ELSEVIER

Applied Numerical Mathematics 15 (1994) 123–131

APPLIED NUMERICAL MATHEMATICS

Extrapolation methods

C. Brezinski [a,*], M. Redivo-Zaglia [b]

[a] Laboratoire d'Analyse Numérique et d'Optimisation, UFR IEEA – M3, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq Cedex, France

[b] Dipartimento di Electronica e Informatica, Università degli Studi di Padova, via Gradenigo 6/a, 35131 Padova, Italy

Abstract

The aim of this paper is to give a short introduction to the most fundamental aspects of extrapolation methods. Such methods are used for accelerating the convergence of sequences.

Keywords: Convergence acceleration; Extrapolation

This paper is based on our book [1] where all the details and the references can be found. This book also contains a diskette with FORTRAN routines for the most important algorithms for accelerating the convergence.

1. Introduction

Let (S_n) be a sequence of (real or complex) numbers which converges to S. We shall transform the sequence (S_n) into another sequence (T_n) and denote by T such a transformation.

For example we can have

$$T_n = \frac{S_n + S_{n+1}}{2}, \quad n = 0, 1, \ldots \tag{1.1}$$

or

$$T_n = \frac{S_n S_{n+2} - S_{n+1}^2}{S_{n+2} - 2S_{n+1} + S_n}, \quad n = 0, 1, \ldots \tag{1.2}$$

which is the well-known Δ^2 process due to Aitken.

* Corresponding author. E-mail: brezinsk@omega@univ-lille1.fr.

0168-9274/94/$07.00 © 1994 Elsevier Science B.V. All rights reserved
SSDI 0168-9274(94)00015-9

In order to be of some practical interest, the new sequence (T_n) must exhibit, at least for some particular classes of convergent sequences (S_n), the following properties:

(1) (T_n) must converge.
(2) (T_n) must converge to the same limit as (S_n).
(3) (T_n) must converge to S faster than (S_n), that is,

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

- If property (2) holds, we say that the transformation T is *regular* for the sequence (S_n).
- If property (3) holds we say that the transformation T *accelerates the convergence* of the sequence (S_n) or that the sequence (T_n) *converges faster* than (S_n).

These properties—in particular, the last one—do not hold for all converging sequence (S_n). But do the first two above-mentioned properties hold for examples (1.1) and (1.2)? The first example is a linear transformation for which it is easy to see that, for each converging sequence (S_n), the sequence (T_n) converges and has the same limit as (S_n). Such linear transformations, called summation processes, have been widely studied and the transformations, named after Euler, Cesaro, Hausdorff, Abel and others, are well known. The positive answer to properties (1) and (2) above for all convergent sequences is a consequence of the so-called Toeplitz theorem which can be found in the literature and whose conditions are very easily checked in practice. Some summation processes are very powerful for some sequences as is the case with Romberg's method for accelerating the trapezoidal rule which is explained in any textbook on numerical analysis. However, let us look again at transformation (1.1) and try to find the class of sequences which it accelerates. We have

$$\frac{T_n - S}{S_n - S} = \frac{1}{2}\left(1 + \frac{S_{n+1} - S}{S_n - S}\right)$$

and thus

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

if and only if

$$\lim_{n\to\infty} (S_{n+1} - S)/(S_n - S) = -1$$

which shows that this transformation is only able to accelerate the convergence of a very restricted class of sequences. This is mainly the case for all summation processes.

On the other hand, let us now look at our sequence transformation (1.2) which is Aitken's Δ^2 process. It can be easily proved that it accelerates the convergence of all the sequences for which there exists a $\lambda \in [-1, +1)$ such that

$$\lim_{n\to\infty} (S_{n+1} - S)/(S_n - S) = \lambda,$$

which is a much wider class than the class of sequences accelerated by our first linear transformation. But, since in mathematics as in life nothing can be obtained without pain, the drawback is that properties (1) and (2) do not hold for all convergent sequences.

In conclusion, nonlinear sequence transformations usually have better acceleration properties than linear summation processes (that is, they accelerate a wider class of sequences) but, on the other hand, they do not always transform a convergent sequence into another converging sequence and, even if they do, the limits may be different.

We shall now exemplify some interesting properties of sequence transformations on our examples (1.1) and (1.2). In the study of a sequence transformation the first question to be asked and solved (before those of convergence and acceleration) is an algebraic one: it concerns the so-called kernel $\mathscr{K}_T$ of the transformation, that is the set of sequences for which there exists an S such that $\forall n$, $T_n = S$ (in the sequel $\forall n$ would eventually mean $\forall n \geqslant N$).

For our linear summation process it is easy to check that its kernel is the set of sequences of the form

$$S_n = S + a(-1)^n,$$

where a is a scalar.

For Aitken's process the kernel is the set of sequences of the form

$$S_n = S + a\lambda^n,$$

where a and λ are scalars with $a \neq 0$ and $\lambda \neq 1$.

Thus, obviously, the kernel of Aitken's process contains the kernel of the first linear summation process.

As we can see, in both cases, the kernel depends on some (almost) arbitrary parameters: S and a in the first case and S, a and $\lambda(\neq 1)$ in the second.

If the sequence (S_n) to be accelerated belongs to the kernel of the transformation used, then, by construction, we shall have $\forall n$, $T_n = S$.

Of course, usually, S is the limit of the sequence (S_n) but this is not always the case and the question needs to be studied. For example, in Aitken's process, S is the limit of (S_n) if $|\lambda| < 1$. If $|\lambda| > 1$, (S_n) diverges and S is called its anti-limit. If $|\lambda| = 1$, (S_n) has no limit at all or it only takes a finite number of distinct values and S is, in this case, their arithmetical mean.

We are now ready to enter into the details and to explain what an extrapolation method is.

2. Extrapolation methods

A sequence transformation $T:(S_n) \mapsto (T_n)$ is said to be an extrapolation method if it is such that $\forall n$, $T_n = S$ if and only if $(S_n) \in \mathscr{K}_T$.

Thus any sequence transformation can be viewed as an extrapolation method. What is the reason for this name? Of course, it comes from interpolation and we shall now explain how a transformation T is built from its kernel $\mathscr{K}_T$, that is from a given relation R that is satisfied $\forall n$.

$S_n, S_{n+1}, \ldots, S_{n+p+q}$ being known, we are looking for the sequence $(u_n) \in \mathscr{K}_T$ satisfying the interpolation conditions

$$u_i = S_i, \quad i = n, \ldots, n+p+q.$$

Then, since (u_n) belongs to $\mathscr{K}_T$, it satisfies the relation R, that is

$$R(u_i, \ldots, u_{i+q}, S) = 0 \quad \forall i.$$

But, thanks to the interpolation conditions we also have

$$R(S_i, \ldots, S_{i+q}, S) = 0, \quad i = n, \ldots, n+p.$$

This is a system of $(p+1)$ equations with $(p+1)$ unknowns S, $a_1, \ldots, a_p$ whose solution depends on n, the index of the first interpolation condition. We shall solve this system to obtain the value of the unknown S which, since it depends on n, will be denoted by T_n. (Sometimes to recall that it also depends on $k = p+q$ it will be called $T_k^{(n)}$.)

Let us give an example to illustrate our purpose. We assume that R has the following form

$$R(u_i, u_{i+1}, S) = a_1(u_i - S) + a_2(u_{i+1} - S) = 0, \quad a_1 \cdot a_2 \neq 0,$$

and thus we have to solve the system

$$\begin{cases} a_1(S_n - S) + a_2(S_{n+1} - S) = 0, \\ a_1(S_{n+1} - S) + a_2(S_{n+2} - S) = 0. \end{cases}$$

Since this system does not change when each equation is multiplied by a nonzero constant, a_1 and a_2 are not independent and the system corresponds to $p = q = 1$. The derivative of R with respect to its last variable is equal to $-(a_1 + a_2)$ which is assumed to be different from zero. Then S is given by

$$S = (a_1 u_i + a_2 u_{i+1})/(a_1 + a_2)$$

and the system to be solved becomes

$$\begin{cases} S = (a_1 S_n + a_2 S_{n+1})/(a_1 + a_2), \\ S = (a_1 S_{n+1} + a_2 S_{n+2})/(a_1 + a_2). \end{cases}$$

Thus we do not restrict the generality if we assume that $a_1 + a_2 = 1$ and the system is written as

$$\begin{cases} S = a_1 S_n + (1 - a_1) S_{n+1}, \\ S = a_1 S_{n+1} + (1 - a_1) S_{n+2}, \end{cases}$$

or

$$0 = a_1 \Delta S_n + (1 - a_1) \Delta S_{n+1},$$

where Δ is the difference operator defined by $\Delta v_n = v_{n+1} - v_n$ and $\Delta^{k+1} v_n = \Delta^k v_{n+1} - \Delta^k v_n$.

The last relation gives

$$a_1 = \Delta S_{n+1} / \Delta^2 S_n$$

($\Delta^2 S_n \neq 0$ since $a_1 + a_2 \neq 0$) and thus we finally obtain

$$S = T_n = \frac{\Delta S_{n+1}}{\Delta^2 S_n} \cdot S_n + \left(1 - \frac{\Delta S_{n+1}}{\Delta^2 S_n}\right) \cdot S_{n+1},$$

that is

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

which is Aitken's Δ^2 process (whose name comes from the Δ^2 in the denominator).

Thus we have shown how to construct a sequence transformation T from a given kernel $\mathscr{K}_T$. By construction $\forall n$, $T_n = S$ if and only if $(S_n) \in \mathscr{K}_T$.

3. Extrapolation algorithms

Let us come back to the example of Aitken's Δ^2 process given in the preceding section. We saw that the system to be solved for constructing T is

$$\begin{cases} T_n = S = a_1 S_n + (1 - a_1) S_{n+1}, \\ T_n = S = a_1 S_{n+1} + (1 - a_1) S_{n+2}. \end{cases}$$

Adding and subtracting S_n to the first equation and S_{n+1} to the second one, leads to the equivalent system

$$\begin{cases} S_n = T_n + b \Delta S_n, \\ S_{n+1} = T_n + b \Delta S_{n+1}, \end{cases}$$

where $b = a_1 - 1$.

We have to solve this system for the unknown T_n. Using the classical determinantal formulae giving the solution of a system of linear equations, we know that T_n can be written as a ratio of determinants

$$T_n = \frac{\begin{vmatrix} S_n & S_{n+1} \\ \Delta S_n & \Delta S_{n+1} \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ \Delta S_n & \Delta S_{n+1} \end{vmatrix}}.$$

Of course the computation of a determinant of dimension two is well known and easy to perform and in the preceding case we obtain

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

which is again the formula of Aitken's process.

Let us now take a more complicated example to illustrate the problems encountered in our approach. We assume now that R has the form

$$a_1(u_i - S) + a_2(u_{i+1} - S) + \cdots + a_{k+1}(u_{i+k} - S) = 0$$

with $a_1 \cdot a_{k+1} \neq 0$. We now set $p = q = k$. For $k = 1$, the kernel of Aitken's process is recovered. Performing the same procedure as above (assuming that $a_1 + \cdots + a_{k+1} = 1$ since this sum has to be different from zero) leads to the system

$$\begin{cases} S_n = T_n + b_1 \Delta S_n + \cdots + b_k \Delta S_{n+k-1}, \\ S_{n+1} = T_n + b_1 \Delta S_{n+1} + \cdots + b_k \Delta S_{n+k}, \\ \vdots \\ S_{n+k} = T_n + b_1 \Delta S_{n+k} + \cdots + b_k \Delta S_{n+2k-1}. \end{cases}$$

Solving this system by the classical determinantal formulae gives for T_n:

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

In that case T_n will be denoted as $e_k(S_n)$. It is the well-known Shanks transformation.

The computation of $e_k(S_n)$ needs the computation of two determinants of dimension $(k+1)$ that is about $2(k+1)(k+1)!$ multiplications. However if the determinants involved in the definition of T_n have some special structures, as is the case of Shanks' transformation, then it is possible to obtain some rules (that is, an algorithm) for computing recursively these ratios of determinants. Such an algorithm will be called an extrapolation algorithm.

For example, Shanks' transformation can be implemented via the ε-algorithm of Wynn whose rules are

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

$$\varepsilon_{k+1}^{(n)} = \varepsilon_{k-1}^{(n+1)} + \frac{1}{\varepsilon_k^{(n+1)} - \varepsilon_k^{(n)}}, \quad k, n = 0, 1, \ldots,$$

and we have

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

the ε's with an odd lower index being intermediate quantities without any interesting meaning.

The ε-algorithm is one of the most important extrapolation algorithms.

Many sequence transformations are defined as a ratio of determinants and thus they need an extrapolation algorithm for their practical implementation. As explained above, such an algorithm usually allows to compute recursively the $T_k^{(n)}$'s.

4. Accelerability and non-accelerability

As explained above, a universal sequence transformation for accelerating the convergence of all convergent sequences cannot exist. More precisely, a universal transformation cannot exist for a set of sequences which is *remanent*. In other words a transformation able to accelerate the convergence of all the sequences of a remanent set cannot exist. This is clearly a very fundamental result in the theory of sequence transformations since it shows the frontier between accelerability and non-accelerability.

A set $\mathscr{S}$ of real convergent sequences is said to possess the property of generalized remanence if and only if

(1) There exists a convergent sequence $(\hat{S}_n)$ with limit $\hat{S}$ such that $\forall n$, $\hat{S}_n \neq \hat{S}$ and such that
 (a) $\exists (S_n^0) \in \mathscr{S}$ such that $\lim_{n\to\infty} S_n^0 = \hat{S}_0$.
 (b) $\forall m_0 \geqslant 0$, $\exists p_0 \geqslant m_0$ and $(S_n^1) \in \mathscr{S}$ such that $\lim_{n\to\infty} S_n^1 = \hat{S}_1$ and $\forall m \leqslant p_0$, $S_m^1 = S_m^0$.
 (c) $\forall m_1 > p_0$, $\exists p_1 \geqslant m_1$ and $(S_n^2) \in \mathscr{S}$ such that $\lim_{n\to\infty} \mathscr{S}_n^2 = \hat{S}_2$ and $\forall m \leqslant p_1$, $S_m^2 = S_m^1$.
 (d) ....

(2) $(S_0^0, \ldots, S_{p_0}^0, S_{p_0+1}^1, \ldots, S_{p_1}^1, S_{p_i+1}^2, \ldots, S_{p_2}^2, S_{p_2+1}^3, \ldots) \in \mathscr{S}$.

The diagram in Fig. 1 makes the property more clear.

The fundamental result is that, if a set of sequences possesses the property of generalized remanence, then a universal transformation able to accelerate the convergence of all its sequences cannot exist.

Such a set of sequences is said to be non-accelerable. Techniques similar but different from remanence can also be used to prove the non-accelerability of some sets of sequences. Actually many sets of sequences were proved to be non-accelerable. They are the following:

- The set of convergent sequences of E, where E is a metric space. This set will be denoted by *conv*(E).
- The set of convergent sequences of E such that $\exists N$, $\forall n \geqslant N$, $S_n \neq \lim_{n\to\infty} S_i$. This set will be denoted *conv*$^*(E)$.
- The subsets of *conv*(E) such that $\forall n$, $S_{n+1} \geqslant S_n$ or $S_{n+1} \leqslant S_n$ or $S_{n+1} > S_n$ or $S_{n+1} < S_n$.
- The subsets of *conv*(E) such that $\forall n$, $(-1)^i \Delta^i S_n \leqslant 0$ for $i = 1, \ldots, k$ or $(-1)^i \Delta^i S_n \geqslant 0$.
- The subsets of *conv*$(\mathbb{R})$ such that $(-1)^n \Delta S_n$ or $(-1)^n (S_n - S)$ has a constant sign.
- The subsets of *conv*$(\mathbb{R})$ such that $(-1)^n \Delta S_n$ or $(-1)^n (S_n - S)$ is monotone with a constant sign.
- The subsets of *conv*$^*(\mathbb{R})$ such that $\forall n \geqslant N$, $0 < \lambda \leqslant (S_{n+1} - S)/(S_n - S) \leqslant \mu < 1$ or $\lambda \leqslant \Delta S_{n+1}/\Delta S_n \leqslant \mu$.
- The subset of *conv*$^*(\mathbb{R})$ such that $\lim_{n\to\infty} (S_{n+1} - S)/(S_n - S) = 0$
- The set of logarithmic sequences, $\lim_{n\to\infty} (S_{n+1} - S)/(S_n - S) = 1$. This set is called LOG.
- The subset of logarithmic sequences such that $\lim_{n\to\infty} (S_{n+1} - S)/(S_n - S) = \lim_{n\to\infty} \Delta S_{n+1}/\Delta S_n = 1$. This set is called LOGSF.

It must be clearly understood that the preceding results do not mean that a particular sequence belonging to a non-accelerable set cannot be accelerated. It means that the same algorithm cannot accelerate all the sequences of the set.

$$
\begin{array}{ccccccc}
S_0^0 & = & S_0^1 & = & S_0^2 & \cdots & \\
S_1^0 & = & S_1^1 & = & S_1^2 & \cdots & \\
\vdots & & \vdots & & \vdots & & \\
S_{p_0}^0 & = & S_{p_0}^1 & = & S_{p_0}^2 & \cdots & \\
S_{p_0+1}^0 & & S_{p_0+1}^1 & = & S_{p_0+1}^2 & \cdots & \\
S_{p_0+2}^0 & & S_{p_0+2}^1 & = & S_{p_0+2}^2 & \cdots & \\
\vdots & & \vdots & & \vdots & & \\
\vdots & & S_{p_1}^1 & = & S_{p_1}^2 & \cdots & \\
\vdots & & S_{p_1+1}^1 & & S_{p_1+1}^2 & \cdots & \\
\vdots & & S_{p_1+2}^1 & & S_{p_1+2}^2 & \cdots & \\
\vdots & & \vdots & & \vdots & & \\
\vdots & & \vdots & & S_{p_2}^2 & \cdots & \\
\vdots & & \vdots & & \vdots & & \\
\downarrow & & \downarrow & & \downarrow & & \\
\hat{S}_0 & & \hat{S}_1 & & \hat{S}_2 & \to & \hat{S}
\end{array}
$$

Fig. 1. The property of generalized remanence.

This property justifies the fact that many extrapolation algorithms have to be discovered and studied in order to be able to accelerate the convergence of as many sets of sequences as necessary. Possible procedures for that purpose consist in using convergence tests for sequences and series, or constructing asymptotic expansions of the error, or building well-adapted extrapolation processes directly from the asymptotic properties of the sequence to be accelerated.

5. General comments

Among the existing extrapolation methods, the most well-known is certainly Aitken's Δ^2 process which is explained in almost all textbooks on numerical analysis. It has been generalized in several directions but the most important one is the ε-algorithm. The most general extrapolation method actually known is the E-algorithm. All these algorithms have been widely studied both from the theoretical and the practical point of view. They have received interesting and powerful applications.

In numerical analysis, one often has to accelerate sequences of vectors. This can be done by applying a scalar extrapolation method to each component of the vectors. However, there exist more interesting procedures, better adapted to the vector case. Among them, one can find the vector and the topological ε-algorithms, a vector E-algorithm and projection algorithms such as the so-called RPA and CRPA.

Apart from these algorithms, there exist special devices which can be very useful in connection with extrapolation methods. For example, it is often interesting in applications to estimate the error or to control it, that is to give a sequence of intervals containing the limit, the bounds of which converge faster than the initial sequence. One can also, before using an extrapolation method, extract a sub-sequence from the initial one and then accelerate it. This could help in enhancing the performance of the acceleration technique. When the user is faced with the choice of an extrapolation procedure, he can also program several of them and then select, among all results obtained at each step, one of them according to some test. Under some assumptions, it can be proved that such an automatic selection procedure provides the best possible answer.

Extrapolation methods have also many applications in the solution of various problems of numerical analysis outside the domain of convergence acceleration. For example, they provide a quadratically convergent method for solving systems of nonlinear equations (under some assumptions) which does not require the computation of any derivatives. The ε-algorithm provides a generalization of the power method which allows to compute simultaneously several eigenvalues of a matrix. Applications to integral and differential equations (and, in particular, boundary value problems), to numerical quadrature and to differentiation have been studied. Finally let us mention that applications to statistics (the jackknife, ARMA models and Monte-Carlo methods) have been developed.

Extrapolation methods have connections with such important topics as projection methods for systems of equations, Padé approximants, continued fractions and orthogonal polynomials, to name a few. It is a domain of numerical analysis presently in full expansion.

Reference

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