Convergence acceleration during the 20th century

C. Brezinski *

In numerical analysis many methods produce sequences, for instance iterative methods for solving systems of equations, methods involving series expansions, 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), perturbation methods, etc. Sometimes, the convergence of these sequences is slow and their effective use is quite limited. Convergence acceleration methods consist of transforming a slowly converging sequence (S_n) into a new sequence (T_n) converging to the same limit faster than the initial one. Among such *sequence transformations*, the most well known are certainly Richardson's extrapolation algorithm and Aitken's Δ^2 process. All known methods are constructed by extrapolation and they are often called *extrapolation methods*. The idea consists of interpolating the terms $S_n, S_{n+1}, \ldots, S_{n+k}$ of the sequence to be transformed by a sequence satisfying a certain relationship depending on parameters. This set of sequences is called *kernel* of the transformation and every sequence of this set is transformed into a constant sequence by the transformation into consideration. For example, as we will see below, the kernel of Aitken's Δ^2 process is the set of sequences satisfying $\forall n, a_0(S_n - S) + a_1(S_{n+1} - S) = 0$, where a_0 and a_1 are parameters such that $a_0 + a_1 \neq 0$. If Aitken's process is applied to such a sequence, then the constant sequence $(T_n = S)$ is obtained. The parameters involved in the definition of the kernel are uniquely determined by the interpolation conditions and then the limit of the interpolating sequence of the kernel is taken as an approximation of the limit of the sequence to be accelerated. Since this limit depends on the index n, it will be denoted by T_n. Effectively, the sequence (S_n) has been transformed into a new sequence (T_n).

This paper, which is based on [31], but includes new developments obtained since 1995, presents my personal views on the historical development of this subject during the 20th century. I do not pretend to be exhaustive nor even to quote every important contribution (if a reference does not appear below, it does not mean that it is less valuable). I refer the interested reader to the literature and, in particular to the recent books [55, 146, 33, 144]. For an extensive bibliography, see [28].

I will begin with scalar sequences and then treat the case of vector ones. As we will see, a sequence transformation able to accelerate the convergence of *all* scalar sequences cannot exist. Thus, it is necessary to obtain many different convergence acceleration methods, each being suitable for a particular class of sequences. Many authors have studied the properties of these procedures and proved some important classes of sequences to be accelerable by a

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

given algorithm. Scalar sequence transformations have also been extensively studied from the theoretical point of view.

The situation is more complicated and more interesting for vector sequences. In the case of a sequence of vectors, it is always possible to apply a scalar acceleration procedure componentwise. However, such a strategy does not take into account connections which may exist between the various components, as in the important case of sequences arising from iterative methods for solving a system of linear or nonlinear equations.

1 Scalar sequences

Let (S_n) be a scalar sequence converging to a limit S. As explained above, an extrapolation method consists of transforming this sequence into a new one, (T_n), by a sequence transformation $T : (S_n) \longrightarrow (T_n)$. The transformation T is said to *accelerate the convergence* of the sequence (S_n) if and only if

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

We can then say that (T_n) *converges* (to S) *faster than* (S_n).

The first methods to have been used were linear transformations

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

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 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. In the case of linear methods, the convergence of the sequence (T_n) to S for any converging sequence (S_n) is governed by the Toeplitz summability theorem; see [115] for a review. Examples of such processes are

$$T_n = \frac{1}{n+1} \sum_{i=0}^{n} S_i$$

or

$$T_n = \frac{1}{k+1} \sum_{i=n}^{n+k} S_i.$$

In the second case, the sequence (T_n) also depends on a second index, k, and the convergence has to be studied either when k is fixed and n tends to infinity, or when n is fixed and k tends to infinity.

With respect to convergence acceleration, summation processes are usually only able to accelerate the convergence of restricted classes of sequences and this is why the numerical analysts of the 20th century turned their efforts to nonlinear transformations. However, there is one exception: Richardson's extrapolation process.

1.1 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). In 1903, Robert Moir Milne (1873– ?) applied the idea of Huygens for computing π [101]. The same idea was exploited again by Karl Kommerell (1871–1948) in his book of 1936 [78]. As explained in [143], Kommerell can be considered as the real discoverer of Romberg's method although he used this scheme in the context of approximating π.

Let us now come to the procedures used for improving the accuracy of the trapezoidal rule for computing approximations to a definite integral. In the case of a sufficiently smooth function, the error of this method is given by the Euler–Maclaurin expansion. In 1742, Colin Maclaurin (1698–1746) [90] showed that its precision could be improved by forming linear combinations of the results obtained with various stepsizes. His procedure can be interpreted as a preliminary version of Romberg's method; see [49] for a discussion.

In 1900, William Fleetwood Sheppard (1863–1936) used an elimination strategy in the Euler–Maclaurin quadrature formula with $h_n = r_n h$ and $1 = r_0 < r_1 < r_2 < \cdots$ to produce a better approximation to the given integral [132].

In 1910, combining the results obtained with the stepsizes h and $2h$, Lewis Fry Richardson (1881–1953) eliminated the first term in a discretization process using central differences [119]. He called this procedure the *deferred approach to the limit* or h^2*–extrapolation*. The transformed sequence (T_n) is given by

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

In a 1927 paper [120] he used the same technique to solve a 6th order differential eigenvalue problem. His process was called (h^2, h^4)*–extrapolation*. Richardson extrapolation consists of 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 $(x_n, S_n), \ldots, (x_{n+k}, S_{n+k})$. Using the Neville–Aitken scheme for these interpolation polynomials, we immediately obtain

$$T_{k+1}^{(n)} = \frac{x_{n+k+1} T_k^{(n)} - x_n T_k^{(n+1)}}{x_{n+k+1} - x_n}$$

with $T_0^{(n)} = S_n$.

Let us mention that Richardson referred to a 1926 paper by Nikolai Nikolaevich Bogolyubov (born in 1909) and Nikolai Mitrofanovich Krylov (1879–1955) where the procedure (often called the *deferred approach to the limit*) can already be found [11].

In 1955, Werner Romberg (born in 1909) was the first to use repeatedly an elimination approach for improving the accuracy of the trapezoidal rule [121]. He himself refers to the book of Lothar Collatz (1910–1990) of 1951 [50]. The procedure became widely known after the rigorous error analysis given in 1961 by Friedrich L. Bauer [3] and the work of Eduard L. Stiefel (1909–1978) [138]. Romberg's derivation of his process was heuristic. It was proved by Pierre–Jean Laurent in 1963 [81] that the process comes out from the Richardson process by choosing $x_n = h_n^2$ 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. Weaker conditions were given by Michel Crouzeix and Alain L. Mignot in [52, pp. 52–55]. As we shall see below, extensions of Romberg's method to non–smooth integrands leads to a method called the E–algorithm.

Applications of extrapolation to the numerical solution of ordinary differential equations were studied by H.C. Bolton and H.I. Scoins in 1956 [12], Roland Bulirsch and Josef Stoer in 1964–1966 [47] and William B. Gragg [65] in 1965. The case of difference methods for partial differential equations was treated by Gurii Ivanovich Marchuk and V.V. Shaidurov [91]. Sturm–Liouville problems are discussed in [117]. Finally, we mention that Heinz Rutishauser (1918–1970) pointed out in 1963 [122] that Romberg's idea can be applied to any sequence as long as the error has 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 [57, 77, 143].

1.2 Aitken's process and the ε–algorithm

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

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

The method was stated by Alexander Craig Aitken (1895–1967) in 1926 [1] who used it to accelerate 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 [104]. The process was also given by James Clerk Maxwell (1831–1879) in his *Treatise on Electricity and Magnetism* of 1873 [95]. However, 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. It is surprising that Aitken's process was known to Takakazu Seki (1642?–1708), often considered the greatest Japanese mathematician. In his book *Katsuyō Sanpō*, vol. IV, he used this process to compute the value of π, the length of a chord and the volume of a sphere. This book was written around 1680 but only published in 1712 by his disciple Murahide Araki. Parts of it can be found in [73]. Let us mention that the Japanese characters corresponding to Takakazu have another pronounciation which is Kōwa. This is the reason why this mathematician is often called, erroneously as in [29, 31], Seki Kōwa.

What makes Aitken's process so popular is that it accelerates the convergence of all linearly converging sequences, that is sequences such that $\exists a \neq 1$

$$\lim_{n\to\infty} \frac{S_{n+1} - S}{S_n - S} = a.$$

It can even accelerate some logarithmic sequences (that is corresponding to $a = 1$) which are those with the slowest convergence and the most difficult to accelerate.

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

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

$a_0 a_k \neq 0$, $a_0 + \cdots + a_k \neq 0$. A particular case of the case $k = 2$ was already treated by Maxwell in his book of 1873 and a particular case of an arbitrary value of k was studied by T.H. O'Beirne

in 1947 [107]. This last work remains almost unknown since it was published only as an internal report. The problem was handled in full generality by Daniel Shanks (1917–1996) in 1949 [130] and again in 1955 [131]. He obtained the sequence transformation defined by

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

When $k = 1$, Shanks transformation reduces to the Aitken's Δ^2 process. It can be proved that $e_k(S_n) = S, \forall n$ if and only if (S_n) belongs to the kernel of the transformation given above. The same ratios of determinants were obtained by R.J. Schmidt in 1941 [127] in his study of a method for solving systems of linear equations.

The determinants involved in the definition of $e_k(S_n)$ have a very special structure. They are called *Hankel determinants* and were studied by Hermann Hankel (1839–1873) in his thesis in 1861 [72]. Such determinants satisfy a five–term recurrence relationship. This relation was used by O'Beirne and Shanks to implement the transformation by computing separately the numerators and the denominators of the $e_k(S_n)$'s. However, numerical analysts know it is difficult to compute determinants (too many arithmetical operations are needed and rounding errors due to the computer often 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. It is called the ε–algorithm [147]. It is as follows. One starts with

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

and then

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

Note that the numbers $\varepsilon_k^{(n)}$'s fill out a two dimensional array. The ε–algorithm 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 auxiliary quantities. They can be eliminated from the algorithm, thus leading to the so–called *cross rule* due to Wynn [153].

When implementing the ε–algorithm or using Wynn's cross rule, division by zero can occur and the algorithm must be stopped. However, if the singularity is confined, a term that will again be used in Section 1.6, that is if it occurs only for some adjacent values of the indexes k and n, one may jump over it by using *singular rules* and continue the computation. If a division by a number close to zero arises, the algorithm becomes numerically unstable due to the cancellation errors. A similar situation holds for the other convergence acceleration algorithms. The study of such problems was initiated by Wynn in 1963 [151] who proposed

particular rules for the ε–algorithm which are more stable than the usual rules. They were extended by Florent Cordellier in 1979 [51]. [151, 51].

The convergence and acceleration properties of the ε–algorithm have only been completely described only for two classes of sequences, namely totally monotonic and totally oscillating sequences [154, 15, 16].

Shanks' transformation and the ε–algorithm have close connections to Padé approximants, continued fractions and formal orthogonal polynomials; see, for example, [18].

1.3 Subsequent developments

The Shanks transformation and the ε–algorithm sparked the rebirth of the study of nonlinear 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, and Padé approximants. They also have applications to the solution of systems of linear and nonlinear equations, the computation of the eigenvalues of a matrix, the solution of systems of linear and nonlinear equations, and many other topics, see [40]. Among other acceleration methods which were obtained and studied, are the W–process of Samuel Lubkin [89], the method of Kjell J. Overholt [110], the ρ–algorithm of Wynn [148], the G–transformation of H.L. Gray, T.A. Atchison and G.V. McWilliams [70], the θ–algorithm of Claude Brezinski [14], the transformations of Bernard Germain–Bonne [63] and the various transformations due to David Levin [85]. To my knowledge, the only known acceleration theorem for the ρ–algorithm was obtained by Naoki Osada [108]. Simultaneously, several applications began to appear. For example, the ε–algorithm provides a quadratically convergent method for solving systems of nonlinear equations and its does not require the knowledge of any derivative. This procedure was proposed simultaneously by Brezinski [13] and Eckhart Gekeler [61]. It has important applications to the solution of boundary value problems for ordinary differential equations [44]. Many other algorithms are given in the work of Ernst Joachim Weniger [145], which also contains applications to physics, or in the book of Brezinski and Michela Redivo Zaglia [40] where applications to various domains of numerical analysis can be found. The authors of this book provide FORTRAN subroutines. The book of Annie Cuyt and Luc Wuytack must also be mentioned [53]. The ε–algorithm has been applied to statistics, see the work of Alain Berlinet [9], and to the acceleration of the convergence of sequences of random variables, considered by Hélène Lavastre [82]. Applications to optimization were proposed by Le Ferrand [84] and Bouchta Rhanizar [118].

Instead of using a quite complicated algorithm, such as the ε–algorithm, it can be interesting to use a simpler one (for instance, Aitken's Δ^2 process) iteratively. Such a use consists of applying the algorithm to (S_n) to produce a new sequence (T_n), then to apply the same algorithm to (T_n), and so on. For example, applying the iterated Δ^2 process to the successive convergents of a periodic continued fraction produces a better acceleration than using the ε–algorithm [24]. In particular, the iterated Δ^2 process transforms a logarithmic sequence into a sequence converging linearly and linear convergence into superlinear, to my knowledge the only known cases of such transformations.

The experience gained during these years lead to a deeper understanding of the subject. Research workers began to study more theoretical and general questions related to the theory of convergence acceleration. The first attempt was made by R. Pennacchi in 1968 [114], who studied rational sequence transformations. His work was generalized by Germain–Bonne in 1973

[62] who proposed a very general framework and showed how to construct new algorithms for accelerating some classes of sequences. However, a ground breaking discovery was made by Jean Paul Delahaye and Germain–Bonne in 1980 [56]. They proved that if a set of sequences satisfies a certain property, called *remanence* (too technical to be explained here), then a universal algorithm, i.e. one able to accelerate *all* sequences of this set, cannot exist. This result shows the limitations of acceleration methods. Many sets of sequences were proved to be remanent, for example, the sets of monotonic or logarithmic sequences. Even some subsets of the set of logarithmic sequences are remanent.

Moulay Driss Benchiboun [5] observed that all the sequence transformations found in the literature could be written as

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

with $D^2 f \equiv 0$, where Df denotes the sum of the partial derivatives of the function f. The reason for that fact was explained by Brezinski [26] who showed that it is related to the translativity property of sequence transformations. Hassane Sadok [123] extended these results to the vector case. Abderrahim Benazzouz [7] proved that quasilinear transformations can be written as the composition of two projections.

In many transformations, such as Shanks', the quantities computed are expressed as a ratio of determinants. This property is related to the existence of a triangular recurrence scheme for their computation as explained by Brezinski and Guido Walz [46].

Herbert Homeier [74] studied a systematic procedure for constructing sequences transformations. He considered iterated transformations which are *hierarchically consistent*, which means that the kernel of the basic transformation is the lowest one in the hierarchy. The application of the basic transformation to a sequence which is higher in the hierarchy leads to a new sequence belonging to a kernel lower in the hierarchy. Homeier wrote several papers on this topics.

Thus, the theory of convergence acceleration methods has progressed impressively. The practical side was not forgotten and authors obtained a number of special devices for improving their efficiency. For example, when a certain sequence is to be accelerated, it is not obvious to know in advance which method will give the best result unless some properties of the sequence are already known. Thus, Delahaye [54] proposed using simultaneously several transformations and selecting, at each step of the procedure, one answer among the answers provided by the various algorithms. He proved that, under some assumptions, some tests are able to find automatically the best answer. The work of Delahaye was extended by Abdelhak Fdil [58, 59]. The various answers could also be combined leading to *composite* transformations [23]. It is 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 [37]. Devices for controlling the error were also constructed [21].

When faced to the problem of accelerating the convergence of a given sequence, two approaches are possible. The first is to use a known extrapolation procedure and to try to prove that it accelerates the convergence of the given sequence. The second possibility is to construct an extrapolation procedure especially for that sequence. Convergence tests for sequences and series can be used for that purpose as explained by Brezinski [25]. This approach was mostly developed by Ana Cristina Matos [92]. 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). Brezinski and Redivo Zaglia did this in [39].

A.H. Bentbib [10] considered the acceleration of sequences of intervals. Mohammed Senhadji [129] defined and studied the condition number of a sequence transformation.

1.4 The E–algorithm

As we saw above, the quantities involved in Shanks transformation are expressed as a ratio of determinants and the ε–algorithm allows one to compute them recursively. It is well known that an interpolation polynomial can be expressed as a ratio of determinants. Thus polynomial extrapolation also leads to such a ratio and the Neville–Aitken scheme can be used to avoid the computation of these determinants which leads to the Richardson extrapolation algorithm. A similar situation arises for many other transformations: in each case, the quantities involved are expressed as a ratio of special determinants and, in each case, one seeks for a special 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 a single general recursive algorithm for 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 [128]. 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 $h_0 > h_1 > \ldots > 0$ and at the point 0 and 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

$$\begin{aligned} & \phi_k^n E = 1 \\ \text{and} \quad & \phi_k^n g_j = 0, \quad j = 1, \ldots, k \end{aligned}$$

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

$$\phi_k^n E = a + b = 1$$

and

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

The values of a and b follow immediately and we have

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

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 must be computed separately by the same scheme using a different initialization. As we shall see below, this algorithm is just 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 [97] on best uniform approximation by functions from $span(g_1, \ldots, g_N) \subset C[a,b]$. They defined the linear functionals L_n^k on $C[a,b]$ by

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

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

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

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

By using Gaussian elimination to solve 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 a 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 from a complete Chebyshev system (a technical concept which insures the existence and uniqueness of the solution). We seek

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

satisfying the interpolation conditions

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

where the x_i's are distinct points and the g_i's given functions. 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 in 1976 [103] 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 $\forall i > 0, g_i(0) = 0$, 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 [2]; see [30] for detailed explanations.

We 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 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 (Romberg emigrated to Norway and came to Trondheim in 1949). In 1978, Håvie wrote a report, published one year later [71], 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 stepsize h to 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 the expansion of $E_1^{(n)}$, and so on. Thus, once again we obtain the E–algorithm

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

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 the Gaussian elimination process 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, but from the point of view of extrapolation [19]. 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

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

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 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 results while, when $g_i(n) = x_n^i$, we obtain the ratio expressing the quantities involved in the Richardson extrapolation process. Other algorithms may be similarly derived.

Now the problem is to find a recursive algorithm for computing the $E_k^{(n)}$'s. Applying Sylvester's determinantal identity, Brezinski obtained the two rules of the above E–algorithm. 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 certain conditions [19]. These results 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 $(g_1(n)), (g_2(n)), \ldots$. The $g_i(n)$ are those to be used in the E–algorithm. 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 [134, 135, 136]. Thus the study of the asymptotic expansion of the error of the sequences to be accelerated is of primary importance, see Walz [144]. For example, Mohammed Kzaz [79, 80] and Pierre Verlinden [142] applied this idea to the problem of accelerating the convergence of Gaussian quadrature formulae [79] and Pedro Lima and Mario Graça to boundary value problems with singularities [88, 87] (see also the works of Lima and Diogo [87], and Lima and Carpentier [86]). Other acceleration results were obtained by Matos and Marc Prévost [94], Prévost [116] and Pascal Mortreux and Prévost [102]. An algorithm, more economical than the E–algorithm, was given by William F. Ford and Avram Sidi [60]. The connection between the E–algorithm and the ε–algorithm was studied by Berhnard Beckermann [4]. A general ε–algorithm connected to the E–algorithm was given by Carsten Carstensen [48]. See [27] for a more detailed review on the E–algorithm.

Convergence acceleration algorithms can also be used for predicting the unknowns terms of a series or a sequence. This idea, introduced by Jacek Gilewicz [64], was studied by Sidi and Levin [137], Brezinski [22] and Denis Vekemans [141].

1.5 A new approach

Over the years, a quite general framework was constructed for the theory of extrapolation algorithms. The situation was quite different for the practical construction of extrapolation algorithms and there was little systematic research in their derivation. However, thanks to a formalism due to Weniger [145], such a construction is now possible, see Brezinski and Matos [38]. 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. 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 means that the sequence obtained by applying L to the sequence (a_n) is identically zero. Such a difference operator is called an *annihilation* operator for the sequence (a_n). We have

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

Applying L and using linearity leads to

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

We solve for S and designate it by the sequence transformation

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

The sequence (T_n) is be such that $\forall n, T_n = S$ if and only if $\forall n, S_n - S = a_n D_n$. This approach is highly versatile.

All the algorithms described above and the related devices such as error control, composite sequence transformations, least squares extrapolation, etc., can be put into this framework. Moreover, many new algorithms can be obtained using this approach. The E–algorithm can also be put into this framework which provides a deeper insight and leads to new properties [41]. Matos [93], using results from the theory of difference equations, obtained new and general convergence and acceleration results when (a_n) has an asymptotic expansion of a certain form.

1.6 Integrable systems

The connection between convergence acceleration algorithms and discrete integrable systems is a subject whose interest is rapidly growing among physicists. When a numerical scheme is used for integrating a partial differential evolution equation, it is important that it conserve the quantities that are conserved by the partial differential equation itself. An important character is the *integrability* of the equation. Although this term has not yet received a completely

satisfactory definition (see [66]), it can be understood as the ability to write the solution explicitly in terms of a finite number of functions or as the confinement of singularities in finite domains. The construction of integrable discrete forms of integrable partial differential equations is highly nontrivial. A major discovery in the field of integrability was the occurrence of a solitary wave (called a *soliton*) in the Korteweg–de Vries (KdV) equation. Integrability is a rare phenomenon and the typical dynamical system is nonintegrable. A test of integrability, called *singularity confinement*, was given by B. Grammaticos, A. Ramani and V. Papageorgiou [67]. It turns out that this test is related to the existence of singular rules for avoiding a division by zero in convergence acceleration algorithms (see Subsection 1.2).

The literature on this topic is vast and we cannot enter into the details of it. We only want to give an indication of the connection between these two subjects since both domains could benefit from it.

In the rule for the ε–algorithm, V. Papageorgiou, B. Grammaticos and A. Ramani set $m = k+n$ and replaced $\varepsilon_k^{(n)}$ by $u(n,m)+mp+nq$, where p and q satisfy $p^2-q^2=1$. They obtained [111]

$$[p-q+u(n,m+1)-u(n+1,m)][p+q+u(n+1,m+1)-u(n,m)] = p^2-q^2.$$

This is the discrete lattice KdV equation. Since this equation is integrable, one can expect integrability to hold also for the ε–algorithm, and, thanks to the singular rules of Wynn and Cordellier mentioned at the end of Subsection 1.2, this is indeed the case.

In the rule of the ε–algorithm, making the change of variable $k = t/\varepsilon^3$ and $n - 1/2 = x/\varepsilon - ct/\varepsilon^3$ and replacing $\varepsilon_k^{(n)}$ by $p+\varepsilon^2 u(x-\varepsilon/2,t)$ where c and p are related by $1-2c = 1/p^2$, A. Nagai and J. Satsuma obtained [105]

$$\varepsilon^2 u(x-\varepsilon/2+c\varepsilon, t+\varepsilon^3) - \varepsilon^2 u(x+\varepsilon/2-c\varepsilon, t-\varepsilon^3) = \frac{1}{p+\varepsilon^2 u(x+\varepsilon/2,t)} - \frac{1}{p+\varepsilon^2 u(x-\varepsilon/2,t)}.$$

We have, to terms of order ε^5, the KdV equation

$$u_t - \frac{1}{p^3} u u_x + \frac{1}{48p^2}(1-p^{-4}) u_{xxx} = 0.$$

Other discrete numerical algorithms, such as the qd, LR, and ρ–algorithms are connected to other discrete or continuous integrable equations (see, for example, [112]). Formal orthogonal polynomials, continued fractions, Padé approximation also play a rôle in this topic [113].

By replacing the integer n in the ε–algorithm by a continuous variable, Wynn derived the *confluent form* of the ε–algorithm [149]

$$\varepsilon_{k+1}(t) = \varepsilon_{k-1}(t) + \frac{1}{\varepsilon_k'(t)}$$

with $\varepsilon_{-1}(t) \equiv 0$ and $\varepsilon_0(t) = f(t)$. This algorithm is the continuous counterpart of the ε–algorithm and its aim is to compute $\lim_{t\to\infty} f(t)$. Setting $N_k(t) = \varepsilon_k'(t)\varepsilon_{k+1}'(t)$, A. Nagai, T. Tokihiro and J. Satsuma [106] obtained

$$N_k'(t) = N_k(t)[N_{k-1}(t) - N_{k+1}(t)].$$

The above equation is the Bäcklund transformation of the discrete Toda molecule equation [139].

So, we see that some properties of integrable systems are related to properties of convergence acceleration algorithms. On the other hand, discretizing integrable partial differential equations leads to new sequence transformations which have to be studied from the point of view of their algebraic and acceleration properties. Replacing the second integer k in the confluent form of the ε–algorithm by a continuous variable, Wynn obtained a partial differential equation [152]. Its relation with integrable systems is an open question.

The connection between integrable systems and convergence acceleration algorithms needs to be investigated in more details to fully understand its meaning which is not clear yet.

2 The vector case

In numerical analysis, many iterative methods lead to vector sequences. To accelerate the convergence of 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 [150] by replacing, in the rule of the scalar ε–algorithm, $1/\Delta\varepsilon_k^{(n)}$ by $\left(\Delta\varepsilon_k^{(n)}\right)^{-1}$ 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. Using Clifford algebra, J.B. McLeod proved in 1971 [96] that $\forall n, \varepsilon_{2k}^{(n)} = S$ if the sequence (S_n) satisfies $a_0(S_n - S) + \cdots + a_k(S_{n+k} - S) = 0, \forall n$ with $a_0 a_k \neq 0$, $a_0 + \cdots + a_k \neq 0$. This result is only valid for real sequences (S_n) and real a_i's. Moreover, contrary to the scalar case, this condition is only sufficient. In 1983, Peter R. Graves–Morris [68] extended this result to the complex case using a quite different approach.

A drawback to the development of the theory of the vector ε–algorithm was that it was not known whether a corresponding generalization of Shanks transformation was underlying the algorithm, that is whether the vectors $\varepsilon_k^{(n)}$ obtained by the algorithm could be expressed as ratios of determinants (or some kind of generalization of determinants). This is why Brezinski [17], following the same path as Shanks, tried 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 determinants. He then had to develop a recursive algorithm for avoiding their computation. This was the so–called topological ε–algorithm. This algorithm has many applications, in particular, to the solution of systems of linear equations (it is related to the biconjugate gradient algorithm [18, pp. 185ff]). In the case of a system of nonlinear equations, it gave rise to a generalization of Steffensen's method [13]. That algorithm has a quadratic convergence under some assumptions as established by Hervé Le Ferrand [83] following the ideas presented by Khalide Jbilou and Sadok [75]. The denominator of the vector $\varepsilon_{2k}^{(n)}$ obtained by the vector ε–algorithm was first written as a determinant of dimension $2k+1$ by Graves–Morris and Chris Jenkins in [69]. The numerator follows immediately by modifying the first row of the denominator, a formula given by Ahmed Salam and Graves–Morris [126]. However, the dimension of the corresponding determinants in the scalar case is only $k+1$. It was proved by Salam [124] that the vectors $\varepsilon_{2k}^{(n)}$ computed by the vector ε–algorithm can be expressed as a ratio of two *designants* of dimension $k+1$. A designant is a generalization of a determinant when solving a system of linear equations in a noncommutative algebra. An algebraic approach to this algorithm was given in [125]. This approach, which involves the use of a Clifford algebra, was used in [45] for extending the mechanism given in [41] to the vector and matrix cases. The vector generalization of the E–algorithm [19] can be explained similarly. This algorithm makes use of

a fixed vector y. Jet Wimp [146, pp. 176–177] generalized it using a sequence (y_n) instead of y. Jeannette van Iseghem [140] gave an algorithm for accelerating vector sequences based on the vector orthogonal polynomials she introduced for generalizing Padé approximants to the vector case. Other vector sequence transformations are due to Osada [109] and Jbilou and Sadok [76]. Benchiboun [6] and Abderrahim Messaoudi [100] studied matrix extrapolation algorithms.

We have seen that, in the scalar case, the kernels of sequence transformations may be expressed as relationships with constant coefficients. This is also the case for the vector and the topological ε–algorithms and the vector E–algorithm. The first (and, to my knowledge, only) transformation treating a relationship with varying coefficients was introduced in [42]. The theory developed there also explains why the case of a relationship with non–constant coefficients is a difficult problem in the scalar case and why it could be solved, on the contrary, in the vector case. The reason is that the number of unknown coefficients appearing in the expression of the kernel must be strictly less than the dimension of the vectors. Brezinski in [34] proposed a general methodology for constructing vector sequence transformations. It leads to a unified presentation of several approaches to the subject and to new results. He also discussed applications to linear systems. In fact, as showed by Sidi [133], and Jbilou and Sadok [75], vector sequence transformations are closely related to projection methods for the solution of systems of equations. In particular, the RPA, a vector sequence transformation defined by Brezinski [20] was extensively studied by Messaoudi who showed its connections to direct and iterative methods for solving systems of linear equations [98, 99].

Vector sequence transformations lead to new methods for the solution of systems of nonlinear equations. They also have other applications. First of all, it is quite important to accelerate the convergence of iterative methods for the solution of systems of linear equations, see [32], [33] and [36]. Special vector extrapolation techniques were designed for the regularization of ill–posed linear systems in [43] and the idea of extrapolation was used in [35] to obtain estimates of the norm of the error when solving a system of linear equations by an arbitrary method, direct or iterative.

General theoretical results similar to those obtained in the scalar case are still lacking in the vector case although some partial results have been obtained. Relevant results on quasilinear transformations are in the papers by Sadok [123] and Benazzouz [8]. The present author proposed a mechanism for vector sequence transformations in [45] and [34].

3 Conclusions and perspectives

In this paper, I have tried to give a survey of the development of convergence acceleration methods for scalar and vector sequences in the 20th century. These methods are based on the idea of extrapolation. Since a universal algorithm for accelerating the convergence of all sequences cannot exist (and this is even true for some restricted classes of sequences), it was necessary to define and study a large variety of algorithms, each of them being appropriate for some special subsets of sequences.

It is, of course, always possible to construct other convergence acceleration methods for scalar sequences. However, to be of interest, such new processes must provide a major improvement over existing ones. For scalar sequence transformations, the emphasis must be placed on the theory rather than on special devices (unless a quite powerful one is found) and on

the application of new methods to particular algorithms in numerical analysis and to various domains of applied sciences. In particular, the connection between convergence acceleration algorithms and continuous and discrete integrable systems brings a different and fresh look to both domains and could be of benefit to them.

An important problem in numerical analysis is the solution of large, sparse systems of linear equations. Most of the methods used nowadays are projection methods. Often the iterates obtained in such problems must be subject to acceleration techniques. However, many of the known vector convergence acceleration algorithms require the storage of too many vectors to be useful. New and cheaper acceleration algorithms are required. This difficult project, in my opinion, offers many opportunities for future research.

In this paper, I only briefly mentioned the confluent algorithms whose aim is the computation of the limit of a function when the variable tends to infinity (the continuous analog of the problem of convergence acceleration for a sequence). This subject and its applications will provide fertile ground for new discoveries.

Acknowledgements: I would like to thank Jet Wimp for his careful reading of the paper. He corrected my English in many places, he asked me to provide more explanations when needed, and suggested many improvements in the presentation. I am also indebted to Naoki Osada for his informations about Takakazu Seki.

References

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

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

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

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

[5] 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.

[6] M.D. Benchiboun, Extension of Henrici's method to matrix sequences, J. Comput. Appl. Math., 75 (1996) 1–21.

[7] A. Benazzouz, Quasilinear sequence transformations, Numer. Algorithms, 15 (1997) 275–285.

[8] A. Benazzouz, GL(E)–quasilinear transformations and acceleration, Appl. Numer. Math., 27 (1998) 109–122.

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

[10] A.H. Bentbib, Acceleration of convergence of interval sequences, J. Comput. Appl. Math., 51 (1994) 395–409.

[11] 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 (in Russian), Acad. Sci. Ukraine (Phys. Math.), 3 (1926) 3–22.

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

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

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

[15] C. Brezinski, Etude sur les ε et ρ–algorithmes, Numer. Math., 17 (1971) 153–162.

[16] C. Brezinski, L'ε–algorithme et les suites totalement monotones et oscillantes, C.R. Acad. Sci. Paris, 276 A (1973) 305–308.

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

[18] C. Brezinski, *Padé–Type Approximation and General Orthogonal Polynomials*, Birkhäuser, Basel, 1980.

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

[20] C. Brezinski, Recursive interpolation, extrapolation and projection, J. Comput. Appl. Math., 9 (1983) 369–376.

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

[22] C. Brezinski, Prediction properties of some extrapolation methods, Appl. Numer. Math., 1 (1985) 457–462.

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

[24] C. Brezinski, A. Lembarki, Acceleration of extended Fibonacci sequences, Appl. Numer. Math., 2 (1986) 1–8.

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

[26] C. Brezinski, Quasi–linear extrapolation processes, in *Numerical Mathematics. Singapore 1988*, R.P. Agarwal et al. eds., ISNM vol.86, Birkhäuser, Basel, 1988, pp.61–78.

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

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

[29] C. Brezinski, *History of Continued Fractions and Padé Approximants*, Springer–Verlag, Berlin, 1991.

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

[31] C. Brezinski, Extrapolation algorithms and Padé approximations: a historical survey, Appl. Numer. Math., 20 (1996) 299–318.

[32] C. Brezinski, Variations on Richardson's method and acceleration, in *Numerical Analysis, A Numerical Analysis Conference in Honour of Jean Meinguet*, Bull. Soc. Math. Belg., 1996, pp. 33-44.

[33] C. Brezinski, *Projection Methods for Systems of Equations*, North–Holland, Amsterdam, 1997.

[34] C. Brezinski, Vector sequence transformations: methodology and applications to linear systems, J. Comput. Appl. Math., 98 (1998) 149–175.

[35] C. Brezinski, Error estimates for the solution of linear systems, SIAM J. Sci. Comput., to appear.

[36] C. Brezinski, Acceleration procedures for matrix iterative methods, Numer. Algorithms, to appear.

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

[38] C. Brezinski, A.C. Matos, A derivation of extrapolation algorithms based on error estimates, J. Comput. Appl. Math., 66 (1996) 5–26.

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

[40] C. Brezinski, M. Redivo Zaglia, *Extrapolation Methods. Theory and Practice*, North–Holland, Amsterdam, 1991.

[41] C. Brezinski, M. Redivo Zaglia, A general extrapolation procedure revisited, Adv. in Comput. Math., 2 (1994) 461–477.

[42] C. Brezinski, M. Redivo Zaglia, Vector and matrix sequence transformations based on biorthogonality, Appl. Numer. Math., 21 (1996) 353–373.

[43] C. Brezinski, M. Redivo Zaglia, G. Rodriguez, S. Seatzu, Extrapolation techniques for ill–conditioned linear systems, Numer. Math., 81 (1998) 1–29.

[44] C. Brezinski, A.C. Rieu, The solution of systems of equations using the vector ε–algorithm, and an application to boundary value problems, Math. Comput., 28 (1974) 731–741.

[45] C. Brezinski, A. Salam, Matrix and vector sequence transformation revisited, Proc. Edinburgh Math. Soc., 38 (1995) 495–510.

[46] C. Brezinski, G.Walz, Sequences of transformations and triangular recursion schemes with applications in numerical analysis, J. Comput. Appl. Math., 34 (1991) 361–383.

[47] R. Bulirsch, J. Stoer, Numerical treatment of ordinary differential equations by extrapolation methods, Numer. Math., 8 (1966) 1–13.

[48] C. Carstensen, On a general epsilon algorithm, in *Numerical and Applied Mathematics*, C. Brezinski ed., Baltzer, Basel, 1989, pp.437–441.

[49] J.L. Chabert et al., *Histoire d'Algorithmes*, Belin, Paris, 1994.

[50] L. Collatz, *Numerische Behandlung von Differentialgleichungen*, Springer–Verlag, Berlin, 1951.

[51] F. Cordellier, 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 vol. 765, Springer–Verlag, Berlin, 1979, pp.36–60.

[52] M. Crouzeix, A.L. Mignot, *Analyse Numérique des Equations Différentielles*, 2nd ed., Masson, Paris, 1989.

[53] A. Cuyt, L. Wuytack, *Nonlinear Methods in Numerical Analysis*, North–Holland, Amsterdam, 1987.

[54] J.P. Delahaye, Automatic selection of sequence transformations, Math. Comput., 37 (1981) 197–204.

[55] J.P. Delahaye, *Sequence Transformations*, Springer–Verlag, Berlin, 1988.

[56] J.P. Delahaye, B. Germain–Bonne, Résultats négatifs en accélération de la convergence, Numer. Math., 35 (1980) 443–457.

[57] J. Dutka, Richardson–extrapolation and Romberg–integration, Historia Math., 11 (1984) 3–21.

[58] A. Fdil, Sélection entre procédés d'accélération de la convergence, M2AN, 30 (1996) 83–101.

[59] A. Fdil, A new technique of selection between sequence transformations, Appl. Numer. Math., 25 (1997) 21–40.

[60] W.F. Ford, A. Sidi, An algorithm for a generalization of the Richardson extrapolation process, SIAM J. Numer. Anal., 24 (1987) 1212–1232.

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

[62] B. Germain–Bonne, Transformations de suites, RAIRO, R1 (1973) 84–90.

[63] 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.

[64] J. Gilewicz, Numerical detection of the best Padé approximant and determination of the Fourier coefficients of insufficiently sampled function, in *Padé Approximants and their Applications*n P.R. Graves–Morris ed., Academic Press, New York, 1973, pp. 99–103.

[65] W.B. Gragg, On extrapolation algorithms for initial–value problems, SIAM J. Numer. Anal., 2 (1965) 384–403.

[66] B. Grammaticos, A. Ramani, Integrability – and how to detect it, in *Integrability of Nonlinear Systems*, Y. Kosmann–Schwarzbach et al. eds., Springer–Verlag, Berlin, 1997, pp. 30–94.

[67] B. Grammaticos, A. Ramani, V.G. Papageorgiou, Do integrable mappings have the Painlevé property?, Phys. Rev. Lett., 67 (1991) 1825–1828.

[68] P.R. Graves–Morris, Vector valued rational interpolants I, Numer. Math., 42 (1983) 331–348.

[69] P.R. Graves–Morris, C.D. Jenkins, Vector valued rational interpolants III, Constr. Approx., 2 (1986) 263–289.

[70] H.L. Gray, T.A. Atchison, G.V. McWilliams, Higher order G–transformations, SIAM J. Numer. Anal., 8 (1971) 365–381.

[71] T. Håvie, Generalized Neville type extrapolation schemes, BIT, 19 (1979) 204–213.

[72] H. Hankel, *Ueber eine besondere Classe der symmetrischen Determinanten*, Inaugural Diss., Universität Göttingen, 1861.

[73] A. Hirayama, K. Shimodaira, H. Hirose, *Takakazu Seki's Collected Works Edited with Explanations*, Osaka Kyoiku Tosho, Osaka, 1974.

[74] H.H.H. Homeier, A hierarchically consistent, iterative sequence transformation, Numer. Algorithms, 8 (1994) 47–81.

[75] K. Jbilou, H. Sadok, Some results about vector extrapolation methods and related fixed–point iterations, J. Comput. Appl. Math., 36 (1991) 385–398.

[76] K. Jbilou, H. Sadok, Hybrid vector transformations, J. Comput. Appl. Math., 81 (1997) 257–267.

[77] D.C. Joyce, Survey of extrapolation processes in numerical analysis, SIAM Rev., 13 (1971) 435–490.

[78] K. Kommerell, *Das Grenzgebiet der Elementaren und Höheren Mathematik*, Verlag Köhler, Leipzig, 1936.

[79] M. Kzaz, Gaussian quadrature and acceleration of convergence, Numer. Algorithms, 15 (1997) 75–89.

[80] M. Kzaz, Convergence acceleration of the Gauss–Laguerre quadrature formula, Appl. Numer. Math., 29 (1999) 201–220.

[81] P.J. Laurent, Un théorème de convergence pour le procédé d'extrapolation de Richardson, C. R. Acad. Sci. Paris, 256 (1963) 1435–1437.

[82] H. Lavastre, On the stochastic acceleration of sequences of random variables, Appl. Numer. Math., 15 (1994) 77–98.

[83] H. Le Ferrand, Convergence of the topological ε–algorithm for solving systems of nonlinear equations, Numerical Algor., 3 (1992) 273–284.

[84] H. Le Ferrand, Recherches d'extrema par des méthodes d'extrapolation, C.R. Acad. Sci. Paris, Sér. I, 318 (1994) 1043–1046.

[85] D. Levin, Development of non–linear transformations for improving convergence of sequences, Int. J. Comput. Math., B3 (1973) 371–388.

[86] P.M. Lima, M.P. Carpentier, Asymptotic expansions and numerical approximation of nonlinear degenerate boundary–value problems, Appl. Numer. Math., 30 (1999) 93–111.

[87] P. Lima, T. Diogo, An extrapolation method for a Volterra integral equation with weakly singular kernel, Appl. Numer. Math., 24 (1997) 131–148.

[88] P.M. Lima, M.M Graça, Convergence acceleration for boundary value problems with singularities using the E–algorithm, J. Comput. Appl. Math., 61 (1995) 139–164.

[89] S. Lubkin, A method of summing infinite series, J. Res. Natl. Bur. Stand., 48 (1952) 228–254.

[90] C. Maclaurin, *Treatise of Fluxions*, Edinburgh, 1742.

[91] G.I. Marchuk, V.V. Shaidurov, *Difference Methods and their Extrapolations*, Springer–Verlag, Berlin, 1983.

[92] A.C. Matos, Acceleration methods based on convergence tests, Numer. Math., 58 (1990) 329–340.

[93] A.C. Matos, Linear difference operators and acceleration methods, submitted.

[94] A.C. Matos, M. Prévost, Acceleration property for the columns of the E–algorithm, Numerical Algor., 2 (1992) 393–408.

[95] J.C. Maxwell, *A Treatise on Electricity and Magnetism*, Oxford University Press, Oxford, 1873.

[96] J.B. McLeod, A note on the ε–algorithm, Computing, 7 (1971) 17–24.

[97] G. Meinardus, G.D. Taylor, Lower estimates for the error of the best uniform approximation, J. Approx. Theory, 16 (1976) 150–161.

[98] A. Messaoudi, Recursive interpolation algorithm: a formalism for solving systems of linear equations–I. Direct methods, J. Comput. Appl. Math., 76 (1996) 13–30.

[99] A. Messaoudi, Recursive interpolation algorithm: a formalism for solving systems of linear equations–II. Iterative methods, J. Comput. Appl. Math., 76 (1996) 31–53.

[100] A. Messaoudi, Matrix extrapolation algorithms, Linear Alg. Appl., 256 (1997) 49–73.

[101] R.M. Milne, Extension of Huygens' approximation to a circular arc, Math. Gaz., 2 (1903) 309–311.

[102] P. Mortreux, M. Prévost, An acceleration property for the E–algorithm for alternate sequences, Adv. Comput. Math., 5 (1996) 443–482.

[103] 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 *Theory of Approximation with Applications*, A.G. Law and B.N. Sahney eds., Academic Press, New York, 1976, pp.200–212.

[104] 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.

[105] A. Nagai, J. Satsuma, Discrete soliton equations and convergence acceleration algorithms, Phys. Letters A, 209 (1995) 305–312.

[106] A. Nagai, T. Tokihiro, J. Satsuma, The Toda molecule equation and the ε–algorithm, Math. Comput., 67 (1998) 1565–1575.

[107] 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, May 1947.

[108] N. Osada, An acceleration theorem for the ρ–algorithm, Numer. Math., 73 (1996) 521–531.

[109] N. Osada, Vector sequence transformations for the acceleration of logarithmic convergence, J. Comput. Appl. Math., 66 (1996) 391–400.

[110] K.J. Overholt, Extended Aitken acceleration, BIT, 5 (1965) 122–132.

[111] V. Papageorgiou, B. Grammaticos, A. Ramani, Integrable difference equations and numerical analysis algorithms, in *Symmetries and Integrability of Difference Equations*, D. Levi et al. eds., CRM Proceedings and Lecture Notes, vol. 9, AMS, Providence, 1996, pp. 269–279.

[112] V. Papageorgiou, B. Grammaticos, A. Ramani, Integrable lattices and convergence acceleration algorithms, Phys. Letters, A, 179 (1993) 111–115.

[113] V. Papageorgiou, B. Grammaticos, A. Ramani, Orthogonal polynomial approach to discrete Lax pairs for initial–boundary value problems of the QD algorithm, Lett. Math. Phys., 34 (1995) 91–101.

[114] R. Pennacchi, Le trasformazioni razionali di una successione, Calcolo, 5 (1968) 37–50.

[115] R. Powell, S.M. Shah, *Summability Theory and its Applications*, Van Nostrand Reinhold, London, 1972.

[116] M. Prévost, Acceleration property for the E–algorithm and an application to the summation of series, Adv. in Comput. Math., 2 (1994) 319–341.

[117] J.D. Pryce, *Numerical Solution of Sturm–Liouville Problems*, Clarendon Press, Oxford, 1993.

[118] B. Rhanizar, On extrapolation methods in optimization, Appl. Numer. Math., 25 (1997) 485–498.

[119] 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.

[120] L.F. Richardson, The deferred approach to the limit. I: Single lattice, Philos. Trans. Roy. Soc. London, ser. A, 226 (1927) 299–349.

[121] W. Romberg, Vereinfachte numerische Integration, Kgl. Norske Vid. Selsk. Forsk., 28 (1955) 30–36.

[122] H. Rutishauser, Ausdehnung des Rombergschen Prinzips, Numer. Math., 5 (1963) 48–54.

[123] H. Sadok, Quasilinear vector extrapolation methods, Linear Alg. Appl., 190 (1993) 71–85.

[124] A. Salam, Non–commutative extrapolation algorithms, Numer. Algorithms, 7 (1994), 225–251.

[125] A. Salam, An algebraic approach to the vector ε–algorithm, Numer. Algorithms, 11 (1996) 327-337.

[126] A. Salam, P.R. Graves–Morris, On the vector ε–algorithm, submitted.

[127] J.R. Schmidt, On the numerical solution of linear simultaneous equations by an iterative method, Phil. Mag., 7 (1941) 369–383.

[128] C. Schneider, Vereinfachte Rekursionen zur Richardson–Extrapolation in Spezialfällen, Numer. Math., 24 (1975) 177–184.

[129] M.N. Senhadji, On condition numbers of some quasi–linear transformations, J. Comput. Appl. Math., 104 (1999) 1–19.

[130] 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, July 1949.

[131] D. Shanks, Non linear transformations of divergent and slowly convergent sequences, J. Math. Phys., 34 (1955) 1–42.

[132] W.F. Sheppard, Some quadrature formulas, Proc. London Math. Soc., 32 (1900) 258–277.

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

[134] A. Sidi, On a generalization of the Richardson extrapolation process, Numer. Math., 57 (1990) 365–377.

[135] A. Sidi, Further results on convergence and stability of a generalization of the Richardson extrapolation process, BIT, 36 (1996) 143–157.

[136] A. Sidi, A complete convergence and stability theory for a generalized Richardson extrapolation process, SIAM J. Numer. Anal., 34 (1997) 1761–1778.

[137] A. Sidi, D. Levin, Prediction properties of the t–transformation, SIAM J. Numer. Anal., 20 (1983) 589–598.

[138] E. Stiefel, Altes und neues über numerische Quadratur, Z. Angew. Math. Mech., 41 (1961) 408–413.

[139] M. Toda, Waves in nonlinear lattice, Prog. Theor. Phys. Suppl., 45 (1970) 174–200.

[140] J. van Iseghem, Convergence of vectorial sequences. Applications, Numer. Math., 68 (1994) 549–562.

[141] D. Vekemans, Algorithm for the E–prediction, J. Comput. Appl. Math., 85 (1997) 181–202.

[142] P. Verlinden, Acceleration of Gauss–Legendre quadrature for an integral with an endpoint singularity, J. Comput. Appl. Math., 77 (1997) 277–287.

[143] G. Walz, The history of extrapolation methods in numerical analysis, Report Nr. 130, Universität Mannheim, Fakultät für Mathematik und Informatik, 1991.

[144] G. Walz, *Asymptotics and Extrapolation*, Akademie Verlag, Berlin, 1996.

[145] E.J. Weniger, Nonlinear sequence transformations for the acceleration of convergence and the summation of divergent series, Comput. Physics Reports, 10 (1989) 189–371.

[146] J. Wimp, *Sequence Transformations and their Applications*, Academic Press, New York, 1981.

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

[148] P. Wynn, On a procrustean technique for the numerical transformation of slowly convergent sequences and series, Proc. Cambridge Phil. Soc., 52 (1956) 663–671.

[149] P. Wynn, Confluent forms of certain nonlinear algorithms, Arch. Math., 11 (1960) 223–234.

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

[151] P. Wynn, Singular rules for certain nonlinear algorithms, BIT, 3 (1963) 175–195.

[152] P. Wynn, Partial differential equations associated with certain non–linear algorithms, ZAMP, 15 (1964) 273–289.

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

[154] P. Wynn, On the convergence and stability of the epsilon algorithm, SIAM J. Numer. Anal., 3 (1966) 91–122.