ISSN: 1017-060X (Print)

ISSN: 1735-8515 (Online)

# Bulletin of the Iranian Mathematical Society

Vol. 41 (2015), No. 4, pp. 981–1001

Title:

## Convergence analysis of the global FOM and GMRES methods for solving matrix equations $AXB = C$ with SPD coefficients

Author(s):

M. Mohseni Moghadam, A. Rivaz, A. Tajaddini and F. Saberi Movahed

Published by Iranian Mathematical Society
http://bims.ims.ir

Bull. Iranian Math. Soc.
Vol. 41 (2015), No. 4, pp. 981–1001
Online ISSN: 1735-8515

# CONVERGENCE ANALYSIS OF THE GLOBAL FOM AND GMRES METHODS FOR SOLVING MATRIX EQUATIONS $AXB = C$ WITH SPD COEFFICIENTS

M. MOHSENI MOGHADAM*, A. RIVAZ, A. TAJADDINI AND F. SABERI MOVAHED

(Communicated by Davod Khojasteh Salkuyeh)

Abstract. In this paper, we study convergence behavior of the global FOM (Gl-FOM) and global GMRES (Gl-GMRES) methods for solving the matrix equation $AXB = C$ where $A$ and $B$ are symmetric positive definite (SPD). We present some new theoretical results of these methods such as computable exact expressions and upper bounds for the norm of the error and residual. In particular, the obtained upper bounds for the Gl-FOM method help us to predict the behavior of the Frobenius norm of the Gl-FOM residual. We also explore the worst-case convergence behavior of these methods. Finally, some numerical experiments are given to show the performance of the theoretical results.
**Keywords:** Convergence analysis; Global FOM; Global GMRES; Global Lanczos algorithm; Worst-case behavior.
**MSC(2010):** Primary: 65F10; Secondary: 15A24.

## 1. Introduction

In this paper, we consider the matrix equation

$$AXB = C, \tag{1.1}$$

where $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{s\times s}$, $C \in \mathbb{R}^{n\times s}$ and the matrices $A$ and $B$ are symmetric positive definite (SPD).

Different methods are devoted to find the special solution structures of the matrix equation $AXB = C$ such as diagonal, triangular, reflexive, symmetric, centro-symmetric, skew-symmetric or least squares solution $X$; see [7,8,13,18] and the references therein.

During the last years, several projection methods have been proposed to solve the matrix equations. The main idea developed in these methods is to

Article electronically published on August 16, 2015.
Received: 18 May 2014, Accepted: 27 June 2014.
*Corresponding author.

construct suitable bases of the Krylov subspace and project the large problem into a smaller one; see [9–11, 20, 21] for more details. For examples, Beik and Salkuyeh [2] have proposed two global Krylov subspace methods for solving the following general coupled linear matrix equations

$$\sum_{j=1}^{p} A_{ij} X_j B_{ij} = C_i, \quad i = 1, \ldots, p. \tag{1.2}$$

Furthermore, using the Schur complement formula, Beik [1] has studied some convergence results of the Gl-GMRES for solving the generalized Sylvester equations and the matrix equations (1.2).

In the case where $B$ is the identity matrix $I_{s\times s}$, Jbilou et al. [14] have presented the properties of the Gl-FOM and Gl-GMRES methods for solving a matrix equation $AX = C$. Then, using the $\diamond$ product, Bouyouli et al. [4] have studied the convergence properties of the Gl-FOM and Gl-GMRES methods. When $A$ is a SPD matrix and $B = I_{s\times s}$ with $s = 1$, the Conjugate Gradient (CG) and the MINRES methods are two well-known Krylov subspace approaches for solving the linear system $Ax = b$. An overview of the convergence analysis of these methods is given in [17]. Also, an interesting work was recently done by Bouyouli et al. [5]. Based on the relationship between the CG and Lanczos methods, they have derived some convergence results for the error and the residual of the CG method.

In this paper, we discuss convergence analysis of the Gl-FOM and Gl-GMRES methods for solving the matrix equation (1.1). Our main interest is to understand the behavior of the error and residual of these methods. These methods use a global Lanczos-based algorithm onto a matrix Krylov subspace. In addition, since these methods are the global orthogonal residual and global minimal residual methods, we call them the G-OR-L and G-MR-L methods, respectively.

This paper is organized as follows. In section 2, we give some preliminary tools. We also recall the matrix Krylov subspaces, then we give a brief description of the global Arnoldi algorithm and its properties. In section 3, we present the G-OR-L and G-MR-L methods and their mathematical properties for solving the matrix equation (1.1). We also prove some convergence results of these methods. In particular, we discuss the importance of the spectral information of $A$ and $B$ on the convergence behavior of the Frobenius norm of the G-OR-L residual. In section 4, we investigate the worst-case convergence behavior of these methods. For the special case $AX = C$ where $A$ is a diagonalizable matrix, we also prove the worst-case convergence behavior of the Gl-GMRES method. Furthermore, in comparison with the proof given in [3], our proof is shorter. Numerical experiments are presented in the last section.

## 2. **Preliminaries**

We use the following notations. The notation $\mathbb{R}^{n\times s}$ denotes the real vector space of $n\times s$ matrices. $e_j$ denotes the $j$th column of an identity matrix of a suitable dimension. Let $X, Y \in \mathbb{R}^{n\times s}$. The vector $\mathrm{vec}(X)$ denotes the vector of $\mathbb{R}^{ns}$ defined by $\mathrm{vec}(X) = [x_1^T, \ldots, x_s^T]^T$ where $x_i$ is the $i$th column of $X$. The Frobenius inner product is defined by $\langle X, Y\rangle_F = \mathrm{trace}(Y^TX)$ where $\mathrm{trace}(\cdot)$ denotes the trace of the square matrix $Y^TX$. The associated norm is the Frobenius norm denoted by $\|\cdot\|_F$. The notation $X\perp_F Y$ means that $\langle X, Y\rangle_F = 0$. The Kronecker product of matrices $A \in \mathbb{R}^{k\times l}$ and $B \in \mathbb{R}^{n\times m}$ is defined by $A\otimes B = [a_{ij}B]$. For this product, we have the following properties. For more details we refer to [4].

(1) $(A\otimes B)(C\otimes D) = (AC\otimes BD)$, provided that $AC$ and $BD$ exist.
(2) $(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$, provided that $A^{-1}$ and $B^{-1}$ exist.
(3) $(A\otimes B)^T = A^T\otimes B^T$.
(4) $\mathrm{vec}(ABC) = (C^T\otimes A)\mathrm{vec}(B)$, provided that $ABC$ exists.
(5) If $A \in \mathbb{R}^{k\times l}$, $B \in \mathbb{R}^{l\times m}$ and $C \in \mathbb{R}^{m\times k}$, then

$$\mathrm{trace}(ABC) = \mathrm{vec}(A^T)^T(I_k\otimes B)\mathrm{vec}(C).$$

We give the following proposition whose proof is straightforward.

**Proposition 2.1.** *Assume that $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{s\times s}$ are SPD. The map $\langle\cdot,\cdot\rangle_{(A,B)} : \mathbb{R}^{n\times s}\times\mathbb{R}^{n\times s} \to \mathbb{R}$ defined as $\langle X, Y\rangle_{(A,B)} = \mathrm{trace}(Y^TAXB)$ is an inner product denoted by the $(A, B)$-inner product. Moreover, the associated norm of the $(A, B)$-inner product is*

$$\|X\|_{(A,B)} = \|\mathrm{vec}(X)\|_{B\otimes A}, \qquad \forall X \in \mathbb{R}^{n\times s}.$$

Let $E = [E_1, E_2, \ldots, E_p] \in \mathbb{R}^{n\times ps}$ and $F = [F_1, F_2, \ldots, F_l] \in \mathbb{R}^{n\times ls}$, where $E_i$ and $F_j$ are $n\times s$ matrices. The matrices $E^T\diamond F$ and $E^T\diamond_{(A,B)} F$ are defined by $(E^T\diamond F)_{ij} = \langle E_i, F_j\rangle_F$ and $(E^T\diamond_{(A,B)} F)_{ij} = \langle E_i, F_j\rangle_{(A,B)}$, respectively. One can see that the $\diamond$ product satisfies the following properties; see [4] for more details.

**Proposition 2.2.** *Let $E, F, M \in \mathbb{R}^{n\times ps}$, $L \in \mathbb{R}^{p\times p}$, $K \in \mathbb{R}^{s\times s}$, $y \in \mathbb{R}^s$ and $\alpha \in \mathbb{R}$. Then we have*

(1) $E^T\diamond(F(L\otimes I_s)(I_p\otimes K)) = (E^T\diamond(F(I_p\otimes K)))L$.
(2) $E^T\diamond(F(I_s\otimes L)(y\otimes I_p)) = (E^T\diamond(F(I_s\otimes L)))y$.
(3) $E^T\diamond(AE(I_p\otimes B)) = E^T\diamond_{(A,B)} E = \tilde{E}^T(B\otimes A)\tilde{E}$,

*where* $\tilde{E} = [\mathrm{vec}(E_1), \mathrm{vec}(E_2), \ldots, \mathrm{vec}(E_p)]$.

2.1. **The generalized matrix Krylov subspace.** Let $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{s\times s}$ be arbitrary matrices. The generalized matrix Krylov subspace associated with the triplet $(A, V, B)$ where $V \in \mathbb{R}^{n\times s}$ is defined by:

$$\mathcal{GK}_i(A, V, B) = \operatorname{span}\{V, AVB, \ldots, A^{i-1}VB^{i-1}\}. \tag{2.1}$$

The following definition of [16, p. 411] is a basic tool to describe the structure of $\mathcal{GK}_i(A, V, B)$.

**Definition 2.3.** *Let $p(x, y) = \sum_{i,j=0}^{k} c_{ij}x^i y^j$ be a polynomial in two variables, $x$ and $y$, with real coefficients $c_{ij}$. $p(C:D)$ is to be defined a matrix of the form $p(C:D) = \sum_{i,j=0}^{k} c_{ij}(C^i \otimes D^j)$ where $C \in \mathbb{R}^{m\times m}$ and $D \in \mathbb{R}^{n\times n}$.*

**Remark 2.4.** *Suppose that $\mathcal{K}_i(G, v) = span\{v, Gv, \ldots, G^{i-1}v\}$ is the classical matrix Krylov subspace where $v = \operatorname{vec}(V)$ and $G = B^T \otimes A$. The map $T : \mathcal{GK}_i(A, V, B) \to \mathcal{K}_i(G, v)$ given by $Z \mapsto \operatorname{vec}(Z)$ is an isomorphism. So, $\mathcal{GK}_i(A, V, B)$ is the set of all matrices $Z$ such that $\operatorname{vec}(Z) = p_k(B^T : A)v$ in which $p_k(x, y) = \sum_{j=0}^{k} c_j x^j y^j$ and $k \le i - 1$.*

To construct an orthonormal basis of $\mathcal{GK}_k(A, V, B)$, we apply the global Arnoldi algorithm which can be summarized in Algorithm 1.

---

Set $\beta = \|V\|_F$ and $V_1 = V/\beta$.
**for** $j = 1, \ldots, k$ **do**
  $W_j = AV_jB$
  **for** $i = 1, \ldots, j$ **do**
    $h_{ij} = \langle W_j, V_i\rangle_F$
    $W_j = W_j - h_{ij}V_i$
  **end for**
  $h_{j+1,j} = \|W_j\|_F$. If $h_{j+1,j} = 0$ then stop.
  $V_{j+1} = W_j/h_{j+1,j}$.
**end for**

---

Let $\mathcal{V}_k = [V_1, \ldots, V_k]$. It is easy to see that $\mathcal{V}_k^T \diamond \mathcal{V}_k = I_k$ and $\mathcal{V}_k^T \diamond V_{k+1} = 0_{k\times 1}$. In addition, we have

$$A\mathcal{V}_k(I_k \otimes B) = [AV_1B, \ldots, AV_kB].$$

From Algorithm 1, it follows that $AV_jB = \sum_{i=1}^{j+1} h_{ij}V_i$, for $j = 1, \ldots, k$. Hence, we get

$$A\mathcal{V}_k(I_k \otimes B) = [\sum_{i=1}^{2} h_{i1}V_i, \ldots, \sum_{i=1}^{k} h_{im}V_i] + h_{k+1,k}[0, \ldots, V_{k+1}].$$

So, it is clear that

$$A\mathcal{V}_k(I_k \otimes B) = \mathcal{V}_k(H_k \otimes I_s) + h_{k+1,k}V_{k+1}(e_k^T \otimes I_s), \tag{2.2}$$

where $H_k = [h_{ij}]$ is an upper Hessenberg matrix whose nonzero entries $h_{ij}$ are defined by the generalized global Arnoldi algorithm.

Using the relation (2.2), the part (3) of Proposition 2.2 implies that $H_k = \tilde{\mathcal{V}}_k^T(B \otimes A)\tilde{\mathcal{V}}_k$, where $\tilde{\mathcal{V}}_k = [\mathrm{vec}(V_1), \dots, \mathrm{vec}(V_k)]$. In the particular case where $A, B$ are SPD, we state an important property of the matrix $H_k$, which is easy to prove.

**Proposition 2.5.** *If $A, B$ are SPD, then the matrix $H_k$ is SPD and tridiagonal. Moreover, all entries of the three diagonals of $H_k$ are positive.*

As a consequence of the above proposition, when $A$ and $B$ are SPD, then the matrices $V_1, V_2, \dots, V_k$ constructed by Algorithm 1 satisfy the three-term recurrence $h_{j+1,j}V_{j+1} = AV_jB - h_{jj}V_j - h_{j-1,j}V_{j-1}$. Therefore, the generalized global Arnoldi algorithm is transformed to the following algorithm. In fact, this algorithm is a version of the global Lanczos algorithm. We call this algorithm the generalized global Lanczos (GGL) algorithm.

**Algorithm 1** The generalized global Lanczos (GGL) algorithm

Set $\beta = \|V\|_F$, $V_1 = V/\beta, h_{0,1} \equiv 0$ and $V_0 \equiv 0$.
**for** $j = 1, \dots, k$ **do**
  $W_j = AV_jB - h_{j-1,j}V_{j-1}$
  $h_{jj} = \langle W_j, V_j \rangle_F$
  $W_j = W_j - h_{jj}V_j$
  $h_{j+1,j} = \|W_j\|_F$. If $h_{j+1,j} = 0$ then stop.
  $V_{j+1} = W_j/h_{j+1,j}$.
**end for**

## 3. Convergence analysis

In this section, we present some convergence results for the Gl-FOM and Gl-GMRES methods for solving the matrix equation (1.1). Both of these methods use the GGL algorithm to construct an orthonormal basis of a generalized matrix Krylov subspace. In addition, since these methods are the global orthogonal residual and global minimal residual methods, we call them the G-OR-L and G-MR-L methods, respectively.

### 3.1. The G-OR-L method.

Assume that $X_0 \in \mathbb{R}^{n \times s}$ is an initial guess and $R_0 = C - AX_0B$ is its corresponding residual. The G-OR-L method consists in generating approximate solutions of the form $X_m^{or} = X_0 + Z_m$ with $Z_m \in \mathcal{GK}_m(A, R_0, B)$ such that

$$R_m^{or} := C - AX_m^{or}B \perp_F \mathcal{GK}_m(A, R_0, B), \qquad (3.1)$$

where $m = 1, 2, \dots$. The matrix $R_m^{or}$ is called the $m$th residual associated with $X_m^{or}$.

Let $\{V_1, \ldots, V_m\}$ be the basis of $\mathcal{GK}_m(A, R_0, B)$ constructed by the GGL algorithm. The approximation $X_m^{or}$ can be written as

$$X_m^{or} = X_0 + \mathcal{V}_m(\alpha_m^{or} \otimes I_s), \tag{3.2}$$

where $\mathcal{V}_m = [V_1, \ldots, V_m]$. It is also easy to see that $\alpha_m^{or}$ solves the linear system

$$H_m \alpha_m^{or} = \|R_0\|_F e_1. \tag{3.3}$$

Moreover, as $H_m$ is SPD, $\alpha_m^{or}$ is the unique solution of (3.3).

Now, we show how to compute the Frobenius norm of the residual without computing explicitly the residual.

**Proposition 3.1.** *Under the same assumptions as in Proposition 2.5, the residual matrix* $R_m^{or}$ *satisfies the following relation*

$$\|R_m^{or}\|_F = \|R_0\|_F \frac{\prod_{i=1}^{m} h_{i+1,i}}{\det(H_m)}. \tag{3.4}$$

*So,* $\|R_m^{or}\|_F = 0$ *if and only if* $h_{m+1,m} = 0$.

*Proof.* We have $R_m^{or} = R_0 - A\mathcal{V}_m(I_m \otimes B)(\alpha_m^{or} \otimes I_s)$. So, using the relations (2.2) and (3.3), we obtain

$$\begin{aligned} R_m^{or} &= R_0 - \mathcal{V}_m(H_m\alpha_m^{or} \otimes I_s) - h_{m+1,m}(e_m^T \alpha_m^{or})V_{m+1} \\ &= -h_{m+1,m}(e_m^T \alpha_m^{or})V_{m+1}. \end{aligned}$$

Therefore, $\|R_m^{or}\|_F = h_{m+1,m}|\alpha_m^{or(m)}|$ where $\alpha_m^{or(m)}$ is the last component of $\alpha_m^{or}$. On the other hand, from (3.3) and the Cramer rule, it follows that

$$\alpha_m^{or(m)} = (-1)^{m+1}\|R_0\|_F \frac{\prod_{i=1}^{m-1} h_{i+1,i}}{\det(H_m)}.$$

Using Proposition 2.5, we have $h_{i+1,i} > 0, i = 1, \ldots, m-1$. Hence $\|R_m^{or}\|_F = 0$ if and only if $h_{m+1,m} = 0$. □

We study further the G-OR-L method by considering the properties of the error associated with $X_m^{or}$, i.e., $X_* - X_m^{or}$ where $X_*$ is the solution of the matrix equation (1.1).

Since the G-OR-L method is an orthogonal projection method onto the generalized matrix Krylov subspace $\mathcal{GK}_i(A, R_0, B)$, we have the minimization property of the error of the G-OR-L method. It is not hard to prove the following theorem.

**Theorem 3.2.** *Assume that* $X_*$ *is the solution of* (1.1) *and* $X_0 \in \mathbb{R}^{n\times s}$. *Then* $X_m^{or}$ *is the mth approximation of the G-OR-L method if and only if*

$$\|X_* - X_m^{or}\|_{(A,B)} = \min_{X \in X_0 + \mathcal{GK}_m(A,R_0,B)} \|X_* - X\|_{(A,B)}.$$

By considering the minimization property of the $(A,B)-$norm of the error of the G-OR-L method, we will try to establish more properties of the error. First, we give two expressions for $\|X_* - X_m^{or}\|_{(A,B)}$. In the following, we assume that $D_0 = X_* - X_0$ and $D_m = X_* - X_m^{or}$.

**Theorem 3.3.** *The error $X_* - X_m^{or}$ associated with $X_m^{or}$ satisfies the following relation*

$$\|X_* - X_m^{or}\|_{(A,B)} = \|R_m^{or}\|_F \sqrt{V_{m+1}^T \diamond_{(A^{-1},B^{-1})} V_{m+1}}. \tag{16-i}$$

*Also, suppose that $K_{m+1} = [R_0, \ldots, A^m R_0 B^m]$. If $\{R_0, \ldots, A^m R_0 B^m\}$ is a linearly independent set, then*

$$\|X_* - X_m^{or}\|_{(A,B)}^2 = \frac{1}{e_1^T (K_{m+1}^T \diamond_{(A^{-1},B^{-1})} K_{m+1})^{-1} e_1}. \tag{16-ii}$$

*Proof.* **(i)** Since $D_m = A^{-1} R_m^{or} B^{-1}$, we have

$$\begin{aligned}
\|X_* - X_m^{or}\|_{(A,B)}^2 &= \mathrm{vec}(D_m)^T (B \otimes A) \mathrm{vec}(D_m) \\
&= \mathrm{vec}(R_m^{or})^T (B^{-1} \otimes A^{-1}) \mathrm{vec}(R_m^{or}) \\
&= \|R_m^{or}\|_F^2 \mathrm{vec}(V_{m+1})^T (B^{-1} \otimes A^{-1}) \mathrm{vec}(V_{m+1}) \\
&= \|R_m^{or}\|_F^2 (V_{m+1}^T \diamond_{(A^{-1},B^{-1})} V_{m+1}).
\end{aligned}$$

**(ii)** We have $X_m^{or} = X_0 + K_m(\alpha_m^{or} \otimes I_s)$ such that $R_m^{or} = R_0 - AK_m(I_m \otimes B)(\alpha_m^{or} \otimes I_s)$, where $K_m = [R_0, \ldots, A^{m-1} R_0 B^{m-1}]$. So,

$$\begin{aligned}
\|X_* - X_m^{or}\|_{(A,B)}^2 &= \mathrm{trace}\left((X_* - X_m^{or})^T A (X_* - X_m^{or}) B\right) \\
&= \langle D_0 - K_m(\alpha_m^{or} \otimes I_s), R_m^{or} \rangle_F \\
&= \langle D_0, R_m^{or} \rangle_F \\
&= \langle D_0, R_0 - AK_m(I_m \otimes B)(\alpha_m^{or} \otimes I_s) \rangle_F.
\end{aligned}$$

From [12, Theorem 7.2.10], the matrix $K_m^T \diamond_{(A,B)} K_m$ is nonsingular. Now, the orthogonality relation (3.1) yields $\alpha_m^{or} = (K_m^T \diamond_{(A,B)} K_m)^{-1} (K_m^T \diamond R_0)$. In addition, we have $R_0 = AD_0B$. Therefore,

$$\begin{aligned}
\|X_* - X_m^{or}\|_{(A,B)}^2 &= D_0^T \diamond_{(A,B)} D_0 - (D_0^T \diamond_{(A,B)} K_m) \alpha_m^{or} \\
&= D_0^T \diamond_{(A,B)} D_0 - \\
&\quad (D_0^T \diamond_{(A,B)} K_m)(K_m^T \diamond_{(A,B)} K_m)^{-1} (K_m^T \diamond_{(A,B)} D_0) \\
&= \left( \begin{bmatrix} D_0^T \diamond_{(A,B)} D_0 & D_0^T \diamond_{(A,B)} K_m \\ K_m^T \diamond_{(A,B)} D_0 & K_m^T \diamond_{(A,B)} K_m \end{bmatrix} \Big/ K_m^T \diamond_{(A,B)} K_m \right),
\end{aligned}$$

where $M/F$ is the Schur complement of $F$ in $M$.
As $K_{m+1} = [AD_0B, AK_m(I_m \otimes B)]$ and $A^{-1} K_{m+1} (I_{m+1} \otimes B^{-1}) = [D_0, K_m]$,

we get

$$\left(A^{-1}K_{m+1}(I_{m+1}\otimes B^{-1})\right)^T \diamond K_{m+1} = \begin{bmatrix} D_0^T \diamond_{(A,B)} D_0 & D_0^T \diamond_{(A,B)} K_m \\ K_m^T \diamond_{(A,B)} D_0 & K_m^T \diamond_{(A,B)} K_m \end{bmatrix}.$$

On the other hand

$$\left(A^{-1}K_{m+1}(I_{m+1}\otimes B^{-1})\right)^T \diamond K_{m+1} = K_{m+1}^T \diamond_{(A^{-1},B^{-1})} K_{m+1},$$

so we obtain

$$\begin{aligned}\|X_* - X_m^{or}\|_{(A,B)}^2 &= \left(K_{m+1}^T \diamond_{(A^{-1},B^{-1})} K_{m+1} \Big/ K_m^T \diamond_{(A,B)} K_m\right) \\ &= \frac{\det(K_{m+1}^T \diamond_{(A^{-1},B^{-1})} K_{m+1})}{\det(K_m^T \diamond_{(A,B)} K_m)} \\ &= \frac{1}{e_1^T(K_{m+1}^T \diamond_{(A^{-1},B^{-1})} K_{m+1})^{-1}e_1}. \qquad \square\end{aligned}$$

**Remark 3.4.**
(1) *Proposition 3.1 together with the relation* (16-i) *follows that* $h_{m+1,m} = 0$ *if and only if* $\|X_* - X_m^{or}\|_{(A,B)} = 0$.
(2) *From [12, Theorem 7.2.10], the matrix* $K_{m+1}^T \diamond_{(A^{-1},B^{-1})} K_{m+1}$ *is SPD, so* $e_1^T(K_{m+1}^T \diamond_{(A^{-1},B^{-1})} K_{m+1})^{-1}e_1 > 0$.

Now, we introduce some upper bounds for $\|X_* - X_m^{or}\|_{(A,B)}$. First, we state the following relations:

$$\|(B\otimes A)^{-1}\|_2 = \frac{1}{\lambda_{min}(B\otimes A)} = \frac{1}{\lambda_{min}(A)\lambda_{min}(B)} = \|A^{-1}\|_2\|B^{-1}\|_2,$$
$$\|B\otimes A\|_2 = \lambda_{max}(B\otimes A) = \lambda_{max}(A)\lambda_{max}(B) = \|A\|_2\|B\|_2.$$

Therefore, $\kappa(B\otimes A) = \kappa(A)\kappa(B)$, where $\kappa(Z) = \|Z\|_2\|Z^{-1}\|_2$. In the following, the expression "UB" is used to denote "Upper Bound".

**Theorem 3.5.** *The error* $X_* - X_m^{or}$ *associated with* $X_m^{or}$ *satisfies the relations*

$$\text{(UB.1)} \qquad \|X_* - X_m^{or}\|_{(A,B)} \le \|R_m^{or}\|_F \sqrt{\frac{\kappa(A)\kappa(B)}{\|A\|_2\|B\|_2}}.$$

$$\text{(UB.2)} \qquad \|X_* - X_m^{or}\|_{(A,B)} \le \frac{\|R_m^{or}\|_F}{2\sqrt{V_{m+1}^T \diamond_{(A,B)} V_{m+1}}} \frac{\kappa(A)\kappa(B)+1}{\sqrt{\kappa(A)\kappa(B)}}.$$

$$\text{(UB.3)} \qquad \|X_* - X_m^{or}\|_{(A,B)} \le \frac{\|R_m^{or}\|_F}{2}\Big(\frac{\kappa(A)\kappa(B)+1}{\sqrt{\|A\|_2\|B\|_2}}\Big).$$

$$\text{(UB.4)} \qquad \|X_* - X_m^{or}\|_{(A,B)} \le \gamma\sqrt{\|R_m^{or}\|_F},$$

*where* $\gamma = \sqrt{\|R_0\|_F \frac{\kappa(A)\kappa(B)}{\|A\|_2\|B\|_2} + \|\alpha_m^{or}\|_2}$.

*Proof.* **(1)**: Let $G = B \otimes A$ and $v = \text{vec}(V_{m+1})$. If we apply the Courant-Fisher theorem [12, Theorem 4.2.11], then

$$v^T G^{-1} v \leq \frac{1}{\lambda_{min}(A)\lambda_{min}(B)} = \frac{\kappa(A)\kappa(B)}{\|A\|_2\|B\|_2}.$$

The relation (16-i) together with the above relation follows the inequality (UB.1).
**(2)**: Since $v^T G^{-1} v = V_{m+1}^T \diamond_{(A^{-1},B^{-1})} V_{m+1}$ and $v^T G v = V_{m+1}^T \diamond_{(A,B)} V_{m+1}$, the Kantorovich inequality [12, Theorem 7.4.41] implies that

$$V_{m+1}^T \diamond_{(A^{-1},B^{-1})} V_{m+1} \leq \frac{1}{4(V_{m+1}^T \diamond_{(A,B)} V_{m+1})} \frac{(\kappa(A)\kappa(B)+1)^2}{\kappa(A)\kappa(B)}. \tag{3.5}$$

Multiplying both sides of the relation (3.5) by $\|R_m^{or}\|_F^2$ follows the inequality (UB.2). On the other hand, using the Courant-Fisher theorem, we obtain

$$\frac{1}{v^T G v} \leq \frac{1}{\lambda_{min}(G)} = \frac{\kappa(A)\kappa(B)}{\|A\|_2\|B\|_2},$$

which follows the inequality (UB.3).
**(3)**: We have

$$\|X_* - X_m^{or}\|_{(A,B)}^2 = \langle G\text{vec}(X_* - X_m^{or}), \text{vec}(X_* - X_m^{or})\rangle, \tag{3.6}$$

and

$$\begin{aligned} \|C - AX_m^{or}B\|_F^2 &= \|A(X_* - X_m^{or})B\|_F^2 \\ &= \|\text{vec}(A(X_* - X_m^{or})B)\|_2^2 \\ &= \langle G^2\text{vec}(X_* - X_m^{or}), \text{vec}(X_* - X_m^{or})\rangle. \end{aligned} \tag{3.7}$$

Let $y = \text{vec}(X_* - X_m^{or})$ and $z = y/\|y\|_2$. Using a result given in [15], we get $\langle Gz, z\rangle^2 \leq \langle G^2 z, z\rangle$, or equivalently, $\langle Gy, y\rangle^2 \leq \|y\|_2^2\langle G^2 y, y\rangle$.
Since $\|\text{vec}(X_* - X_m^{or})\|_2 = \|X_* - X_m^{or}\|_F$, it follows from (3.6) and (3.7) that

$$\|X_* - X_m^{or}\|_{(A,B)} \leq \sqrt{\|X_* - X_m^{or}\|_F}\sqrt{\|C - AX_m^{or}B\|_F}.$$

From the relation $X_* - X_m^{or} = D_0 - \mathcal{V}_m(\alpha_m^{or} \otimes I_s)$, we have

$$\|X_* - X_m^{or}\|_F \leq \|D_0\|_F + \|\mathcal{V}_m(\alpha_m^{or} \otimes I_s)\|_F.$$

Therefore, as $\|\mathcal{V}_m(\alpha_m^{or} \otimes I_s)\|_F = \|\alpha_m^{or}\|_2$ and

$$\|D_0\|_F = \|\text{vec}(D_0)\|_2 = \|(B \otimes A)^{-1}\text{vec}(R_0)\|_2 \leq \|A^{-1}\|_2\|B^{-1}\|_2\|\text{vec}(R_0)\|_2,$$

the proof completes. □

It can be shown that the $(A, B)-$norm of the error at each step of the G-OR-L method satisfies $\|X_* - X_m^{or}\|_{(A,B)} \leq \|X_* - X_0\|_{(A,B)}$. However, the Frobenius norm of the residual at each step of the G-OR-L method may oscillate. In the next result, we show an upper bound for $\|R_m^{or}\|_F$.

**Theorem 3.6.** *Assume that $R_m^{or}$ is the G-OR-L residual obtained at step $m$. Then*

$$\|R_m^{or}\|_F \leq \min\{\alpha_1\|R_0\|_F, \alpha_2\|R_0\|_F\}, \tag{3.8}$$

*where* $\alpha_1 = \sqrt{\kappa(A)\kappa(B)}$ *and* $\alpha_2 = \sqrt{\frac{\|A\|_2\|B\|_2}{(V_1^T \diamond_{(A,B)} V_1)\kappa(A)\kappa(B)}}\left(\frac{\kappa(A)\kappa(B)+1}{2}\right)$.

*Proof.* The Courant-Fisher theorem together with the relation (16-i) implies that

$$\frac{\|R_m^{or}\|_F^2}{\|A\|_2\|B\|_2} \leq \|X_* - X_m^{or}\|_{(A,B)}^2.$$

On the other hand, similar to Theorem 3.5, we obtain

$$\|X_* - X_0\|_{(A,B)} \leq \|R_0\|_F\sqrt{\frac{\kappa(A)\kappa(B)}{\|A\|_2\|B\|_2}},$$

$$\|X_* - X_0\|_{(A,B)} \leq \frac{\|R_0\|_F}{2\sqrt{V_1^T \diamond_{(A,B)} V_1}}\frac{\kappa(A)\kappa(B)+1}{\sqrt{\kappa(A)\kappa(B)}}.$$

Therefore, using the above relations, we have

$$\|X_* - X_0\|_{(A,B)}^2 - \|X_* - X_m^{or}\|_{(A,B)}^2 \leq \frac{\kappa(A)\kappa(B)}{\|A\|_2\|B\|_2}\|R_0\|_F^2 - \frac{1}{\|A\|_2\|B\|_2}\|R_m^{or}\|_F^2,$$

and,

$$\|X_* - X_0\|_{(A,B)}^2 - \|X_* - X_m^{or}\|_{(A,B)}^2 \leq \frac{(\kappa(A)\kappa(B)+1)^2\|R_0\|_F^2}{4(V_1^T \diamond_{(A,B)} V_1)\kappa(A)\kappa(B)} - \frac{\|R_m^{or}\|_F^2}{\|A\|_2\|B\|_2}.$$

Now, since $\|X_* - X_0\|_{(A,B)}^2 - \|X_* - X_m^{or}\|_{(A,B)}^2 \geq 0$, the results follow. □

**Remark 3.7.** *Since $\|R_m^{or}\|_F \leq \min\{\alpha_1\|R_0\|_F, \alpha_2\|R_0\|_F\}$, we expect that the Frobenius norm of the residuals will not oscillate too much when $\alpha_1$ and $\alpha_2$ are small. On the other hand, it is easy to see that $\alpha_2 \leq (\kappa(A)\kappa(B)+1)/2$, so we have*

$$\|R_m^{or}\|_F \leq \min\{\sqrt{\kappa(A)\kappa(B)}\|R_0\|_F, \left(\frac{\kappa(A)\kappa(B)+1}{2}\right)\|R_0\|_F\}.$$

*Hence, we conclude that when $\kappa(A) = 1$ and $\kappa(B) = 1$, then $\|R_m^{or}\|_F \leq \|R_0\|_F$. We also consider the especial case where $\kappa(A)$ and $\kappa(B)$ are close to 1. We will come back to these points in the numerical illustrations section.*

3.2. **The G-MR-L method.** Let $X_0 \in \mathbb{R}^{n\times s}$ be an initial guess and $R_0 = C - AX_0B$ be its corresponding residual. The G-MR-L method consists in generating approximate solutions of the form $X_m^{mr} = X_0 + Z_m$ with $Z_m \in \mathcal{GK}_m(A, R_0, B)$ such that

$$R_m^{mr} := C - AX_m^{mr}B \perp_F A\mathcal{GK}_m(A, R_0, B)B, \tag{3.9}$$

where $m = 1, 2, \ldots$. The matrix $R_m^{mr}$ is called the $m$th residual associated with $X_m^{mr}$.

**Proposition 3.8.** *Assume that $\{V_1, \ldots, V_m\}$ is the basis of $\mathcal{GK}_m(A, R_0, B)$ constructed by the GGL algorithm. The approximation $X_m^{mr}$ can be written as*

$$X_m^{mr} = X_0 + \mathcal{V}_m(\alpha_m^{mr} \otimes I_s), \tag{3.10}$$

*where $\alpha_m^{mr}$ is the unique solution of the linear system*

$$(H_m^2 + h_{m+1,m}^2 e_m e_m^T)\alpha_m^{mr} = \|R_0\|_F H_m e_1. \tag{3.11}$$

*Proof.* We have $X_m^{mr} = X_0 + \mathcal{V}_m(\alpha_m^{mr} \otimes I_s)$ such that $R_m^{mr} = R_0 - A\mathcal{V}_m(I_m \otimes B)(\alpha_m^{mr} \otimes I_s)$ where $\alpha_m^{mr} \in \mathbb{R}^m$. Let $\mathcal{W}_m = A\mathcal{V}_m(I_m \otimes B)$. The orthogonality relation (3.9) yields $(\mathcal{W}_m^T \diamond \mathcal{W}_m)\alpha_m^{mr} = \mathcal{W}_m^T \diamond R_0$. Since $R_0 = \|R_0\|_F V_1$, using (2.2) and Proposition 2.2, we obtain

$$\mathcal{W}_m^T \diamond \mathcal{W}_m = H_m^2 + h_{m+1,m}^2 e_m e_m^T, \qquad \text{and} \qquad \mathcal{W}_m^T \diamond R_0 = \|R_0\|_F H_m e_1.$$

As $H_m^2 + h_{m+1,m}^2 e_m e_m^T$ is SPD, the solution of (3.11) exists and is unique. □

**Proposition 3.9.** *The residual matrix $R_m^{mr}$ satisfies the following relation*

$$R_m^{mr} = h_{m+1,m}\alpha_m^{mr(m)}\Big(h_{m+1,m}\mathcal{V}_m(H_m^{-1}e_m \otimes I_s) - V_{m+1}\Big), \tag{3.12}$$

*where $\alpha_m^{mr(m)}$ is the last component of $\alpha_m^{mr}$. In addition, we get*

$$\|R_m^{mr}\|_F = h_{m+1,m}|\alpha_m^{mr(m)}|\sqrt{h_{m+1,m}^2\|H_m^{-1}e_m\|_2^2 + 1}. \tag{3.13}$$

*So, $\|R_m^{mr}\|_F = 0$ if and only if $h_{m+1,m} = 0$.*

*Proof.* We have $R_m^{mr} = R_0 - A\mathcal{V}_m(I_m \otimes B)(\alpha_m^{mr} \otimes I_s)$. Using the relations (2.2) and (3.11), it follows that

$$\begin{aligned} R_m^{mr} &= R_0 - \mathcal{V}_m(H_m\alpha_m^{mr} \otimes I_s) - h_{m+1,m}(e_m^T\alpha_m^{mr})V_{m+1} \\ &= R_0 - \|R_0\|_F\mathcal{V}_m(e_1 \otimes I_s) + h_{m+1,m}^2(e_m^T\alpha_m^{mr})\mathcal{V}_m(H_m^{-1}e_m \otimes I_s) - \\ &\quad h_{m+1,m}(e_m^T\alpha_m^{mr})V_{m+1} \\ &= h_{m+1,m}(e_m^T\alpha_m^{mr})\Big(h_{m+1,m}\mathcal{V}_m(H_m^{-1}e_m \otimes I_s) - V_{m+1}\Big). \end{aligned}$$

Therefore, $\|R_m^{mr}\|_F = \sqrt{(R_m^{mr})^T \diamond R_m^{mr}} = h_{m+1,m}|\alpha_m^{mr(m)}|\sqrt{h_{m+1,m}^2\|H_m^{-1}e_m\|_2^2 + 1}$. From (3.11) and the Cramer rule, we obtain

$$\alpha_m^{mr(m)} = (-1)^{m+1}\|R_0\|_F \frac{\prod_{i=1}^{m-1} h_{i+1,i}}{\det(H_m + h_{m+1,m}^2 H_m^{-1}e_m e_m^T)}.$$

Using Proposition 2.5, we have $h_{i+1,i} > 0, i = 1, \ldots, m-1$. Hence $\|R_m^{mr}\|_F = 0$ if and only if $h_{m+1,m} = 0$. □

The next result shows that the Frobenius norm of the residual has the minimization property.

**Theorem 3.10.** *Let $X_0 \in \mathbb{R}^{n\times s}$ be an initial guess and $R_0 = C - AX_0B$ be its corresponding residual. Then $X_m^{mr}$ is the mth approximation of the G-MR-L method if and only if*

$$\|C - AX_m^{mr}B\|_F = \min_{X\in X_0+\mathcal{GK}_m(A,R_0,B)} \|C - AXB\|_F.$$

*Proof.* The proof is similar to that of Theorem 3.2 and is omitted. □

## 4. **The worst-case behavior of the G-OR-L and G-MR-L methods**

The worst-case convergence behavior of many well known Krylov subspace methods for normal matrices is described by the min-max approximation problem on the discrete set of the matrix eigenvalues [17],

$$\min_{p_m\in\mathbb{P}_m} \max_{\lambda\in\sigma(A)} |p_m(\lambda)|.$$

We investigate the worst-case convergence behavior of the G-OR-L and G-MR-L methods.

### 4.1. **The worst-case behavior of the G-OR-L method.**

First, we prove the following lemma.

**Lemma 4.1.** *Let $X \in X_0 + \mathcal{GK}_m(A, R_0, B)$ where $X_0 \in \mathbb{R}^{n\times s}$. If $X_*$ is the solution of* (1.1), *then* $\mathrm{vec}(X_* - X) = p_m(B : A)\mathrm{vec}(D_0)$ *where $p_m(x, y) = \sum_{i=0}^{m} c_i x^i y^i$ with $p_m(0, 0) = 1$.*

*Proof.* We have $X = X_0 + \sum_{i=1}^{m-1} d_i A^{i-1} R_0 B^{i-1}$ where $d_i \in \mathbb{R}, i = 1, \ldots, m-1$. As $R_0 = AD_0B$, we get $\mathrm{vec}(X_* - X) = \sum_{i=0}^{m} c_i(B \otimes A)^i \mathrm{vec}(D_0)$ such that $c_0 = 1$ and $c_i = -d_i, i = 1, \ldots, m$. Now, if $p_m(x, y) = \sum_{i=0}^{m} c_i x^i y^i$, then the result follows. □

Let $A = V^TDV$ and $B = U^T\Lambda U$ with $V^TV = I_n$ and $U^TU = I_s$, where $D$ and $\Lambda$ are the diagonal matrices whose elements are the eigenvalues $\mu_1, \ldots, \mu_n$ of $A$ and the eigenvalues $\lambda_1, \ldots, \lambda_s$ of $B$, respectively. If $V = [v_1, \ldots, v_n]$ and $U = [u_1, \ldots, u_s]$, then $\mathrm{vec}(D_0)$ can be written as $\mathrm{vec}(D_0) = \sum_{i=1}^{s}\sum_{j=1}^{n} a_{ij}(u_i \otimes v_j)$ where $a_{ij} \in \mathbb{R}$. From [16, Theorem 1, p. 411], we obtain

$$p_m(B : A)\mathrm{vec}(D_0) = \sum_{i=1}^{s}\sum_{j=1}^{n} a_{ij} p_m(\lambda_i, \mu_j)(u_i \otimes v_j). \tag{4.1}$$

**Theorem 4.2.** *Let $X_*$ be the solution of the matrix equation (1.1). If $X_m^{or}$ is the mth approximation of the G-OR-L method, then*

$$\frac{\|X_* - X_m^{or}\|_{(A,B)}}{\|X_* - X_0\|_{(A,B)}} \leq \min_{\substack{p_m \in \mathbb{P}_m \\ p_m(0,0)=1}} \max_{\substack{\lambda \in \sigma(B) \\ \mu \in \sigma(A)}} |p_m(\lambda, \mu)|, \tag{4.2}$$

*where $\mathbb{P}_m$ denotes the set of polynomials in two variables $x, y$ of the form $p_m(x, y) = \sum_{i=0}^m c_i x^i y^i$ with value one at the origin.*

*Proof.* Using Theorem 3.2 and Lemma 4.1, we obtain

$$\|X_* - X_m^{or}\|_{(A,B)} = \min_{\substack{p_m \in \mathbb{P}_m \\ p_m(0,0)=1}} \|p_m(B:A)\text{vec}(D_0)\|_{B\otimes A}.$$

Let $\langle \cdot, \cdot \rangle$ be the standard inner product in $\mathbb{R}^{ns}$. From (4.1), it is straightforward to verify that

$$\begin{aligned}\|p_m(B:A)\text{vec}(D_0)\|_{B\otimes A}^2 &= \langle p_m(B:A)\text{vec}(D_0), (B\otimes A)p_m(B:A)\text{vec}(D_0)\rangle \\ &= \langle p_m(B:A)\text{vec}(D_0), p_m(B:A)(B\otimes A)\text{vec}(D_0)\rangle \\ &= \sum_{i=1}^{s}\sum_{j=1}^{n} a_{ij}^2 p_m^2(\lambda_i, \mu_j)\lambda_i \mu_j.\end{aligned}$$

As $\|\text{vec}(D_0)\|_{(B\otimes A)}^2 = \sum_{i=1}^s \sum_{j=1}^n a_{ij}^2 \lambda_i \mu_j$, we get

$$\begin{aligned}\|p_m(B:A)\text{vec}(D_0)\|_{B\otimes A} &\leq \max_{i,j} |p_m(\lambda_i, \mu_j)| \|\text{vec}(D_0)\|_{B\otimes A} \\ &= \max_{\substack{\lambda \in \sigma(B) \\ \mu \in \sigma(A)}} |p_m(\lambda, \mu)| \|\text{vec}(D_0)\|_{B\otimes A}.\end{aligned}$$

Finally, since $\|\text{vec}(D_0)\|_{B\otimes A} = \|X_* - X_0\|_{(A,B)}$, the proof completes. □

4.2. **The worst-case behavior of the G-MR-L method.** In this section, we study the worst-case behavior of the G-MR-L method. The main result of this section is derived with the assumptions that the matrices $A$ and $B$ are SPD.

**Theorem 4.3.** *If $X_m^{mr}$ is the mth approximation of the G-MR-L method, then*

$$\frac{\|C - AX_m^{mr}B\|_F}{\|C - AX_0B\|_F} \leq \min_{\substack{p_m \in \mathbb{P}_m \\ p_m(0,0)=1}} \max_{\substack{\lambda \in \sigma(B) \\ \mu \in \sigma(A)}} |p_m(\lambda, \mu)|, \tag{4.3}$$

*where $\mathbb{P}_m$ denotes the set of polynomials in two variables $x, y$ of the form $p_m(x, y) = \sum_{i=0}^m c_i x^i y^i$ with value one at the origin.*

*Proof.* Let $X \in X_0 + \mathcal{GK}_m(A, R_0, B)$ and $R = C - AXB$. Then

$$\begin{aligned}\mathrm{vec}(R) &= \mathrm{vec}(R_0) - \sum_{i=1}^{m} c_i (B \otimes A)^i \mathrm{vec}(R_0) \\ &= (U \otimes V)^T \Big( I - \sum_{i=1}^{m} c_i (\Lambda \otimes D)^i \Big) (U \otimes V) \mathrm{vec}(R_0) \\ &= (U \otimes V)^T p_m(\Lambda : D)(U \otimes V)\mathrm{vec}(R_0),\end{aligned}$$

which $p_m(x, y) = 1 + \sum_{i=1}^{m} c_i x^i y^i$ where $c_i \in \mathbb{R}, i = 1, \ldots, m$. Hence, we have

$$\begin{aligned}\|C - AXB\|_F &= \|\mathrm{vec}(R)\|_2 \\ &\leq \|(U \otimes V)^T\|_2 \|p_m(\Lambda : D)\|_2 \|(U \otimes V)\|_2 \|\mathrm{vec}(R_0)\|_2.\end{aligned}$$

Now, using [16, Theorem 1, p. 411], it follows that

$$\|C - AXB\|_F \leq \max_{\substack{\lambda \in \sigma(B) \\ \mu \in \sigma(A)}} |p_m(\lambda, \mu)| \|R_0\|_F.$$

Together with Theorem 3.10, we conclude

$$\begin{aligned}\|C - AX_m^{mr} B\|_F &= \min_{X \in X_0 + \mathcal{GK}_m(A, R_0, B)} \|C - AXB\|_F \\ &\leq \min_{\substack{p_m \in \mathbb{P}_m \\ p_m(0,0)=1}} \max_{\substack{\lambda \in \sigma(B) \\ \mu \in \sigma(A)}} |p_m(\lambda, \mu)| \|R_0\|_F. \qquad \square\end{aligned}$$

**Remark 4.4.** *In the process of proving Theorem 4.3, we also consider the following cases:*

**(1):** $A$ ***and*** $B$ ***are diagonalizable.***

*When* $A$ *and* $B$ *are diagonalizable, i.e.,* $A = V^{-1} D V$ *and* $B = U^{-1} \Lambda U$ *with* $D = diag(\mu_1, \ldots, \mu_n)$ *and* $\Lambda = diag(\lambda_1, \ldots, \lambda_s)$*, then we obtain the following convergence bound*

$$\frac{\|C - AX_m^{mr} B\|_F}{\|C - AX_0 B\|_F} \leq \kappa(V)\kappa(U) \min_{\substack{p_m \in \mathbb{P}_m \\ p_m(0,0)=1}} \max_{\substack{\lambda \in \sigma(B) \\ \mu \in \sigma(A)}} |p_m(\lambda, \mu)|. \tag{4.4}$$

*Moreover, if we assume that* $B = I_s$*, then we have* $\sigma(B) = \{1\}$ *and* $\kappa(U) = 1$*. In this case, we deduce that Theorem 4.3 coincides with the result of the Gl-GMRES given in [3, Theorem 5].*

**(2):** $A$ ***and*** $B$ ***are normal matrices.***

*If* $A$ *and* $B$ *are normal matrices, then Theorem 4.3 holds. In addition, if* $B = I_s$*, then Theorem 4.3 reduces to Theorem 7 in [3].*

**Remark 4.5.** *Interesting related work was recently done by Bellalij et al. [3] for solving a matrix equation* $AX = C$*. They presented the worst-case convergence behavior of the Gl-GMRES method for diagonalizable matrices. Although all presented results in [3] are correct, as it was discussed in Remark 4.4, when* $B =$

*$I_s$, Theorem 4.3 leads to Theorems 5 and 7 in [3]. Furthermore, a comparison between our proof process with that of [3] shows that Theorem 4.3 presents an easier proof for the worst-case convergence behavior of the Gl-GMRES method.*

## 5. Numerical illustrations

The G-OR-L and G-MR-L methods are summarized in the following algorithm. It is important to notice that when the iterate $m$ becomes large, the computational requirements increase. To remedy this problem, we use the algorithm in a restarted mode, i.e., $X_0 = X_m^{or}(X_0 = X_m^{mr})$ where $X_m^{or}(X_m^{mr})$ is the last computed approximate solution obtained with the G-OR-L (G-MR-L) method.

**Algorithm 2** (The G-OR-L($m$) and G-MR-L($m$) methods)

1: Choose $X_0$.
2: Compute $R_0 = C - AX_0B$ and set $V_1 = R_0/\|R_0\|_F$. Use the generalized global Lanczos algorithm to compute the basis $\{V_1, V_2, \dots, V_m\}$ of $\mathcal{GK}_m(A, R_0, B)$ and the matrix $H_m$.
3: G-OR-L: compute the approximate solution $X_m^{or} = X_0 + \mathcal{V}_m(\alpha_m^{or} \otimes I_s)$ where $\alpha_m^{or}$ solves the equation (3.3).
4: If the approximation $X_m^{or}$ is suitable then stop, else set $X_0 = X_m^{or}$ and goto 2.
5: G-MR-L: compute the approximate solution $X_m^{mr} = X_0 + \mathcal{V}_m(\alpha_m^{mr} \otimes I_s)$ where $\alpha_m^{mr}$ solves the equation (3.11).
6: If the approximation $X_m^{mr}$ is suitable then stop, else set $X_0 = X_m^{mr}$ and goto 2.

In this section, we present some numerical examples to illustrate the quality of the presented theoretical results. To do so, we use the matrices

- $A_1(n) = \text{tridiag}(-1, 10, -1)$ and $B_1(s) = \text{tridiag}(-1, 10, -1)$,
- $A_2(n) = \begin{pmatrix} 4 & -1 & & -1 \\ -1 & 4 & \ddots & \\ & \ddots & \ddots & -1 \\ -1 & & -1 & 4 \end{pmatrix}$ and $B_2(s) = \begin{pmatrix} 8 & -2 & & -2 \\ -2 & 8 & \ddots & \\ & \ddots & \ddots & -2 \\ -2 & & -2 & 8 \end{pmatrix}$,

where $n$ and $s$ are the order of matrices $A$ and $B$, respectively. Also, we use some matrices from the Matrix-Market website at http://math.nist.gov/MatrixMarket.

The entries of the matrix $C$ in all the examples are random values uniformly distributed on $[0, 1]$ and $X_0 = 0$. All numerical tests run in Matlab.

**Example 5.1.** In this example, we seek two goals. First, we examine the influence of the condition numbers of the matrices $A$ and $B$ on the speed of convergence of the G-OR-L(m) and G-MR-L(m) methods. Then, we study the behavior of the Frobenius norm of the residual of the G-OR-L($m$) and G-MR-L($m$) methods. Also, as stopping criteria, we use $\|C - AX_m^{or}B\|_F \leq 10^{-6}$ and $\|C - AX_m^{mr}B\|_F \leq 10^{-6}$ for the G-OR-L($m$) and G-MR-L($m$) methods, respectively.

Tables 1 and 2 report the final Frobenius norm of the residuals, the number of iterations and the CPU time required to convergence for each method. We note that the number of iterations refers to the number of restarts. As shown in Tables 1 and 2, we see that the number of iterations increases when $A$ or $B$ have very large condition numbers. However, if the condition numbers of $A$ and $B$ are small, then the number of iterations decreases.

TABLE 1. Effectiveness of the G-OR-L and G-MR-L methods: $m = 3$.

| $(A, B)$ | | | G-OR-L(3) | G-MR-L(3) |
|---|---|---|---|---|
| | $n = 2000, s = 100$ | Residuals | 5.1214e-008 | 5.1342e-008 |
| $(A_1, B_1)$ | $\kappa(A) = 1.5$ | Number of iterations | 6 | 6 |
| | $\kappa(B) = 1.4997$ | CPU time (sec) | 2.75 | 2.17 |
| | $n = 1000, s = 500$ | Residuals | 9.5702e-007 | 5.9173e-007 |
| $(A_2, B_2)$ | $\kappa(A) = 3$ | Number of iterations | 14 | 14 |
| | $\kappa(B) = 3$ | CPU time (sec) | 15.84 | 14.34 |

TABLE 2. Effectiveness of the G-OR-L and G-MR-L methods: $m = 20$.

| $(A, B)$ | | | G-OR-L(20) | G-MR-L(20) |
|---|---|---|---|---|
| | $n = 900, s = 10$ | Residuals | 4.4148e-007 | 2.3661e-007 |
| (GR3030, $B_1$) | $\kappa(A) = 3.8e+02$ | Number of iterations | 11 | 11 |
| | $\kappa(B) = 1.4749$ | CPU time (sec) | 1.12 | 0.93 |
| | $n = 100, s = 10$ | Residuals | 8.1349e-007 | 9.1156e-007 |
| (NOS4, $B_1$) | $\kappa(A) = 2.7e+03$ | Number of iterations | 53 | 39 |
| | $\kappa(B) = 1.4749$ | CPU time (sec) | 0.25 | 0.18 |
| | $n = 10, s = 468$ | Residuals | 9.5765e-007 | 9.6650e-007 |
| ($A_1$, NOS5) | $\kappa(A) = 1.4749$ | Number of iterations | 331 | 303 |
| | $\kappa(B) = 2.9e+04$ | CPU time (sec) | 20.40 | 19.37 |

According to the aforementioned stopping criteria, we investigate further this example by considering the convergence behavior of the Frobenius norm of the residual at each iteration of the G-OR-L($m$) and G-MR-L($m$) methods, respectively.

• **The G-OR-L method.** In Figures 1–5, the left plots show the Frobenius norm of the G-OR-L residuals and the right plots show the values of $\alpha_1$ and $\alpha_2$. As shown in Figures 1–3, we see that $\alpha_1$ and $\alpha_2$ are small. Moreover, as it is expected from Remark 3.7, we observe that the Frobenius norm of the residuals associated with the matrices $(A_1(2000), B_1(100))$, $(A_2(1000), B_2(500))$ and (GR3030, $B_1(10)$) decreases.

Figures associated with the matrices (NOS4, $B_1(10)$) and ($B_1(10)$, NOS5), especially ($B_1(10)$, NOS5), illustrate that $\alpha_1$ and $\alpha_2$ are large. Therefore, we expect that the rate of changes of the Frobenius norm of the residuals associated with these matrices increases. As we see in Figures 4 and 5, the Frobenius norm of the residuals associated with the matrices (NOS4, $B_1(10)$) and ($B_1(10)$, NOS5), especially ($B_1(10)$, NOS5), oscillates too much.

FIGURE 1. Illustration of $\|R_m^{or}\|_F$ associated with $(A_1(2000), B_1(100))$.

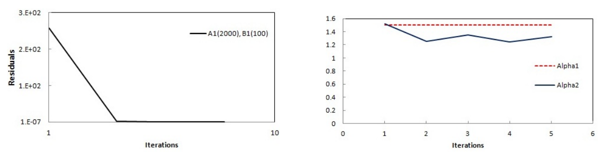



FIGURE 2. Illustration of $\|R_m^{or}\|_F$ associated with $(A_2(1000), B_2(500))$.

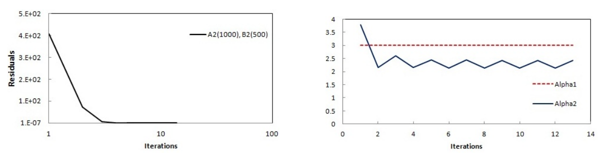



FIGURE 3. Illustration of $\|R_m^{or}\|_F$ associated with (GR3030, $B_1(10)$).

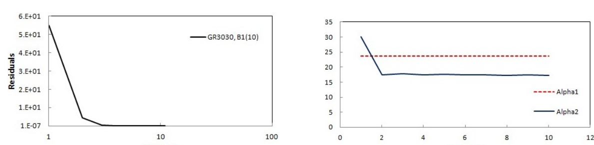
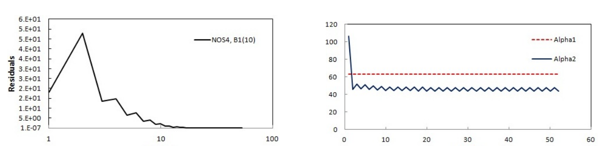
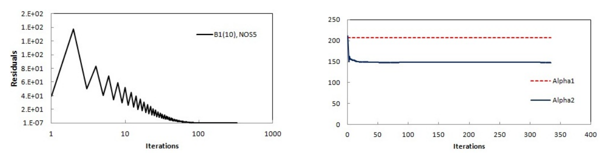



FIGURE 4. Illustration of $\|R_m^{or}\|_F$ associated with (NOS4, $B_1(10)$).



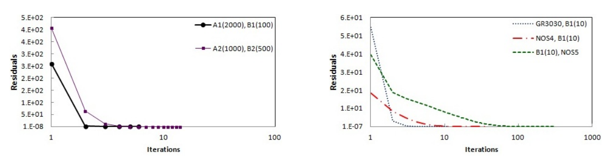

• **The G-MR-L method.** It can be shown that the Frobenius norm of the residual at each step of the G-MR-L method satisfies $\|R_m^{mr}\|_F \leq \|R_0\|_F$.

FIGURE 5. Illustration of $\|R_m^{or}\|_F$ associated with ($B_1(10)$, NOS5).




Therefore, since we use the G-MR-L method in a restarted mode, we expect a "monotonically decreasing" behavior of the Frobenius norm of the residual. The plots of Figure 6 show that the Frobenius norm of the residual is monotonically decreasing.

FIGURE 6. Illustration of $\|R_m^{mr}\|_F$ associated with the matrices $(A_1(2000), B_1(100))$, $(A_2(1000), B_2(500))$, (GR3030, $B_1(10)$), (NOS4, $B_1(10)$) and ($B_1(10)$, NOS5).




We summarize what we have stated so far. As we have seen, the Frobenius norm of the residual of the G-OR-L($m$) may oscillate. In contrast, the Frobenius norm of the residual of the G-MR-L($m$) shows a monotonically decreasing behavior. Therefore, this contrast can be used as one of the reasons for the observed difference in the speed of convergence of the G-OR-L($m$) and G-MR-L($m$) methods.

**Example 5.2.** In this example, we study the behavior of the $(A, B)-$norm of the error of the G-OR-L($m$) method. The test matrices are the same as the previous example. Also, the iterations are stopped when $\|X_* - X_m^{or}\|_{(A,B)} \leq 10^{-6}$.

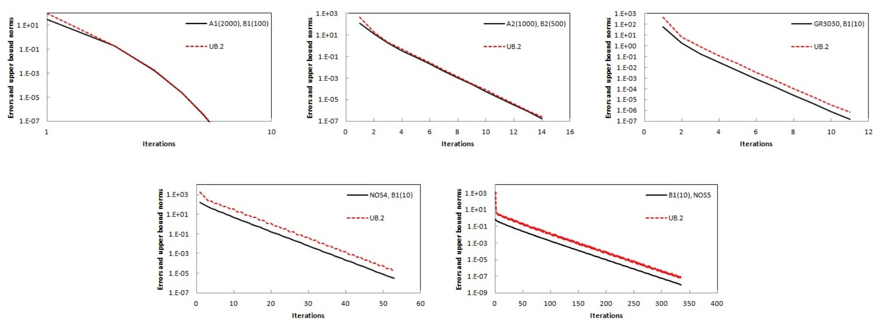
In Table 3, we report the final values of the $(A, B)-$norm of the errors and the final upper bounds derived in Theorem 3.5. The obtained results indicate that these upper bounds are noticeable. However, the upper bound (UB.2) is better than other upper bounds.

Finally, Figure 7 shows the values of the $(A, B)-$norm of the errors and the upper bound (UB.2) together with the number of iterations.

TABLE 3. The final values of $\|X_* - X_m^{or}\|_{(A,B)}$ and the quality of the upper bounds (UB.1)–(UB.4).

| $(A, B)$ | $\|X_* - X_m^{or}\|_{(A,B)}$ | (UB.1) | (UB.2) | (UB.3) | (UB.4) |
|---|---|---|---|---|---|
| $(A_1(2000), B_1(100))$ | 5.0908e-009 | 6.4129e-009 | 5.3203e-009 | 6.9471e-009 | 2.1651e-007 |
| $(A_2(1000), B_2(500))$ | 1.8199e-007 | 3.3881e-007 | 2.4011e-007 | 5.6468e-007 | 1.9907e-006 |
| (GR3030, $B_1(10)$) | 1.4725e-007 | 9.1979e-007 | 6.6144e-007 | 1.0907e-005 | 7.0461e-006 |
| (NOS4, $B_1(10)$) | 2.9270e-006 | 1.9506e-005 | 1.4956e-005 | 6.1563e-004 | 5.1433e-005 |
| ($B_1(10)$, NOS5) | 9.2659e-009 | 7.3097e-008 | 5.1659e-008 | 7.5589e-006 | 4.9907e-006 |

FIGURE 7. Illustration of $\|X_* - X_m^{or}\|_{(A,B)}$ and the values of the upper bound (UB.2).




## 6. Conclusion and further works

We have considered the convergence behavior of the Gl-FOM and Gl-GMRES methods for solving the matrix equation $AXB = C$ where $A$ and $B$ are SPD. More precisely, some new theoretical results of these methods, such as computable exact expressions, upper bounds for the norm of the error and residual and the worst-case convergence behavior of these methods, have been established. In particular, our upper bounds for the norm of the Gl-FOM error depend on the condition numbers of the matrices $A$ and $B$ and the information generated by the Gl-FOM method. In the numerical test section, we have explored the convergence behavior of these methods. The numerical results show the efficiency of the theoretical results.

The relationship between the norm of the Gl-GMRES residual and the spectral information of $A$ and $B$ and the information generated by the Gl-GMRES method can be the subject of further investigations. Furthermore, the theoretical results, presented in this paper, can be used for the Gl-FOM and Gl-GMRES methods in order to solve the generalized Sylvester equations with SPD coefficients.

## References


[1] F. P. A. Beik, Theoretical results on the global GMRES method for solving generalized Sylvester matrix equations, *Bull. Iranian Math. Soc.* **40** (2014), no. 5, 1097–1117.

[2] F. P. A. Beik and D. K. Salkuyeh, On the global Krylov subspace methods for solving general coupled matrix equation, *Comput. Math. Appl.* **62** (2011), no. 12, 4605–4613.

[3] M. Bellalij, K. Jbilou and H. Sadok, New convergence results on the global GMRES method for diagonalizable matrices, *J. Comput. Appl. Math.* **219** (2008), no. 2, 350–358.

[4] R. Bouyouli, K. Jbilou, R. Sadaka and H. Sadok, Convergence properties of some block Krylov subspace methods for multiple linear systems, *J. Comput. Appl. Math.* **196** (2006), no. 2, 498–511.

[5] R. Bouyouli, G. Meurant, L. Smoch and H. Sadok, New results on the convergence of the conjugate gradient method, *Numer. Linear Algebra Appl.* **16** (2009), no. 3, 223–236.

[6] J. B. Conway, A Course in Functional Analysis, Springer-Verlag, New York, 1990.

[7] D. Cvetković-Iliíc, The reflexive solutions of the matrix equations $AXB = C$. *Comput. Math. Appl.* **51** (2006), no. 6-7, 879–902.

[8] H. Dai, On the symmetric solutions of linear matrix equations, *Linear Algebra Appl.* **131** (1990) 1–7.

[9] M. Heyouni, The global Hessenberg and CMRH methods for linear systems with multiple right-hand sides, *Numer. Algorithms* **26** (2001), no. 4, 317–332.

[10] M. Heyouni, Extended Arnoldi methods for large low-rank Sylvester matrix equations, *Appl. Numer. Math.* **60** (2010), no. 11, 1171–1182.

[11] M. Heyouni and A. Essai, Matrix Krylov subspace methods for linear systems with multiple right-hand sides, *Numer. Algorithms* **40** (2005), no. 2, 137–156.

[12] R. A. Horn and C. R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1985.

[13] G. X. Huang, F. Yin and K. Guo, An iterative method for the skew-symmetric solution and the optimal approximate solution of the matrix equation $AXB = C$, *J. Comput. Appl. Math.* **212** (2008), no. 2, 231–244.

[14] K. Jbilou, A. Messaoudi and H. Sadok, Global FOM and GMRES algorithms for matrix equations, *Appl. Numer. Math.* **31** (1999), no. 1, 49–63.

[15] F. Kittaneh, Notes on some inequalities for Hilbert space operators, *Publ. Res. Inst. Math. Sci.* **24** (1988), no. 2, 283–293.

[16] P. Lancaster and M. Tismenetsky, The Theory of Matrices with Applications, Academic Press, New York, 1985.

[17] J. Liesen and P. Tichý, Convergence analysis of Krylov subspace methods, *GAMM Mitt. Ges. Angew. Math. Mech.* **27** (2004), no. 2, 153–173.

[18] Y. A. Peng, X. Hu and L. Zhang, An iteration method for the symmetric solutions and the optimal approximation solution of the matrix equation $AXB = C$, *Appl. Math. Comput.* **160** (2005), no. 3, 763–777.

[19] Y. Saad, Iterative Methods for Sparse Linear Systems, PWS Press, New York, 1995.

[20] K. Zhang and C. Gu, Flexible global generalized Hessenberg methods for linear systems with multiple right-hand sides, *J. Comput. Appl. Math.* **263** (2014) 312–325.

[21] K. Zhang and C. Gu, A global transpose-free method with quasi-minimal residual strategy for the Sylvester equations, *Filomat* **27** (2013), no. 8, 1437–1446.



(Mahmood Mohseni Moghadam) Department of Mathematics and Computer Science, Shahid Bahonar University of Kerman, P.O.Box 761691, Kerman, Iran

*E-mail address*: mohseni@uk.ac.ir

(Azim Rivaz) Department of Mathematics and Computer Science, Shahid Bahonar University of Kerman, P.O.Box 761691 Kerman, Iran
*E-mail address*: arivaz@uk.ac.ir

(Azita Tajaddini) Department of Mathematics and Computer Science, Shahid Bahonar University of Kerman, P.O.Box 761691, Kerman, Iran
*E-mail address*: atajadini@uk.ac.ir

(Farid Saberi Movahed) Department of Mathematics and Computer Science, Shahid Bahonar University of Kerman, P.O.Box 761691, Kerman, Iran
*E-mail address*: fdsaberi@math.uk.ac.ir, fdsaberi@gmail.com