Applied Mathematics and Computation 333 (2018) 246–253

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Iterative methods for finding commuting solutions of the Yang–Baxter-like matrix equation

Ashim Kumar[a], João R. Cardoso[b,c,*]

[a] Department of Mathematics, Panjab University Constituent College, Nihal Singh Wala 142046, India
[b] Polytechnic Institute of Coimbra, ISEC, Rua Pedro Nunes, 3030-199 Coimbra, Portugal
[c] Institute of Systems and Robotics, University of Coimbra, Pólo II, 3030-290 Coimbra, Portugal

ARTICLE INFO

Keywords:
Yang–Baxter-like matrix equation
Fréchet derivative
Iterative methods
Convergence
Stability
Idempotent matrix

ABSTRACT

The main goal of this paper is the numerical computation of solutions of the so-called Yang–Baxter-like matrix equation $AXA = XAX$, where A is a given complex square matrix. Two novel matrix iterations are proposed, both having second-order convergence. A sign modification in one of the iterations gives rise to a third matrix iteration. Strategies for finding starting approximations are discussed as well as a technique for estimating the relative error. One of the methods involves a very small cost per iteration and is shown to be stable. Numerical experiments are carried out to illustrate the effectiveness of the new methods.

1. Introduction

In this work, we are interested in the numerical evaluation of commuting solutions X of the quadratic matrix equation

$$AXA = XAX, \tag{1}$$

where A is a given complex matrix of order n, i. e., $A \in \mathbb{C}^{n\times n}$. This equation is called the Yang–Baxter-like matrix equation (and abbreviated to YB-like equation) because of its connections with the classical Yang–Baxter equation arising in statistical mechanics. More details about those connections can be found, for instance, in [4,6,7] and the references therein.

Note that the YB-like matrix equation has at least two trivial solutions: $X = 0$ and $X = A$. Of course, the interest in solving (1) is to find nontrivial solutions. This has been a hard task because the set of all solutions of (1) is very difficult to characterize for a general matrix A. It can be viewed as being the union of the set of solutions commuting with A with the set of solutions non commuting with A. Some results characterizing the commuting solutions for particular choices of matrices A have been stated by Ding and Rhee and their collaborators. See, for instance, [4] for stochastic matrices and [6,7] for diagonalizable matrices; see also the references therein for other cases of A. Along with those characterizations, some methods for finding commuting solutions have also been proposed. For example, in [4], a method based on Brouwer's fixed point theorem was used for solving (1), under the assumption of A being stochastic; in [5], spectral based methods are proposed, and, in [6], an iterative scheme based on the mean ergodic theorem is derived for a diagonalizable matrix A. All of these papers contain many valuable contributions for understanding the YB-like equation and for computing its solutions. However, their implementation in finite precision environments has not been carried out and an investigation

[*] Corresponding author at: Polytechnic Institute of Coimbra, ISEC, Rua Pedro Nunes, 3030-199 Coimbra, Portugal.
E-mail addresses: ashimsingla1729@gmail.com (A. Kumar), jocar@isec.pt (J.R. Cardoso).

https://doi.org/10.1016/j.amc.2018.03.078
0096-3003/© 2018 Elsevier Inc. All rights reserved.

from a numerical viewpoint is lacking. Moreover, important features of those numerical methods like accuracy, stability and computational cost have not been investigated so far.

Hence, one of the goals of this work is to give a contribution to fill in this gap. We propose two iterative schemes for calculating commuting solutions of YB-like equation (1), regardless of A being or not diagonalizable, and provide a thorough investigation of their numerical behavior. Topics like convergence, stability, relative error, choice of a suitable initial guess, etc., will be investigated in detail. Both iterations are then implemented and compared with the iteration proposed in [6] for diagonalizable matrices A. Our methods are also tested with non-diagonalizable matrices. All the results of the implementations will be discussed. It will become clear that iteration (2) shows a superior numerical performance. Towards our aim, we present our first iterative scheme in what follows:

$$X_{k+1} = X_k^2(2X_k - I)^{-1}, \tag{2}$$

where X_0 is a certain starting matrix and $k = 0, 1, 2, \ldots$. Assuming that X_0 has no eigenvalue with real part equal to 1/2 and commutates with A, it is proven in Section 2 that the sequence (X_k) generated by (2) is well defined and converges quadratically to an idempotent matrix X_* commuting with A, that is, $X_*^2 = X_*$ and $AX_* = X_*A$. Once such an X_* has been computed, it is immediate that

$$X = AX_*$$

is a solution of YB-like equation (1).

A complementary iterative scheme for evaluating the solutions of (1) is provided below.

Given an initial matrix Y_0 commuting with A and such that, for a given subordinate matrix norm,

$$\|AY_0A - Y_0AY_0\| < 1, \tag{3}$$

we will show that the sequence (Y_k) defined by

$$\begin{aligned} \psi_k &= (Y_k(Y_k - A))^2, \\ Y_{k+1} &= \frac{1}{2}\left(A + (A(A - 4\psi_k))^{1/2}\right), \quad k = 0, 1, 2, \ldots, \end{aligned} \tag{4}$$

converges quadratically to a solution of (1). If a matrix $Z \in \mathbb{C}^{n\times n}$ has no eigenvalues on the closed negative real axis, $Z^{1/2}$ stands for the principal matrix square root of Z, that is, $Z^{1/2}$ is the unique square root of Z with eigenvalues having positive real parts [9,10]. For the iteration (4) to be well defined, it shall be assumed that, during the iterative process, the eigenvalues of the matrix $A(A - 4\psi_k)$ do not lie on the closed negative real axis. Choosing the minus sign in the definition of Y_{k+1} instead of the plus one, yields a variant of (4)

$$\begin{aligned} \psi_k &= (Y_k(Y_k - A))^2, \\ Y_{k+1} &= \frac{1}{2}\left(A - (A(A - 4\psi_k))^{1/2}\right), \quad k = 0, 1, 2, \ldots, \end{aligned} \tag{5}$$

which converges to a different solution of (1).

The paper is organized as follows. Section 2 is devoted to the investigation of iteration (2). We show, in particular, that it converges quadratically to an idempotent matrix in a stable fashion, with respect to perturbations of first-order. Then a commuting solution of (1) comes out easily. The second iteration (4) is the topic of Section 3. It is proven that its convergence is quadratic in terms of a residual arising naturally from (1). Termination criteria and error estimation are addressed in Sections 4 and 5, respectively. To summarize the previous results, two practical algorithms are written in pseudo-code in Section 6. Both algorithms are then implemented and tested with several numerical experiments in Section 7. Finally, some conclusions of our work will be drawn in Section 8.

Notation: $\|.\|$ denotes a subordinate matrix norm and $\|.\|_F$ the Frobenius norm; $\sigma(A)$ denotes the spectrum of the matrix A; $\rho(A)$ is the spectral radius of A; $\Re(z)$ is the real part of the complex number z.

2. Analysis of iteration (2)

2.1. Convergence

To investigate the convergence of (2), we start by considering the scalar polynomial $p(z) = z^2 - z$, which has 0 and 1 as roots. Applying Newton's method to the equation $p(z) = 0$, yields the complex scalar iteration

$$z_{k+1} = \frac{z_k^2}{2z_k - 1}, \tag{6}$$

$(k = 1, 2, \ldots)$ which is the scalar counterpart of (2). For $\alpha = 0,\ 1,$ let

$$B(\alpha) := \{z_0 \in \mathbb{C} :\ z_k \to \alpha\},$$

denote the basin of attraction of the root α.

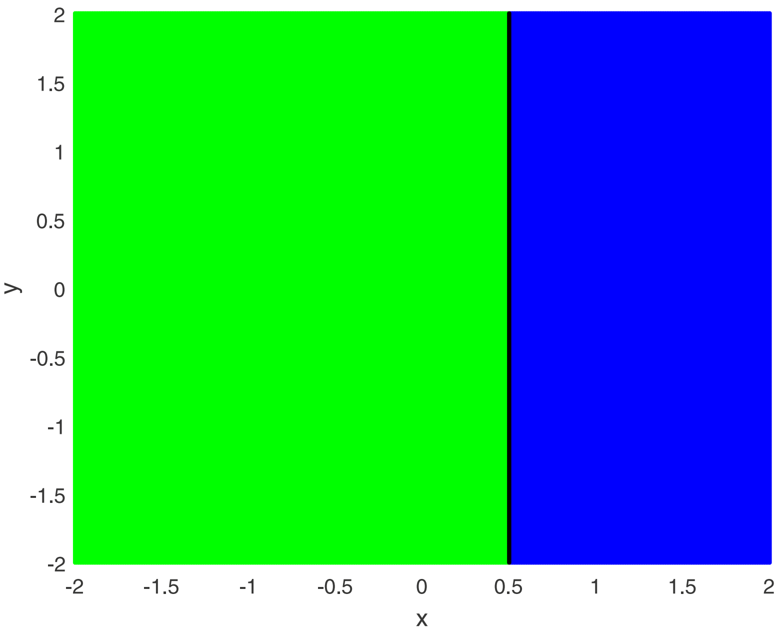

Fig. 1. Basins of attraction of iteration (6) in the square $-2 \le x \le 2,\ -2 \le y \le 2$ of the complex plane. If the initial guess lies on the left-hand (resp., right-hand) side of the line $\Re(z) = 1/2$, then the sequence generated by (6) converges to 0 (resp., 1). Initial guesses on the line $\Re(z) = 1/2$ do not lead to convergence.

The basins of attraction for complex Newton's method were first considered by Cayley [2,3]. In 1879, he showed that for a complex quadratic polynomial $q(z) = az^2 + bz + c$ having two distinct roots z_1 and z_2, the Newton's method applied to $q(z)$ divides the complex plane into two half-planes by means of the perpendicular bisector L of the line segment from z_1 to z_2. The half-plane containing z_1 is the basin of attraction of the root z_1 and the half-plane containing z_2 is the basin of attraction of z_2. Newton's method starting at points on the line L has a chaotical behavior and is divergent.

Applying the Cayley result to our problem, we see that the perpendicular bisector L is now the line $\Re(z) = 1/2$, that is, the set of complex numbers having 1/2 as real part (see Fig. 1). Hence, $B(0)$ is the half-plane defined by $\Re(z) < 1/2$ and $B(1)$ is the half-plane $\Re(z) > 1/2$. This means that starting approximations z_0 on the left-hand side of the line $\Re(z_0) = 1/2$ ensure convergence of the sequence (z_k) to 0, while z_0 on the right-hand side guarantee convergence to 1. It is easy to check that if $\Re(z_0) = 1/2$, then $\Re(z_k) = 1/2$, for all k. Hence, for such an z_0, the Newton iteration (6) does not converge.

We refer the reader to [12,15] for more information on solving nonlinear equations.

Now, we focus our attention on the convergence of matrix iteration (2), which comes from the application of Newton's method to the matrix equation $X^2 - X = 0$. We recall that the solutions of this matrix equation are called *idempotent matrices*. The following lemma shows that (2) converges to an idempotent matrix.

Lemma 1. *If the starting approximation $X_0 \in \mathbb{C}^{n\times n}$ is a matrix with no eigenvalues on the line $\Re(z) = 1/2$, then the sequence (X_k) generated by (2) is well defined (that is, X_k has no eigenvalue on the line $\Re(z) = 1/2$) and converges to an idempotent matrix X_*.*

Proof. Since, for any $\lambda \in \sigma(X_0)$, where $\sigma(X_0)$ denotes the spectrum of X_0, the scalar sequence (z_k) generated by (6) with $z_0 = \lambda$ converges to 0 or 1, the results in [9, Theorem 4.15, Corollary 4.16] (see also [13]) guarantee that the sequence (X_k) converges to a diagonalizable matrix X_* with eigenvalues 0 and 1. Thus X_* is idempotent. We recall that every idempotent matrix is diagonalizable (see, for instance, [11, Problem 3.3.P3]). □

Since (2) is a Newton iteration, the convergence of the sequence (X_k) to an idempotent matrix is quadratic. The following result gives an alternative proof for the quadratic convergence and gives some insight into the behavior of the iteration. In particular, it shows that the convergence may slow down when the norm $\left\|(2X_k - I)^{-1}\right\|$ attains large values.

Lemma 2. *Assume that X_0 has no eigenvalue with real part equal to 1/2 and that the sequence (X_k) generated by (2) converges to the idempotent matrix X_*. Then, for each k,*

$$\|X_{k+1} - X_*\| \le c\|X_k - X_*\|^2,$$

where $c = \max_k \left\|(2X_k - I)^{-1}\right\|$.

Proof. Since $X_*^2 = X_*$, we have, for any k,

$$X_{k+1} - X_* = X_k^2(2X_k - I)^{-1} - X_*^2$$

$$= (X_k - X_*)^2 (2X_k - I)^{-1}.$$

Applying norms yields

$$\begin{aligned} \|X_{k+1} - X_*\| &\leq \|X_k - X_*\|^2 \left\| (2X_k - I)^{-1} \right\| \\ &= c\|X_k - X_*\|^2, \end{aligned}$$

with c given as above. Note that due to the assumptions on X_0, the sequence of positive numbers defined, for every k, by $\left\| (2X_k - I)^{-1} \right\|$ is bounded, so such a constant c exists. □

Theorem 3. *Let $A \in \mathbb{C}^{n\times n}$ be a given matrix and assume that the initial guess X_0 commutes with A and has no eigenvalue on the line $\Re(z) = 1/2$. Then:*

(i) *(X_k) converges to an idempotent matrix X_* commuting with A;*
(ii) *$X = AX_*$ is a commuting solution of YB-like equation (1).*

Proof.

(i) Since X_0 has no eigenvalue on the line $\Re(z) = 1/2$, by Lemma 1, (X_k) converges to an idempotent matrix X_*. From the commutativity between X_0 and A, it follows immediately that, for any k, X_k commutes with A and thus X_* also commutes with A.
(ii) Immediate. □

Now we shall make some practical considerations on how to choose the initial guess X_0 in (2) in order to compute solutions of the YB-like equation. We note that, by Theorem 3, it is mandatory that X_0 commutes with A and has no eigenvalue on the line $\Re(z) = 1/2$. So a possible and practical choice is to take $X_0 = A$. If A has no eigenvalues on the line $\Re(z) = 1/2$, Theorem 3, combined with Lemma 1, ensure that (2) always converges to an idempotent matrix X_*, making easy to find a solution $X = AX_*$ to the YB-like equation. Other possibilities include, for instance, taking X_0 as a polynomial in A, providing that such a polynomial has no eigenvalue λ with $\Re(\lambda) = 1/2$. If $\Re(\lambda) = 1/2$, for some $\lambda \in \sigma(A)$, it is easy to find a positive scalar α such that $X_0 = \alpha A$ or $X_0 = A \pm \alpha I$ do not have eigenvalues with real part equal to 1/2.

2.2. Stability

It is well known that the stability of a matrix iteration is crucial for its success [9, Section 4.9]. In finite precision arithmetic, a loss of commutativity and rounding errors amplification are frequent so that unstable matrix iterations can fail to converge. An important tool to assess the stability of a matrix iteration is the Fréchet derivative. Given a map $f : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$, the Fréchet derivative of f at $X \in \mathbb{C}^{n\times n}$ in the direction of $E \in \mathbb{C}^{n\times n}$ is a linear operator $L_f(X)$ that maps the "direction matrix" E to $L_f(X, E)$ such that

$$\lim_{E\to 0} \frac{\|f(X+E) - f(X) - L_f(X,E)\|}{\|E\|} = 0.$$

The Fréchet derivative of the matrix function f may not exist at X, but if it does it is unique and coincides with the directional (or Gâteaux) derivative of f at X in the direction E. Hence the existence of the Fréchet derivative guarantees that, for any $E \in \mathbb{C}^{n\times n}$,

$$L_f(X,E) = \lim_{t\to 0} \frac{f(X+tE) - f(X)}{t}.$$

Now we recall a necessary and sufficient condition for a matrix iteration to be stable. See [9], p. 97, and also Problem 4.6 and its solution on page 357.

Lemma 4. *Consider the matrix iteration $X_{k+1} = f(X_k)$ with a fixed point X_* and assume that f is Fréchet differentiable at X_*. Then the iteration is stable in a neighborhood of X_* if and only if the Fréchet derivative $L_f(X_*)$ has bounded powers, i.e., if and only if $\rho(L_f(M)) \leq 1$, where $\rho(.)$ denotes the spectral radius, and any eigenvalue λ of $L_f(M)$ such that $|\lambda| = 1$ is semisimple, that is, λ appears only in Jordan blocks of size 1×1.*

Consider the matrix function $f(X) = X^2(2X - I)^{-1}$, where it is assumed that 1/2 is not an eigenvalue of X. By [9, Theorem 3.8], the Fréchet derivative of f at X in the direction of a matrix $E \in \mathbb{C}^{n\times n}$, here denoted by $L_f(X, E)$, exists and is continuous in X and in E. Note that any fixed point X_* of f is an idempotent matrix, that is, $X_*^2 = X_*$. We will show below that iteration (2) is stable in a neighborhood of X_*.

For a given matrix X having no eigenvalue equal to 1/2, some calculation shows that

$$L_f(X,E) = \left(-2X^2(2X-I)^{-1}E + XE + EX\right)(2X-I)^{-1}.$$

Let X_* be a fixed point of f. Then $(2X_* - I)^{-1} = 2X_* - I$ and $L_f(X_*, E)$ simplifies to

$$L_f(X_*, E) = -2X_*EX_* + X_*E + EX_*.$$

It is easy to see that

$$L_f\big(X_*, L_f(X_*, E)\big) = L_f(X_*, E),$$

showing that $L_f(X_*, E)$ is an idempotent linear operator. Now Lemma 4 shows that (2) is stable.

3. Convergence of iteration (4)

Theorem 5. *Given $A \in \mathbb{C}^{n\times n}$, let us consider the sequence (Y_k) generated by (4). Let us denote by*

$$R_k := AY_kA - Y_kAY_k, \tag{7}$$

for any $k = 0, 1, 2, \ldots,$ the residual associated to YB-like equation (1) in the k-th iterate. If Y_0 commutes with A, $\|R_0\| < 1$, and $A(A - 4\psi_k)$ has no eigenvalues on the closed negative real axis, then (R_k) converges quadratically to 0.

Proof. Since Y_0 commutes with A and $A(A - 4\psi_k)$ has no eigenvalues on the closed negative real axis, the properties of the principal matrix square root and mathematical induction ensure that Y_k commutes with A for all $k = 1, 2, \ldots.$ Because of commutativity, we can write $R_k = A^2Y_k - AY_k^2$ and $R_{k+1} = A^2Y_{k+1} - AY_{k+1}^2$. By performing some calculation, one finds that Y_k and Y_{k+1} satisfy the equation

$$A^2Y_{k+1} - AY_{k+1}^2 = -\big(A^2Y_k - AY_k^2\big)^2,$$

meaning that

$$R_{k+1} = -R_k^2. \tag{8}$$

Taking into account that $\|R_0\| < 1$, relation (8) gives

$$\|R_{k+1}\| \le \|R_k\|^2 \le \|R_{k-1}\|^{2^2} \le \cdots \le \|R_0\|^{2^{k+1}}. \tag{9}$$

This shows that (R_k) converges quadratically to 0 as k tends to ∞. □

A practical use of iteration (4) requires to find an initial guess Y_0 satisfying the conditions of Theorem 5. Among the many possibilities of choosing such an Y_0, we may assume that $Y_0 = \alpha I$, for some non-negative real number α. It commutes with A and one just needs to guarantee that condition (3) holds. To find such an α, let us note that

$$\left\|\alpha A^2 - \alpha^2 A\right\| \le \alpha\|A\|^2 + \alpha^2\|A\|,$$

and so imposing

$$\alpha\|A\|^2 + \alpha^2\|A\| < 1$$

will lead us to a suitable choice for α. Solving that inequation with respect to α, it is easy to conclude that, for any α satisfying

$$0 < \alpha < \frac{1}{2\|A\|}\left(-\|A\|^2 + \sqrt{\|A\|^4 + 4\|A\|}\right), \tag{10}$$

then $Y_0 = \alpha I$ verifies condition (3), thus guaranteeing the convergence. Based on (9), one may be led to take α as small as possible to increase the convergence. Care must be taken in doing this, because (Y_k) may converge to 0 or A, which are trivial solutions of (1).

4. Termination criteria

A very important issue arising in the implementation of a matrix iteration is to find a suitable termination criterion. Depending on its nature or order of convergence, some termination criteria may be more appropriate than others. In [9, Section 4.9.2], there is an interesting discussion on how to stop a matrix iteration.

We have carried out many experiments (not reported here) to decide what could be the most adequate to our case. Our conclusion is that the classical criterion of requiring the norm of two successive iterations being smaller than a given tolerance `tol`, that is,

$$\|X_{k+1} - X_k\| < \texttt{tol}, \tag{11}$$

performs very well. Although there has been some criticism on criterion (11), we have observed that choosing $\texttt{tol} = 10^{-10}$ provides, in almost of all the experiments, a relative error close to the unit roundoff (see Section 7). The relative version of (11)

$$\frac{\|X_{k+1} - X_k\|}{\|X_{k+1}\|} < \texttt{tol}$$

has revealed to be unappropriate because $\|X_{k+1}\|$ may be very close to zero, which turns hard to reach the given tolerance.

5. Error estimation

Let $R(X) := AXA - XAX$ and assume that S is an exact solution, that is, $R(S) = 0$, and $\overline{S} = S + E$ is an approximation of S. For a given consistent matrix norm, the absolute and relative errors of $\overline{S}$ are given, respectively, by $\|E\|$ and $\|E\|/\|S\|$. Our aim is to estimate both errors in terms of the residual $R(\overline{S})$.

We have

$$\begin{aligned} R(\overline{S}) &= A(S+E)A - (S+E)A(S+E) \\ &= ASA + AEA - SAS - SAE - EAS - EAE. \end{aligned}$$

Attending that $R(S) = 0$ and dropping off the second order terms in E,

$$R(\overline{S}) \approx AEA - SAE - EAS.$$

Applying the vec(.) operator to the last equation (for background on vec(.) operator and Kronecker products see [9, Appendix B13] for a brief revision and [14] for a detailed approach) yields

$$\operatorname{vec}\left(R(\overline{S})\right) \approx \left(A^T \otimes A - I \otimes (SA) - (AS)^T \otimes I\right) \operatorname{vec}(E). \tag{12}$$

Denoting $M(X) := A^T \otimes A - I \otimes (XA) - (AX)^T \otimes I \in \mathbb{C}^{n^2 \times n^2}$, from (12) follows:

$$\|R(\overline{S})\|_F \approx \|M(S)\|_2 \, \|E\|_F,$$

meaning that

$$\|E\|_F \approx \frac{\|R(\overline{S})\|_F}{\|M(S)\|_2} \tag{13}$$

$|\cdot|_2$ stands for the spectral or 2-norm is an estimation for the absolute error of $\overline{S}$ in terms of the residual $R(\overline{S})$. Since the exact solution S is not available in general, for practical purposes one may use the same norm (e.g., the Frobenius norm) and replace S by $\overline{S}$ in (13). Hence, the estimates we propose to the absolute and relative errors are given, respectively, by

$$\texttt{est}_{\texttt{abs}} \approx \frac{\|R(\overline{S})\|}{\|M(\overline{S})\|} \tag{14}$$

and

$$\texttt{est}_{\texttt{rel}} \approx \frac{\|R(\overline{S})\|}{\|M(\overline{S})\| \, \|\overline{S}\|}. \tag{15}$$

6. Algorithms

Our findings on the previous sections are summarized in the following two algorithms. The first one uses iteration (2) and the second one (4).

Algorithm 6.1. Given $A \in \mathbb{C}^{n\times n}$, a nonzero starting matrix $X_0 \in \mathbb{C}^{n\times n}$ commuting with A, a tolerance `tol` and a maximum number of iterations m, this algorithm computes a commuting solution X of the Yang–Baxter matrix-like equation (1).

```
1. Set $\delta = 1$ and $k = 0$;
2. while $\delta > $ tol and $k \le m$
3.    Solve for $X_1$ the multiple linear system $(2X_0 - I)X_1 = X_0X_0$;
4.    $\delta = \|X_1 - X_0\|_F$;
5.    $k = k + 1$;
6.    $X_0 = X_1$;
7. end
8. $X = AX_1$;
```

Cost estimation: $\frac{20}{3}kn^3$, where k is the number of iterations.

Before stating the second algorithm, it is worth recalling that effective methods for computing the square root of a matrix A with no eigenvalues on $\mathbb{R}_0^-$ involve in general an initial Schur decomposition $A = UTU^*$, where U is unitary and T is upper triangular. This is the case of the MATLAB's function `sqrtm`, which implements the methods proposed in [1,8]. Since iteration (4) requires the computation of one matrix square root per step, Algorithm 6.2 starts with a Schur decomposition of A to reduce the total cost.

Algorithm 6.2. Given $A \in \mathbb{C}^{n\times n}$, a starting matrix $Y_0 \in \mathbb{C}^{n\times n}$ commuting with A such that $\|AY_0A - Y_0AY_0\| < 1$, a tolerance `tol`, and a maximum number of iterations m, this algorithm computes a commuting solution Y of the Yang–Baxter matrix-like equation (1).

Table 1
Comparison of the three algorithms. k is the number of iterations required by the algorithms and $\texttt{est}_{\texttt{rel}}$ is the estimation of the relative error by (15); ∞ means lack of convergence.

	alg-1		alg-2		alg-dr	
A_i	k	$\texttt{est}_{\texttt{rel}}$	k	$\texttt{est}_{\texttt{rel}}$	k	$\texttt{est}_{\texttt{rel}}$
A_1	6	1.4×10^{-16}	7	3.8×10^{-16}	5	1.6×10^{-16}
A_2	10	1.1×10^{-16}	5	2.3×10^{-16}	9	8.5×10^{-14}
A_3	10	6.7×10^{-16}	6	6.9×10^{-16}	7	1.0×10^{-14}
A_4	11	1.9×10^{-16}	7	1.5×10^{-16}	6	4.2×10^{-14}
A_5	8	1.2×10^{-16}	4	5.5×10^{-18}	∞	∞
A_6	7	9.9×10^{-17}	5	6.2×10^{-18}	∞	∞

```
1.  Set $\delta = 1$ and $k = 0$;
2.  Compute the complex Schur decomposition of $A$: $A = UTU^*$, with $U$ unitary and $T$ upper triangular;
3.  while $\delta > \texttt{tol}$ and $k \le m$
4.      $C = (Y_0(Y_0 - T))^2$;
5.      $Y_1 = 1/2\big(T + (T(T - 4C))^{1/2}\big)$;
6.      $\delta = \|Y_1 - Y_0\|_F$;
7.      $k = k + 1$;
8.      $Y_0 = Y_1$;
9.  end
10. $Y = U Y_0 U^*$;
```

Cost estimation: $(25 + 2k)n^3$.

We are assuming that Y_0 is upper triangular (e.g., $Y_0 = T$); if it does not, the cost will be a bit higher. Note that if in Step 5 the plus sign is replaced by the minus one, the algorithm will converge towards a different solution of (1).

7. Numerical experiments

We have implemented Algorithm 6.1, Algorithm 6.2 and an algorithm based on the method of Ding and Rhee [6] in MATLAB, with unit roundoff $u \approx 1.1 \times 10^{-16}$. The three methods will be denoted throughout by alg-1, alg-2 and alg-dr, respectively. We have carried out two experiments. The first one involves 6 diagonalizable matrices with a unique dominant eigenvalue and aims at comparing our two algorithms with the one of [6]; the second experiment illustrates the behavior of alg-1 and alg-2 when A is non-diagonalizable or has eigenvalues equal or close to 1/2. In both the experiments we have considered the tolerance $\texttt{tol} = 10^{-10}$.

7.1. Experiment 1

In this experiment, we consider the following six real and nonreal diagonalizable matrices with a unique dominant eigenvalue (MATLAB style is used):

- $A_1 = \texttt{hilb}(5)$ (Hilbert matrix of order 5);
- $A_2 = \texttt{gallery}('\texttt{lehmer}', 8) + 1\texttt{i} * \texttt{randn}(8)$ (complex matrix built upon a Lehmer and a randomized matrix);
- $A_3 = \texttt{gallery}('\texttt{frank}', 8)$ (Frank matrix of order 8);
- $A_4 = \texttt{gallery}('\texttt{smoke}', 7) + \texttt{rand}(7)$ (complex matrix);
- A_5 and A_6 are diagonalizable, but result from small perturbations of non-diagonalizable matrices.

This experiment shows that alg-dr may not converge for a diagonalizable matrix that is “close” to a non-diagonalizable matrix. Non convergence of alg-dr may also occur if the computed eigenvalue of $\lambda^{-1}A$, where λ stands for the unique dominant eigenvalue of A, is not exactly 1 but a very close number greater than 1 (e.g., $1 + 10^{-p}$, with p positive integer). This may bring some problems when working in finite precision arithmetic. It is worth noticing that the convergence of alg-dr may also fail if the assumptions of [6, Theorem 3.1] are not met. In contrast, alg-1 performs very well in terms of convergence for both diagonalizable and non-diagonalizable matrices. Moreover, there is some freedom in choosing the initial guess X_0. In all the experiments reported in Table 1, we have considered $X_0 = A$, but other options for X_0 are valid, provided that they commute with A. For instance, alg-1 applied to A_4 with $X_0 = A_4/5$ requires just 6 iterations instead of 11, as reported for $X_0 = A$. The results for alg-2 were obtained for $X_0 = \alpha I$, with α satisfying (10). It can be observed that the convergence of this algorithm is faster, but for matrices A_1 and A_3 it converges, respectively, to A_1 and A_3, which are trivial solutions of (1). Due to the strong condition $\|AY_0A - Y_0AY_0\| < 1$ and the commuting requirement (check Theorem 5), in some tests it may not be easy to find an $X_0 \neq \alpha I$ yielding convergence towards a nontrivial solution. More research is needed for finding suitable initial guesses to alg-2.

Table 2
Results for Algorithms 6.1 and 6.2, with two non-diagonalizable matrices (A_7 and A_8) and two matrices (A_9 and A_{10}) with eigenvalues close or equal to 1/2. k is the number of iterations and $\texttt{est}_{\texttt{rel}}$ is the estimation of the relative error by (15).

	alg-1		alg-2	
A_i	k	$\texttt{est}_{\texttt{rel}}$	k	$\texttt{est}_{\texttt{rel}}$
A_7	8	5.5×10^{-17}	5	9.6×10^{-17}
A_8	8	3.7×10^{-17}	6	8.8×10^{-17}
A_9	7	6.7×10^{-17}	5	1.9×10^{-16}
A_{10}	7	1.2×10^{-16}	6	3.5×10^{-16}

7.2. Experiment 2

In this experiment we consider the following four matrices:

- $A_7 = [\texttt{3 -1 1; 7 -5 1; 6 6 -2}]$ (non-diagonalizable);
- $A_8 = [\texttt{3 2 1 0; 0 3 0 0; -1 1 1 0; 0 1 1 3}]$ (non-diagonalizable);
- A_9 is a 7×7 matrix with an eigenvalue equal to 0.4998 which is very close to 1/2;
- A_{10} is a 7×7 matrix with one eigenvalue equal to 1/2.

Our tests (see Table 2) evidence that (4) provides a good alternative to (2) whenever A has at least an eigenvalue equal to 1/2 or very close. Both alg-1 and alg-2 give good results with non-diagonalizable matrices, despite with A_8 where both converge to the trivial solution $X = A_8$. However, if we take $X_0 = A_8/5$ in alg-1, the convergence is towards a nontrivial solution. Since A_9 (resp., A_{10}) has an eigenvalue close (resp., equal) to 1/2, in Table 2 we choose $X_0 = 6A/5$ (resp., $X_0 = 3A/2$) to avoid convergence problems.

As a final remark on the experiments, it becomes evident that alg-1 is the one we recommend to find nontrivial solutions of the Yang–Baxter-like matrix equation (1). It performs very well in the experiments, is stable, has quadratic convergence and there are many possibilities for choosing a starting approximation X_0. It is only required that X_0 commutes with A and has no eigenvalue with real part 1/2.

8. Conclusions

In this paper, we have proposed two iterative schemes to find nontrivial commuting solutions of the Yang–Baxter-like matrix equation. They can be used for either diagonalizable or non-diagonalizable matrices. The numerical features of these two iterations have been investigated in detail, both having quadratic convergence. We have provided techniques for finding suitable starting approximations for both algorithms and an estimative for the relative error of the computed solutions. A comparison between our methods and a previous method valid only for diagonalizable matrices has been included, showing the superiority of our algorithms.

Acknowledgments

The work of the João R. Cardoso was supported by ISR-University of Coimbra (project UID/EEA/00048/2013) funded by “Fundação para a Ciência e a Tecnologia” (FCT).

References

[1] A. Björck, S. Hammarling, A Schur method for the square root of a matrix, Linear Alg. Appl. 52 (53) (1983) 127–140.
[2] A. Cayley, The Newton–Fourier imaginary problem, Am. J. Math. 2 (1) (1879) 97.
[3] A. Cayley, Application of the Newton–Fourier method to an imaginary root of an equation, Q. J. Pure Appl. Math. XVI (1879) 179–185.
[4] J. Ding, N.H. Rhee, A nontrivial solution to a stochastic matrix equation, East Asian J. Appl. Math. 2 (4) (2012) 277–284.
[5] J. Ding, N. Rhee, Spectral solutions of the Yang–Baxter matrix equation, J. Math. Anal. Appl. 402 (2013) 567–573.
[6] J. Ding, N.H. Rhee, Computing solutions of the Yang–Baxter-like matrix equation for diagonalisable matrices, East Asian J. Appl. Math. 5 (1) (2015) 75–84.
[7] Q. Dong, J. Ding, Complete commuting solutions of the Yang–Baxter-like matrix equation for diagonalizable matrices, Comput. Math. Appl. 72 (2016) 194–201.
[8] N.J. Higham, Computing real square roots of a real matrix, Linear Alg. Appl. 88 (89) (1987) 405–430.
[9] N.J. Higham, Functions of Matrices: Theory and Computation, Society for Industrial and Applied Mathematics, Philadelphia, 2008.
[10] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, Cambridge, 1994. Paperback edition.
[11] R.A. Horn, C.R. Johnson, Matrix Analysis, 2nd ed., Cambridge University Press, 2013.
[12] A.S. Householder, The Numerical Treatment of a Single Nonlinear Equation, McGraw-Hill, New York, 1970.
[13] B. Iannazzo, A family of rational iterations and its applications to the computation of the matrix *p*th root, SIAM J. Matrix Anal. Appl. 30 (2008) 1445–1462.
[14] H. Lütkepohl, Handbook of Matrices, John Wiley and Sons, 1996.
[15] J.F. Traub, Iterative Methods for Solution of Equations, Prentice-Hall, Englewood Cliffs, NJ, 1964.