Journal of Computational and Applied Mathematics 36 (1991) 385–398
North-Holland


# Some results about vector extrapolation methods and related fixed-point iterations

Khalide Jbilou and Hassane Sadok

*Laboratoire d'Analyse Numérique et d'Optimisation, UFR IEEA-M3, Université des Sciences et Techniques de Lille Flandres Artois, 59655 Villeneuve d'Ascq Cedex, France*

Received 4 December 1990

*Abstract*

Jbilou, Kh. and H. Sadok, Some results about vector extrapolation methods and related fixed-point iterations, Journal of Computational and Applied Mathematics 36 (1991) 385–398.

In this work we consider polynomial extrapolation methods for vector sequences. We give a unified presentation of these methods and a new recursive algorithm for the implementation of the modified minimal polynomial extrapolation (MMPE) method. Properties, operation counts and storage requirements of the algorithm will be given. The second part of the work is devoted to the application of these methods for solving systems of nonlinear equations; a sufficient condition for quadratic convergence is given.

*Keywords:* Polynomial extrapolation methods, vector sequences, nonlinear equations, quadratic convergence.

## 1. Introduction and notations

Vector extrapolation methods can be divided into two families [27], the polynomial methods and the epsilon algorithms. We consider in this paper the first family which includes the minimal polynomial extrapolation (MPE) method of Cabay and Jackson [8], Vorobyev [29] and Germain-Bonne [11] methods, the reduced rank extrapolation (RRE) method of Eddy [9], Mesina [17] and Kaniel and Stein [14], and the modified minimal polynomial extrapolation (MMPE) method due to Brezinski [3], Pugachev [19] and Sidi, Ford and Smith [25].

Some convergence and stability properties of these methods have been given in [19,22,24,25]. In [23], Sidi showed that the methods considered here, when applied to linearly generated vector sequences, are Krylov conjugate gradient type methods. This problem was also treated by Jbilou [13] and Beuneu [1] who provides another way of studying them in a unified context.

Different recursive techniques for the implementation of the MPE, RRE, MMPE and other related vector extrapolation methods have been derived in [10] and in [7] for the MMPE and other related methods.

In the first section we give a unified presentation of the methods discussed here and derive a new recursive algorithm for the implementation of the MMPE.

The second section is devoted to the application of these methods for solving systems of nonlinear equations. Our contribution to the subject consists in giving a sufficient condition and a full satisfactory proof for their quadratic convergence.

Now, let us give some notations to be used throughout this paper. $(\cdot, \cdot)$ denotes the inner product such that $(\alpha x, \beta y) = \bar{\alpha}\beta(x, y)$ for $\alpha$ and $\beta$ complex numbers and $x$ and $y$ two vectors of $\mathbb{C}^p$.

We shall also use the Euclidean vector norm $\| x \|$, the induced matrix norm $\| A \| = \max_{\| x \| = 1} \| Ax \|$ and the Frobenius matrix norm $\| A \|_F = (\sum_{j=1}^{k} \| a^j \|^2)^{1/2}$, where $A$ is a $p \times k$ matrix whose columns are $a^1, \ldots, a^k$.

## 2. The ($s$-$\beta$)-algorithm

Let $\{s_n\}$ be a sequence of $\mathbb{C}^p$ converging to $s$. For $k$, an integer such that $k \leqslant p$, and $n \in \mathbb{N}$, we define the sequence $\{t_{n,k}\}$ by

$$t_{n,k} = s_n - \Delta S_{n,k}\left(Y_{n,k}^* \, \Delta^2 S_{n,k}\right)^{-1} Y_{n,k}^* \, \Delta s_n, \tag{2.1}$$

where $\Delta^i S_{n,k}$ denotes the $p \times k$ matrix whose columns are $\Delta^i s_n, \ldots, \Delta^i s_{n+k-1}$ for $i = 1, 2$, with $\Delta s_n = s_{n+1} - s_n$ and $\Delta^2 s_n = \Delta s_{n+1} - \Delta s_n$, $Y_{n,k}$ is a $p \times k$ matrix whose columns are $y_1^{(n)}, y_2^{(n)}, \ldots, y_k^{(n)}$, $Y_{n,k}^*$ is the transpose conjugate of $Y_{n,k}$.

Throughout the paper the regularity of the matrices to be inverted is assumed.

The polynomial methods as MPE, RRE and MMPE can be considered as particular cases of (2.1). If we set

(1) $y_i^{(n)} = \Delta s_{n+i-1}$, we obtain the MPE;

(2) $y_i^{(n)} = \Delta^2 s_{n+i-1}$, we have the RRE;

(3) $y_i^{(n)} = y_i$, we get the MMPE.

Now in the last case if we take $k = p$ and $y_i = e_i$ where $\{e_i\}_{i=1,\ldots,p}$ is the canonical basis of $\mathbb{C}^p$, we obtain the Henrici's transformation [12, p.116] studied in [21]. $t_{n,k}$ can be expressed as a ratio of two determinants:

$$t_{n,k} = \frac{\begin{vmatrix} s_n & s_{n+1} & \cdots & s_{n+k} \\ (y_1^{(n)}, \Delta s_n) & (y_1^{(n)}, \Delta s_{n+1}) & \cdots & (y_1^{(n)}, \Delta s_{n+k}) \\ (y_2^{(n)}, \Delta s_n) & (y_2^{(n)}, \Delta s_{n+1}) & \cdots & (y_2^{(n)}, \Delta s_{n+k}) \\ \vdots & \vdots & & \vdots \\ (y_k^{(n)}, \Delta s_n) & (y_k^{(n)}, \Delta s_{n+1}) & \cdots & (y_k^{(n)}, \Delta s_{n+k}) \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ (y_1^{(n)}, \Delta s_n) & (y_1^{(n)}, \Delta s_{n+1}) & \cdots & (y_1^{(n)}, \Delta s_{n+k}) \\ (y_2^{(n)}, \Delta s_n) & (y_2^{(n)}, \Delta s_{n+1}) & \cdots & (y_2^{(n)}, \Delta s_{n+k}) \\ \vdots & \vdots & & \vdots \\ (y_k^{(n)}, \Delta s_n) & (y_k^{(n)}, \Delta s_{n+1}) & \cdots & (y_k^{(n)}, \Delta s_{n+k}) \end{vmatrix}}, \tag{2.2}$$

where the (generalized) determinant in the numerator denotes the vector obtained by expanding it with respect to its first row.

This result is an application of the extension of the Schur complement [5]. We shall construct a new algorithm for computing the $t_{k,n}$'s for the MMPE method and in this case we shall set $s_k^{(n)} = t_{k,n}$. Let us first define in $\mathbb{C}^{2p}$ the linear forms $z_j$, $j = 1, \ldots, k$. For $v = \binom{v_1}{v_2}$ with $v_1 \in \mathbb{C}^p$ and $v_2 \in \mathbb{C}^p$, we define

$$\langle z_j, v\rangle = (y_j, v_2 - v_1), \quad \text{for } j = 1, \ldots, k,$$

where $y_1, \ldots, y_k$ are given vectors of $\mathbb{C}^p$.

Setting $x_n = \binom{s_n}{s_{n+1}}$, we get

$$\langle z_j, x_{n+i}\rangle = (y_j, \Delta s_{n+i}), \quad \text{for } j = 1, \ldots, k \text{ and } i = 0, \ldots, k. \tag{2.3}$$

Let $e_k^{(n)}$ be the vector of $\mathbb{C}^{2p}$ obtained by applying the second variant of the CRPA [4] to $x_n, \ldots, x_{n+k}$ with the linear forms $z_j$, $j = 1, \ldots, k$; then we obtain the algorithm

$$\begin{cases} e_0^{(n)} = \begin{pmatrix} s_n \\ s_{n+1} \end{pmatrix}, \\ e_k^{(n)} = \dfrac{\langle z_k, e_{k-1}^{(n+1)}\rangle e_{k-1}^{(n)} - \langle z_k, e_{k-1}^{(n)}\rangle e_{k-1}^{(n+1)}}{\langle z_k, e_{k-1}^{(n+1)}\rangle - \langle z_k, e_{k-1}^{(n)}\rangle}. \end{cases} \tag{2.4}$$

$e_k^{(n)}$ can be expressed as a ratio of two determinants:

$$e_k^{(n)} = \frac{\begin{vmatrix} x_n & x_{n+1} & \cdots & x_{n+k} \\ \langle z_1, x_n\rangle & \langle z_1, x_{n+1}\rangle & \cdots & \langle z_1, x_{n+k}\rangle \\ \langle z_2, x_n\rangle & \langle z_2, x_{n+1}\rangle & \cdots & \langle z_2, x_{n+k}\rangle \\ \vdots & \vdots & & \vdots \\ \langle z_k, x_n\rangle & \langle z_k, x_{n+1}\rangle & \cdots & \langle z_k, x_{n+k}\rangle \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \langle z_1, x_n\rangle & \langle z_1, x_{n+1}\rangle & \cdots & \langle z_1, x_{n+k}\rangle \\ \langle z_2, x_n\rangle & \langle z_2, x_{n+1}\rangle & \cdots & \langle z_2, x_{n+k}\rangle \\ \vdots & \vdots & & \vdots \\ \langle z_k, x_n\rangle & \langle z_k, x_{n+1}\rangle & \cdots & \langle z_k, x_{n+k}\rangle \end{vmatrix}}. \tag{2.5}$$

It follows from (2.3) and (2.5) that

$$e_k^{(n)} = \begin{pmatrix} s_k^{(n)} \\ \tilde{s}_k^{(n)} \end{pmatrix}$$

where $\tilde{s}_k^{(n)}$ is obtained by replacing the first row in the numerator of (2.2) by $s_{n+1}, \ldots, s_{n+k+1}$.

Let us set

$$\beta_k^{(n)} = \tilde{s}_k^{(n)} - s_k^{(n)}. \tag{2.6}$$

Then it is easy to see that $\beta_k^{(n)}$ can be expressed as a ratio of two determinants:

$$\beta_k^{(n)}=\frac{\begin{vmatrix}\Delta s_n & \Delta s_{n+1} & \cdots & \Delta s_{n+k}\\ (y_1,\Delta s_n) & (y_1,\Delta s_{n+1}) & \cdots & (y_1,\Delta s_{n+k})\\ (y_2,\Delta s_n) & (y_2,\Delta s_{n+1}) & \cdots & (y_2,\Delta s_{n+k})\\ \vdots & \vdots & & \vdots\\ (y_k,\Delta s_n) & (y_k,\Delta s_{n+1}) & \cdots & (y_k,\Delta s_{n+k})\end{vmatrix}}{\begin{vmatrix}1 & 1 & \cdots & 1\\ (y_1,\Delta s_n) & (y_1,\Delta s_{n+1}) & \cdots & (y_1,\Delta s_{n+k})\\ (y_2,\Delta s_n) & (y_2,\Delta s_{n+1}) & \cdots & (y_2,\Delta s_{n+k})\\ \vdots & \vdots & & \vdots\\ (y_k,\Delta s_n) & (y_k,\Delta s_{n+1}) & \cdots & (y_k,\Delta s_{n+k})\end{vmatrix}} \tag{2.7}$$

We shall first derive a recursive algorithm for computing the $\beta_k^{(n)}$'s. Applying again the second variant of CRPA to $\Delta s_n, \ldots, \Delta s_{n+k}$ with the $y_j$'s, we obtain the algorithm:

$$\begin{cases}\beta_0^{(n)}=\Delta s_n,\\ \beta_k^{(n)}=\dfrac{\beta_{k-1}^{(n)}-a_k^{(n)}\beta_{k-1}^{(n+1)}}{1-a_k^{(n)}}, \quad \text{where } a_k^{(n)}=\dfrac{\left(y_k,\beta_{k-1}^{(n)}\right)}{\left(y_k,\beta_{k-1}^{(n+1)}\right)} \text{ and } k\geqslant 1.\end{cases} \tag{2.8}$$

This algorithm has been called the $\beta$-algorithm. The $\beta$-algorithm can be used to compute simultaneously all the eigenvalues of a matrix under certain assumptions; for more details see [13].

Now let us go back to (2.4); we have

$$e_{k-1}^{(n)}=\begin{pmatrix}s_{k-1}^{(n)}\\ \tilde{s}_{k-1}^{(n)}\end{pmatrix},$$

then

$$\left\langle z_k, e_{k-1}^{(n)}\right\rangle=\left(y_k,\tilde{s}_{k-1}^{(n)}-s_{k-1}^{(n)}\right)=\left(y_k,\beta_{k-1}^{(n)}\right).$$

If we use the last equality and consider only the first $p$ components of vectors in (2.4), we obtain the following recursive algorithm:

$$\begin{cases}s_0^{(n)}=s_n, \qquad \beta_0^{(n)}=\Delta s_n,\\ \beta_k^{(n)}=\dfrac{\beta_{k-1}^{(n)}-a_k^{(n)}\beta_{k-1}^{(n+1)}}{1-a_k^{(n)}},\\ s_k^{(n)}=\dfrac{s_{k-1}^{(n)}-a_k^{(n)}s_{k-1}^{(n+1)}}{1-a_k^{(n)}}, \quad \text{where } a_k^{(n)}=\dfrac{\left(y_k,\beta_{k-1}^{(n)}\right)}{\left(y_k,\beta_{k-1}^{(n+1)}\right)} \text{ and } k\geqslant 1.\end{cases} \tag{2.9}$$

This algorithm has been called the $(s\text{-}\beta)$-algorithm.

Computing $s_k^{(n)}$ by the ($s$-$\beta$)-algorithm requires $p(k^2+k)$ multiplications and $p(k^2+k)$ additions. It needs the storage of $2(k+1)$ vectors of $\mathbb{C}^p$.

One of the applications of the ($s$-$\beta$)-algorithm is the implementation of Henrici's method for solving systems of nonlinear equations.

One can see that the computation of $s_k^{(n)}$ from $s_{k-1}^{(n)}$ and $s_{k-1}^{(n+1)}$ is possible only if $1-a_k^{(n)} \neq 0$. If this is not the case, particular rules can be used to jump over this singularity. They are given in the next property.

**Property 1.** *For $m \geqslant 0$ and $k \geqslant 0$ we have*

$$s_{k+m}^{(n)} = \frac{\begin{vmatrix} s_k^{(n)} & \Delta s_k^{(n)} & \cdots & \Delta s_k^{(n+m-1)} \\ \left(y_{k+1}, \beta_k^{(n)}\right) & \left(y_{k+1}, \Delta\beta_k^{(n)}\right) & \cdots & \left(y_{k+1}, \Delta\beta_k^{(n+m-1)}\right) \\ \left(y_{k+2}, \beta_k^{(n)}\right) & \left(y_{k+2}, \Delta\beta_k^{(n)}\right) & \cdots & \left(y_{k+2}, \Delta\beta_k^{(n+m-1)}\right) \\ \vdots & \vdots & & \vdots \\ \left(y_{k+m}, \beta_k^{(n)}\right) & \left(y_{k+m}, \Delta\beta_k^{(n)}\right) & \cdots & \left(y_{k+m}, \Delta\beta_k^{(n+m-1)}\right) \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \left(y_{k+1}, \beta_k^{(n)}\right) & \left(y_{k+1}, \beta_k^{(n+1)}\right) & \cdots & \left(y_{k+1}, \beta_k^{(n+m)}\right) \\ \left(y_{k+2}, \beta_k^{(n)}\right) & \left(y_{k+2}, \beta_k^{(n+1)}\right) & \cdots & \left(y_{k+2}, \beta_k^{(n+m)}\right) \\ \vdots & \vdots & & \vdots \\ \left(y_{k+m}, \beta_k^{(n)}\right) & \left(y_{k+m}, \beta_k^{(n+1)}\right) & \cdots & \left(y_{k+m}, \beta_k^{(n+m)}\right) \end{vmatrix}},$$

*and a similar expression for $\beta_{k+m}^{(n)}$ by replacing the first row in the numerator by $\beta_k^{(n)}, \ldots, \beta_{k+m}^{(n)}$.*

**Proof.** The same as in [6] for the E-algorithm. □

If $k=0$, we get the expression of $s_m^{(n)}$ as a ratio of two determinants. If $m=1$, these relations reduce to the recursive formulas of the ($s$-$\beta$)-algorithm. The preceding formulas for $m>1$ can be used if $a_k^{(n)}=1$ or if $|1-a_k^{(n)}|$ is small. They enable us to jump over the singularity or to improve the numerical stability of the algorithm.

Now applying Schur's formula [5], we obtain $s_{k+m}^{(n)}$ in matrix form:

$$s_{k+m}^{(n)} = s_k^{(n)} - \left(\Delta s_k^{(n)}, \ldots, \Delta s_k^{(n+m-1)}\right)$$
$$\times \left\{ \tilde{Y}_{k,m}^* \left(\Delta\beta_k^{(n)}, \ldots, \Delta\beta_k^{(n+m-1)}\right)\right\}^{-1} \tilde{Y}_{k,m}^* \beta_k^{(n)},$$

where $\tilde{Y}_{k,m}$ is a matrix whose columns are $y_{k+1}, \ldots, y_{k+m}$. If $k=0$, this relation reduces to (2.1) when $y_i^{(n)} = y_i$.

*Convergence properties*

Now we shall give some convergence properties of the ($s$-$\beta$)-algorithm. The proofs are not difficult, so they will be omitted.

**Property 2.** *If* $\lim_{n\to\infty} s_{k-1}^{(n)} = s$ *and if* $\exists a$, $b$ *such that* $0 < a < 1 < b$ *and*

$$\frac{\left(y_k, \beta_{k-1}^{(n+1)}\right)}{\left(y_k, \beta_{k-1}^{(n)}\right)} \notin [a, b], \quad \textit{for } n > N,$$

*then*

$$\lim_{n\to\infty} s_k^{(n)} = s.$$

**Property 3.** *If* $\lim_{n\to\infty} s_{k-1}^{(n)} = s$ *and if*

$$\lim_{n\to\infty} \frac{\left(y_k, \beta_{k-1}^{(n+1)}\right)}{\left(y_k, \beta_{k-1}^{(n)}\right)} = b_k \neq 1,$$

*then* $\lim_{n\to\infty} s_k^{(n)} = s$. *Moreover, if for* $y \in \mathbb{C}^p$

$$\lim_{n\to\infty} \frac{\left(y, s_{k-1}^{(n+1)} - s\right)}{\left(y, s_{k-1}^{(n)} - s\right)} = b_k,$$

*then*

$$\lim_{n\to\infty} \frac{\left(y, s_k^{(n)} - s\right)}{\left(y, s_{k-1}^{(n)} - s\right)} = 0.$$

## 3. Quadratic convergence theorem

We consider the nonlinear system defined by

$$x = G(x), \tag{3.1}$$

where $G: D \subset \mathbb{C}^p \to \mathbb{C}^p$ and $D$ is an open and convex subset of $\mathbb{C}^p$. Let $x^*$ denote a solution of (3.1).

The algorithm which will be considered is the following:

- choose a starting point $x^0$,
- at the iteration $k$, we set $s_0 = x^k$ and $s_{i+1} = G(s_i)$ for $i = 0, \ldots, d_k$, where $d_k$ is the degree of the minimal polynomial of $G'(x^*)$ for the vector $x^k - x^*$, (3.2)
- compute $x^{k+1}$ such that $x^{k+1} = t_{0,d_k}$.

Our contribution to the subject is to give a sufficient condition and, for the first time, a complete and satisfactory proof of the quadratic convergence of the sequence $\{x^k\}$ to $x^*$.

Arguments for the quadratic convergence of $\{x^k\}$ were given by Skelboe [26] for the MPE and by Beuneu [2] for a class of extrapolation methods. Smith, Ford and Sidi [27] have noticed that there was a gap in Skelboe's proof. We have found the same gap in Beuneu's proof.

These authors proved the following relation:

$$\| x^{k+1} - x^* \| \leqslant \alpha_k \| x^k - x^* \|^2.$$

But they did not prove that the sequence $\{\alpha_k\}$ is bounded.

We shall give new arguments to close this gap. We set $G'(x^*)=J$ and assume throughout this section that $G$ satisfies the following conditions.

The matrix $J-I$ is regular. We set $M=\|(J-I)^{-1}\|$. (3.3a)

The Frechet derivative $G'$ of $G$ satisfies the Lipschitz condition
$\|G'(x)-G'(y)\| \leqslant L\|x-y\|$, for $x, y \in D$. (3.3b)

Before the main result, we first give two lemmas.

**Lemma 4.** *Let $A$ be a complex $p \times k$ matrix with $1 \leqslant k \leqslant p$, and* $\operatorname{rank}(A)=k$. *We set $A^+=(A^*A)^{-1}A^*$; then*

$$\|A^+\| \leqslant \frac{\|A\|_F^{k-1}}{\sqrt{\det(A^*A)}}.$$

**Proof.** We have $\|A^+\|^2=\|A^+A^{+*}\|=\|(A^*A)^{-1}\|$ (see [28] for details); moreover, by using [15, Lemma 1] and the fact that $\|A^*\|=\|A\|$ and $\|A\| \leqslant \|A\|_F$, we get the inequalities

$$\|(A^*A)^{-1}\| \leqslant \frac{\|A^*A\|^{k-1}}{|\det(A^*A)|} \leqslant \frac{\|A\|^{2k-2}}{|\det(A^*A)|} \leqslant \frac{\|A\|_F^{2k-2}}{|\det(A^*A)|},$$

which ends the proof. □

**Lemma 5.** *Let $A$ be a complex $p \times k$ matrix whose columns are denoted by $a^i$ for $i=1,\ldots,k$. If $z \in \operatorname{span}\{a^1,\ldots,a^k\}$, then*

$$A(B^*A)^{-1}B^*z=z,$$

*where $B$ is a $p \times k$ matrix such that $B^*A$ is regular.*

**Proof.** Since $z \in \operatorname{span}\{a^1,\ldots,a^k\}$, $z=\sum_{i=1}^k \alpha_i a^i$. Hence

$$A(B^*A)^{-1}B^*z=A(B^*A)^{-1}B^*\left(\sum_{i=1}^k \alpha_i a^i\right)=\sum_{i=1}^k \alpha_i\left[A(B^*A)^{-1}B^*a^i\right].$$

But $A(B^*A)^{-1}B^*A=A$, then $A(B^*A)^{-1}B^*a^i=a^i$, and it follows that $A(B^*A)^{-1}Bz=z$. □

We are now in the position to prove the main theorem for the methods defined by (3.2). We first give some notations to be used in the proof:

$$\tau_k(x)=\frac{(d_k)^{(d_k/2)}}{\alpha_k(x)} ML \max_{0 \leqslant i \leqslant d_k} \|G^i(x)-x^*\|,$$

where

$$\alpha_k(x)=\sqrt{\det(H_k^*(x)H_k(x))}$$

and

$$H_k(x)=\left(\frac{G(x)-x}{\|G(x)-x\|},\ldots,\frac{G^{d_k}(x)-G^{d_k-1}(x)}{\|G^{d_k}(x)-G^{d_k-1}(x)\|}\right).$$

**Theorem 6.** *If $G$ has property* (3.3), *and if*

$$\exists \alpha > 0, \exists K, \ \alpha_k(x^k) > \alpha, \quad \forall k \geqslant K,$$

*then there exists a neighbourhood $U$ of $x^*$ such that $\forall x^0 \in U$*

$$\| x^{k+1} - x^* \| = \mathrm{O}( \| x^k - x^* \|^2), \quad \textit{for } \text{RRE} \textit{ and } \text{MPE}.$$

**Proof.** (a) For the RRE. Let us set $F(x) = G(x) - x$ and $g(x) = F(x) - F'(x^*)(x - x^*)$. We have

$$\| g(x) \| \leqslant \tfrac{1}{2} L \| x - x^* \|^2, \tag{3.4}$$

$$\| g(x) - g(y) \| \leqslant L \| x - y \| \max( \| x - x^* \|, \ \| y - x^* \|), \tag{3.5}$$

for $x, \ y \in D$ (see [18, pp. 70–73]).

We have $\Delta s_0 = G(x^k) - x^k$, $\Delta s_1 = G(G(x^k)) - G(x^k)$ and $\Delta^2 s_0 = F(G(x^k)) - F(x^k)$. Thus for $i = 0, \ldots, d_k$, we get

$$\Delta s_i = G^{i+1}(x^k) - G^i(x^k), \tag{3.6}$$

$$\Delta^2 s_i = F\big(G^{i+1}(x^k)\big) - F\big(G^i(x^k)\big). \tag{3.7}$$

Let

$$C_k = \Delta^2 S_{0,d_k} - F'(x^*) \ \Delta S_{0,d_k}; \tag{3.8}$$

then $C_k$ can be written as:

$$C_k = \big( g\big(G(x^k)\big) - g(x^k), \ldots, g\big(G^{d_k}(x^k)\big) - g\big(G^{d_k - 1}(x^k)\big)\big). \tag{3.9}$$

If we define

$$\tilde{C}_k = \left( \frac{g\big(G(x^k)\big) - g(x^k)}{\| G(x^k) - x^k \|}, \ldots, \frac{g\big(G^{d_k}(x^k)\big) - g\big(G^{d_k - 1}(x^k)\big)}{\| G^{d_k}(x^k) - G^{d_k - 1}(x^k) \|} \right) \tag{3.10}$$

and

$$B_k = F'(x^*)^{-1} C_k \ \Delta S_{0,d_k}^{+}, \tag{3.11}$$

we get

$$B_k = F'(x^*)^{-1} \tilde{C}_k H_k^{+}(x^k).$$

This implies that

$$\| B_k \| \leqslant M \| \tilde{C}_k \| \ \| H_k^{+}(x^k) \| \leqslant M \| \tilde{C}_k \|_{\mathrm{F}} \| H_k^{+}(x^k) \|. \tag{3.12}$$

Using Lemma 4, we have

$$\| H_k^{+}(x^k) \| \leqslant \frac{\| H_k(x^k) \|_{\mathrm{F}}^{d_k - 1}}{\alpha_k(x^k)} = \frac{d_k^{(d_k - 1)/2}}{\alpha_k(x^k)}. \tag{3.13}$$

Now since

$$\| \tilde{C}_k \|_{\mathrm{F}}^2 = \sum_{i=1}^{d_k} \left\| \frac{g\big(G^i(x^k)\big) - g\big(G^{i-1}(x^k)\big)}{\| G^i(x^k) - G^{i-1}(x^k) \|} \right\|,$$

if follows from (3.5) that

$$\|\tilde{C}_k\|_F \leqslant \sqrt{d_k}\, L \max_{0 \leqslant i \leqslant d_k} \| G^i(x^k) - x^* \|. \tag{3.14}$$

From (3.13) and (3.14) we get the inequality

$$\| B_k \|_F \leqslant MLd_k^{d_k/2} \frac{\max_{0 \leqslant i \leqslant d_k} \| G^i(x^k) - x^* \|}{\alpha_k(x^k)}.$$

There exists a neighbourhood of $x^*$ such that $\| B_k \| < 1$; then $I + B_k$ is regular. We set $D_k = I - (I + B_k)^{-1}$. Let us return to the algorithm (3.2):

$$x^{k+1} = x^k - \Delta S_{0,d_k}\left(Y_{0,d_k}^* \, \Delta^2 S_{0,d_k}\right)^{-1} Y_{0,d_k}^* \, \Delta s_0.$$

Now from (3.8) and (3.11) we deduce

$$F'(x^*)(I + B_k)\, \Delta S_{0,d_k} = \Delta^2 S_{0,d_k}. \tag{3.15}$$

So

$$\Delta S_{0,d_k} = (I + B_k)^{-1} F'(x^*)^{-1}\, \Delta^2 S_{0,d_k}.$$

Therefore

$$x^{k+1} = x^k - (I - D_k) F'(x^*)^{-1}\, \Delta^2 S_{0,d_k}\left(Y_{0,d_k}^* \, \Delta^2 S_{0,d_k}\right)^{-1} Y_{0,d_k}^* \, \Delta s_0. \tag{3.16}$$

We have $\Delta s_i = G(s_i) - s_i = G(s_i) - G(x^*) - (s_i - x^*)$, thus

$$\Delta s_i = (J - I)(s_i - x^*) + \phi_{i,1}(s_i - x^*), \quad \text{with } \| \phi_{i,1}(s_i - x^*) \| = O\left(\| s_i - x^* \|^2\right), \tag{3.17}$$

and

$$s_{i+1} - x^* = J(s_i - x^*) + \phi_{i,2}(s_i - x^*), \quad \text{with } \| \phi_{i,2}(s_i - x^*) \| = O\left(\| s_i - x^* \|^2\right). \tag{3.18}$$

Using (3.17) and (3.18), it easily follows that

$$\Delta^2 s_i = (J - I)^2 J^i (x^k - x^*) + \phi_{i,3}(x^k - x^*), \quad \text{with } \| \phi_{i,3}(x^k - x^*) \| = O(\| x^k - x^* \|^2). \tag{3.19}$$

Let $\eta_i^{(k)}$ be the coefficients of the minimal polynomial of $J$ for $x^k - x^*$:

$$\sum_{i=0}^{d_k} \eta_i^{(k)} J^i (x^k - x^*) = 0, \quad \text{with } \eta_{d_k}^{(k)} = 1.$$

Next let us multiply each $\Delta^2 s_i$ by $\eta_i^{(k)}$ and then sum then up; thus from (3.19) we get

$$\begin{aligned} \sum_{i=0}^{d_k} \eta_i^{(k)}\, \Delta^2 s_i &= (J - I)^2 \sum_{i=0}^{d_k} \eta_i^{(k)} J^i (x^k - x^*) + \sum_{i=0}^{d_k} \eta_i^{(k)} \phi_{i,3}(x^k - x^*) \\ &= \sum_{i=0}^{d_k} \eta_i^{(k)} \phi_{i,3}(x^k - x^*). \end{aligned} \tag{3.20}$$

Since

$$\sum_{i=0}^{d_k} \left| \eta_i^{(k)} \right| \leqslant M_0, \tag{3.21}$$

we have

$$\sum_{i=0}^{d_k} \eta_i^{(k)} \, \Delta^2 s_i = \psi_1(x^k - x^*), \quad \text{with } \| \psi_1(x^k - x^*) \| = \mathrm{O}(\| x^k - x^* \|^2),$$

and

$$\sum_{i=0}^{d_k} \eta_i^{(k)} J^i \, \Delta^2 s_0 = (J - I) \psi_2(x^k - x^*), \quad \text{with } \| \psi_2(x^k - x^*) \| = \mathrm{O}(\| x^k - x^* \|^2). \tag{3.22}$$

On the other hand,

$$\sum_{i=0}^{d_k} \eta_i^{(k)} (J^i - I) \, \Delta^2 s_0 = - \sum_{i=0}^{d_k} \eta_i^{(k)} \, \Delta^2 s_0 + (J - I) \psi_2(x^k - x^*). \tag{3.23}$$

As 1 is not a zero of the minimal polynomial of $J$, we have $\Sigma_{i=0}^{d_k} \eta_i^{(k)} \neq 0$. Thus by using (3.23),

$$\begin{aligned} &\frac{\eta_1^{(k)}}{\Sigma_{i=0}^{d_k} \eta_i^{(k)}} \, \Delta^2 s_0 + \sum_{i=2}^{d_k} \frac{\eta_i^{(k)}}{\Sigma_{i=0}^{d_k} \eta_i^{(k)}} (J^{i-1} + \cdots + I) \, \Delta^2 s_0 \\ &\quad = -(J - I)^{-1} \, \Delta^2 s_0 + \psi_2(x^k - x^*). \end{aligned}$$

Then we have

$$\begin{aligned} (J - I)^{-1} \, \Delta^2 s_0 &= \sum_{i=0}^{d_k - 1} a_i^{(k)} J^i \, \Delta^2 s_0 + \psi_2(x^k - x^*) \\ &= \sum_{i=0}^{d_k - 1} a_i^{(k)} \, \Delta^2 s_i + \psi_2(x^k - x^*). \end{aligned}$$

Since

$$\begin{aligned} \Delta s_0 &= (J - I)^{-1} \, \Delta^2 s_0 + \psi_3(x^k - x^*), \quad \text{with } \| \psi_3(x^k - x^*) \| = \mathrm{O}(\| x^k - x^* \|^2), \\ &= \sum_{i=0}^{d_k - 1} a_i^{(k)} \, \Delta^2 s_i + \Psi(x^k - x^*), \quad \text{with } \Psi(x) = \psi_2(x) + \psi_3(x). \end{aligned} \tag{3.24}$$

Using Lemma 5 and (3.24), (3.16) becomes

$$x^{k+1} = x^k - (I - D_k) F'(x^*)^{-1} \big( F(x^k) - \Psi(x^k - x^*) \big) + \Phi(x^k - x^*),$$

and

$$\begin{aligned} x^{k+1} - x^* = {} & (D_k - I) F'(x^*)^{-1} \big( g(x^k) - \Psi(x^k - x^*) \big) \\ & + D_k (x^k - x^*) + \Phi(x^k - x^*), \end{aligned}$$

with

$$\Phi(x^k - x^*) = (D_k - I) F'(x^*)^{-1} \, \Delta^2 S_{0,d_k} \left( Y_{0,d_k}^* \, \Delta^2 S_{0,d_k} \right)^{-1} Y_{0,d_k}^* \Psi(x^k - x^*). \tag{3.25}$$

Since for the RRE $Y_{0,d_k} = \Delta^2 S_{0,d_k}$, we have $\| \Delta^2 S_{0,d_k}(Y_{0,d_k}^* \, \Delta^2 S_{0,d_k})^{-1} Y_{0,d_k}^* \| = 1$; then

$$\| \Phi(x^k - x^*) \| \leqslant M_1 \| x^k - x^* \|^2. \tag{3.26}$$

Thus

$$\begin{aligned} \| x^{k+1} - x^* \| \leqslant & \| (D_k - I) \| M( \| g(x^k) \| + \| \Psi(x^k - x^*) \| ) \\ & + \| D_k \| \| x^k - x^* \| + \| \Phi(x^k - x^*) \|. \end{aligned} \tag{3.27}$$

But

$$\| D_k \| \leqslant \frac{\| B_k \|}{1 - \| B_k \|} \quad \text{and} \quad \| D_k - I \| \leqslant \frac{1}{1 - \| B_k \|}.$$

Now from (3.4) it follows that

$$\begin{aligned} & \| x^{k+1} - x^* \| \leqslant \frac{3}{2} \left( \frac{\tau_k(x^k)}{1 - \tau_k(x^k)} \right) \| x^k - x^* \| + \| \Theta(x^k - x^*) \|, \\ & \quad \text{with } \| \Theta(x^k - x^*) \| = O( \| x^k - x^* \|^2). \end{aligned} \tag{3.28}$$

From the definition of $\tau_k(x)$ and the assumptions we get

$$\tau_k(x^k) \leqslant \frac{M L p^{p/2} M_2}{\alpha} \| x^k - x^* \|. \tag{3.29}$$

Then $\lim_{k \to \infty} \tau_k(x^k) = 0$. This implies that there exists a number $M_3$ such that

$$\left| \frac{1}{1 - \tau_k(x^k)} \right| < M_3, \quad \text{for } k > N_0. \tag{3.30}$$

Returning to (3.28) and using (3.29) and (3.30) we finally obtain

$$\| x^{k+1} - x^* \| = O( \| x^k - x^* \|^2).$$

(b) For the MPE we shall use the presentation given in [27]. Let $c^{(k)} = (c_0^{(k)}, \ldots, c_{d_k - 1}^{(k)})$ be the vector of $\mathbb{C}^{d_k}$ defined by $c^{(k)} = -\Delta S_{0,d_k}^+ \, \Delta s_{d_k}$; then $x^{k+1}$ can be expressed as

$$\left( \sum_{i=0}^{d_k} c_i^{(k)} \right) x^{k+1} = \sum_{i=0}^{d_k} c_i^{(k)} G^i(x^k), \quad \text{with } c_{d_k}^{(k)} = 1. \tag{3.31}$$

From (3.17) and (3.18) we get

$$\sum_{i=0}^{d_k} \eta_i^{(k)} \, \Delta s_i = \psi_4(x^k - x^*), \quad \text{with } \| \psi_4(x^k - x^*) \| = O( \| x^k - x^* \|^2).$$

Since $\eta_{d_k}^{(k)} = 1$ we have

$$\Delta s_{d_k} = - \sum_{i=0}^{d_k - 1} \eta_i^{(k)} \, \Delta s_i + \psi_4(x^k - x^*).$$

Then

$$c^{(k)} = \Delta S_{0,d_k}^+ \left( \sum_{i=0}^{d_k - 1} \eta_i^{(k)} \, \Delta s_i - \psi_4(x^k - x^*) \right).$$

If $\eta^{(k)}=(\eta_0^{(k)},\dots,\eta_{d_k-1}^{(k)})$, we have

$$c^{(k)}=\eta^{(k)}-\Delta S_{0,d_k}^{+}\psi_4(x^k-x^*). \tag{3.32}$$

Let $E_k$ be the diagonal matrix whose diagonal elements are $\|\Delta s_0\|,\dots,\|\Delta s_{d_k-1}\|$, we have $\Delta S_{0,d_k}=H_k(x^k)E_k$.

From (3.13) and the fact that $\Delta S_{0,d_k}^{+}=E_k^{-1}H_k(x^k)^{+}$ it follows that

$$\|\Delta S_{0,d_k}^{+}\|\leqslant\frac{M_4}{\|x^k-x^*\|}, \tag{3.33}$$

and consequently

$$c^{(k)}=\eta^{(k)}+\psi_5(x^k-x^*),\quad \text{with } \|\psi_5(x^k-x^*)\|=\mathrm{O}(\|x^k-x^*\|). \tag{3.34}$$

Now let us go back to (3.31), we have

$$x^{k+1}-x^*=\frac{\sum_{i=0}^{d_k}c_i^{(k)}\left(G^i(x^k)-x^*\right)}{\sum_{i=0}^{d_k}c_i^{(k)}}. \tag{3.35}$$

From the relation (3.32) we get

$$x^{k+1}-x^*=\frac{\sum_{i=0}^{d_k}\eta_i^{(k)}\left(G^i(x^k)-x^*\right)}{\sum_{i=0}^{d_k}\eta_i^{(k)}}+\psi_6(x^k-x^*),$$

$$\text{with } \|\psi_6(x^k-x^*)\|=\mathrm{O}(\|x^k-x^*\|^2).$$

Since the $\eta_i^{(k)}$'s are the coefficients of the minimal polynomial of $J$ for $x^k-x^*$, and since $G^i(x^k)-x^*=J^i(x^k-x^*)+\mathrm{O}(\|x^k-x^*\|^2)$, we finally get

$$x^{k+1}-x^*=\mathrm{O}(\|x^k-x^*\|^2).\qquad\square$$

**Example 7.** Let $x=(x_1,x_2)^{\mathrm{T}}$; we define the functions $F$, $G$ and $H$ by

$$F(x)=G(x)-x,$$

$$G(x)=\begin{pmatrix}\frac{7}{27}x_1-\frac{16}{27}x_2-x_1^2-\frac{1}{2}x_1x_2\\ -\frac{32}{27}x_1+\frac{23}{27}x_2-x_1x_2\end{pmatrix},$$

$$H(x)=x-\left(G(x)-x,\ G^2(x)-G(x)\right)$$
$$\times\left(F(G(x))-F(x),\ F(G^2(x))-F(G(x))\right)^{-1}F(x),$$

$$\alpha(x)=\frac{|\det(G(x)-x,\ G^2(x)-G(x))|}{\|G(x)-x\|\ \|G^2(x)-G(x)\|}.$$

Let $x(e)=(e,\ e-e^3)$, the degree of the minimal polynomial of $G'(0)$ for the vector $x(e)$ is two. We also have

$$\alpha(x(e))\sim\tfrac{1}{2}e,\quad e\to 0, \tag{3.36}$$

$$\lim_{e\to 0}\frac{\|H(x(e))\|}{\|x(e)\|}=\frac{12\sqrt{221}}{475\sqrt{2}}. \tag{3.37}$$

One can see that the condition of the theorem is not satisfied for such an $x(e)$, since $\lim_{e\to 0}\alpha(x(e))=0$.

The relation (3.37) shows that the convergence to $x^*=0$ of the sequence defined by $x^{k+1}=H(x^k)$, which is the sequence given in (3.2) when $d_k=2$, is not quadratic in general.

**Remarks.** (1) The quadratic convergence result stated in Theorem 6, under the condition $\alpha_k(x^k)>\alpha$, has been given for the RRE and the MPE. In the general case this result is still true if we add the condition.

$$\exists\beta>0 \text{ such that } \left\|\Delta^2 S_{0,d_k}\left(Y_{0,d_k}^* \;\Delta^2 S_{0,d_k}\right)^{-1} Y_{0,d_k}^*\right\| < \beta, \quad \forall k \geqslant K_1. \tag{3.38}$$

Indeed, in the proof for the RRE, we have used the fact that $Y_{0,i}=\Delta^2 S_{0,i}$ and then $\| \Delta^2 S_{0,d_k}(Y_{0,d_k}^* \;\Delta^2 S_{0,d_k})^{-1} Y_{0,d_k}^* \| = 1$. But it is sufficient to have (3.38) in order to get the relation (3.26). The remaining part of the proof is the same as for the RRE.

(2) If $d_k=p$, the algorithm (3.2) reduces to Henrici's method [12, p.116] for solving systems of nonlinear equations. In this case Ortega and Rheinbolt [18, p.373] showed that $\{x^k\}$ converge quadratically to $x^*$. The proof that we gave for the theorem can be considered as a generalization of theirs.

(3) Since there exists a neighbourhood of $x^*$ such that $\Delta s_0,\ldots,\Delta s_{d_k-1}$ are linearly independent and $\Delta s_0,\ldots,\Delta s_{d_k}$ are almost linearly dependent and since $d_k$ is not known in practice, we can replace, in the method (3.2), $d_k$ by $l_k$ where $l_k$ is defined by

$$l_k = \max_{i\in\{1,\ldots,p\}} \{\mathrm{Rank}(\Delta S_{0,i})=i\},$$

and hence we obtain a variant of the method which has the same convergence property and which is more interesting in practice. Some other methods of this type can be found in [20].

## References


[1] J. Beuneu, Méthodes de projection à convergence finie; remarques sur leur forme incomplète, ANO 88, Univ. Lille, 1982.

[2] J. Beuneu, Some projection methods for nonlinear problems, ANO 134, Univ. Lille, 1984.

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

[4] C. Brezinksi, Recursive interpolation, extrapolation and projection, *J. Comput. Appl. Math.* **9** (4) (1983) 369–376.

[5] C. Brezinski, Other manifestations of the Schur complement, *Linear Algebra Appl.* **24** (1988) 231–247.

[6] C. Brezinski, Algebraic properties of the E-transformation, in: *Numerical Analysis and Mathematical Modelling*, Banach Center Publ. **24** (PWN, Warsaw, 1990) 85–90.

[7] C. Brezinski and H. Sadok, Vector sequence transformations and fixed point methods, in: C. Taylor et al., Eds., *Numerical Methods in Laminar and Turbulent Flows* (Pineridge, Swansea, 1987) 3–11.

[8] S. Cabay and L.W. Jackson, A polynomial extrapolation method for finding limits and antilimits of vector sequences, *SIAM J. Numer. Anal.* **13** (1976) 734–752.

[9] R.P. Eddy, Extrapolating to the limit of a vector sequence, in: P.C.C. Wang, Ed., *Information Linkage between Applied Mathematics and Industry* (Academic Press, New York, 1979) 387–396.

[10] W.F. Ford and A. Sidi, Recursive algorithms for vector extrapolation methods, *Appl. Numer. Math.* **4** (6) (1988) 477–489.

[11] B. Germain-Bonne, Estimation de la limite de suites et formalisation de procédés d'accélération de la convergence, Thèse d'Etat, Univ. Lille, 1978.

[12] P. Henrici, *Elements of Numerical Analysis* (Wiley, New York, 1964).
[13] Kh. Jbilou, Methodes d'extrapolation et de projection. Application aux suites de vecteurs, Thèse de 3ème cycle, Univ. Lille, 1988.
[14] S. Kaniel and J. Stein, Least-square acceleration of iterative methods for linear equations, *J. Optim. Theory Appl.* **14** (1974) 431–437.
[15] T. Kato, Estimation of iterated matrices with application to the von Neumann condition, *Numer. Math.* **2** (1960) 22–29.
[16] C.L. Lawson and R.J. Hanson, *Solving Least Squares Problems* (Prentice-Hall, Englewood Cliffs, NJ, 1974).
[17] M. Mešina, Convergence acceleration for the iterative solution of the equations $X = AX + f$, *Comput. Methods Appl. Mech. Engrg.* **10** (2) (1977) 165–173.
[18] J.M. Ortega and W.C. Rheinbolt, *Iterative Solution of Nonlinear Equations in Several Variables* (Academic Press, New York, 1970).
[19] B.P. Pugachev, Acceleration of the convergence of iterative processes and a method of solving systems of nonlinear equations, *U.S.S.R. Comput. Math. and Math. Phys.* **17** (1978) 199–207.
[20] H. Sadok, Accélération de la convergence de suites vectorielles et méthodes de point fixe, Thèse, Univ. Lille, 1988.
[21] H. Sadok, About Henrici's transformation for accelerating vector sequences, *J. Comput. Appl. Math.* **29** (1) (1990) 101–110.
[22] A. Sidi, Convergence and stability properties of minimal polynomial and reduced rank extrapolation algorithms, *SIAM J. Numer. Anal.* **23** (1986) 197–209.
[23] A. Sidi, Extrapolation vs. projection methods for linear systems of equations, *J. Comput. Appl. Math.* **22** (1) (1988) 71–88.
[24] A. Sidi and J. Bridger, Convergence and stability analyses for some vector extrapolation methods in the presence of defective iteration matrices, *J. Comput. Appl. Math.* **22** (1) (1988) 35–61.
[25] A. Sidi, W.F. Ford and D.A. Smith, Acceleration of convergence of vector sequences, *SIAM J. Numer. Anal.* **23** (1986) 178–196.
[26] S. Skelboe, Computation of the periodic steady-state reponse of nonlinear networks by extrapolation methods, *IEEE Trans. Circuits and Systems* **27** (1980) 161–175.
[27] D.A. Smith, W.F. Ford and A. Sidi, Extrapolation methods for vector sequences, *SIAM Rev.* **29** (1987) 199–233; Correction, *SIAM Rev.* **30** (1988) 623–624.
[28] G.W. Stewart, *Introduction to Matrix Computations* (Academic Press, New York, 1973).
[29] Y.V. Vorobyev, *Method of Moments in Applied Mathematics Variables* (Gordon and Breach, New York, 1965).