frac{c_k z^k}{1 - c_{k+1} z / c_k}.$$

When k is odd, $[k/1]$ has a simple pole at $z = 1/10$ while f has no pole. Thus the sequence $([k/1])$ cannot converge to f in $|z| < 1$.

This example shows that the poles of the Padé approximants can prevent convergence and that a sequence of approximants can be non-convergent in a domain where the series is. In order to prove the convergence of a sequence of Padé approximants in a domain D of the complex plane, it must be proved that the spurious poles of the approximants (that is, the poles that do not approximate poles of f) move out of D when the degree(s) of the approximants tends to infinity.

Another more paradoxical situation can arise: the zeros of the Padé approximants can also prevent convergence. Let us take the reciprocal series g of the series f of the preceding example

$$g(z) = \frac{1 - z^2}{10 + z}.$$

It converges in $|z| < 10$. We have

$$[1/k]_g(z) = 1/[k/1]_f(z).$$

Since $[1/2k + 1]_g(0.1) = 0$ and $g(0.1) \neq 0$ the sequence $([1/k]_g)$ cannot converge in $|z| < 10$ where the series g does.

Another counterexample is due to Perron (1957). Let an arbitrary sequence (z_n) of points of $\mathbb{C}$ be given, and let us define the following function

$$f(z) = \sum_{i=0}^{\infty} c_i z^i,$$

$$\begin{aligned} \text{if } |z_n| \leq 1, \quad & c_{3n} = z_n/(3n+2)!, \\ & c_{3n+1} = c_{3n+2} = 1/(3n+2)!, \\ \text{if } |z_n| > 1, \quad & c_{3n} = c_{3n+1} = 1/(3n+2)!, \\ & c_{3n+2} = z_n^{-1}/(3n+2)!. \end{aligned}$$

We have $|c_i| < 1/i, \forall i \geq 0$. Thus f is an entire function and either $[3n/1]$ or $[3n+1/1]$ has a pole at z_n. The sequence (z_n) is a subsequence of the poles of $([m/1])$ and if (z_n) is dense in $\mathbb{C}$, the sequence $([m/1])$ cannot converge in any open set of the complex plane.

2.2. Meromorphic functions

Let us first consider the convergence of the columns $([m/n])_m$ of the Padé table.

The most famous theorem is that of de Montessus de Ballore (1902). It is concerned with meromorphic functions with a fixed known number of poles in the disc of radius R centered at the origin.

An extension of this result has been given by Saff (1972) for the case of interpolating rational functions instead of Padé approximants. Montessus's theorem is then a particular case of it when all the interpolation points coincide at zero.

Theorem 2.3 (Montessus de Ballore's theorem). Let f be analytic at $z = 0$ and meromorphic with exactly n poles $\alpha_1, \ldots, \alpha_n$, counted with their multiplicities, in the disc $D_R = \{z, |z| < R\}$. Let D be the domain $D_R - \{\alpha_i\}_{i=1,\ldots,n}$.

The sequence $([m/n])_{m\geq 0}$ converges to f uniformly on every compact subset of D. The poles of $[m/n]$ approach the poles of f as m tends to infinity.

This result is optimal since, if there is a pole on the boundary of D_R, then divergence of the sequence $R_{m,n}$ occurs outside $\mathbb{C} - D$ as proved by Wallin (1987).

Let us now have a look at some simple examples to illustrate the different aspects of the result. The computations have been conducted with *Mathematica* and they are given up to the first inexact digit with at most 9 digits.

First of all, the convergence to the poles is obtained with a speed of convergence that is $\mathcal{O}(r/R)$, where r is the modulus of the pole to be computed and R is the radius of the largest disc of meromorphy of the function, or $\mathcal{O}(r)$ if the function is meromorphic in the whole plane.

For the function $\sin z/(z-1)(z-2)(z-3)$, we obtain the following results, where the various columns represent the three zeros of the Padé approximant $[n/3]$ (up to the first inexact digit).

$[n/3]$			
3	0.991	2.3	−0.5
5	1.0006	1.92	0.2
7	1.000001	1.996	−4.
9	0.99999991	2.001	2.6
11	1.000000000	1.99997	3.02
13	1.000000000	2.0000004	2.9989
15	1.000000000	1.999999994	3.00003
17	1.000000000	2.000000000	2.9999990

Now, let us consider the function $\log(a + z)/(1 - z^3)$. Here $R = |a|$ and we take $a = 1.1$ and $a = 3$. The two conjugate poles are in the same column. The convergence is, of course, much better for the last two columns which correspond to $a = 3$.

$[n/3]$					
1	1.1	$-.59 \pm.71\,i$	1.01	$-.48$	$\pm.85\,i$
3	1.06	$-.52 \pm.91\,i$	1.0005	$-.501$	$\pm.865\,i$
5	1.04	$-.44 \pm.84\,i$	1.00004	$-.499992$	$\pm.8661\,i$
7	1.02	$-.52 \pm.84\,i$	1.000003	$-.499994$	$\pm.866021\,i$
9	1.01	$-.504 \pm.88\,i$	1.0000003	$-.5000005$	$\pm.8660251\,i$
11	1.01	$-.49 \pm.860\,i$	1.00000003	$-.499999993$	$\pm.86602546\,i$
13	1.007	$-.508 \pm.858\,i$	1.000000003	$-.499999995$	$\pm.866025400\,i$

Let us finally consider the function $1/\cos z$, which has an infinite number of simple poles. Although the theorem concerns the columns, this example shows that it is possible, in some cases, to obtain approximations of all the poles using the diagonals (only the positive poles are given).

$[n/n]$				
2	1.54			
4	1.57082	4.4		
6	1.570796320	4.72	6.9	
8	1.570796327	4.71231	8.0	9.0
exact	1.570796327	4.712388981	7.8539	10.9955

Montessus's theorem provides a result only if the exact number n of poles is known and only for the column sequence $([m/n])_{m\geq 0}$. The poles of f serve as *attractors* for the poles of $[m/n]$. But if f has less than n poles, then only some of the poles of $[m/n]$ are attracted and the other ones may go anywhere, destroying the convergence. If another column is considered, no result can be obtained as can be seen from counterexamples. The first one, due to Bender and Orszag (1978) and already quoted above, concerns the series $f(z) = (10+z)/(1-z^2)$ where the first column $([m/1])_{m\geq 0}$ cannot converge.

Taking into account the last counterexample (due to Perron (1957)), and coming back to a meromorphic function with n poles, it is now obvious that it is impossible to obtain a convergence result for all the sequences $([m/k])_m$ with k smaller or greater than n. It is a conjecture, made by Baker and Graves-Morris (1977), that at least a subsequence of $([m/k])_{m\geq 0}$ converges for $k \geq n$. Such a result was proved by Beardon (1968) for the column sequence $([m/1])_{m\geq 0}$:

Theorem 2.4 Let f be analytic in $|z| < R$. Then, for $r < R$, there exists a subsequence of $([m/1])_{m\geq 0}$ converging uniformly to f in the disc $|z| \leq r$.

The same result has been proved by Baker and Graves-Morris (1977) for the second and third columns. Buslaev, Gončar and Suetin (1984) established the conjecture for entire functions. For $R < \infty$ they showed that the conjecture is still true in a neighborhood of zero and they gave a counterexample for the whole disc.

2.3. Stieltjes series

The main references for this section are Baker (1975) and Baker and Graves-Morris (1981). The complete proofs are given in the last reference. A study of the subject can also be found in Karlson and Von Sydow (1976).

A Stieltjes series is a power series of the form

$$S(z) = \sum_{i=0}^{\infty} f_i(-z)^i \quad \text{with } f_i = \int_0^{\infty} x^i \, \mathrm{d}\varphi(x),$$

where φ is a positive, bounded, non-decreasing measure. The Stieltjes function associated to the Stieltjes series is

$$f(z) = \int_0^{\infty} \frac{\mathrm{d}\varphi(x)}{1+xz}.$$

So, the series S is the formal expansion of f into a power series, although this series may not converge except for $z = 0$ while the function is analytic in the cut plane $\mathbb{C} - (-\infty, 0)$. This is, for example, the case for the Euler series

$$f(z) = \int_0^{\infty} \frac{\mathrm{e}^{-t}\,\mathrm{d}t}{1+tz}, \qquad S(z) = \sum_{n=0}^{\infty} n!(-z)^n.$$

An important question is the *moment problem*, i.e. the existence and uniqueness of f corresponding to the moments f_i. If φ takes only a finite number of values, it is a step function: φ is constant on (u_i, u_{i+1}) for a finite number of u_i and so

$$f(z) = \sum_{1}^{p} \frac{\lambda_i}{1+zu_i}.$$

To avoid this too-simple case, φ is assumed in the sequel to take an infinite number of different values.

The particularity of Stieltjes series is that the special form of the coefficients f_i allows us to study the corresponding Hankel determinants and thus to locate the zeros of the orthogonal polynomials $P_n^{(m)}$ where

$$D_{mn}(z) = z^n P_n^{(m)}(1/z)$$

is the denominator of $[m+n-1/n]$.

So the most natural sequences to be considered are the paradiagonal sequences $([n+J/n])_n$, $J \geq -1$. From the first section we know that for each J the $P_n^{n+J} (\equiv P_n)$ satisfy the three-term recurrence relation

$$P_{n+1}(x) = (x - \beta_n^{n+J})P_n(x) - \gamma_n^{n+J} P_{n-1}(x),$$

$$\gamma_n^{n+J} = h_n/h_{n-1}, h_n = c^{(J-1)}(P_n^2) = c^{(J-1)}(x^n P_n) = H_{n+1}^{(J-1)}/H_n^{(J-1)}.$$

It can be proved that all the Padé approximants exist for $m \geq n-1$. So the Padé table is normal. Then, in the recurrence relation of $P_n^{(m)}$, all the γ_n^m are positive. From the theory of orthogonal polynomials, it means that each diagonal sequence $(P_n^{(m)})_n$, m fixed, is orthogonal with respect to a positive-definite functional. So $P_n^{(m)}$ has n real distinct zeros and the zeros of $P_n^{(m)}$ and $P_{n+1}^{(m)}$ interlace. We have the following theorem:

Theorem 2.5 All the zeros of $P_n^{(m)}$ are real, distinct and negative, for $m \geq n-1 \geq 0$.

Let us now consider the convergence of paradiagonal sequences. As $P_n^{(m)}$ has n simple negative zeros α_i, the denominators D_{mn} have also n simple negative zeros $1/\alpha_i$. Thus, all the poles of the Padé approximants $([n+J/n])_n$ lie on the cut and there is no obstacle to the convergence in the cut plane $\mathbb{C} - (-\infty, 0]$.

We have the following theorem:

Theorem 2.6 Let $D(\Delta, r) = \{z \in \mathbb{C}, |z| \leq r \text{ and } \forall x \leq 0, d(z,x) \geq \Delta\}$. Then, for each $J \geq -1$, the sequence $([n+J/n])_n$ converges uniformly on $D(\Delta, r)$ to a function f^J analytic in the cut plane $\mathbb{C} - (-\infty, 0]$.

If the moment problem is determinate (i.e. if there exists a measure φ such that for all i the coefficients f_i of f are given by $f_i = \int_0^\infty x^i \, \mathrm{d}\varphi(x)$), then all the f^J are identical to f. The problem is known to be determinate if the Stieltjes series has a nonzero radius of convergence R, or if $R = 0$ and the f_i's satisfy Carleman's condition: $\sum_{i\geq 1}(f_i)^{-1/2i}$ diverges.

For the Euler series, Carleman's condition is satisfied since

$$\sum_{n\geq 1}(1/n!)^{1/2n} \text{ is equivalent to } \sum_{n\geq 1}(1/n),$$

which diverges, and so the last theorem holds for the Euler function

$$f(z) = \int_0^\infty \frac{\mathrm{e}^{-t}\,\mathrm{d}t}{1+tz}.$$

For such sequences, Padé approximants can be useful for reconstructing the function from its power series expansion.

In the case of convergent Stieltjes series of radius R, the last theorem can be put into a more precise form due to Markov (1948):

Theorem 2.7 $f(z) = \displaystyle\int_0^{1/R} \frac{\mathrm{d}\varphi(u)}{1+uz}$ is analytic in the cut plane $\mathbb{C} - (-\infty, -R]$. All the poles of $[n+J/n]$ lie in $(-\infty, -R]$. The convergence of the sequence $([n+J/n])_n$ is uniform in $D^+(\Delta, r)$, where

$$D^+(\Delta, r) = \{z, |z| \leq r, \forall x \in (-\infty, -R], d(x,z) \geq \Delta\}.$$

A convergence result has also been proved by Prévost (1990) for the product of two Stieltjes functions:

Theorem 2.8 Let $f(z) = \int_0^a \frac{\mathrm{d}\alpha(x)}{1-xz}$ and $g(z) = \int_{-b}^0 \frac{\mathrm{d}\beta(x)}{1-xz}$, where a and b are finite and positive, and let α and β be positive, bounded and non-decreasing measures. Let also the integral $\int_{-b}^0 \int_0^a \frac{z}{z-x} \mathrm{d}\alpha(x)\, \mathrm{d}\beta(z)$ be assumed to exist. Then $f \cdot g(z) = \int_{-b}^a \frac{\mathrm{d}\gamma(x)}{1-xz}$ is a Stieltjes function, and the sequence $([m_k + J, m_k]_{f\cdot g})_{m_k}$ ($J \geq -1$ and $\lim_{k\to\infty} m_k = +\infty$) of Padé approximants of $f \cdot g$ converges uniformly on every compact subset of $\mathbb{C} - ((-\infty, -b^{-1}] \cup [a^{-1}, +\infty))$.

3. Generalizations

There exist many generalizations of Padé approximants. First of all, it is possible to define rational approximants to formal power series where the denominator is arbitrarily chosen and the numerator is then defined in order to achieve the maximum order of approximation. Such approximants are called *Padé-type approximants.* Their definition was given in the first section and we shall study below some of their convergence properties. In rational approximants, it is also possible to choose only a part of the denominator or a part of the numerator and the denominator. These are the *partial Padé approximants.* Such generalizations allow us to include into the construction of the approximant the information that is known about the zeros and the poles of the function being approximated, thus often leading to better convergence properties. *Multipoint Padé approximants* have expansions around several points which agree with the expansion of the function around the same points up to given orders. Padé approximants for series of functions have also received much attention. Other generalizations deal with Padé approximants for double series. Another important generalization is the vector case, which will be studied below. Series with coefficients in a non-commutative algebra have also received much interest, in particular the matrix case, due to their applications. Other types of approximants, such as the *Cauchy-type approximants* or the *Padé–Hermite approximants*, have been defined. It is, of course, possible to study combinations of these various generalizations such as the multipoint Padé–type approximants for multiple series of functions with matrix coefficients. Due to the space limitation of this article, we shall only present Padé–type approximants and the vector case and refer the interested reader to Brezinski and Van Iseghem

(1994) where more details about these generalizations can be found with the relevant references to the literature.

3.1. Padé-type approximants

As we saw in the first section, the function $f(z) = \sum_{n\geq 0} c_n z^n$ completely defines the linear functional c. It always has an integral representation in the complex field as stated by the following theorem:

Theorem 3.1 Let R be the radius of convergence of the series f and $\mathcal{H}_\alpha$ the space of holomorphic functions in the disc $D_{1/\alpha}$. Then c has the following representation

$$\forall g \in \mathcal{H}_\alpha, \quad c(g) = \frac{1}{2\pi \mathrm{i}} \int_{|x|=r} f(x) g(1/x) \frac{\mathrm{d}x}{x}, \qquad \alpha < r < R.$$

In fact, in practical situations, the contour can be transformed continuously in such a way that $f(x)g(1/x)$ remains holomorphic in a neighborhood of it. An application of this result is the representation of the remainder term of Padé or Padé-type approximants $f(z) - P_{n-1}/Q_n(z)$.

$$\begin{aligned}
f(z) &= \sum_{n\geq 0} c_n z^n, \quad \tilde{v}_n(z) = z^n v_n(z^{-1}), \\
f(z) &= c\left(\frac{1}{1-xz}\right), \qquad c(g) = \frac{1}{2\pi \mathrm{i}} \int_C f(x) g(1/x) \frac{\mathrm{d}x}{x}, \\
f(z) - c(P) &= \frac{z^n}{\tilde{v}_n(z)} c\left(\frac{v(x)}{1-xz}\right) = \frac{1}{2\pi \mathrm{i}} \frac{z^n}{\tilde{v}_n(z)} \int_C f(x) \frac{v_n(x^{-1})}{x-z} \mathrm{d}x \\
&= \frac{1}{2\pi \mathrm{i}} \int_{-C^{-1}} \frac{v_n(x)}{v_n(z^{-1})} \frac{f(x^{-1})}{x(1-xz)} \mathrm{d}x
\end{aligned}$$

Finally, using the notation of Cala Rodriguez and Wallin (1992), conditions on $\Gamma = -C^{-1}, f$ and v_n are now summarized, with $v_n(z) = \tilde{Q}_n(z) = \prod_{j=1}^n (1 - \beta_{jn} z)$, which means that the β_{jn} are the poles of the approximant whose denominator is $Q_n(z) = \prod_{j=1}^n (z - \beta_{jn})$ and the following error formula is obtained.

Theorem 3.2 Let f be analytic in a domain D containing zero, let $\beta_{jn}, 1 \leq j \leq n, n \geq 1$, be given complex numbers and let $z \in D - \{\beta_{jn}^{-1}\}$. Let Γ be a contour in D^{-1} consisting of a finite number of piecewise continuously differentiable closed curves with index

$$\mathrm{ind}_\Gamma(a) = \begin{cases} 1 & \text{if } a \in \bar{\mathbb{C}} - D^{-1}, \\ 0 & \text{if } a = z^{-1}, z \neq 0. \end{cases}$$

Finally, let P_{n-1}/Q_n be the $(n-1/n)$ Padé-type approximant of f with preassigned poles at the zeros of Q_n. Then

$$f(z) - \frac{P_{n-1}(z)}{Q_n(z)} = \frac{1}{\mathrm{i}} \int_\Gamma \frac{\tilde{Q}_n(t)\, f(t^{-1})}{\tilde{Q}_n(z^{-1})\, t(1-zt)} \mathrm{d}t.$$

Let us now formulate the theorem of Eiermann (1984) in the form he proved it (for a complete and detailed proof, see Eiermann (1984) or Cala Rodriguez and Wallin (1992)).

Theorem 3.3 Let f be analytic in a domain $D \subset \mathbb{C}$ containing zero. Let β_{jn}, $j = 1, \ldots, n$, be the given zeros of Q_n, and let $\Lambda \subset \mathbb{C}^2$ containing $\mathbb{C} \times \{0\}$ and such that, uniformly for (x, z) in compact subsets of Λ,

$$\lim_{n\to\infty} \frac{\tilde{Q}_n(x)}{\tilde{Q}_n(z^{-1})} = 0.$$

Then, the sequence of Padé-type approximants P_{n-1}/Q_n converges to f uniformly on compact subsets of $A = \{z,\ \forall\xi \in \bar{\mathbb{C}}/D,\ (\xi^{-1}, z) \in \Lambda\}$.

If f is a Stieltjes function defined on $\mathbb{C} - (-\infty, -R]$, then $K = \mathbb{C} - D^{-1} = [-1/R, 0]$, and Γ is any contour containing K. The best choice is to take Γ as small as possible, and the assumption on Q_n becomes, for z in some compact subset F of D, $\lim_{n\to\infty} \sup_{z\in F} \sup_{x\in K} \left| \frac{\tilde{Q}_n(x)}{\tilde{Q}_n(z^{-1})} \right| = 0$. This remark leads to the following alternative results. With the same notation as before, we get:

Theorem 3.4 Let $K = \bar{\mathbb{C}}/D^{-1}, O$ a neighborhood of K and F some compact subset of D. If

$$\lim_{n\to\infty} \Big(\sup_{(t,z)\in \bar{O}\times F} \Big| \frac{\tilde{Q}_n(t)}{\tilde{Q}_n(z^{-1})} \Big| \Big) = 0, \text{ then } \lim_{n\to\infty} \Big(\max_{z\in F} \Big| f(z) - \frac{P_{n-1}(z)}{Q_n(z)} \Big| \Big) = 0.$$

Similarly, we have the following theorem:

Theorem 3.5 Let $K = \bar{\mathbb{C}}/D^{-1}, O$ a neighborhood of K and F some compact subset of D. Then for the sup norm on F ($z \in F$), if

$$\overline{\lim_{n\to\infty}} \Big(\sup_{t\in\bar{O}} \Big\| \frac{\tilde{Q}_n(t)}{\tilde{Q}_n(z^{-1})} \Big\|_F^{1/n} \Big) \le r < 1, \text{then } \overline{\lim_{n\to\infty}} \Big(\Big\| f(z) - \frac{P_{n-1}(z)}{Q_n(z)} \Big\|_F^{1/n} \Big) \le r.$$

If the $\tilde{Q}_n$'s are some orthogonal polynomials for which asymptotic formulae are known, the preceding theorem leads to interesting results (Prévost, 1983).

Another idea is to take Q_n with one multiple zero, so that $|Q_n(x)|^{1/n} = |x - \beta_n|$. Two cases are to be considered: the first one with $\beta_n = \beta$ or $\lim_n \beta_n = \beta$ (which gives the same rate of convergence), the second one with

β_n depending on n (and which may tend to ∞) in the case of entire functions having one singularity at $+\infty$ and none at $-\infty$ such as the exponential function (Van Iseghem, 1992; Le Ferrand, 1992b).

3.2. The vector case

• *Introduction*

There are at least two ways for obtaining approximations for vector problems.

The first one consists in considering simultaneously d scalar functions defined by their power series expansions in a neighborhood of zero, organizing them as one power series with vector coefficients in $\mathbb{C}^d$ and then looking for a rational approximation following the idea of Padé approximation, that is, finding a *best* approximant. This approach has been developed through the simultaneous approximants by de Bruin (1984) and through the vector approximants by Van Iseghem (1985, 1987b). In each case, the result is a rational approximant $(P_1/Q, \dots, P_d/Q)$. In the vector case, all the P_i's have the same degree while they have to satisfy some constraints in the simultaneous case. Although the ideas are rather similar, the approximants are not the same, except when $\deg Q = nd$ and $\deg P_i = n,\ \mathrm{i} = 1, \dots, d$. In both cases, the scalar Padé approximants are recovered if the dimension d of the vectors is one. Only the vector case, which seems to be simpler, will be explained here.

The second way for obtaining vector approximations is through extrapolation and acceleration of vector sequences. In the scalar case, the ε-algorithm provides a way for computing Padé approximants and such a link can also be developed through the vector ε-algorithm or the topological ε-algorithm, as we shall see below.

• *Vector Padé approximants*

For vector Padé approximants, giving m as the common degree of all the numerators P_i and n as the degree of the common denominator Q defines completely the approximant. The vector Padé approximant $\mathbf{R}(t) = (P_\alpha/Q)_{\alpha=1,\dots,d}$ is the best in the sense that it is impossible to improve simultaneously the order of approximation of all the components.

Let $\mathbf{F} = (f_1, \dots, f_d)$. If, for each $\alpha = 1, \dots, d$, we write

$$f_\alpha(z) = \sum_{i\geq 0} c_i^\alpha z^i,$$

then we set $\Gamma_i = (c_i^1, \dots, c_i^d)^T \in \mathbb{C}^d$ and we define the series $\mathbf{F}$ by

$$\mathbf{F}(z) = \sum_{i\geq 0} \Gamma_i z^i.$$

Let $\Gamma : \mathbb{C}[[x]] \to \mathbb{C}^d$ be the linear functional defined by

$$\Gamma(x^i) = \Gamma_i.$$

Taking the components of the vector approximant as the Padé-type approximants of the f_α's, we get the following theorem:

Theorem 3.6 Let P be the Hermite interpolation polynomial of $1/(1-xz)$ at $x_1, \ldots, x_n$. Setting

$$v(x) = \prod_{i=1}^{n}(x - x_i), \qquad \tilde{v}(t) = t^n v(t^{-1}), \qquad \mathbf{W}(z) = \Gamma\left(\frac{v(z) - v(x)}{z - x}\right),$$

where the functional Γ acts on x, we have $\Gamma(P) = \tilde{\mathbf{W}}(z)/\tilde{v}(z)$, and

$$\mathbf{F}(z) - \Gamma(P) = \frac{z^n}{\tilde{v}(z)}\Gamma\left(\frac{v(x)}{1 - xz}\right) = \frac{z^n}{\tilde{v}(z)}\sum_{i\geq 0} D_i z^i, \qquad D_i = \Gamma(x^i v(x)).$$

$\tilde{\mathbf{W}}(z)/\tilde{v}(z)$ is called the *Padé-type approximant* $(n-1/n)$ of $\mathbf{F}$.

The proof of this result is similar to that of the scalar case because each component $(\tilde{W}(z)/\tilde{V}(z))_\alpha$ is the Padé-type approximant of f_α for $\alpha = 1, \ldots, d$.

In order to improve the order of approximation on all the components, we have to choose the polynomial v such that a maximum number of D_i are zero. D_i is a vector of $\mathbb{C}^d$ and so $D_i = 0$ represents d scalar equations with the coefficients of v as unknowns. Thus, n being the degree of the denominator, the best order of approximation by rational functions of type $(n-1, n)$ is $n + [n/d]$, where $[n/d]$ is the integer part of n/d.

Padé-type approximants (s/r) for arbitrary degrees s and r can be defined as for the scalar case. For any integer h, positive or not, we get

$$\mathbf{F}(z) = \sum_{i=0}^{h-1}\Gamma_i z^i + \mathbf{F}_h(z), \qquad \Gamma_i = 0 \quad \text{if } i < 0,$$

$$(r + h - 1/r)_{\mathbf{F}}(z) = \sum_{i=0}^{h-1}\Gamma_i z^i + (r - 1/r)_{\mathbf{F}_h}(z).$$

The order of approximation is $r + h - 1$. It can be increased up to $r + h - 1 + [r/d]$ by choosing the generating polynomials v of the vector Padé–type approximants. Let r and h be arbitrary integers ($r = nd + k$, $0 \leq k < d$); let us denote by $\Gamma^{(h)}$ the linear functional defined by $\Gamma^{(h)}(x^i) = \Gamma_{i+h}$ and by $P^{(h)}_{nd+k}$ the polynomial defined by the following equations:

$$\left.\begin{array}{ll} \Gamma^{(h)}(x^i P^{(h)}_{nd+k}(x)) = 0, & i = 0, \ldots, n-1, \\ c^\alpha(x^{n+h} P^{(h)}_{nd+k}(x)) = 0, & \alpha = 1, \ldots, k. \end{array}\right\} \tag{3.1}$$

The Padé-type approximant $(r+h-1/r)$ generated by $P^{(h)}_{nd+k}$ will have the maximal order of approximation: the order of approximation is $r+h+n-1$ at least, $r+h+n$ for the first k components. This approximant will be called the *vector Padé approximant* $[r+h-1/r]_{\mathbf{F}}$.

Writing the conditions (3.1) as a linear system, we get an expression of $P^{(h)}_r$ as a ratio of two determinants. As usual the determinantal expression for $P^{(h)}_r$ gives rise to a determinantal expression for the vector Padé approximants where only the last row of the numerator is a vector, all the other rows $(\Gamma_i,\ldots,\Gamma_{r+i})$ being put for d scalar rows and the last one $(\Gamma^{(k)}_{n+h},\ldots)$ representing the first k components of $(\Gamma_{n+h},\ldots)$. Setting $s=r+h$ and $\sum_k=\sum_{i=0}^k \Gamma_i z^i$, we have

$$P^{(h)}_r(x)=\frac{\begin{vmatrix}\Gamma_h & \cdots & \Gamma_s\\ \vdots & & \vdots\\ \Gamma_{n-1+h} & \cdots & \Gamma_{n+s-1}\\ \Gamma^{(k)}_{n+h} & \cdots & \Gamma^{(k)}_{n+s}\\ 1 & \cdots & x^r\end{vmatrix}}{\begin{vmatrix}\Gamma_h & \cdots & \Gamma_{s-1}\\ \vdots & & \vdots\\ \Gamma_{h+n-1} & \cdots & \Gamma_{n+s-1}\\ \Gamma^{(k)}_{n+h} & \cdots & \Gamma^{(k)}_{n+s-1}\end{vmatrix}},\quad [s-1/r]_{\mathbf{F}}(z)=\frac{\begin{vmatrix}\Gamma_h & \cdots & \Gamma_s\\ \vdots & & \vdots\\ \Gamma_{h+n-1} & \cdots & \Gamma_{n+s-1}\\ \Gamma^{(k)}_{h+n} & \cdots & \Gamma^{(k)}_{n+s}\\ z^r\Sigma_{h-1} & \cdots & \Sigma_{s-1}\end{vmatrix}}{\begin{vmatrix}\Gamma_h & \cdots & \Gamma_s\\ \vdots & & \vdots\\ \Gamma^{(k)}_{h+n} & \cdots & \Gamma^{(k)}_{n+s}\\ z^r & \cdots & 1\end{vmatrix}}.$$

Similarly to the scalar case, the polynomials $(P^{(h)}_r)_{r\geq 0}$ are the generating polynomials of $[h+r-1/r]$, and for each h they satisfy a recurrence formula, which, here, is of order $d+1$ (i.e. with $d+2$ terms):

$$P^s_{r+1}(x)=(x-\beta^s_r)P^s_r(x)-\sum_{\mu=1}^{d}\gamma^s_\mu P^s_{r-\mu}(x). \tag{3.2}$$

If all the $P^{(h)}_r$ exist, then the last coefficient γ^h_d is not zero. Since a theory analogous to the theory of orthogonal polynomials can be developed, these polynomials have been called *vector orthogonal polynomials* (or of dimension d) (Van Iseghem, 1985, 1987b, 1989). A Shohat–Favard theorem can be proved: given a family $(P_r)_{r\geq 0}$ satisfying the the relation (3.2), there exist d functionals $c^1,\ldots,c^d$ such that the P_r's satisfy the relation (3.1) with respect to $\Gamma=(c^1,\ldots,c^d)$. The space of all possible Γ's is a vector space of dimension $(d!)$. And so there is, as in the scalar case, an equivalence between the vector orthogonality defined by the relations (3.1) and the family of polynomials defined by the recurrence relations (3.2).

From an algorithmic point of view, a QD-type algorithm linking two diagonals $(P^{(h)}_r)_r$ and $(P^{(h+1)}_r)_r$ can be defined. It allows one to move in all directions either in the table of the polynomials $(P_r^{(h)})$ or in the table of the vector Padé approximants. The approximants can be also computed by algorithms such as the recursive projection algorithm (R.P.A.) or the compact recursive projection algorithm (C.R.P.A.) of Brezinski (1983b).

An example of an algorithm deduced from the recurrence relations in the table of the polynomials is the generalization of the cross rule of Wynn obtained by Van Iseghem (1986). In the scalar case, this cross rule involves five approximants in the following array:

$$\begin{array}{ccc} & N & \\ W & C & E \\ & S & \end{array}$$

and it can be written in two different forms

$$\begin{aligned} (E-N)(C-W)(S-C) &= (C-N)(E-C)(S-W), \\ \frac{1}{C-N}+\frac{1}{C-S} &= \frac{1}{C-E}+\frac{1}{C-W}. \end{aligned}$$

The proof can be extended to the vector case. The approximants involved are the following: W_i and N_i lying on diagonals and

$$\begin{array}{cccc} W_d & N_d & & \\ & \ddots & & \\ & N_1 & N & \\ & W_1 & C & E \\ & & S & \end{array}$$

The explicit form is given for all the components $(E-C)^\alpha, \alpha = 1, \ldots, d$, the D_i being $d \times d$ determinants and the vectors indicated being the columns of the determinants

$$D_1 = |C-W_1, W_d-W_{d-1}, \ldots, W_2-W_1|, D_2 = |S-C, N_{d-1}-N_{d-2}, \ldots, N_1-C|,$$

$$D_3 = |S-W_1, W_d-W_{d-1}, \ldots, W_2-W_1|, D_4 = |C-N, N_{d-1}-N_{d-2}, \ldots, N_1-C|,$$

$$\frac{1}{(E-C)^\alpha}+\frac{1}{(C-N)^\alpha} = \frac{1}{(C-N)^\alpha}\frac{D_3D_4}{D_1D_2}, \qquad \alpha = 1, \ldots, d.$$

As in the scalar case, symbolic negative columns are defined so that the algorithm can be used with the center term C in the column with subscript zero. So this algorithm can be used for computing E from the first column and it gives all the approximants $[p/q]$ with $p \geq q-1$.

Vector Padé approximants can be used for accelerating the convergence of vector sequences. Let $\mathbf{F} = \sum_{i\geq 0} \Gamma_i z^i$. Then, a vector sequence (S_n) can be canonically associated to $\mathbf{F}$ by taking $\Gamma_i = \Delta S_i$. Thus S_n is the nth partial sum of the series $\mathbf{F}(1)$ and the vector Padé approximants are associated to a transformation of sequences. By combining the rows in the determinantal expression of the vector approximant of $\mathbf{F}$ it follows that

$$[r+h/r]_{\mathbf{F}}(1) = \psi_r(S_h) = \frac{\begin{vmatrix} S_h & \cdots & S_{h+r} \\ \Delta S_h & \cdots & \Delta S_{h+r} \\ \vdots & & \vdots \\ \Delta S_{h+n-1} & \cdots & \Delta S_{h+r+n-1} \\ \Delta S_{h+n}^{(k)} & \cdots & \Delta S_{h+r+n}^{(k)} \end{vmatrix}}{\begin{vmatrix} 1 & \cdots & 1 \\ \Delta S_h & \cdots & \Delta S_{h+r} \\ \vdots & & \vdots \\ \Delta S_{h+n-1} & \cdots & \Delta S_{h+r+n-1} \\ \Delta S_{h+n}^{(k)} & \cdots & \Delta S_{h+r+n}^{(k)} \end{vmatrix}}.$$

The first row is formed by vectors and the following rows stand for the d rows of their components except, as usual, the last one which contains only the first k components.

The basic result is the following theorem.

Theorem 3.7 A necessary and sufficient condition for $\psi_r(S_n) = S\,\forall n$ is that the sequence (S_n) satisfies a linear recurrence relationship, that is, $\forall n$

$$\sum_{i=0}^{r} a_i(S_{n+i} - S) = 0, \quad \text{with} \quad \sum_{i=0}^{r} a_i \neq 0, \quad a_i \in \mathbb{C}.$$

In the scalar case, Aitken's Δ^2 process is recovered for $r = 2$ and the Padé table, Shanks transformation and ε-algorithm for $r \geq 2$. If $d \neq 1$ and if $r < d$, we recover the MPE (Minimal Polynomial Extrapolation) algorithm studied by Sidi, Ford and Smith (1986). For $r = d$, a transformation due to Henrici (1964) is obtained. For $r > d$, the transformation does not seem to have been studied yet independently from the vector Padé approximants (Van Iseghem, 1994). Graves-Morris (1994) defined another kind of approximants, also called vector Padé approximants, different from those studied below, which are generated by the vector ε-algorithm.

From the relation of the theorem above, it is clear that the results of this extrapolation method are to be linked with those of other extrapolation methods such as the vector ε-algorithm or the topological ε-algorithm.

- *The vector ε-algorithm*

The rule of the ε-algorithm is

$$\varepsilon_{k+1}^{(n)} = \varepsilon_{k-1}^{(n+1)} + \left[\varepsilon_k^{(n+1)} - \varepsilon_k^{(n)}\right]^{-1}$$

with $\varepsilon_{-1}^{(n)} = 0$ and $\varepsilon_0^{(n)} = S_n$.

Now, if S_n is a vector of $\mathbb{C}^p$, the preceding rule can be applied if the inverse of a nonzero vector y is defined. Using the pseudo-inverse of the rectangular matrix y, Wynn (1962) took $y^{-1} = y/(y,y)$. For this algorithm, the following result holds.

Theorem 3.8 A sufficient condition for $\varepsilon_{2k}^{(n)} = S\,\forall n$ is that $\forall n$,

$$a_0(S_n - S) + \cdots + a_k(S_{n+k} - S) = 0$$

with $a_0 + \cdots + a_k \neq 0$ and $a_0 a_k \neq 0$.

This theorem was first given by McLeod (1971) in the case where the a_i's are real numbers. The full proof was obtained by Graves-Morris (1983) for complex coefficients. Contrary to the scalar case, only the sufficiency has been proved. The theory of this vector ε-algorithm and its applications were based only on this theorem and other results about it were quite difficult to obtain (see, for example, Cordellier (1977)). This was due to the nonexistence of determinantal formulae for the vectors $\varepsilon_k^{(n)}$. It has been proved by Salam (1994) that these vectors can be expressed as a ratio of two designants, a notion generalizing that of determinants in a non-commutative algebra. This approach should lead to new theoretical results about the vector ε-algorithm.

- *The topological ε-algorithm*

A drawback of the vector ε-algorithm was its lack of determinantal expressions for the $\varepsilon_k^{(n)}$, due to the fact that this algorithm was obtained directly from the rule of the scalar ε-algorithm, by defining the inverse of a vector. Thus a possible remedy was to construct a vector-sequence transformation following the ideas of Shanks (1955) for the scalar case and then to obtain a recursive algorithm for its implementation following Wynn (1956). We start by assuming that the sequence (S_n) satisfies $\forall n$,

$$a_0(S_n - S) + \cdots + a_k(S_{n+k} - S) = 0.$$

Since, as above, it is assumed that $a_0 + \cdots + a_k \neq 0$, it is not a restriction to set this sum to 1. Thus we have $\forall n$,

$$S = a_0 S_n + \cdots + a_k S_{n+k}.$$

For computing the a_i's, we subtract this relation from the next one and multiply it scalarly by an arbitrary vector y. Thus for $i = 0, \ldots, k-1$, we have

$$a_0(y, \Delta S_{n+i}) + \cdots + a_k(y, \Delta S_{n+i+k}) = 0.$$

Solving this system provides the a_i's and S. If we set

$$e_k(S_n) = \frac{\begin{vmatrix} S_n & \cdots & S_{n+k} \\ (y, \Delta S_n) & \cdots & (y, \Delta S_{n+k}) \\ \vdots & & \vdots \\ (y, \Delta S_{n+k-1}) & \cdots & (y, \Delta S_{n+2k-1}) \end{vmatrix}}{\begin{vmatrix} 1 & \cdots & 1 \\ (y, \Delta S_n) & \cdots & (y, \Delta S_{n+k}) \\ \vdots & & \vdots \\ (y, \Delta S_{n+k-1}) & \cdots & (y, \Delta S_{n+2k-1}) \end{vmatrix}},$$

then, by construction, we have $\forall n, e_k(S_n) = S$ if the sequence (S_n) satisfies the relation given above. If the sequence does not satisfy such a relation, then the determinants appearing in the expression of $e_k(S_n)$ can, however, be computed (the determinant in the numerator denotes the vector obtained as the linear combination of the vectors in its first row given by using the classical rules for expanding a determinant), and $e_k(S_n)$ is a generalization of the Shanks transformation and Padé approximants. Now, from a practical point of view, it is necessary to find an algorithm for computing recursively the vectors $e_k(S_n)$ without computing explicitly the determinants involved in the formula. This algorithm was called the topological ε-algorithm. Its rules are the following (Brezinski, 1975), with $\varepsilon_{-1}^{(n)} = 0$ and $\varepsilon_0^{(n)} = S_n$:

$$\begin{aligned} \varepsilon_{2k+1}^{(n)} &= \varepsilon_{2k-1}^{(n+1)} + \frac{y}{(y, \varepsilon_{2k}^{(n+1)} - \varepsilon_{2k}^{(n)})}, \\ \varepsilon_{2k+2}^{(n)} &= \varepsilon_{2k}^{(n+1)} + \frac{\varepsilon_{2k}^{(n+1)} - \varepsilon_{2k}^{(n)}}{(\varepsilon_{2k+1}^{(n+1)} - \varepsilon_{2k+1}^{(n)}, \varepsilon_{2k}^{(n+1)} - \varepsilon_{2k}^{(n)})}. \end{aligned}$$

The topological ε-algorithm can be considered as the construction of Padé approximants in the direction of y. If d independent directions are chosen and if the rows $((y, \Delta S_{n+i}), \ldots), i = 0, \ldots, k-1$, are replaced by the d rows $((y_i, \Delta S_n), \ldots), i = 1, \ldots, d$, then the vector Padé approximants (or more exactly $\psi_k(S_n)$) are recovered. For $k = d + 1$, Henrici's transformation is obtained.

All these algorithms are, in fact, constructed from the same idea which is that of Padé approximation of achieving the maximum degree of approximation at zero. So, as we shall see in the next section, they have similar properties for their applications, for example, in solving systems of linear and nonlinear equations.

4. Applications

Other applications can be found in Cuyt and Wuytack (1987).

4.1. A-acceptable approximations to the exponential function

Let us consider the differential equation $y'(x) = -\lambda y(x)$ where λ is a complex number whose real part is strictly positive. Thus the solution will satisfy $\lim_{x\to\infty} y(x) = 0$. This differential equation (with the initial condition $y(0) = y_0$) is integrated by a numerical method that computes approximations y_n of the exact solution $y(nh)$, where h is the step size. This numerical method is said to be A-stable if $\forall h\lambda$ such that $\mathrm{Re}(h\lambda) > 0, \lim_{n\to\infty} y_n = 0$, which means that both the exact and the approximate solutions tend to zero at infinity.

Of course, since the exact solution is $y(x) = y_0 \mathrm{e}^{-\lambda x}$, we have $y(x_{n+1}) = \mathrm{e}^{-h\lambda} y(x_n)$ with $x_n = nh$. When using either a one-step or a multistep method, it can be proved that the approximate solution satisfies

$$y_{n+1} = r(h\lambda) y_n,$$

where r is a rational function. Thus, if the numerical method has order p, we have

$$r(z) = \mathrm{e}^{-z} + \mathcal{O}\left(z^{p+1}\right).$$

Moreover, if the method is A-stable we must have, $\forall z$ such that $\mathrm{Re}(z) > 0, |r(z)| < 1$ since $y_n = [r(h\lambda)]^n y_0$.

Such a rational approximation to the exponential function is called A-acceptable and, of course, Padé, Padé-type and partial Padé approximants are candidates for such an r.

Using the maximum-modulus principle it can be shown that r is A-acceptable if and only if $\forall t \in \mathbb{R}, |r(it)| \le 1, \lim_{|z|\to\infty} |r(z)| \le 1$ and r is analytic in the right half part of the complex plane; see Alt (1972).

The A-acceptability of Padé approximants to the exponential function was studied by Ehle (1973) who proved:

Theorem 4.1 The Padé approximants $[n/n], [n-1/n]$ and $[n-2/n]$ of e^{-z} are A-acceptable for all n.

Let us now turn to Padé-type approximants. The following result is an adaptation of that of Crouzeix and Ruamps (1977) for rational approximants to the exponential function.

Theorem 4.2 Let r be a Padé-type approximant of e^{-z} with real coefficients, whose numerator has degree k and whose denominator has degree $n+k \quad (n \ge 0)$. Let

$$|r(\mathrm{i}t)|^2 = \frac{1 + \beta_1 t^2 + \cdots + \beta_k t^{2k}}{1 + \alpha_1 t^2 + \cdots + \alpha_{n+k} t^{2(n+k)}}.$$

If the zeros of the denominator of r have negative reals parts, if $\beta_i \le \alpha_i$

for $i = [k/2]+1,\dots,k$ and if $0 \le \alpha_i$ for $i = k+1,\dots,k+n$, then r is A-acceptable. ($[x]$ denotes the integer part of the real number x.)

When solving a parabolic partial differential equation of the second order, one obtains, after discretization of the space variable, a differential system of the form

$$\begin{aligned} Cu'(t) &= -Au(t) + v(t), \\ u(0) &= u_0, \end{aligned}$$

where C and A are real square matrices whose elements are independent of the time t. Using a one-step method for integrating this differential equation leads to $Q_k(Bh)u_{n+1} = P_m(Bh)u_n + T_n$, where, $B = C^{-1}A$ T_n, is a matrix depending on k and m, u_n is an approximation of the exact solution $u(t_n)$ at the point t_n and Q_k and P_m are matrix polynomials of the respective degrees k and m. As before, $[Q_k(Bh)]^{-1}P_m(Bh)$ must be an approximation of e^{-Bh} and the order of the method is determined by that of the approximation. This approximation must be A-acceptable if an A-stable one-step method is needed. The computation of u_{n+1} from u_n requires the computation of the inverse of the matrix $Q_k(Bh)$. This computation is greatly simplified if $Q_k(z) = (1+\alpha_k z)^k$. Indeed, in that case, the computation of u_{n+1} reduces to the solution of k systems of linear equations with the same matrix

$$(I + \alpha_k Bh)v_{p+1} = v_p,\ p = 0,\dots,k-1,\ v_0 = P_m(Bh)u_n + T_n,\ v_k = u_{n+1}.$$

Of course, such a simplification is impossible with Padé approximants but it becomes possible with Padé-type approximants. For convergence reasons, since $\lim_{k\to\infty}(1+z/k)^k = \mathrm{e}^z$ we shall make the choice $\alpha_k = 1/k$ which corresponds to the generating polynomials $v_k(x) = (x+1/k)^k$. The following result can be proved:

Theorem 4.3 The Padé-type approximants $(k-1/k)$ of e^{-z} constructed with the generating polynomials $v_k(x) = (x+1/k)^k$ are A-acceptable for $k = 1,2,3$. The Padé-type approximants (k/k) constructed with the same generating polynomials are A-acceptable for $k = 1,\dots,4$. $(6/6)$ is not A-acceptable.

The study of the convergence of these approximants is due to Van Iseghem (1984) who proved the following results:

Theorem 4.4 The sequences $((k-1/k))$ and $((k/k))$ of Padé-type approximants to $\exp(-z)$ constructed with the generating polynomials $v_k(x) = (x+1/k)^k$ converge to $\exp(-z)$ uniformly and geometrically on every compact subset of the complex plane.

Complementary results on the A-acceptability of Padé-type approximants with a single pole were given by González-Concepción (1987). See Wanner

(1987) for a review on the A-acceptability of Padé approximants to the exponential function.

4.2. Laplace and other transforms

In this section, we shall show how Padé approximants can be used in the numerical solution of problems related to the Laplace, the Borel and the z-transforms.

The Laplace transform of a function f is defined by

$$\bar{f}(p) = \int_0^\infty e^{-pt} f(t)dt.$$

The Borel transform is a Laplace transform, $B(z) = \bar{f}(1/z)$, and the z-transform is a discrete equivalent of the Laplace transform, as will be shown at the end of this section.

The Laplace transform is used in several cases: if f satisfies a functional equation, $\bar{f}$ satisfies a simpler one which can be solved more easily. For example, a linear ordinary differential equation is replaced by an algebraic equation, a linear partial differential equation is replaced by an ordinary differential equation and so forth. For example, the partial differential equation

$$\left(\frac{\partial^2}{\partial r^2} - 2\frac{\partial^2}{\partial r \partial x} + \frac{1}{r}\left(\frac{\partial}{\partial r} - \frac{\partial}{\partial x}\right) - \frac{m^2}{r^2}\right) R(x,r) = 0$$

is transformed, with respect to x, into

$$\left(\frac{\partial}{\partial r^2} + \left(\frac{1}{r} - 2p\right)\frac{\partial}{\partial r} - \left(\frac{p}{r} + \frac{m^2}{r^2}\right)\right) \bar{R}(p,r) = 0.$$

The problem is to find $f(t)$ from $\bar{f}(p)$. It requires the construction of approximate methods for computing inverse Laplace transforms that permit us to find the original function in a broad class of cases.

It must be noticed at once that the problem is always unstable: if $\bar{f}(p) = 1/(p-\alpha)$, then $f(t) = e^{\alpha t}$. So if there is an error in α, of say ε, then there is an amplification of the error: $f_\varepsilon = e^{\varepsilon t} f$. Conversely, if f is modified on a small interval, then $\bar{f}$ will have a very smooth change.

An extensive literature exists on the subject and on the different methods for inverting the Laplace transform. A review can be found, for example, in Luke (1969). We will only discuss here the problem of inversion by use of rational approximants.

In many applications, one knows explicitly $\bar{f}(p)$ and one wants to find $f(t)$ numerically. One way of obtaining approximations to the inverse $f(t)$ of $\bar{f}(p)$ is by approximating $\bar{f}(p)$ by a sequence of rational functions $\bar{f}_n(p)$, $n \geq 1$, and then inverting the $\bar{f}_n(p)$ exactly to obtain the sequence $f_n(t)$, $n \geq 1$. The hope is that, if the sequence $(\bar{f}(p))_n$ converges to $\bar{f}(p)$ quickly, then the sequence $(f_n(t))_n$ will converge to $f(t)$ also, more or less quickly. There are several ways of obtaining rational approximations to a given function, one of them being by expanding it into a Taylor series, and then forming the Padé table associated with the Taylor series. Detailed discussions and references to various applications can be found in the paper by Longman (1973).

The link with Prony's method (interpolation by a sum of exponential functions) has been generalized by Sidi (1981).

If $\bar{f}_n(p)$ has n poles, not necessarily distinct but of multiplicity n_i, with $\sum_1^m n_i \leq n$, then

$$\bar{f}_n(p) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} \frac{A_{ij}}{(p-\alpha_i)^j} \Leftrightarrow f_n(t) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} A_{ij}t^{j-1}\mathrm{e}^{\alpha_i t}.$$

So the problem is the approximation of $f(t)$ by functions of the same type as $f_n(t)$. The result proved by Sidi is as follows:

Theorem 4.5 Define the set G_n as

$$G_n = \left\{ g(t) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} B_{ij}t^{j-1}\mathrm{e}^{\alpha_i t}, \quad \alpha_i \neq \alpha_j, \ \sum_{1}^{m} n_i \leq n, B_{ij} \in \mathbb{C} \right\}.$$

Now, let $g_n(t)$ be the function, if it exists, belonging to G_n that approximates $f(t)$ in $[0,\infty)$ in the following weak sense:

$$\int_0^\infty t^N \mathrm{e}^{-wt}(f(t) - g_n(t))t^i \mathrm{d}t = 0, \quad i = 0, \dots, 2n-1.$$

Then $\bar{g}_n(p)$, the Laplace transform of $g_n(t)$, is the Padé approximant $[n-1/n]$ of $\bar{f}(p-w)$. Furthermore, $g_n(t)$ is a real function of t if $\bar{f}(p)$ is real for real p.

As seen before, Padé-type approximants can give better numerical results than Padé approximants if some information about the poles is known. Various investigations have been conducted in the past ten years concerning the convergence of methods using Padé or Padé-type approximants.

One research path, followed by Van Iseghem (1987a), is through orthogonal polynomials and Padé-type approximants with one multiple pole. The basic remark is the following one, due to Tricomi (see Sneddon (1972)), about the Laguerre polynomials of order zero:

$$f(t) = \mathrm{e}^{\lambda t}L_k(2\lambda t) \Leftrightarrow \bar{f}(p) = \frac{(p-\lambda)^k}{(p+\lambda)^{k+1}}.$$

So, the Laplace transform formally achieves a correspondence between a series in powers of $(p-\lambda)/(p+\lambda)$ and an expansion in Laguerre polynomials. Convergence in the least-squares sense is to be expected for (f_n), but better results are obtained. We have

$$\bar{f}(p) = \frac{1}{p+\lambda} \sum_{k\geq 0} a_k \left(\frac{p-\lambda}{p+\lambda}\right)^k .$$

The partial sum $\bar{f}_n(p)$ of this series is the $(n/n+1)$ Padé-type approximant of $\bar{f}(p)$ with denominator $(p+\lambda)^{n+1}$. If $\bar{f}(p) = \sum_{i\geq 0} c_i (p-\lambda)^i$, then the a_k's are easily computed:

$$a_k = \sum_{i=0}^{k} \binom{k}{i} c_i (2\lambda)^i.$$

The following results are obtained:

Theorem 4.6 Let us assume that $\bar{f}$ is analytic in the half plane $\mathrm{Re}(p) > 0$, $f(t)$ exists and $\int_0^\infty f^2(t)\mathrm{d}t < \infty$. Then the sequence $(\bar{f}(p))_n$ converges to $\bar{f}$, uniformly on every compact set of the half plane $\mathrm{Re}(p) > 0$. The sequence $(f_n(t))_n$ converges to f in the least-squares sense. Furthermore if $(p+\lambda)\bar{f}(p)$ is analytic at infinity and if $\bar{f}(p)$ is analytic in $\mathrm{Re}(p) \geq 0$, then the sequence $(f_n(t))_n$ converges to $f(t)$ uniformly on compact sets of $\mathbb{R}^+$.

This theorem can be improved in two directions:
- $\bar{f}$ is analytic in $\mathrm{Re}(p) > w$ (instead of $w = 0$);
- $(p+\lambda)\bar{f}^{(k)}(p)$ is analytic at infinity (instead of $k = 0$).

With the first assumption, we get the following sequence $(h_n(t))$:

$$h_n(t) = \frac{1}{t^k} \mathrm{e}^{(w-\lambda)t} \sum_0^n a_i L_i(2\lambda t), \quad \lim_{n\to\infty} \int_0^\infty t^{2k} \mathrm{e}^{-2wt} (f(t) - h_n(t))^2 \mathrm{d}t = 0.$$

With the second one, the same sequence converges uniformly to f on compact sets of $\mathbb{R}^+$.

The idea is to obtain a quickly decreasing sequence $a_i(\lambda)$, and the computations are very sensitive to that choice, even if the theoretical results are not. The choice of λ must be made in order to speed up the convergence of the power series $\sum_{i\geq 0} a_i u^i$ (i.e. the convergence of $\bar{f}_n(p)$ to $\bar{f}(p)$). Let

$$\varphi(u) = \sum_i a_i u^i, \quad (p+\lambda)\bar{f}(p) = \varphi(u), \quad u = \frac{p-\lambda}{p+\lambda}.$$

The singularities of φ are $(\alpha-\lambda)/\alpha+\lambda)$ with α a singularity of $\bar{f}$. So R_λ, the radius of convergence of φ, is $R_\lambda = 1/(\max_i |LA_i|)$ with L being the point represented by the complex number 1, and A_i being represented by $2\lambda/(\lambda - \alpha_i)$. So, the best λ is obtained by minimizing $\max_i |LA_i|$. Computations have been made and compared with those of Longman (1973) using Padé approximants. For the examples studied (with poles, branch points or isolated essential singularities) they give better results.

Let us now summarize two examples: the first one shows the improvement due to the choice of λ for the approximant $F_7(t)$ (the case $\lambda = 1$ has been obtained by Ward in Sneddon (1972) and the exact inverse $F(t) = e^{-t} - e^{-2t}$ is known); for the second example the results are obtained with the weight function, depending on k, $t^{2k}e^{-2kt}$, the choice of λ being optimal. The last line of each array contains the exact results when known.

$\bar{f}(p) = 1/(p+1)(p+2)$

λ	$t=4$	$t=1$	$t=0.5$
1	.01797	.23263	.23866
1.2	.0179797	.2325430	.2386533
1.4	.0179802	.2325440	.2386513
	.0179801	.2325441	.2386512

$\bar{f}(p) = (1/p)\ln(1+p)$

k	$t=4$	$t=2$	$t=0.5$
1	.013179	.049133	.558368
3	.013044	.048899	.559835
5	.013048	.048900	.559787
7	.013054	.048897	.559662
		.048900	.559773

It is obvious that for such unstable problems, no single method will give optimal results for all purposes and all occasions.

As the Borel transform is also a Laplace transform, Padé and Padé-type approximants can be used for its inversion, as explained by Marziani (1987). Let us first recall the basis of the Borel method, namely, the Watson–Nevanlinna theorem:

Theorem 4.7 Let $\alpha > 0$, $R > 0$ and $A > 0$ be given. We set $D_{\alpha,R} = \{z \in \mathbb{C}, 0 < |z| < R, |\arg z| \le \alpha + \pi/2\}$, $T_{\alpha,A} = \{z \in \mathbb{C}, |z| < 1/A\} \cup \{z \in \mathbb{C}, |\arg z| < \alpha\}$. Let f be analytic in $D_{\alpha,R}$ and continuous on $\bar{D}_{\alpha,R}$ and have there the asymptotic expansion $f(z) = \sum_{n=0}^{\infty} c_n z^n \quad (z \to 0)$. We assume that there exists $C > 0$ such that $\forall z \in D_{\alpha,R}$ and $\forall N$

$$\left| f(z) - \sum_{n=0}^{N} c_n z^n \right| \le C(N+1)! A^{N+1} |z|^{N+1}.$$

Then the Borel transform series $B(z) = \displaystyle\sum_{n=0}^{\infty} \frac{c_n}{n!} z^n$ converges in $\{z \in \mathbb{C}, |z| < 1/A\}$, $B(z)$ has an analytic continuation $g(z)$ in $T_{\alpha,A}$, the integral $F(z) = \displaystyle\frac{1}{z}\int_0^{\infty} e^{-t/z} g(t)dt$ is absolutely convergent $\forall z \in \{z \in \mathbb{C}, |z| < R, |\arg z| < \alpha\}$ and $F(z) = f(z)$.

Thus the integral $F(z)$ provides a formal sum for the asymptotic series $f(z)$ and the Taylor expansion of F around the origin coincides with the series f.

The main drawback of this method is that usually g is not known since, in practice, only a finite set of numerically computed coefficients c_n is available. The series B cannot be used either since it converges only for $|z| < 1/A$. Thus, the idea was to replace $g(t)$ in the definition of F by $[n+k/k]_B(t)$

with $n \geq -1$, giving rise to the Borel–Padé approximants

$$F_B^{[n+k/k]}(z) = \frac{1}{z}\int_0^\infty e^{-t/z}[n+k/k]_B(t)\mathrm{d}t.$$

To prove that these Borel–Padé approximants tend to $f(z)$ when k tends to infinity, one has first to prove that $[n+k/k]_B$ tends to B uniformly when $k \to \infty$. Usually this is not possible and this was the reason why Marziani (1987) replaced the Padé approximant $[n+k/k]_B$ by the Padé-type approximant $(n+k/k)$ thus obtaining the so-called Borel–Padé-type approximant denoted by $F_B^{(n+k/k)}$. Using the convergence results for Padé-type approximants, he was able to prove the following theorem:

Theorem 4.8 Let f be an analytic function satisfying the assumptions of the preceding theorem with α arbitrarily close to π. If the generating polynomials v_k are the Chebyshev polynomials of the first kind $v_k(x) = T_k(2x/A+1)$, then $\forall n \geq -1$

$$f(z) = \lim_{k\to\infty} F_B^{(n+k/k)}(z)$$

for every z in the half plane $\{z, \mathrm{Re}(z) > 0\}$.

The numerical results given by Marziani (1987) show that the Borel–Padé-type approximants converge with almost the same rate as the Borel-Padé approximants. The main advantage is that one has complete control of the poles when using the Padé-type approximants and thus a proof of the convergence of the method can be obtained.

Let us end this section with the z-transform. It is a functional transformation of sequences that can be considered as equivalent to the Laplace transform for functions. While the Laplace transform is useful in solving differential equations, the z-transform plays a central rôle in the solution of difference equations. If one changes z into z^{-1}, it is identical to the method of generating functions introduced by the French mathematician François Nicole (1683–1758) and developed by Joseph Louis Lagrange (1736–1813). It has many applications in digital filtering and in signal processing as exemplified by Vich (1987). By signal processing we mean the transformation of a function f of the time t, called the input signal, into an output signal h. This transformation is realized via a system G called a digital filter. f can be known for all values of t, and in that case we speak of a continuous signal and a continuous filter, or it can only be known at equally spaced values of t, $t_n = nT$ for $n = 0, 1, \ldots$, where T is the period, and, in that case, we speak of a discrete signal and a discrete filter. The z-transform of a discrete signal is given by $F(z) = \sum_{n=0}^\infty f_n z^{-n}$, where $f_n = f(nT)$. Corresponding to the input sequence (f_n) is the output sequence $(h_n = h(nT))$. If we set $H(z) = \sum_{n=0}^\infty h_n z^{-n}$ then the system G can be represented by its so-called transfer function $G(z)$ such that $H(z) = G(z)F(z)$. In other words, if we write $G(z) = \sum_{n=0}^{\infty} g_n z^{-n}$ then $h_n = \sum_{k=0}^{n} f_k g_{n-k}$, $n = 0, 1, \ldots$. Thus if (f_n) and (h_n) are known, then (g_n) can be computed. An important problem in the analysis of digital filters is the identification of the transfer function when (f_n) and (h_n) are known. If the filter is linear then G is a rational function of z and, if not, its transfer function can be approximated by a rational function $R(z) = P(z)/Q(z)$ which is in fact the Padé approximant $[s/s]_G(z)$.

4.3. Systems of equations

As explained in Section 1.6, the scalar ε-algorithm (Wynn, 1956) is a recursive method for computing Padé approximants. It is also a powerful convergence-acceleration process (see, for example, Brezinski and Redivo-Zaglia (1991)). Since, in numerical analysis, one often has to deal with vector sequences, the ε-algorithm was generalized to the vector case. Also as usual, when generalizing from one dimension to several, several possible generalizations exist. However, in our case, they all have some properties in common since they were all built in order to compute exactly the vector S for sequences of vectors (S_n) such that, $\forall n$,

$$a_0(S_n - S) + \cdots + a_k(S_{n+k} - S) = 0,$$

where $a_0, \ldots, a_k$ are scalars such that $a_0 + \cdots + a_k \neq 0$.

Due to this property, the various generalizations of the ε-algorithm (which obviously give rise to the corresponding generalizations of Padé approximants for series with vector coefficients) have applications in linear algebra. Indeed, let us consider the sequence of vectors (x_n) generated by

$$x_{n+1} = Bx_n + b,$$

where B is a square matrix such that $A = I - B$ is regular and b is a vector. Let x be the unique solution of the system $Ax = b$. Then $x_n - x = B^n(x_0 - x)$, where x_0 is the initial vector of the sequence (x_n). Let $P_k(t) = a_0 + a_1 t + \cdots + a_k t^k$ be the minimal polynomial of the matrix B for the vector $x_0 - x$, that is, the polynomial of the minimal degree such that $P_k(A)(x_0 - x) = 0$. Then, $\forall n, A^n P_k(A)(x_0 - x) = a_0(x_n - x) + \cdots + a_k(x_{n+k} - x) = 0$. Moreover, since A is assumed to be regular, 1 is not an eigenvalue of B, that is, $P_k(1) = a_0 + \cdots + a_k \neq 0$. It follows, from the aforementioned property shared by the various generalizations of the ε-algorithm, that, when applied to the sequence (x_n), they will all lead to $\varepsilon_{2k}^{(n)} = x \forall n$. Thus, all these generalizations (and also the scalar ε-algorithm applied componentwise) are direct methods for solving systems of linear equations.

As we shall see below, the connection between extrapolation methods and linear algebra is still deeper. Moreover, such algorithms can also be used for solving systems of nonlinear equations. They are derivative-free and exhibit a quadratic convergence under the usual assumptions.

• *The vector ε-algorithm*

Let us give an application of Theorem 3.8 to the solution of systems of linear equations. We have the following theorem (Brezinski, 1974):

Theorem 4.9 Let us apply the vector ε-algorithm to the sequence

$$x_{n+1} = Bx_n + b, \qquad n = 0, 1, \ldots,$$

where x_0 is a given arbitrary vector.

If $A = I - B$ is regular, if m is the degree of the minimal polynomial of B for the vector $x_0 - x$ and if 0 is a zero of multiplicity r (possibly $= 0$) of this polynomial, then $\forall n \geq 0$

$$\varepsilon_{2(m-r)}^{(n+r)} = x.$$

If A is singular, if b belongs to its range (which means that the system has infinitely many solutions), if m and r are defined as above and if q denotes the multiplicity of the zero equal to 1 of this polynomial, then, if $q = 1$ we have $\forall n \geq 0$

$$\varepsilon_{2(m-r)-2}^{(n+r)} = x,$$

where x is one of the solutions of the system. If $q = 2$, then $\forall n \geq 0$

$$\varepsilon_{2(m-r)-3}^{(n)} = y,$$

where y is a constant vector independent of n.

If A is singular, if b does not belong to its range (which means that the system has no solution), if m is the degree of the minimal polynomial of B for the vector $x_1 - x_0$, if 0 is a zero of multiplicity r (possibly $= 0$) of this polynomial and if q denotes the multiplicity of its zero equal to 1, then, if $q = 1$, we have $\forall n \geq 0$

$$\varepsilon_{2(m-r)-1}^{(n+r)} = z,$$

where z is a constant vector independent of n.

This theorem was generalized to the case where the sequence (x_n) is generated by

$$x_{n+1} = \sum_{i=0}^{k} B_i x_{n-i} + b,$$

where the B_i's are square matrices (Brezinski, 1974).

• *The topological ε-algorithm*

We saw above that the coefficients a_i appearing in the recurrence relation assumed to be satisfied by the S_n's are obtained by writing

$$a_0(y, \Delta S_{n+i}) + \cdots + a_k(y, \Delta S_{n+k+i}) = 0$$

for $i = 0, \ldots, k-1$, where y is an arbitrary vector. Another possibility consists in taking $i = 0$ in this relation and choosing several linearly independent vectors y_i instead of y. Again, solving the system of equations (together with $a_0 + \cdots + a_k = 1$) provides the a_i's and thus S. A recursive algorithm for implementing this procedure (Brezinski, 1975) was obtained by Jbilou (1988). It is called the $S\beta$-algorithm. When $y_i = e_i$, the method proposed by Henrici (1964) is recovered. Its theory was developed in detail by Sadok (1990) and a recursive algorithm for its implementation, called the H-algorithm, was given by Brezinski and Sadok (1987).

Thus, from Theorem 3.8, the scalar (applied component-wise), the vector and the topological ε-algorithms and the $S\beta$-algorithm are direct methods for solving systems of linear equations. However, it must be noticed that, due to their storage requirements, the ε-algorithms are not competitive with other direct methods from the practical point of view.

The topological ε-algorithm is also related to the method due to Lanczos (1952) for solving a system of linear equations $Ax = b$. This method consists in constructing a sequence of vectors (x_k) such that

- $x_k - x_0 \in \text{span}(r_0, Ar_0, \ldots, A^{k-1}r_0)$,
- $r_k = b - Ax_k \perp \text{span}(y, A^*y, \ldots, A^{*^{k-1}}y)$,

where x_0 and y are arbitrary vectors and $r_0 = b - Ax_0$. These relations completely define the vectors x_k if they exist. An important property of the Lanczos method is its finite termination, namely, that $\exists k \leq p$ (the dimension of the system) such that $x_k = x$.

The vectors x_k and r_k can be recursively computed by several algorithms, the most well known being the biconjugate-gradient method due to Fletcher (1976) which becomes the conjugate gradient of Hestenes and Stiefel (1952) when the matrix A is symmetric and positive definite. The other algorithms for implementing the Lanczos method can be deduced from the theory of formal orthogonal polynomials (Brezinski and Sadok, 1993) thus showing the link with Padé approximants as studied by Gutknecht (1990). Thanks to the theory of formal orthogonal polynomials, the vectors r_k of the Lanczos method can be expressed as the ratios of two determinants. After some manipulations on the rows and the columns of these determinants and using the relation $B = I - A$, it can be proved (Brezinski, 1980) that the vectors x_k generated by the Lanczos method are identical to the vectors $\varepsilon_{2k}^{(0)}$ obtained by applying the topological ε-algorithm to the sequence $y_{n+1} = By_n + b$ with $y_0 = x_0$. With the determinantal formula of the topological ε-algorithm, a determinantal expression for the iterates of the CGS algorithm of Sonneveld (1989), which consists in squaring the formal orthogonal polynomials involved in the Lanczos method, was obtained by Brezinski and Sadok (1993).

There are many more connections between methods of numerical linear algebra and extrapolation algorithms but it is not our purpose here to emphasize this point. We shall refer the interested reader to Brezinski (1980), Sidi (1988), Brezinski and Redivo-Zaglia (1991), Brezinski and Sadok (1992) and Brezinski (1993).

• *Systems of nonlinear equations*

Let us consider the nonlinear fixed-point problem $x = F(x)$, where F is an application of $\mathbb{R}^p$ into itself. This problem can be solved by Newton's method which constructs a sequence of vectors converging quadratically to x under some assumptions. The main drawback of Newton's method is that it needs the knowledge of the Jacobian matrix of F, which is not always easily available. Quasi-Newton methods provide an alternative but with a slower rate of convergence. When $p = 1$, a well-known method is Steffensen's, which has a quadratic convergence under the same assumptions as Newton's method. Steffensen's method is based on Aitken's Δ^2 process and it does not need the knowledge of the derivative of F. Since the ε-algorithm generalizes Aitken's process, the problem arises of finding a generalization of Steffensen's method for solving a system of p nonlinear equations written in the form $x = F(x)$. This algorithm (Brezinski, 1970; Gekeler, 1972) is as follows:

- choose x_0,
- for $n = 0, 1, \ldots$ until convergence
- set $u_0 = x_n$,
- compute $u_{i+1} = F(u_i)$ for $i = 0, \ldots, 2p_n - 1$, where p_n is the degree of the minimal polynomial of $F'(x)$ for the vector $x_n - x$,
- apply the ε-algorithm to the vectors $u_0, \ldots, u_{2p_n}$,
- set $x_{n+1} = \varepsilon_{2p_n}^{(0)}$.

In this method, either the scalar (component-wise), the vector or the topological ε-algorithm could be used. However, the following theorem (Le Ferrand, 1992a) was only proved in the case of the topological ε-algorithm although all the numerical experiments show that it might also be true for the two other ε-algorithms. The proof is based on the determinantal expression of the vectors computed by the topological ε-algorithm. A similar result was also proved to hold for the vector Padé approximants (Van Iseghem, 1994). Let H_n be the matrix

$$H_n = \frac{1}{\|\Delta u_0\|} \begin{pmatrix} 1 & \cdots & 1 \\ (y, \Delta u_0) & \cdots & (y, \Delta u_{p_n}) \\ \vdots & & \vdots \\ (y, \Delta u_{p_n - 1}) & \cdots & (y, \Delta u_{2p_n - 1}) \end{pmatrix}.$$

Theorem 4.10 If the matrix $I - F'(x)$ is regular, if F' satisfies a Lipschitz condition and if $\exists N, \exists \alpha > 0$ such that $\forall n > N, |\det H_n| > \alpha$, then the sequence (x_n) generated by the previous algorithm converges quadratically to x for any x_0 in a neighborhood of x.

Henrici's method is also a method with a quadratic convergence for systems of nonlinear equations (Ortega and Rheinboldt, 1970). It is equivalent to the vector Padé approximants when $k = d + 1$. It has the same general structure as the algorithm given above after replacing $2p_n$ by $p_n + 1$ and using either the $S\beta$-algorithm or the H-algorithm instead of one of the ε-algorithms.

REFERENCES

R. Alt (1972),'Deux théorèmes sur la A-stabilité des schémas de Runge–Kutta simplement implicites', *RAIRO* **R3**, 99–104.

G. A. Baker, Jr. (1970), 'The Padé approximant method and some related generalizations', in *The Padé Approximant in Theoretical Physics* (G. A. Baker, Jr., and J. L. Gammel, eds.), Academic Press (New York), 1–39.

G. A. Baker, Jr. (1975), *Essentials of Padé Approximants*, Academic Press (New York).

G. A. Baker, Jr. and P. R. Graves-Morris (1977),'Convergence of rows of the Padé table', *J. Math. Anal. Appl.* **57**, 323–339.

G. A. Baker, Jr. and P. R. Graves-Morris (1981), *Padé Approximants*, Addison-Wesley (Reading).

A. F. Beardon (1968), 'The convergence of Padé approximants', *J. Math. Anal Appl.* **21**, 344–346.

C. M. Bender and S. A. Orszag (1978), *Advanced Mathematical Methods for Scientists and Engineers*, McGraw–Hill (New York).

C. Brezinski (1970), 'Application de l'ε-algorithme à la résolution des systèmes non linéaires', *C.R. Acad. Sci. Paris* **271 A**, 1174–1177.

C. Brezinski (1974), 'Some results in the theory of the vector ε-algorithm', *Linear Algebra Appl.* **8**, 77–86.

C. Brezinski (1975), 'Généralisation de la transformation de Shanks, de la table de Padé et de l'ε-algorithme', *Calcolo* **12**, 317–360.

C. Brezinski (1977), *Accélération de la Convergence en Analyse Numérique*, LNM 584, Springer-Verlag (Berlin).

C. Brezinski (1980), *Padé-Type Approximation and General Orthogonal Polynomials*, ISNM vol. 50, Birkhäuser (Basel).

C. Brezinski (1983a), 'Outlines of Padé approximation', in *Computational Aspects of Complex Analysis* (H. Werner et al., eds.), Reidel (Dordrecht), 1–50.

C. Brezinski (1983b), 'Recursive interpolation, extrapolation and projection', *J. Comput. Appl. Math.* **9**, 369–376.

C. Brezinski (1991), *History of Continued Fractions and Padé Approximants*, Springer-Verlag (Berlin).

C. Brezinski (1991a), *A Bibliography on Continued Fractions, Padé Approximation, Extrapolation and Related Subjects*, Prensas Universitarias de Zaragoza (Zaragoza).

C. Brezinski (1991b), *Biorthogonality and Its Applications to Numerical Analysis*, Dekker (New York).

C. Brezinski (1993), 'Biorthogonality and conjugate gradient-type algorithms', in *Contributions in Numerical Mathematics* (R. P. Agarwal ed.), World Scientific (Singapore), 55–70.

C. Brezinski (1994), 'The methods of Vorobyev and Lanczos', *Linear Algebra Appl.*, to appear.

C. Brezinski and M. Redivo-Zaglia (1991), *Extrapolation Methods: Theory and Practice*, North-Holland (Amsterdam).

C. Brezinski and H. Sadok (1987), 'Vector sequence transformations and fixed point methods', in *Numerical Methods in Laminar and Turbulent Flows*, vol. 1 (C. Taylor et al., eds.), Pineridge Press (Swansea), 3–11.

C. Brezinski and H. Sadok (1992), 'Some vector sequence transformations with applications to systems of equations', *Numerical Algor.* **3**, 75–80.

C. Brezinski and H. Sadok (1993), 'Lanczos-type algorithms for solving systems of linear equations', *Appl. Numer. Math.* **11**, 443–473.

C. Brezinski and J. Van Iseghem (1994), 'Padé approximations', in *Handbook of Numerical Analysis*, vol. 3 (P. G. Ciarlet and J. L. Lions, eds.), North-Holland (Amsterdam).

M. G. de Bruin (1984), 'Simultaneous Padé approximation and orthogonality', in *Polynômes Orthogonaux et Applications* (C. Brezinski et al., eds.), LNM 1171, Springer-Verlag (Berlin), 74–83.

M. G. de Bruin and H. Van Rossum (1975), 'Formal Padé approximation', *Nieuw Arch. Wiskd.* Ser. 23, **3** 115–130.

V. I. Buslaev, A. A. Gončar and S. P. Suetin (1984), 'On the convergence of subsequences of the mth row of the Padé Table', *Math. USSR Sb.* **48**, 535–540.

F. Cala Rodriguez and H. Wallin (1992), 'Padé-type approximants and a summability theorem by Eiermann', *J. Comput. Appl. Math.* **39**, 15–21.

F. Cordellier (1977), 'Particular rules for the vector ε-algorithm', *Numer. Math.* **27**, 203–207.

F. Cordellier (1979), 'Démonstration algébrique de l'extension de l'identité de Wynn aux tables de Padé non normales', in *Padé Approximation and Its Applications* (L. Wuytack ed.), LNM 765, Springer-Verlag (Berlin), 36–60.

M. Crouzeix and F. Ruamps (1977), 'On rational approximations to the exponential', *RAIRO Anal. Numer.* **R11**, 241–243.

A. A. M. Cuyt and L. Wuytack (1987), *Nonlinear Methods in Numerical Analysis*, North-Holland (Amsterdam).

A. Draux (1983), *Polynômes Orthogonaux Formels: Applications*, LNM 974, Springer-Verlag (Berlin).

A. Draux and P. Van Ingelandt (1986), *Polynômes Orthogonaux et Approximants de Padé: Logiciels*, Technip (Paris).

B. L. Ehle (1973), 'A-stable methods and Padé approximation to the exponential', *SIAM J. Math. Anal.* **4**, 671–680.

M. Eiermann (1984), 'On the convergence of Padé-type approximants to analytic functions', *J. Comput. Appl. Math.* **10**, 219–227.

J. Favard (1935), 'Sur les polynômes de Tchebicheff', *C.R. Acad. Sci. Paris* **200**, 2052–2053.

R. Fletcher (1976), 'Conjugate gradient methods for indefinite systems', in *Numerical Analysis, Dundee 1975*, LNM 506, Springer-Verlag (Berlin), 73–89.

E. Gekeler (1972), 'On the solution of systems of equations by the epsilon algorithm of Wynn', *Math. Comput.* **26**, 427–436.

J. Gilewicz (1978), *Approximants de Padé*, LNM 667, Springer-Verlag (Berlin).

C. González-Concepción (1987), 'On the A-acceptability of Padé-type approximants to the exponential with a single pole', *J. Comput. Appl. Math.* **19**, 133–140.

W. B. Gragg (1972), 'The Padé table and its relation to certain algorithms of numerical analysis', *SIAM Rev.* **14**, 1–62.

P. R. Graves-Morris (1983), 'Vector-valued rational interpolants I', *Numer. Math.* **42**, 331–348.

P. R. Graves-Morris (1994), 'A review of Padé methods for the acceleration of convergence of a sequence of vectors', *Appl. Numer. Math.*, to appear.

M. H. Gutknecht (1990), 'The unsymmetric Lanczos algorithms and their relations to Padé approximation, continued fractions, and the QD algorithm', in *Proceedings of the Copper Mountain Conference on Iterative Methods.*

M. H. Gutknecht (1992), 'A completed theory of the unsymmetric Lanczos process and related algorithms, Part I', *SIAM J. Matrix Anal. Appl.* **13**, 594–639.

A. J. Guttmann (1989), 'Asymptotic analysis of power-series expansions', in *Phase Transitions and Critical Phenomena*, vol. 13 (C. Domb and J. L. Lebowitz, eds.), Academic Press (New York), 1–234.

P. Henrici (1964), *Elements of Numerical Analysis*, Wiley (New York).

P. Henrici (1974), *Applied and Computational Complex Analysis*, vol. 1, Wiley (New York).

M. R. Hestenes and E. Stiefel (1952), 'Methods of conjugate gradients for solving linear sytems', *J. Res. Natl. Bur. Stand.* **49**, 409–436.

K. Jbilou (1988), 'Méthodes d'Extrapolation et de Projection: Applications aux Suites de Vecteurs', Thesis, Université des Sciences et Technologies de Lille.

W. B. Jones and W. J. Thron (1988), 'Continued fractions in numerical analysis', *Appl. Numer. Math.* **4**, 143–230.

J. Karlson and B. Von Sydow (1976), 'The convergence of Padé approximants to series of Stieltjes', *Arkiv for Math.* **14**, 44–53.

C. Lanczos (1952), 'Solution of systems of linear equations by minimized iterations', *J. Res. Natl. Bur. Stand.* **49**, 33–53.

H. Le Ferrand (1992a), 'The quadratic convergence of the topological epsilon algorithm for systems of nonlinear equations', *Numerical Algor.* **3**, 273–284.

H. Le Ferrand (1992b), 'Convergence et Applications d'Approximations Rationnelles Vectorielles', Thesis, Université des Sciences et Technologies de Lille.

I. Longman (1971), 'Computation of the Padé table', *Int. J. Comput. Math.* **3B**, 53–64.

I. Longman (1973), 'On the generation of rational function approximations for Laplace transform inversion with an application to viscoelasticity', *SIAM J. Appl. Math.* **24**, 429–440.

L. Lorentzen and H. Waadeland (1992), *Continued Fractions with Applications*, North-Holland (Amsterdam).

Y. L. Luke (1969), *The Special Functions and Their Approximation, vol. 2*, Academic Press (New York).

A. P. Magnus (1988), *Approximations Complexes Quantitatives*, Cours de Questions Spéciales en Mathématiques, Facultés Universitaires Notre Dame de la Paix, Namur.

A. A. Markov (1948), *Selected Papers on Continued Fractions and the Theory of Functions Deviating Least from Zero* (in Russian) (N. I. Achyeser, ed.), Ogiz (Moscow).

M. F. Marziani (1987), 'Convergence of a class of Borel-Padé type approximants', *Nuovo Cimento* **99 B**, 145–154.

J. B. McLeod (1971), 'A note on the ε-algorithm', *Computing* **7**, 17–24.

R. de Montessus de Ballore (1902), 'Sur les fractions continues algébriques', *Bull. Soc. Math. France* **30**, 28–36.

J. M. Ortega and W. C. Rheinboldt (1970), *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press (New York).

J. Nuttall (1967), 'Convergence of Padé approximants for the Bethe–Salpeter amplitude', *Phys. Rev.* 157, 1312–1316

H. Padé (1984), *Oeuvres* (C. Brezinski, ed.), Librairie Scientifique et Technique Albert Blanchard (Paris).

O. Perron (1957), *Die Lehre von den Kettenbrüchen*, vol. 2, Teubner (Stuttgart).

M. Prévost (1983), 'Padé-type approximants with orthogonal generating polynomials', *J. Comput. Appl. Math.* **9**, 333–346.

M. Prévost (1990), 'Cauchy product of distributions with applications to Padé approximants', Note ANO–229, Université des Sciences et Technologies de Lille.

H. Rutishauser (1954), 'Der Quotienten–Differenzen–Algorithms', *Z. Angew. Math. Physik* **5**, 233–251.

H. Sadok (1990), 'About Henrici's transformation for accelerating vector sequences', *J. Comput. Appl. Math.* **29**, 101–110.

E. B. Saff (1972), 'An extension of Montessus de Ballore's theorem on the convergence of interpolating rational functions', *J. Approx. Theory* **6**, 63–67.

A. Salam (1994), 'Non-commutative extrapolation algorithms', *Numerical Algor.* **7**, 225–251.

D. Shanks (1955), 'Nonlinear transformations of divergent and slowly convergent sequences', *J. Math. Phys.* **34**, 1–42 .

J. Shohat (1938), 'Sur les polynômes orthogonaux généralisés', *C.R. Acad. Sci. Paris* **207**, 556–558.

A. Sidi (1981), 'The Padé table and its connection with some weak exponential function approximations to Laplace transform inversion', in *Padé Approximation and Its Applications, Amsterdam 1980* (M. G. De Bruin and H. Van Rossum, eds.), LNM 888, Springer-Verlag (Berlin), 352–362.

A. Sidi (1988), 'Extrapolation vs. projection methods for linear systems of equations', *J. Comput. Appl. Math.* **22**, 71–88.

A. Sidi, W. F. Ford and D. A. Smith (1986), 'Acceleration of convergence of vector sequences', *SIAM J. Numer. Anal.* **23**, 178–196. ff

I. N. Sneddon (1972), *The Use of Integral Transforms*, McGraw–Hill (New York).

P. Sonneveld (1989), 'CGS, a fast Lanczos-type solver for nonsymmetric linear systems', *SIAM J. Sci. Stat. Comput.* **10**, 36–52.

J. Van Iseghem (1984), 'Padé-type approximants of $\exp(-z)$ whose denominators are $(1+z/n)^n$', *Numer. Math.* **43**, 282–292.

J. Van Iseghem (1985), 'Vector Padé approximants', in *Numerical Mathematics and Applications* (R. Vichnevetsky and J. Vignes, eds.), North-Holland (Amsterdam), 73–77.

J. Van Iseghem (1986), 'An extended cross rule for vector Padé approximants', *Appl. Numer. Math.* **2**, 143–155.

J. Van Iseghem (1987a), 'Laplace transform inversion and Padé-type approximants', *Appl. Numer. Math.* **3** , 529–538.

J. Van Iseghem (1987b), 'Approximants de Padé Vectoriels', Thesis, Université des Sciences et Technologies de Lille.

J. Van Iseghem (1989), 'Convergence of the vector QD algorithm. Zeros of vector orthogonal polynomials', *J. Comput. Appl. Math.* **25**, 33–46. ff

J. Van Iseghem (1992), 'Best choice of the poles for the Padé-type approximant of a Stieltjes function', *Numerical Algor.* **3**, 463–476.

J. Van Iseghem (1994), 'Convergence of vectorial sequences: applications', *Numer. Math.*, to appear.

H. Van Rossum (1953), 'A Theory of Orthogonal Polynomials Based on the Padé Table', Thesis, University of Utrecht, Van Gorcum (Assen).

R. Vich (1987), *Z-Transform: Theory and Applications*, Reidel (Dordrecht).

H. Wallin (1974), 'The convergence of Padé approximants and the size of the power series coefficients', *Applicable Analysis* **4**, 235–251.

H. Wallin (1987), 'The divergence problem for multipoint Padé approximants of meromorphic functions', *J. Comput. Appl. Math.* **19**, 61–68.

G. Wanner (1987), 'Order stars and stability', in *The State of the Art in Numerical Analysis* (A. Iserles and M. J. D. Powell, eds.), Clarendon Press (Oxford), 451–471.

P. J. S. Watson (1973), 'Algorithms for differentiation and integration', in *Padé Approximants and Their Applications* (P. R. Graves-Morris, ed.), Academic Press (New York), 93–97.

P. Wynn (1956), 'On a device for computing the $e_m(S_n)$ transformation', *MTAC* **10**, 91–96.

P. Wynn (1962), 'Acceleration techniques for iterated vector and matrix problems', *Math. Comput.* **16**, 301–322.

P. Wynn (1966), 'Upon systems of recursions which obtain among the quotients of the Padé table', *Numer. Math.* **8**, 264–269.