Numer Algor (2015) 68:107–119
DOI 10.1007/s11075-014-9879-z

ORIGINAL PAPER

Matrix polynomial and epsilon-type extrapolation methods with applications

K. Jbilou · H. Sadok

Received: 4 February 2014 / Accepted: 4 June 2014 / Published online: 14 June 2014

Abstract In the present paper we introduce new matrix extrapolation methods as generalizations of well known methods such as polynomial vector extrapolation methods or ϵ-type algorithms. We give expressions of the obtained approximations via the Schur complement, the Kronecker product and also by using a new matrix product. We apply these methods to linearly generated sequences especially those arising in control or in ill-posed problems.

Keywords Matrix extrapolation · Projection · Sequence transformation · Ill-posed

Mathematics Subject Classifications (2010) MSC 65F · MSC 15A

1 Introduction

The most popular vector sequence transformation methods are the minimal polynomial extrapolation (MPE) method of Cabay and Jackson [10], the reduced rank extrapolation (RRE) method of Eddy [11] and Mesina [18] and the modified minimal polynomial extrapolation (MMPE) method of Sidi, Ford and Smith [24], Brezinski [4] and Pugatchev [20]. Analysis and computational procedures of these vector extrapolation methods could be found in [3, 14, 15, 23, 24]. A second class of vector sequence transformations contains the topological ϵ-algorithm (TEA) of Brezinski [3, 4, 8] and the vector ϵ-algorithm (VEA) of Wynn [27]. Block generalizations of these methods are given in [9, 17]. These vector extrapolation methods have been

K. Jbilou (✉) · H. Sadok
Laboratoire LMPA, Université du Littoral Côte d'Opale Batiment H. Poincarré, 50 Rue F. Buisson, 62280 Calais, France
e-mail: jbilou@lmpa.univ-littoral.fr

H. Sadok
e-mail: sadok@lmpa.univ-littoral.fr

applied successfully for solving linear and nonlinear systems of equations; see [1, 14, 15, 25].

In this paper, we introduce new matrix sequence transformations generalizing to the matrix case the vector extrapolation methods above. Then, we apply those matrix sequence transformations to some linear matrix equation arising in control or in ill-posed problems.

The paper is organized as follows. In Section 2, we remind the definitions of the Kronecker product together with a recently introduced $\diamond$ matrix product. In Section 3, we define the matrix polynomial extrapolation methods and the topological ϵ-transformation. We use the Kronecker and the $\diamond$ products to give new expressions of the obtained sequence transformations. Applications of the matrix polynomial extrapolation methods to Lyapunov equations and to ill-posed problems are given in Section 4. In the last section, we give some conclusions.

2 The Kronecker product and the $\diamond$ product

For two matrices Y and Z in $\mathbb{R}^{N\times s}$, we define the inner product $\langle Y, Z\rangle_F = \mathrm{tr}\left(Y^T Z\right)$ where $\mathrm{tr}(Y^T Z)$ denotes the trace of the matrix $Y^T Z$. The associated norm is the Frobenius norm denoted by $\|.\|_F$. A system of vectors (matrices) of $\mathbb{R}^{N\times s}$ is said to be F-orthonormal if it is orthonormal with respect to $\langle ., .\rangle_F$. For $Y = [y_{i,j}] \in \mathbb{R}^{N\times s}$, we denote by $\mathrm{vec}(Y)$ the vector of $\mathbb{R}^{Ns}$ defined by $\mathrm{vec}(Y) = [y(., 1)^T,\ y(., 2)^T,\ \ldots\ , y(., s)^T]^T$ where $y(., j)$, $j = 1, \ldots, s$, is the j-th column of Y. The Kronecker product of the matrices A and B is defined by $A \otimes B = [a_{i,j}B]$. For this product, we have the following properties:

1. $(A \otimes B)^T = A^T \otimes B^T$,
2. $(A \otimes B)(C \otimes D) = (AC \otimes BD)$.
3. If the matrices A and B are square and nonsingular, then $(A \otimes B)$ is invertible and $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

In the following we we recall the definition of the operator $\diamond$ that was introduced [2] .

Definition 1 Let $A = [A_1, A_2, \ldots, A_p]$ and $B = [B_1, B_2, \ldots, B_l]$ be matrices of dimension $n \times ps$ and $n \times ls$ respectively where A_i and B_j ($i = 1, \ldots, p$; $j = 1, \ldots, l$) are $N \times s$ matrices. Then the $p \times l$ matrix $A^T \diamond B$ is defined by:

$$A^T \diamond B = \begin{pmatrix} \langle A_1, B_1\rangle_F & \langle A_1, B_2\rangle_F & \ldots & \langle A_1, B_l\rangle_F \\ \langle A_2, B_1\rangle_F & \langle A_2, B_2\rangle_F & \ldots & \langle A_2, B_l\rangle_F \\ \vdots & \vdots & \vdots & \vdots \\ \langle A_p, B_1\rangle_F & \langle A_p, B_2\rangle_F & \ldots & \langle A_p, B_l\rangle_F \end{pmatrix}.$$

Remarks

1. If $s = 1$ then $A^T \diamond B = A^T B$.
2. If $s = 1$, $p = 1$ and $l = 1$, then setting $A = u \in \mathbb{R}^N$ and $B = v \in \mathbb{R}^N$, we have $A^T \diamond B = u^T v \in \mathbb{R}$.
3. The matrix $A = [A_1, A_2, \ldots, A_p]$ is F-orthonormal if and only $A^T \diamond A = I_p$.
4. If $X \in \mathbb{R}^{N\times s}$, then $X^T \diamond X = \|X\|_F^2$.

It is not difficult to show the following properties given in [2] and satisfied by the product $\diamond$. □

Proposition 1 *Let* $A, B, C \in \mathbb{R}^{N\times ps}$, $D \in \mathbb{R}^{N\times N}$, $L \in \mathbb{R}^{p\times p}$ *and* $\alpha \in \mathbb{R}$. *Then we have*

1. $(\alpha A)^T \diamond C = \alpha \left(A^T \diamond C\right)$.
2. $(A^T \diamond B)^T = B^T \diamond A$.
3. $(DA)^T \diamond B = A^T \diamond \left(D^T B\right)$.
4. $A^T \diamond (B(L \otimes I_s)) = \left(A^T \diamond B\right) L$.

Before closing this section, we also recall the definition of the Schur complement [6, 17, 19] to be used later. If M is the matrix partitioned as $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, then the Schur complement of D in M, where D is square and nonsingular, is given by

$$(M/D) = A - BD^{-1}C.$$

3 The matrix extrapolation methods

3.1 The polynomial matrix extrapolation methods

Let (S_n) be a sequence of matrices of $\mathbb{R}^{N\times s}$ and consider the transformation T_k from $\mathbb{R}^{N\times s}$ to $\mathbb{R}^{N\times s}$ defined by

$$T_k : \ S_n \rightarrow T_k^{(n)} = S_n + \sum_{i=1}^{k} \alpha_i^{(n)} G_i(n), \ n \geq 0, \tag{1}$$

where the auxiliary $N \times s$ matrix sequences $(G_i(n))_n; i = 1, \ldots, k$ are given and the scalar coefficients $\alpha_i^{(n)}$ will be determined . Let $\widetilde{T}_k$ denotes the new transformation obtained from T_k as follows

$$\widetilde{T}_k^{(n)} = S_{n+1} + \sum_{i=1}^{k} \alpha_i^{(n)} G_i(n+1), \ n \geq 0. \tag{2}$$

We define the generalized residual of $T_k^{(n)}$ by $\widetilde{R}(T_k^{(n)}) = \widetilde{T}_k^{(n)} - T_k^{(n)}$ also given by

$$\widetilde{R}(T_k^{(n)}) = \Delta S_n + \sum_{i=1}^{k} \alpha_i^{(n)} \Delta G_i(n). \tag{3}$$

The forward difference operator Δ acts on the index n, i.e., $\Delta G_i(n) = G_i(n+1) - G_i(n)$, $i = 1, \ldots, k$. We will see later that when the sequence (S_n) is generated by a linear process, the generalized residual coincides with the classical residual.
The scalar coefficients $\alpha_i^{(n)}$ are obtained from the orthogonality relation

$$\widetilde{R}(T_k^{(n)}) \in (span\{Y_1^{(n)}, \ldots, Y_k^{(n)}\})^{\perp}, \tag{4}$$

where $Y_1^{(n)}, \ldots, Y_k^{(n)}$ are given $N \times s$ matrices. Here, the orthogonality is with the inner Kronecker product $\langle ., .\rangle_F$ and $span\left\{Y_1^{(n)}, \ldots, Y_k^{(n)}\right\}$ is the matrix subspace generated by the matrices $Y_1^{(n)}, \ldots, Y_k^{(n)}$.

If $\widetilde{\mathcal{W}}_{k,n}$ and $\widetilde{\mathcal{L}}_{k,n}$ denote the matrix subspaces $\widetilde{\mathcal{W}}_{k,n} = span\{\Delta G_1(n), \ldots, \Delta G_k(n)\}$ (the matrix subspace generated by $\Delta G_1(n), \ldots, \Delta G_k(n)$) and $\widetilde{\mathcal{L}}_{k,n} = span\{Y_1^{(n)}, \ldots, Y_k^{(n)}\}$, then from (3) and (4) , the generalized residual satisfies the following relations

$$\widetilde{R}\left(T_k^{(n)}\right) - \Delta S_n \in \widetilde{\mathcal{W}}_{k,n} \tag{5}$$

and

$$\widetilde{R}\left(T_k^{(n)}\right) \in \widetilde{\mathcal{L}}_{k,n}^{\perp}. \tag{6}$$

The relations (5) and (6) show that the generalized residual $\widetilde{R}\left(T_k^{(n)}\right)$ is obtained by projecting, orthogonaly to $\widetilde{\mathcal{L}}_{k,n}$, the matrix ΔS_n onto the subspace $\widetilde{\mathcal{W}}_{k,n}$.

Let $\Delta\mathcal{G}_{k,n}$ and $L_{k,n}$ be the $N \times ks$ matrices defined by $\Delta\mathcal{G}_{k,n} = [\Delta G_1(n), \ldots, \Delta G_k(n)]$ and $L_{k,n} = [Y_1^{(n)}, \ldots, Y_k^{(n)}]$, respectively. Then the relations (5) and (6) could be expressed as follows

$$\widetilde{R}(T_k^{(n)}) - \Delta S_n = \Delta\mathcal{G}_{k,n}\left(\alpha^{(n)} \otimes I_s\right); \quad \text{with } \alpha^{(n)} = \left(\alpha_1^{(n)}, \ldots, \alpha_k^{(n)}\right)^T, \tag{7}$$

and

$$L_{k,n}^T \diamond \widetilde{R}\left(T_k^{(n)}\right) = 0. \tag{8}$$

Assuming that $L_{k,n}^T \diamond \Delta\mathcal{G}_{k,n}$ is nonsingular, using (8) and the relation 4 of Proposition 1, the vector $\alpha^{(n)}$ appearing in the expression (7) of the generalized residual $\widetilde{R}(T_k^{(n)})$ is given as

$$\alpha^{(n)} = -\left(L_{k,n}^T \diamond \Delta\mathcal{G}_{k,n}\right)^{-1} \left(L_{k,n}^T \diamond \Delta S_n\right). \tag{9}$$

The approximation $T_k^{(n)}$ is given by

$$T_k^{(n)} = S_n + \mathcal{G}_{k,n}(\alpha^{(n)} \otimes I_s), \tag{10}$$

where $\mathcal{G}_{k,n}$ is the $N \times ks$ block matrix $\mathcal{G}_{k,n} = [G_1(n), \ldots, G_k(n)]$ and the vector $\alpha^{(n)}$ is given by (9).
Let $\mathcal{T}_{k,n}$ be the matrix given by

$$\mathcal{T}_{k,n} = \begin{pmatrix} S_n & \mathcal{G}_{k,n} \\ (L_{k,n}^T \diamond \Delta S_n) \otimes I_s & (L_{k,n}^T \diamond \Delta\mathcal{G}_{k,n}) \otimes I_s . \end{pmatrix}. \tag{11}$$

The approximation $T_k^{(n)}$ is then expressed as the following Schur complement

$$T_k^{(n)} = \left(\mathcal{T}_{k,n} \;/\; (L_{k,n}^T \diamond \Delta\mathcal{G}_{k,n}) \otimes I_s)\right). \tag{12}$$

If we set $G_i(n) = \Delta S_{n+i-1}; i = 1, \ldots, k$ and $Y_i(n) = \Delta G_i(n); i = 1, \ldots, k$ then we obtain the matrix Reduced Rank Extrapolation (M-RRE) method. In this case, the approximation $T_k^{(n)}$ is given by

$$T_{k,M-RRE}^{(n)} = \left(\mathcal{T}_{k,n} \ / \ \left(\Delta^2 \mathcal{S}_{k,n}^T \diamond \Delta^2 \mathcal{S}_{k,n} \right) \otimes I_s \right), \tag{13}$$

where $\Delta^2 \mathcal{S}_{k,n} = \left[\Delta^2 S_n, \ldots, \Delta^2 S_{n+k-1} \right]$. As M-RRE is an orthogonal projection method (ΔS_n is projected orthogonally onto the matrix subspace $\widetilde{W}_{k,n}$), the corresponding generalized residual satisfies the following minimization property $\| \widetilde{R} \left(T_{k,M-RRE}^{(n)} \right) \|_F = \min\limits_{Z \in \widetilde{W}_{k,n}} \| \Delta S_n - Z \|_F$.
We notice that when $k = 2$, the obtained M-RRE transformation is a matrix version of the well known Aitken process.

If $G_i(n) = \Delta S_{n+i-1}; i = 1, \ldots, k$ and $Y_i(n) = G_i(n); i = 1, \ldots, k$ we get the Matrix Minimal Polynomial Extrapolation (M-MPE) method and then

$$T_{k,M-MPE}^{(n)} = \left(\mathcal{T}_{k,n} \ / \ \left(\Delta \mathcal{S}_{k,n}^T \diamond \Delta^2 \mathcal{S}_{k,n} \right) \otimes I_s \right). \tag{14}$$

Finally if $G_i(n) = \Delta S_{n+i-1}; i = 1, \ldots, k$ and $Y_i(n) = Y_i; i = 1, \ldots, k$ (arbitrary $N \times s$ matrices) we obtain the Matrix Modified Minimal Polynomial Extrapolation (M-MMPE) method.

3.2 The matrix topological ϵ-transformation

In [4], Brezinski proposed a generalization of the scalar ϵ-algorithm [27] for vector sequences called the topological ϵ-algorithm (TEA). In this section we introduce the matrix topological ϵ-transformation (M-TET).
We consider approximations $E_k(S_n) = E_k^{(n)}$ of the limit or the anti-limit of the sequence $(S_n)_n$ such that

$$E_k^{(n)} = S_n + \sum_{i=1}^{k} \beta_i^{(n)} \Delta S_{n+i-1}, \ n \geq 0. \tag{15}$$

We introduce the new transformations $\widetilde{E}_{k,j}, \ j = 1, \ldots, k$ defined by

$$\widetilde{E}_{k,j}^{(n)} = S_{n+j} + \sum_{i=1}^{k} \beta_i^{(n)} \Delta S_{n+i+j-1}, \ j = 1, \ldots, k. \tag{16}$$

We set $\widetilde{E}_{k,0}^{(n)} = E_k^{(n)}$ and define the j-th matrix generalized residual as follows

$$\widetilde{R}_j \left(E_k^{(n)} \right) = \widetilde{E}_{k,j}^{(n)} - \widetilde{E}_{k,j-1}^{(n)}.$$

Therefore the coefficients $\beta_i^{(n)}$'s involved in the expression (15) of $E_k^{(n)}$ are computed such that each j-th generalized residual is F-orthogonal to some chosen $N \times s$ matrix Y, that is

$$\left\langle \widetilde{R}_j(E_k^{(n)}), Y \right\rangle_F = 0; \quad j = 1, \ldots, k. \tag{17}$$

Let $\mathcal{D}_{k,n}$ denotes the following matrix

$$\mathcal{D}_{k,n} = \begin{bmatrix} \left\langle Y, \Delta^2 S_n \right\rangle_F & \left\langle Y, \Delta^2 S_{n+1} \right\rangle_F & \dots & \left\langle Y, \Delta^2 S_{n+k-1} \right\rangle_F \\ \left\langle Y, \Delta^2 S_{n+1} \right\rangle_F & \left\langle Y, \Delta^2 S_{n+2} \right\rangle_F & \dots & \left\langle Y, \Delta^2 S_{n+k} \right\rangle_F \\ \vdots & \vdots & \dots & \vdots \\ \left\langle Y, \Delta^2 S_{n+k-1} \right\rangle_F & \left\langle Y, \Delta^2 S_{n+k} \right\rangle_F & \dots & \left\langle Y, \Delta^2 S_{n+2k-2} \right\rangle_F . \end{bmatrix} \tag{18}$$

Therefore, from the orthogonality relation (17), the vector $\beta^{(n)} = \left(\beta_1^{(n)}, \dots, \beta_k^{(n)}\right)^T$ is given by

$$\beta^{(n)} = -\,\mathcal{D}_{k,n}^{-1}\, z^{(n)}, \tag{19}$$

where $z^{(n)}$ is the vector given by $z^{(n)} = (\langle Y, \Delta S_n\rangle_F , \dots, \langle Y, \Delta S_{n+k-1}\rangle_F)^T$. We assume here that the matrix $\mathcal{D}_{k,n}$ is nonsingular. Hence, the approximation $E_k^{(n)}$ exists, is unique and is expressed as

$$E_k^{(n)} = S_n + \Delta\mathcal{S}_{k,n}\left(\beta^{(n)} \otimes I_s\right) \tag{20}$$

where $\Delta\mathcal{S}_{k,n} = [\Delta S_n, \dots, \Delta S_{n+k-1}]$. We can also express $E_k^{(n)}$ as the following Schur complement

$$E_k^{(n)} = (\mathcal{M}_{k,n}/\mathcal{D}_{k,n} \otimes I_s);\ \text{where } \mathcal{M}_{k,n} = \begin{bmatrix} S_n & \Delta\mathcal{S}_{k,n} \\ z^{(n)} \otimes I_s & \mathcal{D}_{k,n} \otimes I_s \end{bmatrix}.$$

We notice that the Schur complement presentation could allow us to determine iterative algorithms for computing the approximations at each iteration.

4 The matrix polynomial extrapolation methods for linear problems

4.1 The general case

Consider the linear matrix equation

$$\mathcal{M}(X) = B \tag{21}$$

where $\mathcal{M}$ is a nonsingular linear matrix operator from $\mathbb{R}^{N\times s}$ onto $\mathbb{R}^{N\times s}$, B is a matrix of $\mathbb{R}^{N\times s}$ and X^* denotes the unique solution of (21). Starting from an initial guess S_0, we construct the matrix sequence $(S_n)_n$ by

$$S_{n+1} = \mathcal{C}(S_n) + B; \quad n = 0, 1, \dots, \tag{22}$$

with $\mathcal{C} = I - \mathcal{M}$. Note that if the sequence (S_n) is convergent, it's limit $S = X^*$ is the solution of the linear matrix (21).

From (21), we have $\Delta S_n = B - \mathcal{M}\,(S_n) = R(S_n)$ the residual of the vector S_n. Therefore the generalized residual of the approximation $T_k^{(n)}$ (defined earlier), becomes the true residual $\widetilde{R}(T_k^{(n)}) = R(T_k^{(n)})$, $\forall n,\ k$.

For simplicity and until specified, we set $n = 0$, denote $T_k^{(0)} = X_k$ and drop the index n in all our notations.

When applied to the sequence generated by the linear relation (22), the matrix polynomial extrapolation methods, (M-RRE, M-MPE and M-MMPE) produce approximations X_k to the exact solution X^* of (21) such that the corresponding residuals $R_k = B - A\,X_k$ satisfies the relations

$$X_k - X_0 \in \widetilde{\mathcal{W}}_k = \mathcal{M}(\widetilde{\mathcal{V}}_k) \text{ and } R_k \perp \widetilde{\mathcal{L}}_k.$$

Remark that $\widetilde{\mathcal{V}}_k = \mathcal{K}_k(\mathcal{M}, R_0)$ where $(\mathcal{K}_k(\mathcal{M}, R_0)$ is the matrix Krylov subspace generated by the matrices R_0, $\mathcal{M}(R_0)$,..., $\mathcal{M}^{k-1}(R_0)$ and $\widetilde{\mathcal{W}}_k = \mathcal{K}_k(\mathcal{M}, \mathcal{M}(R_0))$. Then we have $\widetilde{\mathcal{L}}_k \equiv \widetilde{\mathcal{W}}_k$ for M-RRE, $\widetilde{\mathcal{L}}_k \equiv \widetilde{\mathcal{V}}_k$ for M-MPE and $\widetilde{\mathcal{L}}_k \equiv \widetilde{Y}_k = span\{Y_1, \ldots, Y_k\}$ for M-MMPE where $Y_1, \ldots, Y_k$ are chosen $N \times s$ matrices. Then, we can state the following result which is a generalisation of the vecor case studied in [13, 23].

Theorem 1 *When applied to sequence generated by (22) the M-RRE, and the M-MPE are matrix Krylov subspace methods and are mathematically equivalent to the global GMRES and the global FOM methods, respectively* [16].

When the linear process (22) is convergent, it is more useful in practice to apply the extrapolation methods after a fixed number p of basic iterations. We also notice that, when these methods are used in their complete form, the required work and storage grow linearly with the iteration step. To overcome this drawback we use them in a cycling mode and this means that we have to restart the algorithms after a chosen number m of iterations.

The algorithm is summarized as follows

1. Set $k = 0$, choose X_0 and the numbers p and m.
2. Basic iteration
   - $T_0 = X_0$
   - $Z_0 = T_0$
   - $Z_{j+1} = \mathcal{C}(Z_j) + B,\ j = 0, \ldots, p - 1.$
3. Extrapolation scheme
   - $S_0 = Z_p$
   - $S_{j+1} = \mathcal{C}(S_j) + B,\ j = 0, \ldots, m.$
4. Compute the approximation T_m by M-RRE, M-MPE or M-MMPE.
5. Set $X_0 = T_m$, $k = k + 1$ and go to step 2.

4.2 Some particular cases

As examples of the application of the matrix extrapolation methods for linear problem, we will consider here two examples.

4.2.1 Lyapunov and Stein matrix equations

We first consider the following symmetric Stein matrix equation (also called the discrete-time Lyapunov equation)

$$\mathcal{M}(X) = X - AXA^T = C \tag{23}$$

where A and C are real $N \times N$ matrices and we assume that $\rho(A) < 1$. Equation (23) plays an important role in model reduction for linear discrete time dynamical systems. We assume that the eigenvalues of A lie inside the unit disk. It is well known that under this assumption, the matrix (23) has the unique solution

$$X = \sum_{i=0}^{\infty} A^i C (A^T)^i. \tag{24}$$

The solution X^* is given by $X^* = \lim\limits_{j \longrightarrow \infty} S_j$ with

$$S_j = \sum_{i=0}^{j} A^i C (A^T)^i. \tag{25}$$

The matrix sequence $(S_j)_j$ can be generated, by the following Smith iteration [26]

$$S_0 = 0, \qquad S_j = C + \mathcal{C}(S_{j-1}) \quad \text{with} \quad \mathcal{C}(S_{j-1}) = A\, S_{j-1} A^T; \; j \geq 1. \tag{26}$$

The Smith method is convergent but has very slow convergence and hence the application of the matrix polynomial extrapolation methods could be effective to speed up the convergence. We notice that instead of using the Smith iteration, one can consider the squared Smith iteration which have a rapid convergence. The squared Smith iteration is defined as follows

$$\begin{array}{l} S_0 = C,\, A_0 = A, \\ S_j = S_{j-1} + A_{j-1} S_{j-1} A_{j-1}^T, \;\; A_j = A_{j-1}^2, \; j \geq 1. \end{array} \tag{27}$$

Consider now the following Lyapunov matrix equation

$$AX + XA^T + FF^T = 0, \tag{28}$$

where A is assumed to be a stable matrix (all it's eigenvalues have negative real parts). Then, for any real $\mu < 0$, (28) is equivalent to the following symmetric Stein equation

$$X - A_\mu X A_\mu^T = F_\mu F_\mu^T, \tag{29}$$

where $A_\mu = (A - \mu I)(A + \mu I)^{-1}$ and $F_\mu = \sqrt{-2\mu}(A + \mu I)^{-1} F$. As the matrix A is stable, it can be proved that the spectral radius $\rho(A_\mu) < 1$. The parameter μ is chosen to minimize $\rho(A_\mu)$ to get a good convergence of the Smith iteration.

As a first numerical example, we consider the symmetric Stein matrix (23) where the matrix A is obtained from the 5-point discretization of the operator

$$L_u = \Delta u - f_1(x, y)\frac{\partial u}{\partial x} + f_2(x, y)\frac{\partial u}{\partial y} + g(x, y), \tag{30}$$

Table 1 Results for Stein matrix equations

Sequence	Error Norms
$\bar{X}_j$; $j = 3, nc = 3$	1.2×10^{-2}
$T_{2,M-RRE}$; $nc = 2$	1.5×10^{-4}
$\bar{X}_j$; $j = 3, nc = 4$	1.2×10^{-11}
$T_{2,M-RRE}$; $nc = 4$	2.5×10^{-14}

on the unit square $[0, 1] \times [0, 1]$ with homogeneous Dirichlet boundary conditions. The number of inner grid points in each direction was $n_0 = 30$ and the dimension of the matrix A is $n = n_0^2$ which yields a matrix of size 900. The matrix A was divided by $\|A\|$ so that $\rho(A) < 1$. Here we set $f_1(x, y) = 10xy$, $f_2(x, y) = e^{x^2 y}$ and $g(x, y) = 20y$. The computations were carried out using MATLAB 8.0 . The matrix C appearing in the right hand side of (23) is chosen such that the exact solution X^* is some fixed random $N \times N$ matrix generated by the matlab function `rand`. In Table 1, we listed the norm of the errors $\|\bar{X}_j - X^*\|_F$ and $\|T_{2,M-RRE} - X^*\|_F$, where $T_{2,M-RRE}$ is the approximate solution obtained by cycling the matrix RRE method (with $k = 2$) and $(\bar{X}_j)$ is the sequence obtained by cycling the squared Smith algorithm. The number of cycles is denoted by nc.

As observed from the results of Table 1, the matrix RRE method with $k = 2$ returns good results. We also notice that iterating the squared Smith method gives better results than the classical squared Smith method and this is the reason why we presented results obtained by the sequence $(\bar{X}_j)$. For large problems and when the right hand side has a low rank $C = FF^T$ where F has s columns ($s \ll N$), the exact solution has always a numerical low rank and then the approximate solution should be given in a low rank form: $X_k \approx Z_k Z_k^T$ with $\texttt{rank}(Z_k) << N$. Such a decomposition allows one to save memory and time by computing and saving only the matrix Z_k. This point should be developed in a future work to implement efficiently the matrix extrapolation methods when used for solving Lyapunov and Stein matrix equations.

4.2.2 Matrix linear ill-posed problems

We consider linear discrete ill-posed problems of the form

$$\mathcal{C}(X) = A_1\, X\, A_2^T = B, \tag{31}$$

where at least one of the matrices $A_1,\ A_2 \in \mathbb{R}^{N\times N}$ is of ill-determined rank. This makes the solution $X^* \in \mathbb{R}^{N\times N}$ of (31) very sensitive to perturbations in the right-hand side $B \in \mathbb{R}^{N\times N}$. The right hand side matrix B represents observations that are contaminated by measurement errors of unknown size, i.e.,

$$B = \widetilde{B} + E, \tag{32}$$

where $\widetilde{B}$ denotes the unavailable error-free right-hand side and the matrix $E \in \mathbb{R}^{N\times N}$ represents the error.

Ill-posed problems of the form (31) arise from the discretization of Fredholm integral equations of the first kind in two space-dimensions,

$$\iint_{\Omega} K(x, y, s, t) f(s, t) ds dt = g(x, y), \qquad (x, y) \in \Omega', \tag{33}$$

where Ω and Ω' are rectangles in $\mathbb{R}^2$ and the kernel is separable,

$$K(x, y, s, t) = k_1(x, s)\, k_2(y, t), \qquad (x, y) \in \Omega', \qquad (s, t) \in \Omega.$$

Discretization of (33) gives a matrix equation of the form (31).

Consider the singular value decompositions $A_1 = U^{(1)}\Sigma^{(1)}(V^{(1)})^T$ and $A_2 = U^{(2)}\Sigma^{(2)}V^{(2)})^T$, of the matrices A_1 and A_2 in (31) with the orthonormal matrices $U^{(k)} = [u_{1,k}, u_{2,k}, \ldots, u_{n,k}] \in \mathbb{R}^{N\times N}$ and $V^{(k)} = [v_{1,k}, v_{2,k}, \ldots, v_{n,k}] \in \mathbb{R}^{N\times N}$; $k = 1, 2$. The diagonal entries of the matrix $\Sigma^{(k)} = \text{diag}[\sigma_{1,k}, \sigma_{2,k}, \ldots, \sigma_{n,k}] \in \mathbb{R}^{N\times N}$ are the singular values of A_k and are given in the decreasing order $\sigma_{1,k} \geq \sigma_{2,k} \geq \ldots \geq \sigma_{n,k} \geq 0$, $k = 1, 2$. The singular value decomposition of the matrix A, is given by

$$A = U\Sigma V^T \ \text{ with } U = U^{(2)} \otimes U^{(1)},\ V = V^{(2)} \otimes V^{(1)} \text{ and } \ \Sigma = \Sigma^{(2)} \otimes \Sigma^{(1)}.$$

The singular values of A, i.e., the diagonal entries of $\Sigma = \text{diag}[\sigma_1, \sigma_2, \ldots, \sigma_{N^2}] \in \mathbb{R}^{N^2\times N^2}$ are ordered so that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\ell_0} > \sigma_{\ell_0+1} = \ldots = \sigma_{N^2} = 0$. They are an enumeration of the diagonal entries of $\Sigma^{(2)} \otimes \Sigma^{(1)}$, i.e.,

$$\sigma_\ell = \sigma_{j(\ell),2}\sigma_{i(\ell),1}, \qquad 1 \leq \ell \leq N^2, \tag{34}$$

where $i(\ell)$ and $j(\ell)$ are nondecreasing functions of ℓ with range $\{1, 2, \ldots, N\}$. These functions, defined in [1], allow us to compute the SVD of the matrix A without using directly the Kronecker product . Using this enumeration, the columns of the orthogonal matrices U and V are such that

$$u_\ell = u_{j(\ell),2} \otimes u_{i(\ell),1}, \qquad v_\ell = v_{j(\ell),2} \otimes v_{i(\ell),1}, \qquad 1 \leq \ell \leq N^2. \tag{35}$$

The Truncated Singular Value Decomposition (TSVD) sequence $\{S_n\}_{n\geq 0}$, approaching the solution of the matrix least squares problem

$$\min_X \|A_1 X A_2 - B|_F \tag{36}$$

is given by [1]

$$S_n = \sum_{l=1}^{n} \frac{u_{i(\ell),1}^T\ B\ u_{j(\ell),2}}{\sigma_{i(\ell),1}\sigma_{j(\ell),2}}\, v_{i(\ell),1}v_{j(\ell),2}^T = \sum_{l=1}^{l} \delta_l\ V_l, \tag{37}$$

with

$$S_0 = 0,\ \delta_l = \frac{u_{i(\ell),1}^T\ B\ u_{j(\ell),2}}{\sigma_{i(\ell),1}\sigma_{j(\ell),2}},\ \text{ and }\ V_l = v_{i(\ell),1}v_{j(\ell),2}^T. \tag{38}$$

Therefore, a matrix extrapolation method could be applied to the sequence $\{S_n\}_{n\geq 0}$ to get a new sequence (T_k) with a better convergence.

Using the expressions (9) and (10), we get

$$T_k = \sum_{i=1}^{k} \Delta\mathcal{S}_{i-1}(\alpha_i \otimes I_s), \tag{39}$$

where $\Delta\mathcal{S}_{i-1} = [\Delta S_0, \ldots, \Delta S_{i-1}]$. The vector $\alpha = (\alpha_1, \ldots, \alpha_k)^T$ solves the linear system

$$\left(\Delta^2\mathcal{S}_{k-1}{}^T \diamond \Delta^2\mathcal{S}_{k-1}\right)\alpha = -\Delta^2\mathcal{S}_{k-1}{}^T \diamond \Delta S_0. \tag{40}$$

Using the fact that $\Delta S_j = \delta_j V_j$, it follows that the matrix $\left(\Delta^2\mathcal{S}_{k-1}{}^T \diamond \Delta^2\mathcal{S}_{k-1}\right)$ is a tridiagonal matrix given by

$$\left(\Delta^2\mathcal{S}_{k-1}{}^T \diamond \Delta^2\mathcal{S}_{k-1}\right) = \left[\text{Tridiag}\left(-\delta_i^2, \delta_i^2 + \delta_{i+1}^2, -\delta_{i+1}^2\right)\right].$$

We also have $\Delta^2\mathcal{S}_{k-1}{}^T \diamond \Delta S_0 = -\delta_1^2 e_1$ where e_1 is the first unit vector of the canonical basis of $\mathbb{R}^k$. Using all these relations, the Matrix RRE approximation T_k can be expressed as

$$T_k = \sum_{l=1}^{k} \alpha_l \frac{u_{i(\ell),1}^T\ B\ u_{j(\ell),2}}{\sigma_{i(\ell),1}\sigma_{j(\ell),2}}\ v_{i(\ell),1} v_{j(\ell),2}^T, \tag{41}$$

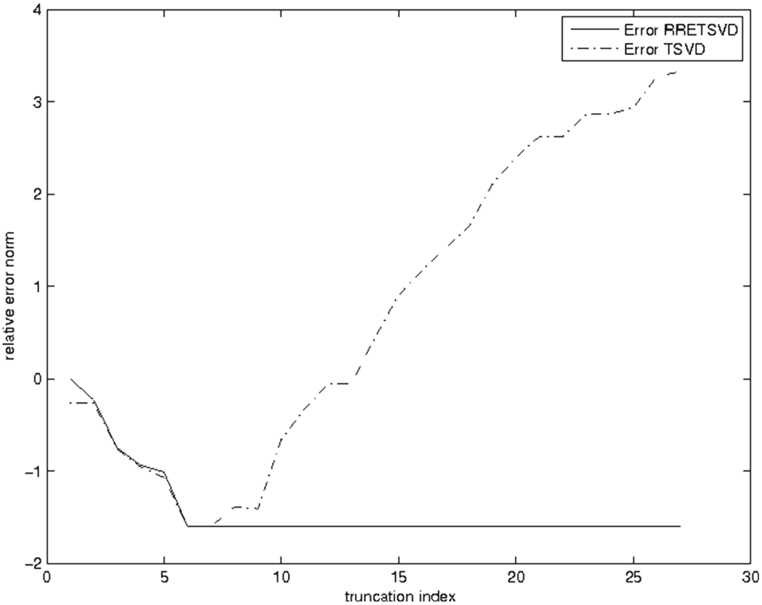

Fig. 1 Relative errors: $\|S_k - X^*\|_F/\|X^*\|_F$ for TSVD (dashed graph) and $\|T_k - X^*\|_F/\|X^*\|_F$ for M-RRE-TSVD (solid graph)

where $\alpha_l = \left(\sum_{j=l+1}^{k} \frac{1}{\delta_{j+1}^2}\right) / \left(\sum_{p=0}^{k} \frac{1}{\delta_{p+1}^2}\right)$; $l = 1, \ldots, k$. From the expressions (37) and (41), the M-RRE transformation acts as a filter on the TSVD sequence.

To illustrate this, we consider an example given in [1]. The computations were carried out using MATLAB 8.0 with machine epsilon about $2 \cdot 10^{-16}$.
The nonsymmetric matrices $A_1 = \texttt{baart(1500)}$ and $A_2 = \texttt{foxgood(1500)}$ are from [12]. Specifically, we let $A_1 = \texttt{baart(1500)}$ and $A_2 = \texttt{foxgood(1500)}$. The computed condition numbers of these matrices are $\kappa(A_1) = 2 \cdot 10^{18}$ and $\kappa(A_2) = 3 \cdot 10^{13}$, where $\kappa(A_i) = \|A_i\|_2 \, \|A_i^{-1}\|_2$. Since $\kappa(A) = \kappa(A_1)\kappa(A_2)$, the matrix A is numerically singular. The desired solution X^* is the matrix with all entries unity. The error-free right-hand side is given by $\widetilde{B} = A_1 X^* A_2^T$ and the associated error-contaminated right-hand side B is determined by $B = \widetilde{B} + E$, where the error-matrix E has normally distributed entries with zero mean and is normalized to correspond to a specific noise-level $\frac{\|E\|_F}{\|\widetilde{B}\|_F} = 1.2 \cdot 10^{-2}$. Neither $\widetilde{B}$ nor ν are assumed to be known. Here we apply the matrix RRE method (Fig. 1).

5 Conclusion

We presented in this work new matrix versions of the most popular vector extrapolation methods namely RRE, MPE, MMPA and TEA. Using the Kronecker inner product, the defined generalized residuals are obtained as orthogonal or oblique projections of the initial generalized residual. The use of the $\diamond$ product allows more compact forms of the obtained matrix sequence transformations . These matrix extrapolation methods were applied to some linear matrix problems and the obtained results are promising. Other applications to large linear and nonlinear matrix problems are under investigation.

Acknowledgments We would like to thank the two referees for their valuable remarks and suggestions.

References

1. Bouhamidi, A., Jbilou, K., Reichel, L., Sadok, H.: An extrapolated TSVD method for linear discrete ill-posed problems with Kronecker structure. J. Comput. Appl. Math. **236**, 2078–2089 (2012)
2. Bouyouli, R., Jbilou, K., Sadaka, R., Sadok, H.: Convergence properties of some block Krylov subspace methods for multiple linear systems. J. Comput. Appl. Math. **196**, 498–511 (2006)
3. Brezinski, C., Redivo Zaglia, M.: Extrapolation Methods. Theory and Practice. North-Holland, Amsterdam (1991)
4. Brezinski, C.: Généralisation de la transformation de Shanks, de la table de la Table de Padé et l'epsilon-algorithm. Calcolo **12**, 317–360 (1975)
5. Brezinski, C., Redivo Zaglia, M.: Vector and matrix sequence transformations based on biorthogonality. Appl. Numer. Math. **21**, 353–373 (1996)
6. Brezinski, C.: Other manifestations of the Schur complement. Lin. Alg. Appl. **111**, 231–247 (1988)
7. Brezinski, C., Redivo-Zaglia, M., A Schur complement approach to a general extrapolation algorithm: Lin. Alg. Appl. **279-301**, 368 (2003)

8. Brezinski, C., Redivo-Zaglia, M.: The simplified topological epsilon-algorithms for accelerating sequences in a vector space. http://arxiv.org/abs/1402.2473
9. Brezinski, C.: Block descent methods and hybrid procedures for linear systems. Numer. Algo. **29**, 21–32 (2002)
10. Cabay, S., Jackson, L.W.: A polynomial extrapolation method for finding limits and antilimits for vector sequences. SIAM J. Numer. Anal. **13**, 734–752 (1976)
11. Eddy, R.P.: Extrapolation to the limit of a vector sequence. In: Wang, P.C.C. (ed.): Information Linkage Between Applied Mathematics and Industry, pp. 387–396. Academic Press, New-York (1979)
12. Hansen, P.C.: Regularization tools version 4.0 for MATLAB 7.3. Numer. Algo. **46**, 189–194 (2007)
13. Jbilou, K., Sadok, H.: Vector extrapolation methods. Application and numerical comparison. J. Comp. Appl. Math. **122**, 149–165 (2000)
14. Jbilou, K., Sadok, H.: Analysis of some vector extrapolation methods for linear systems. Numer. Math. **70**, 73–89 (1995)
15. Jbilou, K., Sadok, H.: LU-implementation of the modified minimal polynomial extrapolation method. IMA J. Numer. Anal. **19**, 549–561 (1999)
16. Jbilou, K., Messaoudi, A., Sadok, H.: Global FOM and GMRES algorithms for multiple right-hand sides. Appl. Num. Math. **31**, 49–63 (1999)
17. Jbilou, K., Messaoudi, A., Tabaa, K.: Some Schur complement identities to matrix extrapolation methods. Lin. Alg. Appl. **392**, 195–210 (2004)
18. Meš̀ina, M.: Convergence acceleration for the iterative solution of x=Ax+f. Comput. Meth. Appl. Mech. Eng. **10**(2), 165–173 (1977)
19. Ouellette, D.V.: Schur complements and statistics. Lin. Alg. Appl. **36**, 1870–295 (1981)
20. Pugatchev, B.P.: Acceleration of the convergence of iterative processes and a method for solving systems of nonlinear equations, U.S.S.R. Comput. Math. and Math. Phys. **17**, 199–207 (1978)
21. Saad, Y.: Iterative methods for sparse linear systems. PWS Press, New York (1995)
22. Shanks, D.: Nonlinear transformations of divergent and slowly convergent sequences. J. Math. Phys. **34**, 1–42 (1955)
23. Sidi, A.: Convergence and stability of minimal polynomial and reduced rank extrapolation algorithms. SIAM J. Numer. Anal. **23**, 197–209 (1986)
24. Sidi, A., Ford, W.F., Smith, D.A.: Acceleration of convergence of vector sequences. SIAM J. Numer. Anal. **23**, 178–196 (1986)
25. Sidi, A.: Efficient implementation of minimal polynomial and reduced rank extrapolation methods. J. Comput. Appl. Math. **36**, 305–337 (1991)
26. Smith, R.: Matrix equation $XA + BX = C$. SIAM J. Appl. Math. **16**, 198–201 (1968)
27. Wynn, P.: Acceleration technique for iterated vector and matrix problems. Math. Comp. **16**, 301–322 (1962)
28. Zhang, F.-Z.: The Schur Complement and its applications. Springer, New York (2005)