Numer. Math. 61, 475–487 (1992)

Numerische Mathematik

Extrapolation methods for vector sequences

P.R. Graves-Morris
Department of Mathematics, University of Bradford, Bradford, West Yorkshire BD7 1DP, UK

Received October 16, 1990

Summary. An analogue of Aitken's Δ^2 method, suitable for vector sequences, is proposed. Aspects of the numerical performance of the vector ε-algorithm, based on using the Moore-Penrose inverse, are investigated. The fact that the denominator polynomial associated with a vector Padé approximant is the square of its equivalent in the scalar case is shown to be a source of approximation error. In cases where the convergence of the vector sequence is dominated by real eigenvalues, a hybrid form of the vector Padé approximant, having a denominator polynomial of minimal degree, is proposed and its effectiveness is demonstrated on several standard examples.

Mathematics Subject Classification (1991): 65305

1 Introduction

Wynn (1962) discovered that the ε-algorithm can be used successfully to accelerate the convergence of vector sequences. To implement the ε-algorithm for vector-valued quantities, generalised inverses are used to define the reciprocals of vectors.

Suppose that a sequence

$$S := \{s_0, s_1, s_2, \ldots : s_i \in \mathbb{C}^d\} \tag{1.1}$$

of vectors is given. The vector ε-algorithm is initialised by taking

$$\varepsilon_0^{(j)} := s_j, \quad j = 0, 1, 2, \ldots, \tag{1.2}$$

and using the artificial values

$$\varepsilon_{-1}^{(j)} := \mathbf{0}, \quad j = 1, 2, 3, \ldots. \tag{1.3}$$

Its iterative steps are defined by

$$\varepsilon_{k+1}^{(j)} := \varepsilon_{k-1}^{(j+1)} + [\varepsilon_k^{(j+1)} - \varepsilon_k^{(j)}]^{-1}, \quad j = 1, 2, 3, \ldots, \quad k = 0, 1, 2, \ldots. \tag{1.4}$$

Generalised Moore-Penrose inverses are defined by

$$\boldsymbol{v}^{-1} := \boldsymbol{v}^*/(\boldsymbol{v}\cdot\boldsymbol{v}^*) \tag{1.5}$$

and are used in the right-hand side of (1.4); we have used the conventions that

$$\boldsymbol{a}\cdot\boldsymbol{b} := \sum_{i=1}^{d} a_i b_i \tag{1.6}$$

for vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{C}^d$, and an asterisk denotes complex conjugation. The elements $\{\boldsymbol{\varepsilon}_k^{(j)}, j = 0, 1, 2, \ldots\}$ are said to form the column of index k of the vector ε-table. The ε-table is constructed column by column, using (1.4). Brezinski (1975) and Cordellier (1989) report more recent developments.

A sequence is defined to be *generalised geometric* if its coefficients obey a recurrence relation with constant coefficients. McLeod (1971) proved Wynn's conjecture that the appropriate column of the vector ε-table consists of identical elements which are the limit values of a convergent generalised geometric sequence of vectors (subject to the technical conditions detailed in Theorem 2.2). The importance of McLeod's theorem is that it showed the possibility of obtaining analytical results governing the convergence of the vector ε-algorithm.

The formalism of vector Padé approximants is summarised in Sect. 2, where the connection between the vector ε-algorithm and vector Padé approximants is given explicitly. We also state the two main theorems governing convergence of the vector ε-algorithm in this section.

Section 3 contains a new approach to the old question of formulating a satisfactory analogue of Aitken's Δ^2 process, for vector sequences. There are a number of contenders for the best natural analogue, but here we specifically single out

$$\boldsymbol{t}_n := \boldsymbol{s}_{n+1} - \Delta\boldsymbol{s}_{n+1}\frac{(\Delta\boldsymbol{s}_n)^2}{\Delta\boldsymbol{s}_n\cdot\Delta^2\boldsymbol{s}_n} \tag{1.7}$$

for $\boldsymbol{s}_n \in \mathbb{R}^d$ as the most obvious generalisation supported by a mathematical framework. The result (1.7) is *neither* an entry in the vector ε-table *nor* a vector Padé approximant but a hybrid development which is motivated both by a theoretical derivation and by its performance. There are a number of generalisations of Aitken's formula for vector sequences (Smith et al. 1987; Macleod 1986), but (1.7) appears to be different from all of them.

Section 4 contains a construction of the vector Padé approximant of the generating function associated with the recurrence

$$\boldsymbol{s}_{i+1} := E\boldsymbol{s}_i + \boldsymbol{e}, \quad i = 0, 1, 2, \ldots$$

where $\boldsymbol{s}_i$, $\boldsymbol{e} \in \mathbb{C}^d$ and E is a complex valued $d \times d$ matrix. The degrees of the numerator and denominator of this generating function dictate which elements in which column of the vector ε-algorithm are equal to the fixed point of the recurrence. These considerations are an important supplement to previous observations on the applicability of McLeod's theorem in this context, as is borne out by the examples in Sect. 5.

In Sect. 5, we review the numerical examples originally considered by Wynn (1963) and subsequently in the excellent review by Smith et al. (1987). Hitherto, it does not seem to have been noticed that the known asymptotic properties of

certain columns of the vector ε-table are noticeably absent in low order. We have no fundamental understanding of this fact, but, by accepting it, we have a partial explanation of why the vector ε-algorithm does not work as well in low order as the major theorems prove that it does in high order. We also draw attention to the impact of rounding error. We note that the denominator of a vector Padé approximant is the square of the corresponding one for a scalar Padé approximant when such a correspondence can be made [see (2.16), (2.17), (3.3), (3.4), (4.12)], and this circumstance is usually associated with loss of precision if the denominators are numerically small. As Wynn (1962) implies, rounding error seems to be a more serious problem in the vector than in the scalar case.

In Sect. 6, we consider a numerical example due to Varga (1962). It originates in the solution by spatial discretisation of an equation for neutron diffusion in a two dimensional model reactor. Varga describes several different SOR methods and illustrates their effectiveness in the context of this model. Here we show how the number of iterations can be substantially reduced by using acceleration of convergence of a chosen vector sequence.

2 Vector-valued Padé approximants

In this section, the vector ε-algorithm and its connection with vector-valued Padé approximants are briefly reviewed and the relevant theorems are summarised. The theory is developed in terms of a generating function formally defined as

$$\boldsymbol{f}(z) := \sum_{j=0}^{\infty} \boldsymbol{c}_j z^j , \tag{2.1}$$

where $\boldsymbol{c}_0 := \boldsymbol{s}_0$ and $\boldsymbol{c}_j := \boldsymbol{s}_j - \boldsymbol{s}_{j-1}, j = 1, 2, 3, \ldots$. Vector Padé approximants of the form

$$\boldsymbol{r}(z) = \boldsymbol{p}(z)/q(z) \tag{2.2}$$

(previously called generalised inverse Padé approximants by Graves-Morris and Jenkins 1986) are associated with the series (2.1). The right-hand side of (2.2) is a vector Padé representation of $\boldsymbol{r}(z)$ of type $[n/2k]$ if

$$\partial\{\boldsymbol{p}\} \leqq n, \qquad \partial\{q\} = 2k \tag{2.3}$$

$$q(z) \mid \boldsymbol{p}(z) \cdot \boldsymbol{p}^*(z) \tag{2.4}$$

$$q(z) = q^*(z) \tag{2.5}$$

and

$$\boldsymbol{r}(z) - \boldsymbol{f}(z) = O(z^{n+1}) . \tag{2.6}$$

The factorisation property (2.4) means that a polynomial $\dot{q}(z)$ must exist such that

$$q(z)\dot{q}(z) = \boldsymbol{p}(z) \cdot \boldsymbol{p}^*(z) .$$

These four axioms suffice to determine $\boldsymbol{r}(z)$ uniquely and it is the analogue of the $[n - k/k]$ Padé approximant of a scalar function. We have ignored degenerate cases in (2.6). These cases are treated by Graves-Morris and Jenkins (1989), who

show how vector Padé polynomials $(\boldsymbol{p}(z), q(z))$ are constructed in the general case. (The existence of these polynomials is assumed in Theorem 2.1).

Reverting to non-degenerate cases, the denominator polynomial is given by

$$q(z) = \begin{vmatrix} 0 & M_{01} & \dots & M_{0,2k-1} & M_{0,2k} \\ -M_{01} & 0 & \dots & M_{1,2k-1} & M_{1,2k} \\ \vdots & \vdots & & \vdots & \vdots \\ -M_{0,2k-1} & -M_{1,2k-1} & \dots & 0 & M_{2k-1,2k} \\ z^{2k} & z^{2k-1} & \dots & z & 1 \end{vmatrix} \tag{2.7}$$

and

$$M_{ij} := \sum_{l=0}^{j-i-1} \boldsymbol{c}_{l+i+n-2k+1} \cdot \boldsymbol{c}^*_{j-l+n-2k}, \quad j > i\,. \tag{2.8}$$

The numerator polynomial is obtained from (2.6)

$$\boldsymbol{p}(z) = [\boldsymbol{f}(z)q(z)]_0^n \tag{2.9}$$

using Nuttall's truncation notation in which

$$[\varphi(z)]_\alpha^\beta := \sum_{i=\alpha}^{\beta} \varphi_i z^i\,.$$

It is also true that

$$q(x) \geqq 0, \quad -\infty < x < \infty\ ; \tag{2.10}$$

in non-degenerate cases, which follows as an immediate Corollary of (2.4) and (2.7). The connection between vector Padé approximants and the vector ε-algorithm is given by the formula

$$\boldsymbol{\varepsilon}_{2k}^{(j)} = [2k + j/2k](1) \tag{2.11}$$

for all $j, k \geqq 0$ whenever either side is well defined. Because of this correspondence, we can interpret the behaviour of the columns of the vector epsilon table in terms of the corresponding rows of the vector Padé table, for which the following theorem has been established.

Theorem 2.1 *(Graves-Morris and Saff 1988). Let*

$$\boldsymbol{f}(z) = \boldsymbol{g}(z)/Q(z) \tag{2.12}$$

be a vector function which is analytic in a disk except for poles of total multiplicity k. More specifically, we require that

(2.13)
(i) *Q is a monic real polynomial of degree k with roots $\{z_1, z_2, \dots, z_k\}$ and $0 < |z_i| < \rho$, $i = 1, 2, \dots, k$,*
(ii) *$\boldsymbol{g}(z)$ is analytic on $|z| < \rho$,*
(iii) *$\boldsymbol{g}(z_i) \cdot \boldsymbol{g}^*(z_i) \neq 0$, $i = 1, 2, \dots, k$.*

Define $D_\rho^- := \{z : |z| < \rho\} - \bigcup_{i=1}^k \{z_i\}$ *and let* K *be any compact subset of* D_ρ^-. *Let* E *be any compact subset of* $\mathbb{C}$. *Let* $(\boldsymbol{P}_n(z), Q_n(z))$ *be vector Padé polynomials of type* $[n/2k]$ *for which* $Q_n(z)$ *has leading coefficient equal to unity. Then*

$$\lim_{n\to\infty} \boldsymbol{P}_n(z)/Q_n(z) = \boldsymbol{f}(z), \quad z \in D_\rho^- \tag{2.14}$$

and the rate of convergence is governed by

$$\limsup_{n\to\infty} |\boldsymbol{f}(z) - \boldsymbol{P}_n(z)/Q_n(z)|^{1/n} \leqq |z|/\rho \tag{2.15}$$

for $z \in K$. *The denominators converge according to*

$$\lim_{n\to\infty} Q_n(z) = Q^2(z) \tag{2.16}$$

and the rate is given by

$$\limsup_{n\to\infty} \| Q_n - Q^2 \|_E^{1/n} \leqq \max_{1\leqq i\leqq k} \frac{|z_i|}{\rho} \tag{2.17}$$

where $\| \cdot \|_E$ *denotes the supremum norm on the set* E.

In Sect. 4, we will see applications of these results in the context of convergence of columns of the ε-table in several examples. If the generating function (2.1) is rational, the appropriate vector Padé approximants (2.2) are exact, and this too will be apparent from the examples in Sect. 4. The theorem which governs these cases is

Theorem 2.2 *(McLeod 1971; Graves-Morris 1983). Suppose that the vector sequence* S *of* (1.1) *satisfies a non-trivial recurrence relation*

$$\sum_{i=0}^{k} \beta_i \boldsymbol{s}_{i+j} = \left[\sum_{i=0}^{k} \beta_i\right] \boldsymbol{a}, \quad j = 0, 1, 2, \ldots \tag{2.18}$$

with $\beta_i \in \mathbb{C}$. *Then the vector* ε*-algorithm* (1.3) *leads to*

$$\boldsymbol{\varepsilon}_{2k}^{(j)} = \boldsymbol{a} \tag{2.19}$$

provided that zero divisors are not encountered in the construction.

The condition (2.18) ensures that $\boldsymbol{s}_i$ satisfy a $k+1$ term recurrence relation, and thus that S is a generalised geometric sequence. If $\beta_0 = 0$, or $\beta_k = 0$ or $\sum_{i=0}^k \beta_i = 0$, the vectors $\boldsymbol{s}_i$ would, in fact, satisfy a k term recurrence relation, and then zero divisors would be encountered before the column of index $2k$ of the vector ε-table could be constructed.

3 Aitken's Δ^2 process for sequences of real vectors

Aitken's (1926) Δ^2 algorithm for a scalar sequence $s_0, s_1, s_2, \ldots$, takes the standard form

$$t_n := s_n - (\Delta s_n)^2/\Delta^2 s_n, \quad n = 0, 1, 2, \ldots, \tag{3.1}$$

using the convention that $\Delta s_n := s_{n+1} - s_n$, etc. The idea is that, if $\{s_n\}$ converges to a limit s_∞ and the convergence is dominated by a single geometric component, then $\{t_n\}$ converges more quickly to s_∞. It is well known that (3.1) corresponds to the use of the second row of the Padé table, namely the sequence of Padé approximants of type $[n + 1/1]$ (Baker and Graves-Morris 1981, Chap. 3). It is elementary to prove that (3.1) may also be expressed as

$$t_n = s_{n+1} - (\Delta s_n)(\Delta s_{n+1})/\Delta^2 s_n, \quad n = 0, 1, 2, \ldots ; \tag{3.2}$$

Weniger (1989) has compiled six other similar formulae, all equivalent to (3.1). Since (3.2) and the other six formulae lack the aesthetic appeal of (3.1), they are rarely used.

How should (3.1) or (3.2) be generalised for vectors $s_i \in \mathbb{R}^d$? Smith et al. (1987, Sect. 10) and Macleod (1986) review various alternatives previously given. An obvious answer is to use the column $\varepsilon_2^{(j)}$ of the vector ε-table corresponding to entries in the second row of the table of vector Padé approximants. From (2.7)–(2.9), we find that we should define

$$\hat{\boldsymbol{t}}_n := \boldsymbol{s}_{n+1} - [\Delta \boldsymbol{s}_n (\Delta \boldsymbol{s}_{n+1})^2 - \Delta \boldsymbol{s}_{n+1} (\Delta \boldsymbol{s}_n)^2]/(\Delta^2 \boldsymbol{s}_n)^2, \quad n = 0, 1, 2, \ldots , \tag{3.3}$$

so that $\hat{\boldsymbol{t}}_n = \boldsymbol{r}^{[n/2]}(1)$. This formula is clearly related to (3.2), and it reduces to (3.2) for the case of $\boldsymbol{s}_n \in \mathbb{R}^1$.

Let us reconsider the theory underlying (3.1)–(3.3). The denominator of (3.3) arises from use of the formula

$$q^{[n/2]}(z) = z^2 |\boldsymbol{c}_n|^2 - 2z \boldsymbol{c}_n \cdot \boldsymbol{c}_{n-1} + |\boldsymbol{c}_{n-1}|^2 , \tag{3.4}$$

[using a more convenient normalisation than that of (2.7)]. Because the theoretical basis for Aitken's algorithm uses the supposition that convergence of the given sequence is dominated by a single geometric component, we will suppose that the dominant contribution to the vector generating function $\boldsymbol{f}(z)$ at $z = 1$ to be that of a nearby real simple pole. We would expect the denominators of the vector Padé approximants (given by (3.4)) to have a pair of complex-conjugate zeros near the dominant pole of $\boldsymbol{f}(z)$, according to (2.16) and (2.17). Because we are really seeking an estimate of the position of a real pole, it seems better to do this by using the zero of the first derivative of $q^{[n/2]}(z)$. We use

$$\tilde{q}^{[n/2]}(z) := z|\boldsymbol{c}_n|^2 - \boldsymbol{c}_n \cdot \boldsymbol{c}_{n-1} \tag{3.5}$$

as a hybrid denominator, whose single simple real zero approximates the mean position of the complex-conjugate pair of zeros of (3.4). We define a corresponding hybrid numerator polynomial by

$$\tilde{\boldsymbol{p}}^{[n/2]}(z) := [\boldsymbol{f}(z)\tilde{q}^{[n/2]}(z)]_0^n . \tag{3.6}$$

We obtain

$$\boldsymbol{t}_n := \frac{\tilde{\boldsymbol{p}}^{[n/2]}(1)}{\tilde{q}^{[n/2]}(1)} = \boldsymbol{s}_{n+1} - \Delta \boldsymbol{s}_{n+1} \frac{(\Delta \boldsymbol{s}_n)^2}{\Delta \boldsymbol{s}_n \cdot \Delta^2 \boldsymbol{s}_n} \tag{3.7}$$

as the appropriate generalisation of Aitken's formula for $\boldsymbol{s}_n \in \mathbb{R}^d$. Numerical advantages of this formula have been reported by Graves-Morris (1990) for the case

$d = \infty$, and another is reported in Sect. 6 for $d = 16$. Its closest relative amongst the published generalisations of Aitken's formula appears to be that of Zienkiewicz and Löhner (1985).

4 Representation of the generating functions

In this section, we establish conditions under which a certain column of the vector ε-table contains identical elements which are the limiting value (or possibly the anti-limit) of a given sequence of vectors. We begin by assuming that the sequence S is generated by

$$\boldsymbol{s}_{i+1} := E\boldsymbol{s}_i + \boldsymbol{e}, \quad i = 0, 1, 2, \ldots, \tag{4.1}$$

where $\boldsymbol{s}_i, \boldsymbol{e} \in \mathbb{C}^d$ and E is a complex-valued $d \times d$ matrix. We derive the generating function S and its minimal Padé representation. Our approach is directly related to that of Smith et al. [1987], who establish that the $\boldsymbol{s}_i$ satisfy a recurrence relation and hence that McLeod's theorem is applicable. However, they omit a full discussion of minimality, and this discussion is necessary to determine which column of the ε-table contains the limiting values. Obviously, any further iteration with (4.1) would encounter zero divisors, and McLeod's theorem would not apply to such cases.

If the matrix $I - E$ is non-singular, the recurrence (4.1) has a unique fixed point $\boldsymbol{s}_\infty$ satisfying

$$\boldsymbol{s}_\infty = E\boldsymbol{s}_\infty + \boldsymbol{e} . \tag{4.2}$$

A necessary and sufficient condition for $I - E$ to be non-singular is that no eigenvalue of E is equal to unity.

Let $q^{(M)}(z)$ be the minimal polynomial of E (Wilkinson, 1965, p. 37), and suppose that its degree is k_M. We have $k_M \leqq d$ and, by definition,

$$q^{(M)}(E) = 0 . \tag{4.3}$$

If $q^{(M)}(0) = 0$, let μ be the multiplicity of this zero of $q^{(M)}(z)$; otherwise let $\mu = 0$ when $q^{(M)}(0) \neq 0$. Let

$$q^{(B)}(z) := \sum_{i=0}^{k_B} q_i^{(B)} z^i := z^{k_M} q^{(M)}(z^{-1}) \tag{4.4}$$

so that

$$k_B := \partial\{q^{(B)}(z)\} = k_M - \mu . \tag{4.5}$$

We have the property that $q^{(B)}(0) \neq 0$, by virtue of the definition of k_M. As before, we define $\boldsymbol{c}_0 := \boldsymbol{s}_0$, $\boldsymbol{c}_{i+1} = \Delta\boldsymbol{s}_i$, $i = 0, 1, 2, \ldots$, but (4.1) gives the extra property that

$$\boldsymbol{c}_{i+1} = E\boldsymbol{c}_i, \quad i = 1, 2, \ldots . \tag{4.6}$$

It is familiar that the series (2.1) for $\boldsymbol{f}(z)$ converges if $|z| < \|E\|^{-1}$. By elementary algebra, we find that

$$q^{(B)}(z)\boldsymbol{f}(z) = \boldsymbol{p}^{(B)}(z) \tag{4.7}$$

where

$$\boldsymbol{p}^{(B)}(z) := \sum_{i=0}^{k_M} z^i \sum_{j=0}^{\min(i,k_B)} q_j \boldsymbol{c}_{i-j} \,. \tag{4.8}$$

From (4.7), we have (trivially),

$$\boldsymbol{f}(z) = \boldsymbol{p}^{(B)}(z)/q^{(B)}(z) \tag{4.9}$$

and (4.9) is useful because it extends the domain of definition of $\boldsymbol{f}(z)$ by analytic continuation to all z for which $q^{(B)}(z) \neq 0$.

From (2.1) and (4.6), we have

$$(zE - I)\boldsymbol{f}(z) = (z-1)\boldsymbol{c}_0 - z\boldsymbol{e} \,. \tag{4.10}$$

On putting $z = 1$, we obtain

$$\boldsymbol{f}(1) = E\boldsymbol{f}(1) + \boldsymbol{e} \,, \tag{4.11}$$

showing that $\boldsymbol{s}_\infty = \boldsymbol{f}(1)$ is a fixed point of (4.1), which is unique if $I - E$ is non-singular.

The expression (4.9) is not normally a vector Padé representation of $\boldsymbol{f}(z)$. We express it as

$$\boldsymbol{f}(z) = \boldsymbol{p}^{(B)}(z)q^{(B)*}(z)/(q^{(B)}(z)q^{(B)*}(z)) \tag{4.12}$$

which is a vector Padé representation of type $[k_M + k_B/2k_B]$ in which the numerator and denominator polynomials satisfy the factorisation property (2.4). Using the methods of Graves-Morris and Jenkins (1989), a unique minimal vector Padé representation of precise type $[\hat{n}/2\hat{k}]$ can be found and it takes the form

$$\boldsymbol{f}(z) = \hat{\boldsymbol{p}}(z)/\hat{q}(z) \tag{4.13}$$

with

$$\hat{n} \leqq 2d - \mu, \qquad \hat{k} \leqq d - \mu \,. \tag{4.14}$$

Using the identification (2.11) and the uniqueness theorem of Graves-Morris and Jenkins (1989), we have the result

$$\varepsilon_{2\hat{k}}^{(j)} = \boldsymbol{f}(1) = \boldsymbol{s}_\infty, \quad j = \mu, \mu + 1, \ldots \,. \tag{4.15}$$

The examples of Sect. 5 show that equality in (4.14) is possible; in fact, equality is the normal occurrence. Example 5.2 shows that the equality of (4.15) does not normally extend to $j < \mu$.

Because $q^{(M)}(z)$ is the minimal polynomial of E, the roots of $z^{-k_M}q^{(M)}(z)$ are non-null eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{k_M - \mu}$ of E. Therefore the roots z_i of $q^{(B)}(z)$ are their reciprocals:

$$z_i := \lambda_i^{-1}, \quad i = 1, 2, \ldots, k_M - \mu \,.$$

We apply Theorem 2.1 to $\boldsymbol{f}(z)$ and consider convergence of row sequences of vector Padé approximants for $\boldsymbol{f}(z)$ in the disks $|z| < |z_i|$, $i = 1, 2, \ldots, k_M - \mu$. Convergence of the vector ε-algorithm is associated with convergence of this sequence at $z = 1$. Let us suppose that

$$|\lambda_1| \geqq |\lambda_2| \geqq \ldots \geqq |\lambda_K| > 1 > |\lambda_{K+1}| > |\lambda_{K+2}| > \ldots > |\lambda_{\hat{k}}| \,. \tag{4.16}$$

Then the even index columns of the vector ε-table with index greater than or equal to $2K$ are convergent, and the rate of convergence is governed by

$$\limsup_{j\to\infty} |f(1) - \varepsilon_{2k}^{(j)}|^{1/j} \leqq |\lambda_{k+1}|, \quad k = K, K+1, \ldots, \hat{k} - 1 . \tag{4.17}$$

In Sects. 5 and 6 we give numerical examples which illustrate the results (4.15), (4.17), but only when sufficient numerical precision is used, and sufficiently large values of j are used in (4.17).

5 Numerical convergence in standard examples

In this section, we review two standard examples in the light of preceding theoretical results. Both originate from the solution of

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} , \tag{5.1}$$

where A and $\boldsymbol{b}$ are given by

$$A := \begin{bmatrix} 5 & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix}, \qquad \boldsymbol{b} := \begin{bmatrix} 23 \\ 32 \\ 33 \\ 31 \end{bmatrix} . \tag{5.2}$$

The examples are purely illustrative, because realistic applications would involve matrices of high dimension. To generate an iterative solution of (5.1), A is decomposed as

$$A = B + C \tag{5.3}$$

and the recurrence is

$$\boldsymbol{s}_{i+1} = E\boldsymbol{s}_i + \boldsymbol{d} \tag{5.4}$$

where

$$E := -B^{-1}C \qquad \text{and} \qquad \boldsymbol{d} := B^{-1}\boldsymbol{b} . \tag{5.5}$$

Example 5.1. This is a simple example of Jacobi iteration (diagonal relaxation). The iteration expressed by (5.2)–(5.5) is implemented with

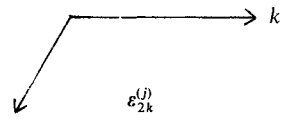
$$B := \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 10 \end{bmatrix}, \qquad \boldsymbol{s}_0 := \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} .$$

Some of the entries in the corresponding vector epsilon table are shown in Table 1. The eigenvalues of the iteration matrix E are

$$\lambda_1 = -2.4758, \qquad \lambda_2 = 0.9985, \qquad \lambda_3 = 0.9151, \qquad \lambda_4 = 0.5622 .$$

Table 1. Entries in even index columns of the vector epsilon table for Example 5.1. For brevity, only six vectors of the initialising sequence are shown explicitly

0.000				
0.000				
0.000				
0.000				
4.600	1.311			
3.200	0.912			
3.300	0.954			
3.100	0.905			
− 6.940	1.273	1.222		
− 4.830	0.885	0.855		
− 4.810	0.969	0.980		
− 4.410	0.947	1.007		
21.54	1.251	1.221	1.207	
14.99	0.871	0.857	0.874	
15.30	0.956	0.977	0.948	
14.28	0.972	1.008	1.031	
− 49.03	1.238	1.221	1.207	1.000
− 34.11	0.863	0.858	0.875	1.000
− 34.47	0.978	0.974	0.948	1.000
− 32.93	0.987	1.010	1.031	1.000
125.66	1.231	1.220	1.207	1.000
87.45	0.860	0.860	0.875	1.000
88.75	0.978	0.971	0.948	1.000
82.52	0.996	1.011	1.031	1.000
$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$



We know, therefore, that the columns $2k = 4,6$ of the ε-table converge at rates $\sim (\lambda_3)^j$ and $\sim (\lambda_4)^j$, but these rates of convergence are not evident from Table 1. As occurred in the previous example, there is substantial loss of numerical precision due to rounding, and only one significant figure of decimal precision in the $k = 4$ column was obtained using single precision on the Cyber 180. The results shown in Table 1 were derived using double precision working. The asymptotic rate of convergence of $\varepsilon_6^{(j)}$ was attained for $j \geqq 27$. In practice, $\varepsilon_2^{(j)}$ and $\varepsilon_4^{(j)}$ never showed asymptotic behaviour, because rounding error obscured its onset. Apart from the differences in character of the denominator polynomial (as expressed by (2.16)) and its immediate consequences, we note that the asymptotic behaviour and its associated error analysis for the vector case are virtually identical to that for the scalar case, when the initial sequence satisfies a given linear recurrence relation. Similar observations were made by Wynn (1962). Estimates of precision should only be based on a running error analysis.

Example 5.2. This is a simple example of Gauss-Seidel iteration. The iteration expressed by (5.3)–(5.5) is implemented with

$$B := \begin{bmatrix} 5 & 0 & 0 & 0 \\ 7 & 10 & 0 & 0 \\ 6 & 8 & 10 & 0 \\ 5 & 7 & 9 & 10 \end{bmatrix}, \qquad s_0 := \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

In fact the eigenvalues of the iteration matrix $E = B^{-1}C$ are

$$\lambda_1 = 0.9969, \quad \lambda_2 = 0.8373, \quad \lambda_3 = 0.6038, \quad \lambda_4 = 0 . \tag{5.6}$$

Because one eigenvalue is zero, the index μ in (4.5) takes the value $\mu = 1$, and from (4.12), we expect that $f(\lambda)$ has a GIPA representation of type (7/6). From (4.15), we obtain

$$\varepsilon_6^{(j)} = f(1), \quad j = 1, 2, 3, \ldots \tag{5.7}$$

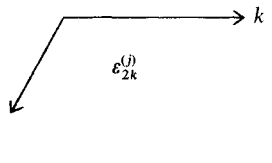
which is the exact solution of (5.1), (5.2). Some of the entries of the vector epsilon table are shown in Table 2, and we see that (5.7) holds good in this case. Because the possibility that

$$\varepsilon_6^{(0)} \neq \varepsilon_6^{(j)}, \quad j = 1, 2, 3, \ldots$$

does not occur in the equivalent scalar case, its occurrence in this example appears to have been overlooked by previous authors. This is not surprising, because the size of the rounding error encountered using single precision arithmetic obscures the effect (see Table 3 of Wynn 1962). Because the dominant eigenvalue of E is $\lambda_1 = 0.9969$, we expect that the column $\varepsilon_0^{(j)}$ converges at a rate $\sim(0.9969)^j$, $\varepsilon_2^{(j)}$ at a rate $\sim(0.84)^j$ and $\varepsilon_4^{(j)}$ at a rate $\sim(0.6)^j$. These expectations are borne out in practice, although they are not evident from the figures shown in Table 2. In fact

Table 2. Entries in even index columns of the vector epsilon table for Example 5.2. For brevity, only six vectors of the initialising sequence are shown

0.000				
0.000				
0.000				
0.000				
4.600	3.781			
− 0.020	− 0.018			
0.556	0.738			
0.314	0.454			
3.647	2.290	2.310		
− 0.017	0.059	0.078		
0.843	1.037	0.978		
0.530	0.980	1.010		
3.083	2.310	2.328	2.324	
− 0.003	0.078	0.201	0.203	
0.976	0.978	0.664	0.666	
0.682	1.010	1.197	1.196	
2.751	2.329	2.324	1.000	1.000
0.016	0.093	0.203	1.000	1.000
1.023	0.929	0.666	1.000	1.000
0.793	1.034	1.196	1.000	1.000
2.557	2.347	2.320	1.000	
0.036	0.107	0.206	1.000	indeterminate
1.023	0.886	0.667	1.000	
0.875	1.054	1.196	1.000	
$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$



$\varepsilon_2^{(j)}$ starts to display its asymptotic convergence rate at $j \sim 50$ and $\varepsilon_4^{(j)}$ at $j \sim 30$. The figures in Table 2 show that the initial convergence rate appears more logarithmic than geometric. Why this should be so is not known yet, partly because this phenomenon does not occur in the scalar case. A consequence of this poor convergence rate is that further substantial rounding error is introduced by the epsilon algorithm, and about 24 bits of precision have been lost in forming $\varepsilon_6^{(j)}$. The single precision working used here is based on 47 bits of mantissa precision from a Cyber 180, and this is quite sufficient to justify the four decimal places shown in Table 2.

6 Numerical convergence in SOR iteration

For a long time, there has been speculation that acceleration of convergence of a sequence of vectors would find application to the SOR iterates (Smith et al. 1987). We consider here an example due to Varga (1962), involving the sequence generated

$$\boldsymbol{s}_{i+1} = E\boldsymbol{s}_i \,. \tag{6.1}$$

where $\boldsymbol{s}_i \in \mathbb{R}^{16}$, $E \in \mathbb{R}^{16\times 16}$. It is initialised by

$$\boldsymbol{s}_0 = (10^4, 10^4, \ldots, 10^4) \,, \tag{6.2}$$

and arises from an investigation of the numerical solution of

$$A\boldsymbol{x} = \boldsymbol{0} \tag{6.3}$$

The values of the entries of A are given in Appendix B of Varga (1962). In the point SOR method, A is decomposed as

$$A = D - B - C \tag{6.4}$$

where D is diagonal and B, C are strictly lower and upper triangular respectively. We then define

$$L := D^{-1}B, \qquad U := D^{-1}C \tag{6.5}$$

and the point successive relaxation matrix by

$$E := (1 - \omega L)^{-1}\{(1-\omega)I + \omega U\} \tag{6.6}$$

following Varga (1962, Chap. 3). Equations (6.1), (6.2) and (6.6) specify the recurrence, subject to provision of a value for ω. Iteration is continued until

$$|(\boldsymbol{s}_i)_j| < 1.0, \quad j = 1, 2, \ldots, 16 \,. \tag{6.7}$$

The optimal value of ω for point SOR is $\omega = 1.9177$ and 139 iterations of (6.1) suffice for all components of $\boldsymbol{s}_{139}$ to be less than unity.

We will consider extrapolation of the vector sequence $\{\boldsymbol{s}_i\}$ using column n of index 2 of the vector ε-table. As Theorem 2.1 and (4.16) indicate, convergence will be fastest if the effects of one dominant pole are taken into account by the algorithm. This is achieved by taking $\omega = 1.52$. The eigenvalues of E then satisfy

$$\lambda_1 = 0.9942,$$

$$|\lambda_i| \leqq 0.52, \quad i = 2, 3, \ldots, 16 \,.$$

It was found that 26 iterations sufficed for (6.7) to hold.

Bearing in mind the analysis of Sect. 3, it is natural to ask if Aitken's Δ^2 method as expressed by (3.7) reduces the number of iterations. Using the same value, $\omega = 1.52$, it was found that 20 iterations of (6.1) yields $\boldsymbol{t}_{20}$ which satisfies the convergence criterion of (6.7). It will be interesting to see if the improved rates of convergence in the model problem of dimension $d = 16$ survive in full-scale applications.

Acknowledgement. I am grateful to David Roberts for some helpful comments.

References

Aitken, A.C. (1926): On Bernoulli's numerical solution of algebraic equations. Proc. Roy. Soc. Edin. **46**, 289–305

Baker, G.A. Jr., Graves-Morris, P.R. (1981): Padé approximants. Addison Wesley, Cambridge

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

Cordellier, F. (1989): Thesis, Univ. Lille

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

Graves-Morris, P.R. (1990): Solution of integral equations using generalised inverse, function-valued Padé approximants I. J. Comput. Appl. Math. **32**, 117–124

Graves-Morris, P.R., Jenkins, C.D. (1986): Vector-valued rational interpolants III. Constr. Approx. **2**, 263–289

Graves-Morris, P.R., Jenkins, C.D. (1989): Degeneracies of generalised inverse, vector-valued Padé approximants, Constr. Approx. **5**, 463–485

Graves-Morris, P.R., Saff, E.B. (1988): Row convergence theorems for generalised inverse vector-valued Padé approximants. J. Comput. Appl. Math. **23**, 63–85

Macleod, A.J. (1986): Acceleration of vector sequence by multidimensional Δ^2 methods. Comm. Appl. Numer. Meth. **2**, 385–392

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

Smith, D.A., Ford, W.F., Sidi, A. (1987): Extrapolation methods for vector sequences. SIAM Rev. **29**, 199–233

Varga, R.S. (1962): Matrix iterative analysis. Prentice-Hall, Englewood Cliffs, N.J.

Weniger, E.J. (1989): Non-linear sequence transformations for the acceleration of convergence and the summation of divergent series. Comput. Phys. Rep. **10**, 189–371

Wilkinson, J.H. (1965): The algebraic eigenvalue problem. Oxford, Oxford University Press

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

Wynn, P. (1963): Continued fractions whose coefficients obey a non-commutative law of multiplication. Arch. Rat. Mech. Anal. **12**, 273–312

Zienkiewicz, O.C., Löhner, R. (1985): Accelerated 'relaxation' or direct solution? Future prospects for FEM. Int. J. Numer. Meth. Eng. **21**, 1–11