ELSEVIER

Applied Numerical Mathematics 20 (1996) 299–318

APPLIED NUMERICAL MATHEMATICS

Extrapolation algorithms and Padé approximations: a historical survey

C. Brezinski [1]

Laboratoire d'Analyse Numérique et d'Optimisation, UFR IEEA, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq cedex, France

In numerical analysis there are many methods producing sequences. Such is the case of iterative methods, of methods involving series expansions, of discretization methods, that is methods depending on a parameter such that the approximate solution tends to the exact one when the parameter tends to zero, of perturbation methods, etc. Sometimes, the convergence of these sequences is so slow that their effective use is quite limited. The aim of extrapolation methods is to construct a new sequence converging to the same limit faster than the initial one. Among these methods, the most well known are certainly Richardson's extrapolation algorithm and Aitken's Δ^2 process.

In many branches of applied sciences, the solution of a given problem is often obtained as a power series expansion. The question is then trying to approximate the function from its series expansion. A possible answer is to construct a rational function whose series expansion matches the original one as far as possible. Such rational functions are called *Padé approximants*.

These two subjects, which have some connections, go quite deep and far into the history of mathematics. They are related to continued fractions (a field which goes back to the Greek antiquity), orthogonal polynomials, the moment problem, etc., they played an important role in the development of mathematics (such as the transcendence of e and π) and they have many applications.

This paper will give a short historical overview of these two subjects. Of course, we do not pretend to be exhaustive nor even to quote every important contribution. We refer the interested reader to the literature and, in particular, to the recent books [5,22,29,24,38,46,48,68,78,131]. For an extensive bibliography, see [23].

1. Extrapolation methods

Let (S_n) be the sequence to be accelerated. It is assumed to converge to a limit S. An extrapolation method consists in transforming this sequence into a new one, (T_n), by a sequence transformation

[1] E-mail: brezinsk@omega.univ-lille1.fr.

SSDI 0168-9274(95)00110-7

T: $(S_n) \to (T_n)$. The transformation T is said to accelerate the convergence of the sequence (S_n) if and only if

$$\lim_{n\to\infty} \frac{T_n - S}{S_n - S} = 0.$$

In that case, one can also say that (T_n) converges to S faster than (S_n).

1.1. The early history

The first methods to have been used were linear transformations, that is

$$T_n = \sum_{i=0}^{\infty} a_{ni} S_i, \quad n = 0, 1, \ldots,$$

where the numbers a_{ni} are constants independent of the terms of the sequence (S_n). Such a linear transformation is usually called a *summation process* and its properties are completely determined by the matrix $A = (a_{ni})$. For obvious practical reasons, only a finite number of the coefficients a_{ni} are different from zero for each n. Among such processes are those named after Euler, Cesaro and Hölder. The convergence of the sequence (T_n) to S for any converging sequence (S_n) is governed by the Toeplitz theorem; see [101] for a review. With respect to convergence acceleration, summation processes are usually only able to accelerate the convergence of restricted classes of sequences and this is why numerical analysts soon turned their efforts to nonlinear transformations. However, there is one exception which is Richardson's extrapolation process.

1.2. Richardson's process

It seems that the first appearance of a particular case of what is now called the Richardson extrapolation process is due to Christian Huygens (1629–1695). The value of π is usually computed by considering inscribed and circumscribed polygons with respect to the unit circle. If S'_n and S_n denote respectively the perimeters of these polygons with n sides, then

$$S'_n = n \sin \pi/n < \pi < S_n = n \tan \pi/n.$$

Replacing n by $1/h_n$, we have

$$S_n = S(h_n) = \pi + \frac{\pi^3}{3} h_n^2 + \frac{2\pi^5}{15} h_n^4 + \cdots.$$

In 1654, Huygens [65] had the idea of combining $S(h_n)$ and $S(h_n/2)$ to get a better approximation. He proved that

$$T_n = T(h_n) = \frac{4S(h_n/2) - S(h_n)}{3} = \pi - \frac{\pi^5}{30} h_n^4 + \frac{\pi^7}{252} h_n^6 - \cdots$$

and he obtained 35 exact digits for π with $h_n = 2^{-30}$. As we shall see below, this T_n corresponds to the Richardson process. Huygens' idea was extended by Jacques Frédéric Saigey (1797–1871) in 1856 and 1859 [111] who considered the approximations

$$S'_n = S_{2n} + \frac{1}{3}(S_{2n} - S_n),$$
$$S''_n = S'_{2n} + \frac{1}{15}(S'_{2n} - S'_n),$$
$$S'''_n = S''_{2n} + \frac{1}{63}(S''_{2n} - S''_n),$$

which are similar to those that will be given later by Romberg in the context of accelerating the convergence of the trapezoidal rule (see below). In 1903, Robert Moir Milne (1873– ?) applied the idea of Huygens for computing π [90]. Finally the same idea was exploited again by Karl Kommerell (1871–1948) in his book of 1936 [70]. As explained in [128], Kommerell can be considered as the real discoverer of Romberg's method although he also used his scheme in the context of approximating π.

Thus, let us now come to the procedures used for improving the accuracy of the trapezoidal rule for computing approximations to a definite integral. The error of this method is given by the so-called Euler–Maclaurin expansion. In 1742, Colin Maclaurin (1698–1746) [82] showed that its precision could be improved by linear combinations of the results obtained with various stepsizes. His procedure can be interpreted as a preliminary version of Romberg's method; see [41] for commentaries.

In 1900, William Fleetwood Sheppard (1863–1936) used an elimination strategy in the Euler–Maclaurin quadrature formula with $h_n = r_n h$ and $1 = r_0 < r_1 < r_2 < \cdots$ for producing a better approximation [117].

In 1910, Lewis Fry Richardson (1881–1953) eliminated the first term in a discretization process using central differences by combining the results obtained with the stepsizes h and $2h$ [104]. He called this procedure the *deferred approach to the limit* or h^2*-extrapolation*. It is

$$T_n = \frac{h_{n+1}^2 S(h_n) - h_n^2 S(h_{n+1})}{h_{n+1}^2 - h_n^2}.$$

In another paper of 1927 [105], he used the same technique for solving a 6th order differential eigenvalue problem. His process was called (h^2, h^4)*-extrapolation*. Richardson extrapolation consists in fact in computing the value at 0, denoted by $T_k^{(n)}$, of the interpolation polynomial of the degree at most k, which passes through the points $(x_n, S_n), \ldots, (x_{n+k}, S_{n+k})$. Thus, using the Neville–Aitken scheme for these interpolation polynomials, we obtain

$$T_{k+1}^{(n)} = \frac{x_{n+k+1} T_k^{(n)} - x_n T_k^{(n+1)}}{x_{n+k+1} - x_n} \quad \text{with } T_0^{(n)} = S_n.$$

Let us mention that Richardson referred to a paper by Nikolai Nikolaevich Bogolyubov (born 1909) and Nikolai Mitrofanovich Krylov (1879–1955) of 1926 where the deferred approach to the limit can already be found [10].

In 1955, Werner Romberg (born 1909) was the first to use repeatedly an elimination approach for improving the accuracy of the trapezoidal rule [106]. He himself refers to the book of Lothar Collatz (1910–1990) of 1951 [42]. The procedure became widely known after the rigorous error analysis given in 1961 by Friedrich L. Bauer [6] and the synthesis of Eduard L. Stiefel (1909–1978) [121]. Romberg's derivation of his process was mainly heuristic. It was proved by Pierre-Jean Laurent in 1963 [73] that the process follows from the Richardson process when taking $x_n = h_n^2$ and $h_n = h_0/2^n$. Laurent also gave conditions on the choice of the sequence (x_n) in order that the sequences $(T_k^{(n)})$ tend to S either

when k or n tends to infinity. As we shall see below, extensions of Romberg's method to nonsmooth integrands will lead to the E-algorithm.

Applications to the extrapolation of the numerical solutions of ordinary differential equations was studied by H.C. Bolton and H.I. Scoins in 1956 [11], Roland Bulirsch and Josef Stoer in 1964–1966 [37] and William B. Gragg [56] in 1965. The case of difference methods for partial differential equations was treated by Gurii Ivanovich Marchuk (born 1925) and V.V. Shaidurov [83]. Sturm–Liouville problems are discussed in [103]. Finally, let us mention that Heinz Rutishauser (1918–1970) pointed out in 1963 [109] that the idea of Romberg can be applied to any sequence as long as its error possesses an asymptotic expansion of a form similar to the Euler–Maclaurin's.

For a detailed history of the Richardson method, its developments and applications, see [50,69,128].

1.3. Aitken's process and the ε-algorithm

The most popular nonlinear acceleration method is certainly Aitken's Δ^2 process which is

$$T_n = \frac{S_n S_{n+2} - S_{n+1}^2}{S_{n+2} - 2S_{n+1} + S_n}, \quad n = 0, 1, \ldots .$$

It was given by Alexander Craig Aitken (1895–1967) in 1926 [2] for accelerating the convergence of Bernoulli's method for computing the dominant zero of a polynomial. Aitken pointed out that the same method was obtained by Hans von Naegelsbach (1838– ?) in 1876 in his study of Furstenau's method for solving nonlinear equations [94]. The process was also given by James Clerk Maxwell (1831–1879) in his *Treatise on Electricity and Magnetism* of 1892 [87] but neither Naegelsbach nor Maxwell used it for the purpose of acceleration. Maxwell wanted to find the equilibrium position of a pointer oscillating with an exponentially damped simple harmonic motion from three experimental measurements.

In fact, Aitken's process was known to Seki Kowa (1642?–1708) who is considered as the greatest Japanese mathematician. In his book *Hatubi Sanpō* of 1674, he used this process for computing the value of π. Let c_i be the perimeter of the polygon with 2^i sides inscribed in the circle of diameter 1. To obtain a better approximation of π, he made use of a method called *Enri*, which means *principle of the circle*, which consists in computing S by the formula

$$S = c_{16} + \frac{(c_{16} - c_{15})(c_{17} - c_{16})}{(c_{16} - c_{15}) - (c_{17} - c_{16})},$$

which is exactly Aitken's process as given above. With

$$c_{15} = 3.1415926487769856708,$$
$$c_{16} = 3.1415926523565913571,$$
$$c_{17} = 3.1415926532889027755,$$

he obtained $S = 3.14159265359$.

Aitken's Δ^2 process is exact (which means that $\forall n$, $T_n = S$) for sequences such that, $\forall n$, $a_0(S_n - S) + a_1(S_{n+1} - S) = 0$ with $a_0 a_1 \neq 0$ and $a_0 + a_1 \neq 0$. Such sequences form the *kernel* of Aitken's process. Then, the idea came of finding a transformation with the kernel

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

with $a_0 a_k \neq 0$ and $a_0 + \cdots + a_k \neq 0$. A particular case of the case $k = 2$ was already treated by Maxwell in his book of 1892 and a particular case of an arbitrary value of k was studied by T.H. O'Beirne in 1947 [97]. However, this last work remains almost unknown since it was only published as an internal report. The problem was handled in its full generality by Daniel Shanks (born 1917) in 1949 [115] and again in 1955 [116]. He considered the sequence transformation defined by

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

When $k = 1$, Shanks transformation reduces to Aitken's Δ^2 process. It can be proved that $\forall n$, $e_k(S_n) = S$ if and only if (S_n) belongs to the kernel of the transformation as given above. Let us also mention that the same ratios of determinants were obtained by R.J. Schmidt in 1941 [113] while studying a method for solving systems of linear equations.

The determinants involved in the definition of $e_k(S_n)$ have a very special structure and they are called *Hankel determinants* since they were studied by Hermann Hankel (1839–1873) in his thesis in 1861 [62]. Such determinants satisfy a five-term recurrence relationship and this relation was used by O'Beirne and Shanks for implementing the transformation by computing separately the numerators and the denominators of the $e_k(S_n)$'s. But, since a numerical analyst is unable to compute determinants (because it needs too many arithmetical operations and because rounding errors due to the computer will usually lead to a completely wrong result), a recursive procedure for computing the $e_k(S_n)$'s without computing the determinants involved in their definition was needed. This algorithm was obtained in 1956 by Peter Wynn (born 1932). It is called the ε-algorithm [132] and it is as follows

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

$$\varepsilon_{k+1}^{(n)} = \varepsilon_{k-1}^{(n+1)} + \frac{1}{\varepsilon_k^{(n+1)} - \varepsilon_k^{(n)}}.$$

It is related to Shanks transformation by

$$\varepsilon_{2k}^{(n)} = e_k(S_n) \quad \text{and} \quad \varepsilon_{2k+1}^{(n)} = 1/e_k(\Delta S_n).$$

Thus, the ε's with an odd lower index are only intermediate quantities.

1.4. Subsequent developments

Shanks transformation and the ε-algorithm were really the spark for the rebirth of the study of convergence acceleration processes. They now form an independent chapter in numerical analysis

with connections to other important topics such as orthogonal and biorthogonal polynomials, continued fractions, Padé approximants and so on. They also have applications in the solution of systems of linear and nonlinear equations, the computation of the eigenelements of a matrix and many other questions; see [29]. Among the other acceleration methods which were obtained and studied, one could mention the W-process of S. Lubkin [81], the method of Kjell J. Overholt [98], the ρ-algorithm of Wynn [133], the G-transformation of H.L. Gray, T.A. Atchison and G.V. McWilliams [58], the θ-algorithm of Claude Brezinski (born 1941) [13], the transformations of Bernard Germain-Bonne (born 1940) [55] and the various transformations due to David Levin [76]. In parallel, several applications began to appear. For example, the ε-algorithm provides a quadratically convergent method for solving systems of nonlinear equations without needing the knowledge of any derivative. This procedure was proposed simultaneously by Brezinski [12] and Eckhart Gekeler [53]. It has important applications such as the solution of boundary value problems for ordinary differential equations [34]. Many other algorithms are given in the work of Ernst Joachim Weniger [130] which also contains applications to physics or in the book of Brezinski and Michela Redivo Zaglia [29] where applications to various questions of numerical analysis and FORTRAN subroutines can be found. The book of Annie Cuyt and Luc Wuytack must also be mentioned [44]. The ε-algorithm has also received applications in statistics, see the work of Alain Berlinet [9], and the acceleration of the convergence of sequences of random variables was considered by Hélène Lavastre [74].

The implementation of convergence acceleration methods leads to important questions due to the cancellation errors coming out from the finite precision of the computers. The study of such problems was initiated by Wynn in 1963 [135] who proposed particular rules for the ε-algorithm which are more stable than the usual rule. They were extended by Florent Cordellier in 1979 [43].

The experience gained during this period led to a deeper understanding of the subject and more theoretical and general questions related to the theory of convergence acceleration methods began to be studied. The first attempt was made by R. Pennacchi in 1968 [100] who studied rational sequence transformations. His work was generalized by Germain-Bonne in 1973 [54] who proposed a framework including many methods and showed how to construct new algorithms for accelerating some classes of sequences. However, the main step was achieved by Jean Paul Delahaye and Germain-Bonne in 1980 [47] when they proved that if a set of sequences has a certain property, called *remanence*, then a universal algorithm able to accelerate *all* the sequences it contains cannot exist. This result shows the limitations of acceleration methods.

It was remarked by Moulay Driss Benchiboun [8] that all the sequence transformations in the literature can be written as

$$T_n = \frac{f(S_n, \ldots, S_{n+k})}{\mathrm{D}f(S_n, \ldots, S_{n+k})}$$

with $\mathrm{D}^2 f \equiv 0$, where $\mathrm{D}f$ denotes the sum of the partial derivatives of the function f. The reason for that property was explained by Brezinski [20] who showed that it is very much related to the translativity property of sequence transformations. These results were extended to the vector case by Hassane Sadok [110].

In many transformations, such as Shanks', the quantities computed are expressed as a ratio of two determinants. This property is very much related to the existence of a triangular recurrence scheme for their computation as explained by Brezinski and Guido Walz [36].

Thus, along the years, the theory of convergence acceleration methods was gradually built up. But the practical side was not forgotten and special devices to be used in connection with extrapolation methods in order to improve their efficiency were obtained. For example, when a certain sequence has to be accelerated, it is not obvious to know in advance which method will give the best result unless some theoretical results have been proved. Thus, Delahaye [45] proposed to program simultaneously several transformations and to select, at each step of the procedure, one answer among all the answers provided by the various algorithms. He proved that, under some assumptions, some tests are able to find automatically the best answer. The various answers could also be combined together thus leading to *composite* transformations [18]. It is also possible, in some cases, to extract a linear subsequence from the original one and then to accelerate it, for example, by Aitken's Δ^2 process [26]. Devices for controlling the error were also constructed [17].

When faced to the problem of accelerating the convergence of a given sequence, two attitudes are possible. The first one is to use a known extrapolation procedure and try to prove that it will accelerate the sequence under consideration. The second possibility consists in constructing a well adapted extrapolation method for that sequence. Convergence tests for sequences and series can be used for that purpose as explained by Brezinski [19]. This approach was mostly developed by Ana Cristina Matos [84]. Special extrapolation procedures for sequences such that $\forall n,\ S_n - S = a_n D_n$ where (D_n) is a known sequence and (a_n) an unknown one, can also be constructed from the asymptotic properties of the sequences (a_n) and (D_n) as done by Brezinski and Redivo Zaglia [28].

1.5. The E-algorithm

As we saw above, the quantities involved in Shanks transformation are expressed as a ratio of two determinants and the ε-algorithm allows to compute them recursively. It is well known that the interpolation polynomial can be expressed as a ratio of two determinants. Thus polynomial extrapolation can also be expressed by such a ratio and the Neville–Aitken scheme can be used for avoiding the computation of these determinants thus leading to the Richardson extrapolation algorithm. A similar situation arises for many other transformations: in each particular case, the quantities involved are expressed as a ratio of two determinants and, then, in each particular case, one has to look for a recursive algorithm for the practical implementation of the transformation. Thus, there was a real need for a general theory of such sequence transformations and for finding one single general algorithm allowing their implementation. This work was performed independently between 1973 and 1980 by five different people. It is now known as the E-algorithm.

It seems that the first appearance of this algorithm is due to Claus Schneider in a paper received on December 21, 1973 [114]. The quantities $S(h_i)$ being given for $i = 0, 1, \ldots$, Schneider looked for $S'(h) = S' + a_1 g_1(h) + \cdots + a_k g_k(h)$ satisfying the interpolation conditions $S'(h_i) = S(h_i)$ for $i = n, \ldots, n+k$, where the g_j's are given functions of h. Of course, the value of the unknown S' thus obtained will depend on the indexes k and n. Assuming that $\forall j,\ g_j(0) = 0$, we have $S' = S'(0)$. Denoting by ϕ_k^n the extrapolation functional on the space of functions f defined at the points $h_0 > h_1 > \cdots > 0$ and at the point 0 such that $\phi_k^n f = f(0)$, we have

$$\phi_k^n S' = c_0 S(h_n) + \cdots + c_k S(h_{n+k})$$

with $c_0 + \cdots + c_k = 1$. The interpolation conditions become

$$\phi_k^n E = 1 \quad \text{and}$$

$$\phi_k^n g_j = 0 \quad \text{for } j = 1, \ldots, k$$

with $E(h) \equiv 1$. Schneider wanted to express the functional ϕ_k^n under the form $\phi_k^n = a\phi_{k-1}^n + b\phi_{k-1}^{n+1}$. Thus

$$\phi_k^n E = a + b = 1 \quad \text{and} \quad \phi_k^n g_k = a\phi_{k-1}^n g_k + b\phi_{k-1}^{n+1} g_k = 0.$$

The values of a and b follow immediately and he finally obtained

$$\phi_k^n = \frac{[\phi_{k-1}^{n+1} g_k]\phi_{k-1}^n - [\phi_{k-1}^n g_k]\phi_{k-1}^{n+1}}{[\phi_{k-1}^{n+1} g_k] - [\phi_{k-1}^n g_k]}.$$

Thus, the quantities $\phi_k^n S'$ can be recursively computed by this scheme. The auxiliary quantities $\phi_k^n g_j$ needed in this formula will have to be computed separately by the same scheme with a different initialization. As we shall see below, this algorithm is exactly the E-algorithm. In a footnote, Schneider mentioned that this representation for ϕ_k^n was suggested by Prof. Börsch-Supan from Johannes Gutenberg Universität in Mainz.

In 1976, Günter Meinardus and G.D. Taylor wrote a paper [89] on best uniform approximation from the linear subspace $\operatorname{span}(g_1, \ldots, g_N)$ of $C[a, b]$. They defined the linear functional L_n^k on $C[a, b]$ by

$$L_n^k(f) = \sum_{i=n}^{n+k} c_i f(h_i),$$

where $a \leqslant h_1 < h_2 < \cdots < h_{N+1} \leqslant b$ and where the coefficients c_i, which depend on n and k, are such that $c_n > 0$, $c_i \neq 0$ for $i = n, \ldots, n+k$, $\operatorname{sign} c_i = (-1)^{i-n}$ and

$$\sum_{i=n}^{n+k} |c_i| = 1,$$

$$\sum_{i=n}^{n+k} c_i g_j(h_i) = 0 \quad \text{for } j = 1, \ldots, k.$$

By using Gaussian elimination for solving the system of linear equations

$$\sum_{i=n}^{N} a_i g_i(h_j) + (-1)^j \lambda = f(h_j), \quad j = 1, \ldots, k,$$

Meinardus and Taylor obtained the recursive scheme

$$L_i^k(f) = \frac{L_{i+1}^{k-1}(g_k) L_i^{k-1}(f) - L_i^{k-1}(g_k) L_{i+1}^{k-1}(f)}{L_{i+1}^{k-1}(g_k) - L_i^{k-1}(g_k)}$$

with $L_i^0(f) = f(h_i)$, $i = n, \ldots, n+k$. This is the same scheme as above.

Newton's formula for computing the interpolation polynomial is well known. It is based on divided differences. One can try to generalize these formulae to the case of interpolation by a linear combination of functions forming a complete Chebyshev system (a technical condition insuring the existence and uniqueness of the solution). So, let us look for

$$P_k^{(n)}(x) = a_0 g_0(x) + \cdots + a_k g_k(x)$$

satisfying the interpolation conditions

$$P_k^{(n)}(x_i) = f(x_i) \quad \text{for } i = n, \ldots, n+k,$$

where the x_i's are distinct points. The $P_k^{(n)}$ can be recursively computed by an algorithm which generalizes the Neville–Aitken scheme for polynomial interpolation. This algorithm was obtained by Günter Mühlbach (born 1941) in 1976 [93] from a generalization of the notion of divided differences and their recurrence relationship. This algorithm was called the Mühlbach–Neville–Aitken algorithm, for short the MNA. It is as follows

$$P_k^{(n)}(x) = \frac{g_{k-1,k}^{(n+1)}(x)P_{k-1}^{(n)}(x) - g_{k-1,k}^{(n)}(x)P_{k-1}^{(n+1)}(x)}{g_{k-1,k}^{(n+1)}(x) - g_{k-1,k}^{(n)}(x)}$$

with $P_0^{(n)}(x) = f(x_n)g_0(x)/g_0(x_n)$. The $g_{k,i}^{(n)}$'s can be recursively computed by a quite similar relationship

$$g_{k,i}^{(n)}(x) = \frac{g_{k-1,k}^{(n+1)}(x)g_{k-1,i}^{(n)}(x) - g_{k-1,k}^{(n)}(x)g_{k-1,i}^{(n+1)}(x)}{g_{k-1,k}^{(n+1)}(x) - g_{k-1,k}^{(n)}(x)}$$

with $g_{0,i}^{(n)}(x) = g_i(x_n)g_0(x)/g_0(x_n) - g_i(x)$. If $g_0(x) \equiv 1$, if it is assumed that $\forall i > 0$, $g_i(0) = 0$, then the quantities $P_k^{(n)}(0)$ are the same as those obtained by the E-algorithm and the MNA reduces to it. Let us mention that, in fact, the MNA is closely related to the work of Henri Marie Andoyer (1862–1929) which goes back to 1906 [3]; see [25] for detailed explanations.

Let us now come to the work of Tore Håvie. We already mention Romberg's method for accelerating the convergence of the trapezoidal rule. The success of this procedure is based on the existence of the Euler–Maclaurin expansion for the error. This expansion only holds if the function to be integrated has no singularity in the interval. In the presence of singularities, the expansion of the error is no longer a series in h^2 (the stepsize) but a more complicated one depending on the singularity. Thus, Romberg's scheme has to be modified accordingly in order to incorporate the various terms appearing in the expansion of the error. Several authors worked on this question, treating several types of singularities. In particular, Håvie began to study this question under Romberg who emigrated to Norway and came to Trondheim in 1949. In 1978, Håvie wrote a report, published one year later [60], where he treated the most general case of an error expansion of the form

$$S(h) - S = a_1 g_1(h) + a_2 g_2(h) + \cdots,$$

where $S(h)$ denotes the approximation obtained by the trapezoidal rule with the stepsize h of the definite integral S and the g_i are the known functions (forming an asymptotic sequence when h tends to zero) appearing in the expansion of the error. Let $h_0 > h_1 > \cdots > 0$, $S_n = S(h_n)$ and $g_i(n) = g_i(h_n)$. Håvie set

$$E_1^{(n)} = \frac{g_1(n+1)S_n - g_1(n)S_{n+1}}{g_1(n+1) - g_1(n)}.$$

Replacing S_n and S_{n+1} by their expansions, he obtained

$$E_1^{(n)} = S + a_2 g_{1,2}^{(n)} + a_3 g_{1,3}^{(n)} + \cdots$$

with

$$g_{1,i}^{(n)} = \frac{g_1(n+1)g_i(n) - g_1(n)g_i(n+1)}{g_1(n+1) - g_1(n)}.$$

The same process can be repeated for eliminating $g_{1,2}^{(n)}$ in the expansion of $E_1^{(n)}$, and so on. Thus, the E-algorithm is obtained

$$E_k^{(n)} = \frac{g_{k-1,k}^{(n+1)} E_{k-1}^{(n)} - g_{k-1,k}^{(n)} E_{k-1}^{(n+1)}}{g_{k-1,k}^{(n+1)} - g_{k-1,k}^{(n)}}$$

with $E_0^{(n)} = S_n$ and $g_{0,i}^{(n)} = g_i(n)$. The auxiliary quantities $g_{k,i}^{(n)}$ are recursively computed by the quite similar rule

$$g_{k,i}^{(n)} = \frac{g_{k-1,k}^{(n+1)} g_{k-1,i}^{(n)} - g_{k-1,k}^{(n)} g_{k-1,i}^{(n+1)}}{g_{k-1,k}^{(n+1)} - g_{k-1,k}^{(n)}}$$

with $g_{0,i}^{(n)} = g_i(n)$.

Håvie gave an interpretation of this algorithm in terms of Gaussian elimination for solving the system

$$E_k^{(n)} + b_1 g_1(n+i) + \cdots + b_k g_k(n+i) = S_{n+i}, \quad i = 0, \ldots, k$$

for the unknown $E_k^{(n)}$.

In 1980, Brezinski took up the same problem from the point of view of extrapolation [15]. Let (S_n) be the sequence to be accelerated. Interpolating it by a sequence of the form $S_n' = S + a_1 g_1(n) + \cdots + a_k g_k(n)$ where the g_i's are known sequences which can depend on the sequence (S_n) itself, leads to

$$S_{n+i} = S_{n+i}' \quad \text{for } i = 0, \ldots, k.$$

Solving this system directly for the unknown S which, since it depends on n and k, will be denoted by $E_k^{(n)}$ gives

$$E_k^{(n)} = \frac{\begin{vmatrix} S_n & \cdots & S_{n+k} \\ g_1(n) & \cdots & g_1(n+k) \\ \vdots & & \vdots \\ g_k(n) & \cdots & g_k(n+k) \end{vmatrix}}{\begin{vmatrix} 1 & \cdots & 1 \\ g_1(n) & \cdots & g_1(n+k) \\ \vdots & & \vdots \\ g_k(n) & \cdots & g_k(n+k) \end{vmatrix}}.$$

Thus $E_k^{(n)}$ is given as a ratio of two determinants which is very similar to the ratios previously mentioned. Indeed, for the choice $g_i(n) = \Delta S_{n+i}$, the ratio appearing in Shanks transformation

is recovered while, when $g_i(n) = x_n^i$, we find the ratio expressing the quantities involved in the Richardson extrapolation process, and so on.

Now the problem was to find a recursive algorithm for computing the $E_k^{(n)}$'s. Applying Sylvester's determinantal identity, Brezinski obtained the two rules of the E-algorithm as given above. His derivation of the E-algorithm is closely related to Håvie's since Sylvester's identity can be proved by using Gaussian elimination. Brezinski also gave convergence and acceleration results for this algorithm when the $(g_i(n))$ satisfy some conditions [15]. They show that, for accelerating the convergence of a sequence, it is necessary to know the expansion of the error $S_n - S$ with respect to some asymptotic sequence $(g_1(n))$, $(g_2(n))$, ... Then, these $g_i(n)$ are those to be used in the E-algorithm and it can be proved that, $\forall k$

$$\lim_{n\to\infty} \frac{E_{k+1}^{(n)} - S}{E_k^{(n)} - S} = 0.$$

These results were refined by Avram Sidi [119]. Thus the study of the asymptotic expansion of the error of the sequences to be accelerated is of primary importance as explained by Walz [129]. Other acceleration results were obtained by Matos and Marc Prévost [86] and Prévost [102]. A more economical algorithm for the implementation of this transformation was given by William F. Ford and Sidi [52]. The connection between the E-algorithm and the ε-algorithm was studied by Berhnard Beckermann [7]. A general ε-algorithm connected to the E-algorithm was given by Carsten Carstensen [40].

See [21] for a more detailed review on the E-algorithm.

1.6. A new approach

As we saw above, a quite general framework was constructed along the years for the theory of extrapolation algorithms which now lies on a firm basis. The situation was quite different for the practical construction of extrapolation algorithms and there was no systematic way for their derivation. However, thanks to a formalism due to Weniger [130], such a construction is now possible as explained by Brezinski and Matos [27]. It is as follows.

Let us assume that the sequence (S_n) to be accelerated satisfies, $\forall n$, $S_n - S = a_n D_n$ where (D_n) is a known sequence, called a *remainder* (or error) *estimate* for the sequence (S_n), and (a_n) an unknown sequence. Let us see how it is possible to construct a sequence transformation such that its kernel is precisely this set of sequences. For that purpose, we have to assume that a difference operator L (that is a linear mapping of the set of sequences into itself) exists such that $\forall n$, $L(a_n) = 0$, this meaning that the sequence obtained by applying L to (a_n) is identically zero. Such a difference operator is called an *annihilation* operator for the sequence (a_n). We have

$$\frac{S_n}{D_n} - \frac{S}{D_n} = a_n.$$

Thus, applying L and using its linearity property leads to

$$L\left(\frac{S_n}{D_n}\right) - SL\left(\frac{1}{D_n}\right) = L(a_n) = 0.$$

Then, the sequence transformation defined by

$$T_n = \frac{L(S_n/D_n)}{L(1/D_n)}$$

will be such that $\forall n,\ T_n = S$ if and only if $\forall n,\ S_n - S = a_n D_n$.

All the algorithms described above and the devices such as the error control, composite sequence transformations, least squares extrapolation, and so on, can be put into this framework which shows its usefulness. Moreover, many new algorithms can be obtained from this approach. The E-algorithm can be put into this framework thus providing a deeper insight which leads to new properties [31]. Using results from the theory of difference equations, this approach leads to new and general convergence and acceleration results as obtained by Matos [85] when (a_n) possesses some asymptotic expansion.

1.7. The vector case

In numerical analysis, many iterative methods lead to vector sequences. Of course, for accelerating such sequences, it is always possible to apply a scalar algorithm componentwise. However, vector sequence transformations, specially built for that purpose, are usually more powerful. The first vector algorithm to be studied was the vector ε-algorithm. It was obtained by Wynn [134] by replacing, in the rule of the scalar ε-algorithm, $1/\Delta\varepsilon_k^{(n)}$ by $(\Delta\varepsilon_k^{(n)})^{-1}$ where the inverse y^{-1} of a vector y is defined by $y^{-1} = y/(y,y)$. Thus, with this definition, the rule of the ε-algorithm can be applied to vector sequences. It was proved, with difficulties, by J.B. McLeod in 1971 [88] that $\forall n,\ \varepsilon_{2k}^{(n)} = S$ if the sequence (S_n) satisfies $\forall n,\ a_0(S_n - S) + \cdots + a_k(S_{n+k} - S) = 0$ with $a_0 a_k \neq 0$ and $a_0 + \cdots + a_k \neq 0$. This result is only valid for real sequences (S_n) and real a_i's. Moreover, contrarily to the scalar case, this condition is only sufficient. This result was extended to the complex case by Peter R. Graves-Morris in 1983 [57] by a quite different approach.

A drawback of the vector ε-algorithm, which was an obstacle to the development of its theory, was that it was not known if a corresponding generalization of Shanks transformation was underlying the algorithm, that is, to be more precise, if the vectors $\varepsilon_k^{(n)}$ obtained by the algorithm could be expressed as ratios of two determinants (or some kind of generalization). This was the reason why Brezinski [14] followed the same path as Shanks, trying to construct a vector sequence transformation with the kernel $a_0(S_n - S) + \cdots + a_k(S_{n+k} - S) = 0$. He obtained a transformation expressed as a ratio of two determinants and then had to find out a recursive algorithm for avoiding their computation. This was the so-called topological ε-algorithm. This algorithm has many applications, in particular in the solution of systems of linear equations since it is related to the biconjugate gradient algorithm. In the case of a system of nonlinear equations, it gave rise to a generalization of Steffensen's method [12] which has a quadratic convergence under some assumptions as proved by Hervé Le Ferrand [75]. A vector generalization of the E-algorithm can be obtained similarly. Vector sequence transformations are very much related to projection methods used in the solution of systems of equations as showed by Sidi [118] and Khalide Jbilou and Sadok [67].

It was recently proved by Ahmed Salam [112] that the vectors computed by the vector ε-algorithm can be expressed as a ratio of two *designants*, a notion replacing the notion of determinants when solving a system of linear equations in a noncommutative algebra.

2. Padé approximations

Let f be a formal power series

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

Let us consider a rational function with a numerator of the degree p at most and a denominator of the degree q at most such that its power series expansion (obtained by dividing the numerator by the denominator in ascending powers of the variable z) agrees with that of f as far as possible. Such a rational function is called a *Padé approximant* of f and it is usually denoted by $[p/q]_f(z)$. Its numerator has $p+1$ coefficients and its denominator $q+1$. But, since a rational function is defined apart from a multiplying factor (which will be taken so that the constant term of the denominator is equal to 1), there are only $p+q+1$ unknown coefficients in this Padé approximant. Thus, if $[p/q]$ exists, the two power series will agree at least up to the term of the degree $p+q$ inclusively, that is

$$f(z) - [p/q]_f(z) = O(z^{p+q+1}), \quad z \to 0.$$

It is not our purpose here to give a full account of the history of Padé approximation (which can be found in [22]) but to give mainly its relations to the Shanks transformation and the ε-algorithm.

Padé approximants were obtained by two different ways by two prominent mathematicians of the 18th century. The first of them is Johan Heinrich Lambert (1728–1777) who introduced them exactly by the definition given above in 1758 [72]. The second one is Joseph Louis Lagrange (1736–1813) who, in a paper of 1776 [71], gave a method for finding the solution of an ordinary differential equation as a continued fraction which has the advantage of converging often faster than the expansion of the solution into a power series. Since the successive convergents of this continued fraction were not easy to manipulate, he reduced them to ordinary fractions by using the recurrence relationship for their numerators and denominators. Then, he observed that their series expansions agree with that of the original series up to the term whose degree is the sum of the degrees of the numerator and the denominator inclusively. This is again the definition of Padé approximants as given above. These approximants can be written as a ratio of two determinants, a result due to Carl Gustav Jacob Jacobi (1804–1851) in 1845 [66].

The systematic study of these approximants is due to the French mathematician Henri Eugène Padé (1863–1963) in his thesis in 1892 [99] under the supervision of Charles Hermite (1822–1901). He had the idea of displaying the approximants in a two dimensional array now called the *Padé table* and he studied the case of the exponential function in detail. He also showed that approximants can only be identical if they are adjacent in the table and located in square blocks were all the approximants are identical. This property is known as the *block structure* of the Padé table. A Padé table where all the approximants are different is called *normal*. Otherwise, it is called *nonnormal*.

Many other mathematicians were interested in Padé approximants and in the related subjects of continued fractions and orthogonal polynomials. It is with the help of a kind of simultaneous Padé approximants that Hermite was able to prove the transcendence of the number e in 1873 [64]. His proof was extended in 1882 by Carl Louis Ferdinand Lindemann (1852–1939) who showed that π is a transcendental number [77] thus ending by a negative answer an open problem for more than 2000 years! Continued fractions were also a main ingredient in the construction of the new definition of an integral with respect to a general measure by Thomas Jan Stieltjes (1856–1894) in 1894 [123].

Finally, let us mention their role in the proof by Georg Cantor (1845–1918) in 1878 [39] that $\mathbb{R}$ and $\mathbb{R}^2$ have the same cardinality. Thus, during this period, continued fractions and Padé approximants were mainly used in questions relating to pure mathematics. The interest of numerical analysts in these questions was raised by the necessity of constructing approximations to special functions after the appearance of the first computers around the end of World War II and culminating with the *Handbook of Mathematical Functions* edited by Milton Abramowitz (1915–1958) and Irene A. Stegun [1].

Padé approximants are also closely related to the Shanks transformation, which is not a surprising result and can be proved just by looking at the respective ratios of determinants involved in both subjects. If the Shanks transformation, or the ε-algorithm, is applied to the sequence (S_n) of the partial sums of a power series f then

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

Thus, the ε-algorithm provides a recursive procedure for computing half of the Padé table.

Around 1960, Padé approximants were rediscovered by theoretical physicists and, in particular, by George Alan Baker Jr who used them in the investigation of some magnetic properties of the Ising model [4]. Physicists were interested in computing the locations and the residues of the poles of analytic functions and, thus, Padé approximants were well adapted to the solution of this problem. They also soon realized that they could also be used for the study of the resolvant and the spectrum of the self-adjoint operators which arise in quantum mechanics and in quantum field theory. A review of these questions can be found in [5]; see also [95]. A historical account on the use of Padé in theoretical physics is given in [92]. Since then, Padé approximants have received very many applications in various branches of science because of their quite interesting approximation properties and, in particular, their possible convergence outside the domain of convergence of the series they approximate thus leading to analytic continuation; for a recent synthesis, see [35] where the main theoretical results obtained recently, and in particular those on the convergence of sequences of Padé approximants, can be found. These questions are connected to the asymptotics for orthogonal polynomials as studied by Doron S. Lubinsky [79], Lubinsky and Edward B. Saff [80], Walter van Assche [124] and Herbert Stahl and Vilmos Totik [120]; see also [96]. They are also related to the zeros of the partial sums of the power series f as explained by Albert Edrei, Saff and Richard S. Varga [51]. Let us mention that one of the first important theorem on the convergence of a sequence of Padé approximants is due to Robert de Montessus de Ballore in 1902 [91]. It is concerned with the case when f is a meromorphic function with a known fixed number of poles in a disc.

Numerical analysts also became very much interested in Padé approximants for obvious reasons. The first problem was to compute them recursively. Although several recurrence relations were derived separately by various individuals, the derivation of all the possible relations between adjacent approximants in the Padé table was only made possible through the systematical use of formal orthogonal polynomials by Brezinski in 1980 [16]. These relations were only valid in the case of a normal table. They were extended to the nonnormal case by André Draux in his thesis in 1981 [48]. A universal subroutine, able to compute, in exact arithmetic, any sequence of Padé approximants in a table with any block structure was given by Draux and Patrick van Ingelandt [49]. Related to the recursive computation of Padé approximants, is also the QD-algorithm. It is due to Heinz Rutishauser [107] but it was already contained is some earlier work by Stieltjes dating 1889 [122]. This algorithm was developed Peter Henrici (1923–1987) [63]. It was the QD-algorithm who gave Rutishauser the idea for his LR-algorithm for computing the eigenvalues of a matrix [108]. Recently, the block structure

of the Padé table and, more simply, the related structure of the table of adjacent families of formal orthogonal polynomials served as the basic tool for solving the problems of *breakdown* (that is division by zero) in the algorithms for implementing the Lanczos method for solving nonsymmetric systems of linear equations and that of *near-breakdown* (division by a quantity close to zero). These results were obtained by Martin H. Gutknecht [59] and Brezinski, Sadok and Redivo Zaglia [32,33,30]. In numerical analysis, Padé approximants also have applications in the theory of A-stable methods for the integration of stiff ordinary differential equations [61].

Many generalizations of Padé approximants were also introduced these last 25 years. They are either dealing with different kinds of series, such as multivariate series, series with coefficients in a non-commutative algebra, series with vector coefficients, series of functions, Laurent series or with related kinds of approximations, such as the Padé-type approximants, the Cauchy approximants, the partial Padé approximants, the multipoint approximants, the Padé–Hermite approximants. Of course, these generalizations can also be combined together. Many of them can be understood within the framework of the theory of biorthogonality as explained in its most generality in [24]. The E-algorithm and some vector extrapolation algorithms can also be connected to this theory. Among these generalizations are the vector Padé approximants of Jeannette van Iseghem for approximating simultaneously (in the sense of Padé) several series by several rational approximants with a common denominator [125]. These vector approximants are based on a corresponding theory of vector orthogonal polynomials and a generalization of the QD-algorithm [126]. They can also be used for accelerating the convergence of vector sequences [127] and are related to the topological ε-algorithm.

3. Conclusion

As we tried to show in this paper, the fields of extrapolation methods and Padé approximations both go quite far and deep into the history of mathematics. They are closely related to such subjects as continued fractions and orthogonal polynomials to name only two of the main ones. They had, in the past, several quite important implications and applications and they are still very active. They are not only used for the purpose of convergence acceleration and the approximation of functions, but they also had a great influence in the solution of various questions in numerical analysis: linear algebra, A-stability of methods for integrating differential equations, the inversion of the Laplace transform and the Borel transform, numerical quadrature, systems of nonlinear equations, integral equations, optimization, Monte Carlo methods, ... They were used in the solution of many problems in theoretical physics, chemistry, engineering, statistics, electronics, ...

But, above all, they are beautiful subjects embodying many unsolved and interesting problems.

References

[1] M. Abramowitz and I.A. Stegun, eds., *Handbook of Mathematical Functions* (Dover, New York, 1965).

[2] A.C. Aitken, On Bernoulli's numerical solution of algebraic equations, *Proc. R. Soc. Edinb.* 46 (1926) 289–305.

[3] H. Andoyer, Interpolation, in: J. Molk, ed., *Encyclopédie des Sciences Mathématiques Pures et Appliquées*, Tome I, vol. 4, Fasc. 1, I–21 (Gauthier-Villars, Paris, 1904–1912) 127–160; reprinted by (Editions Gabay, Paris, 1993).

[4] G.A. Baker Jr, Application of the Padé method to the investigation of some magnetic properties of the Ising model, *Phys. Rev.* 124 (1961) 768–774.

[5] G.A. Baker Jr and P.R. Graves-Morris, *Padé Approximants*, 2 vols. (Addison-Wesley, Reading, MA, 1981); also: (Cambridge University Press, Cambridge, 2nd ed., 1996).

[6] F.L. Bauer, La méthode d'intégration numérique de Romberg, in: *Colloque sur l'Analyse Numérique* (Librairie Universitaire, Louvain, 1961) 119–129.

[7] B. Beckermann, A connection between the E-algorithm and the epsilon-algorithm, in: C. Brezinski, ed., *Numerical and Applied Mathematics* (Baltzer, Basel, 1989) 443–446.

[8] M.D. Benchiboun, Etude de Certaines Généralisations du Δ^2 d'Aitken et Comparaison de Procédés d'Accélération de la Convergence, Thèse 3ème cycle, Université de Lille I (1987).

[9] A. Berlinet, Sequence transformations as statistical tools, *Appl. Numer. Math.* 1 (1985) 531–544.

[10] N. Bogolyubov, N. Krylov, On Rayleigh's principle in the theory of differential equations of mathematical physics and upon Euler's method in the calculus of variation, *Acad. Sci. Ukraine (Phys. Math.)* 3 (1926) 3–22 (in Russian).

[11] H.C. Bolton, H.I. Scoins, Eigenvalues of differential equations by finite-difference methods, *Proc. Cambridge Phil. Soc.* 52 (1956) 215–229.

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

[13] C. Brezinski, Accélération de suites à convergence logarithmique, *C. R. Acad. Sci. Paris* 273 A (1971) 727–730.

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

[15] C. Brezinski, A general extrapolation algorithm, *Numer. Math.* 35 (1980) 175–187.

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

[17] C. Brezinski, Error control in convergence acceleration processes, *IMA J. Numer. Anal.* 3 (1983) 65–80.

[18] C. Brezinski, Composite sequence transformations, *Numer. Math.* 46 (1985) 311–321.

[19] C. Brezinski, A new approach to convergence acceleration methods, in: A. Cuyt, ed., *Nonlinear Numerical Methods and Rational Approximation* (Reidel, Dordrecht, 1988) 373–405.

[20] C. Brezinski, Quasi-linear extrapolation processes, in: R.P. Agarwal et al., eds., *Numerical Mathematics. Singapore 1988*, ISNM 86 (Birkhäuser, Basel, 1988) 61–78.

[21] C. Brezinski, A survey of iterative extrapolation by the E-algorithm, *Det Kong. Norske Vid. Selsk. Skr.* 2 (1989) 1–26.

[22] C. Brezinski, *History of Continued Fractions and Padé Approximants* (Springer, Berlin, 1990).

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

[24] C. Brezinski, *Biorthogonality and its Applications to Numerical Analysis* (Marcel Dekker, New York, 1991).

[25] C. Brezinski, The generalizations of Newton's interpolation formula due to Mühlbach and Andoyer, *Elect. Trans. Numer. Anal.* 2 (1994) 130–137.

[26] C. Brezinski, J.P. Delahaye and B. Germain-Bonne, Convergence acceleration by extraction of linear subsequences, *SIAM J. Numer. Anal.* 20 (1983) 1099–1105.

[27] C. Brezinski and A.C. Matos, A derivation of extrapolation algorithms based on error estimates, *J. Comput. Appl. Math.*, to appear.

[28] C. Brezinski and M. Redivo Zaglia, Construction of extrapolation processes, *Appl. Numer. Math.* 8 (1991) 11–23.

[29] C. Brezinski and M. Redivo Zaglia, *Extrapolation Methods. Theory and Practice* (North-Holland, Amsterdam, 1991).
[30] C. Brezinski and M. Redivo Zaglia, Treatment of near-breakdown in the CGS algorithm, *Numerical Algor.* 7 (1994) 33–73.
[31] C. Brezinski and M. Redivo Zaglia, A general extrapolation procedure revisited, *Adv. Comput. Math.* 2 (1994) 461–477.
[32] C. Brezinski, M. Redivo Zaglia and H. Sadok, Avoiding breakdown and near-breakdown in Lanczos type algorithms, *Numerical Algor.* 1 (1991) 261–274.
[33] C. Brezinski, M. Redivo Zaglia and H. Sadok, A breakdown-free Lanczos type algorithm for solving linear systems, *Numer. Math.* 63 (1992) 29–38.
[34] C. Brezinski and A.C. Rieu, The solution of systems of equations using the vector ε-algorithm, and an application to boundary value problems, *Math. Comp.* 28 (1974) 731–741.
[35] C. Brezinski and J. van Iseghem, Padé approximations, in: P.G. Ciarlet and J.L. Lions, eds., *Handbook of Numerical Analysis*, Vol. III (North-Holland, Amsterdam, 1994) 47–222.
[36] C. Brezinski and G. Walz, Sequences of transformations and triangular recursion schemes with applications in numerical analysis, *J. Comput. Appl. Math.* 34 (1991) 361–383.
[37] R. Bulirsch and J. Stoer, Numerical treatment of ordinary differential equations by extrapolation methods, *Numer. Math.* 8 (1966) 1–13.
[38] A. Bultheel, *Laurent Series and their Padé Approximations* (Birkhäuser, Basel, 1987).
[39] G. Cantor, Ein Beitrag zur Mannigfaltigkeitslehre, *J. Reine Angew. Math.* 84 (1878) 242–258.
[40] C. Carstensen, On a general epsilon algorithm, in: C. Brezinski, ed., *Numerical and Applied Mathematics* (Baltzer, Basel, 1989) 437–441.
[41] J.L. Chabert et al., *Histoire d'Algorithmes* (Belin, Paris, 1994).
[42] L. Collatz, *Numerische Behandlung von Differentialgleichungen* (Springer, Berlin, 1951).
[43] F. Cordellier, Démonstration algébrique de l'extension de l'identité de Wynn aux tables de Padé non normales, in: L. Wuytack, ed., *Padé Approximation and its Applications*, Lecture Notes in Mathematics 765 (Springer, Berlin, 1979) 36–60.
[44] A. Cuyt and L. Wuytack, *Nonlinear Methods in Numerical Analysis* (North-Holland, Amsterdam, 1987).
[45] J.P. Delahaye, Automatic selection of sequence transformations, *Math. Comp.* 37 (1981) 197–204.
[46] J.P. Delahaye, *Sequence Transformations* (Springer, Berlin, 1988).
[47] J.P. Delahaye and B. Germain-Bonne, Résultats négatifs en accélération de la convergence, *Numer. Math.* 35 (1980) 443–457.
[48] A. Draux, *Polynômes Orthogonaux Formels. Applications*, Lecture Notes in Mathematics 974 (Springer, Berlin, 1983).
[49] A. Draux and P. van Ingelandt, *Polynômes Orthogonaux et Approximants de Padé, Logiciels* (Editions Technip, Paris, 1987).
[50] J. Dutka, Richardson-extrapolation and Romberg-integration, *Historia Math.* 11 (1984) 3–21.
[51] A. Edrei, E.B. Saff and R.S. Varga, *Zeros of Sections of Power Series*, Lecture Notes in Mathematics 1002 (Springer, Berlin, 1983).
[52] W.F. Ford and A. Sidi, An algorithm for a generalization of the Richardson extrapolation process, *SIAM J. Numer. Anal.* 24 (1987) 1212–1232.
[53] E. Gekeler, On the solution of systems of equations by the epsilon algorithm of Wynn, *Math. Comp.* 26 (1972) 427–436.
[54] B. Germain-Bonne, Transformations de suites, RAIRO, R1 (1973) 84–90.
[55] B. Germain-Bonne, Estimation de la Limite de Suites et Formalisation de Procédés d'Accélération de la Convergence, Thèse d'Etat, Université de Lille I (1978).

[56] W.B. Gragg, On extrapolation algorithms for initial-value problems, *SIAM J. Numer. Anal.* 2 (1965) 384–403.
[57] P.R. Graves-Morris, Vector valued rational interpolants I, *Numer. Math.* 42 (1983) 331–348.
[58] H.L. Gray, T.A. Atchison and G.V. McWilliams, Higher order G-transformations, *SIAM J. Numer. Anal.* 8 (1971) 365–381.
[59] M.H. Gutknecht, 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* (1990).
[60] T. Håvie, Generalized Neville type extrapolation schemes, *BIT* 19 (1979) 204–213.
[61] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations* II: *Stiff and Differential-Algebraic Problems* (Springer, Berlin, 1991).
[62] H. Hankel, Ueber eine besondere Classe der symmetrischen Determinanten, Inaugural Diss., Universität Göttingen (1861).
[63] P. Henrici, The quotient-difference algorithm, *Nat. Bur. Stand. Appl. Math. Ser.* 49 (1958) 23–46.
[64] C. Hermite, Sur la fonction exponentielle, *C. R. Acad. Sci. Paris* 77 (1873) 18–24, 74–79, 226–233, 285–293.
[65] C. Huygens, De circuli magnitudine inventa, in: *Œuvres*, t. XII (1654).
[66] C.G.J. Jacobi, Über die Darstellung einer Reihe gegebener Werthe durch eine gebrochen rationale Funktion, *J. Reine Angew. Math.* 30 (1845) 127–156.
[67] K. Jbilou and H. Sadok, Some results about vector extrapolation methods and related fixed-point iterations, *J. Comput. Appl. Math.* 36 (1991) 385–398.
[68] W.B. Jones and W.J. Thron, *Continued Fractions, Analytic Theory and Applications* (Addison-Wesley, Reading, 1980).
[69] D.C. Joyce, Survey of extrapolation processes in numerical analysis, *SIAM Rev.* 13 (1971) 435–490.
[70] K. Kommerell, *Das Grenzgebiet der Elementaren und Höheren Mathematik* (Verlag Köhler, Leipzig, 1936).
[71] J.L. Lagrange, Sur l'usage des fractions continues dans le calcul intégral, *Nouveaux Mém. Acad. Sci. Berlin* 7 (1776) 236–264; *Oeuvres* 4 (1869) 301–322.
[72] J.H. Lambert, Observationes variae in mathesin puram, *Acta Helvetica* 3 (1758) 128–168.
[73] P.J. Laurent, Un théorème de convergence pour le procédé d'extrapolation de Richardson, *C. R. Acad. Sci. Paris* 256 (1963) 1435–1437.
[74] H. Lavastre, On the stochastic acceleration of sequences of random variables, *Appl. Numer. Math.* 15 (1994) 77–98.
[75] H. Le Ferrand, Convergence of the topological ε-algorithm for solving systems of nonlinear equations, *Numerical Algor.* 3 (1992) 273–284.
[76] D. Levin, Development of nonlinear transformations for improving convergence of sequences, *Int. J. Comput. Math.* B3 (1973) 371–388.
[77] F. Lindemann, Ueber die Zahl π, *Math. Ann.* 20 (1882) 213–225.
[78] L. Lorentzen and H. Waadeland, *Continued Fractions with Applications* (North-Holland, Amsterdam, 1992).
[79] D.S. Lubinsky, *Strong Asymptotics for Extremal Errors and Polynomials Associated with Erdős-type Weights*, Pitman Research Notes 202 (Longman House, Harlow, 1988).
[80] D.S. Lubinsky and E.B. Saff, *Strong Asymptotics for Extremal Polynomials Associated with Weights on* $\mathbb{R}$, Lecture Notes in Mathematics 1305 (Springer, Berlin, 1988).
[81] S. Lubkin, A method of summing infinite series, *J. Res. Nat. Bur. Stand.* 48 (1952) 228–254.
[82] C. Maclaurin, Treatise of Fluxions, Edinburgh, 1742.
[83] G.I. Marchuk and V.V. Shaidurov, *Difference Methods and their Extrapolations* (Springer, Berlin, 1983).
[84] A.C. Matos, Acceleration methods based on convergence tests, *Numer. Math.* 58 (1990) 329–340.

[85] A.C. Matos, Linear difference operators and acceleration methods, submitted.
[86] A.C. Matos and M. Prévost, Acceleration property for the columns of the E-algorithm, *Numerical Algor.* 2 (1992) 393–408.
[87] J.C. Maxwell, *A Treatise on Electricity and Magnetism* (Oxford University Press, Oxford, 1892).
[88] J.B. McLeod, A note on the ε-algorithm, *Computing* 7 (1971) 17–24.
[89] G. Meinardus and G.D. Taylor, Lower estimates for the error of the best uniform approximation, *J. Approx. Theory* 16 (1976) 150–161.
[90] R.M. Milne, Extension of Huygens' approximation to a circular arc, *Math. Gaz.* 2 (1903) 309–311.
[91] R. de Montessus de Ballore, Sur les fractions continues algébriques, *Bull. Soc. Math. France* 30 (1902) 28–36.
[92] P. Moussa, Physique théorique: les approximants de Padé retrouvés, in: *Approximants de Padé* (IREM de Lille, Lille, 1994).
[93] G. Mühlbach, Neville–Aitken algorithms for interpolating by functions of Čebyšev-systems in the sense of Newton and in a generalized sense of Hermite, in: A.G. Law and B.N. Sahney, eds., *Theory of Approximation with Applications* (Academic Press, New York, 1976) 200–212.
[94] H. Naegelsbach, Studien zu Fürstenau's neuer Methode der Darstellung und Berechnung der Wurzeln algebraischer Gleichungen durch Determinanten der Coefficienten, *Arch. Math. Phys.* 59 (1876) 147–192; 61 (1877) 19–85.
[95] P. Nevai, ed., *Orthogonal Polynomials: Theory and Practice* (Kluwer, Dordrecht, 1990).
[96] E.M. Nikishin and V.N. Sorokin, *Rational Approximations and Orthogonality*, Trans. Math. Monographs 92 (American Mathematical Society, Providence, RI, 1991).
[97] T.H. O'Beirne, On linear iterative processes and on methods of improving the convergence of certain types of iterated sequences, Technical Report, Torpedo Experimental Establishment, Greenock (1947).
[98] K.J. Overholt, Extended Aitken acceleration, *BIT* 5 (1965) 122–132.
[99] H. Padé, Sur la représentation approchée d'une fonction par des fractions rationnelles, *Ann. Sci. Ec. Norm. Sup.* 9 (1892) 1–93; also in: C. Brezinski, ed., *Oeuvres* (Librairie Scientifique et Technique Albert Blanchard, Paris, 1984) 72–166.
[100] R. Pennacchi, Le trasformazioni razionali di una successione, *Calcolo* 5 (1968) 37–50.
[101] R. Powell and S.M. Shah, *Summability Theory and its Applications* (Nostrand Reinhold, London, 1972).
[102] M. Prévost, Acceleration property for the E-algorithm and an application to the summation of series, *Adv. Comput. Math.* 2 (1994) 319–341.
[103] J.D. Pryce, *Numerical Solution of Sturm–Liouville Problems* (Clarendon Press, Oxford, 1993).
[104] L.F. Richardson, The approximate arithmetical solution by finite difference of physical problems involving differential equations, with an application to the stress in a masonry dam, *Philos. Trans. Roy. Soc. London Ser. A* 210 (1910) 307–357.
[105] L.F. Richardson, The deferred approach to the limit, I: Single lattice, *Philos. Trans. Roy. Soc. London Ser. A* 226 (1927) 299–349.
[106] W. Romberg, Vereinfachte numerische Integration, *Kgl. Norske Vid. Selsk. Forsk.* 28 (1955) 30–36.
[107] H. Rutishauser, Der Quotienten-Differenzen-Algorithmus, *Z. Angew. Math.* Phys. 5 (1954) 233–251.
[108] H. Rutishauser, Bestimmung der Eigenwerte und Eigenvektoren einer Matrix mit Hilfe des Quotienten-Differenzen-Algorithmus, *Z. Angew. Math. Phys.* 6 (1955) 387–401.
[109] H. Rutishauser, Ausdehnung des Rombergschen Prinzips, *Numer. Math.* 5 (1963) 48–54.
[110] H. Sadok, Quasilinear vector extrapolation methods, *Linear Alg. Appl.* 190 (1993) 71–85.
[111] J.F. Saigey, *Problèmes d'Arithmétique et Exercices de Calcul du Second Degré avec les Solutions Raisonnées* (Hachette, Paris, 1859).
[112] A. Salam, Non-commutative extrapolation algorithms, *Numerical Algor.* 7 (1994) 225–251.

[113] J.R. Schmidt, On the numerical solution of linear simultaneous equations by an iterative method, *Phil. Mag.* 7 (1941) 369–383.
[114] C. Schneider, Vereinfachte Rekursionen zur Richardson-Extrapolation in Spezialfällen, *Numer. Math.* 24 (1975) 177–184.
[115] D. Shanks, An analogy between transient and mathematical sequences and some nonlinear sequence-to-sequence transforms suggested by it, Part I, Memorandum 9994, Naval Ordnance Laboratory, White Oak (1949).
[116] D. Shanks, Non linear transformations of divergent and slowly convergent sequences, *J. Math. Phys.* 34 (1955) 1–42.
[117] W.F. Sheppard, Some quadrature formulas, *Proc. London Math. Soc.* 32 (1900) 258–277.
[118] A. Sidi, Extrapolation vs. projection methods for linear systems of equations, *J. Comput. Appl. Math.* 22 (1988) 71–88.
[119] A. Sidi, On a generalization of the Richardson extrapolation process, *Numer. Math.* 57 (1990) 365–377.
[120] H. Stahl and V. Totik, *General Orthogonal Polynomials* (Cambridge University Press, Cambridge, 1992).
[121] E. Stiefel, Altes und neues über numerische Quadratur, *Z. Angew. Math. Mech.* 41 (1961) 408–413.
[122] T.J. Stieltjes, Sur la réduction en fraction continue d'une série procédant suivant les puissances descendantes d'une variable, *Ann. Fac. Sci. Toulouse* 3 (1889) H.1–17.
[123] T.J. Stieltjes, Recherches sur les fractions continues, *Ann. Fac. Sci. Toulouse* 8 (1894) 1–122.
[124] W. van Assche, *Asymptotics for Orthogonal Polynomials*, Lecture Notes in Mathematics 1265 (Springer, Berlin, 1987).
[125] J. van Iseghem, Vector Padé approximants, in: R. Vichnevetsky and J. Vignes, eds., *Numerical Mathematics and Applications* (North-Holland, Amsterdam, 1985) 73–77.
[126] J. van Iseghem, Vector orthogonal relations. Vector QD algorithm, *J. Comput. Appl. Math.* 19 (1987) 141–150.
[127] J. van Iseghem, Convergence of vectorial sequences. Applications, *Numer. Math.* 68 (1994) 549–562.
[128] G. Walz, The history of extrapolation methods in numerical analysis, Report No. 130, Universität Mannheim, Fakultät für Mathematik und Informatik (1991).
[129] G. Walz, *Asymptotic Expansions in Numerical Analysis*, to appear.
[130] E.J. Weniger, Nonlinear sequence transformations for the acceleration of convergence and the summation of divergent series, *Comput. Physics Reports* 10 (1989) 189–371.
[131] J. Wimp, *Sequence Transformations and their Applications* (Academic Press, New York, 1981).
[132] P. Wynn, On a device for computing the $e_m(S_n)$ transformation, *MTAC* 10 (1956) 91–96.
[133] P. Wynn, On a procrustean technique for the numerical transformation of slowly convergent sequences and series, *Proc. Cambridge Phil. Soc.* 52 (1956) 663–671.
[134] P. Wynn, Acceleration techniques for iterated vector and matrix problems, *Math. Comp.* 16 (1962) 301–322.
[135] P. Wynn, Singular rules for certain nonlinear algorithms, *BIT* 3 (1963) 175–195.