On Some Extended Block Krylov Based Methods for Large Scale Nonsymmetric Stein Matrix Equations

Abdeslem Hafid Bentbib 1, Khalide Jbilou 2 and EL Mostafa Sadek 3,*

1 Laboratory LAMAI, University of Cadi Ayyad, Marrakesh 40000, Morocco; a.bentbib@uca.ac.ma
2 LMPA, 50 rue F. Buisson, ULCO Calais, Calais 62228, France; jbilou@lmpa.univ-littoral.fr
3 ENSA d’EL Jadida, University Chouaib Doukkali, EL Jadida 24002, Morocco
* Correspondence: sadek.maths@gmail.com; Tel.: +212-523-344-822

Academic Editor: Lokenath Debnath
Received: 22 December 2016; Accepted: 17 March 2017; Published: 27 March 2017

Abstract: In the present paper, we consider the large scale Stein matrix equation with a low-rank constant term $AXB - X + EF^T = 0$. These matrix equations appear in many applications in discrete-time control problems, filtering and image restoration and others. The proposed methods are based on projection onto the extended block Krylov subspace with a Galerkin approach (GA) or with the minimization of the norm of the residual. We give some results on the residual and error norms and report some numerical experiments.

Keywords: extended block Krylov subspaces; low-rank approximation; Stein matrix equation; Galerkin approach (GA); minimal residual (MR) methods

1. Introduction

In this paper, we are interested in the numerical solution of large scale nonsymmetric Stein matrix equations of the form:

$$AXB - X + EF^T = 0$$

(1)

where $A$ and $B$ are real, sparse and square matrices of size $n \times n$ and $s \times s$, respectively, and $E$ and $F$ are matrices of size $n \times r$ and $s \times r$, respectively.

Stein matrix equations play an important role in many problems in control and filtering theory for discrete-time large-scale dynamical systems, in each step of Newton’s method for discrete-time algebraic Riccati equations, model reduction problems, image restoration techniques and other problems [1–10].

Direct methods for solving the matrix Equation (1), such as those proposed by Bartels–Stewart [11] and the Hessenberg–Schur [12] algorithms, are attractive if the matrices are of small size. For a general overview of numerical methods for solving the Stein matrix equation [1,2,13].

The Stein matrix Equation (1) can be formulated as an $ns \times ns$ large linear system using the Kronecker formulation:

$$(B^T \otimes A - I_n \otimes I_n)\text{vec}(X) = -\text{vec}(EF^T)$$

(2)

where $\text{vec}(X)$ is the vector obtained by stacking all the columns of the matrix $X$, $I_n$ is the $n$-by-$n$ identity matrix, and the Kronecker product of two matrices $A$ and $B$ is defined by $A \otimes B = [a_{ij}B]$, where $A = [a_{ij}]$. This product satisfies the properties $(A \otimes B)(C \otimes D) = (AC \otimes BD)$, $(A \otimes B)^T = A^T \otimes B^T$ and $\text{vec}(AXB) = (B^T \otimes A)\text{vec}(X)$. Then, the matrix Equation (1) has a unique solution if and only if $\lambda \mu \neq 1$ for all $\lambda \in \sigma(A)$ and $\mu \in \sigma(B)$, where $\sigma(A)$ denotes the spectrum of the matrix $A$. Throughout the paper, we assume that this condition is satisfied. Moreover, if both $A$ and $B$ are Schur stable, i.e.,
\(\sigma(A)\) and \(\sigma(B)\) lie in the open unit disc, and then the solution of Equation (1) can be expressed as the following infinite matrix series:

\[X = \sum_{i=0}^{\infty} A^i E F^T B^i\]

To solve large linear matrix equations, several Krylov subspace projection methods have been proposed (see, e.g., [1,13–24] and the references therein). The main idea developed in these methods is to use a block Krylov subspace or an extended block Krylov subspace and then project the original large matrix equation onto these Krylov subspaces using a Galerkin condition or a minimization property of the obtained residual. Hence, we will be interested in these two procedures to get approximate solutions to the solution of the Stein matrix Equation (1). The rest of the paper is organized as follows. In the next section, we recall the extended block Krylov subspace and the extended block Arnoldi (EBA) algorithm with some properties. In Section 3, we will apply the Galerkin approach (GA) to Stein matrix equations by using the extended Krylov subspaces. In Section 4, we define the minimal residual (MR) method for Stein matrix equations by using the extended Krylov subspaces. We finally present some numerical experiments in Section 5.

2. The Extended Block Krylov Subspace Algorithm

In this section, we recall the EBA algorithm applied to \((A, V)\), where \(V \in \mathbb{R}^{n \times r}\). The block Krylov subspace associated with \((A, V)\) is defined as:

\[\mathcal{K}_m(A, V) = \text{Range}\{V, AV, A^2V, \ldots, A^{m-1}V\}\]

The extended block Krylov subspace associated with the pair \((A, V)\) is given as:

\[\mathcal{K}_m^e(A, V) = \mathcal{K}_m(A, V) + \mathcal{K}_m(A^{-1}, A^{-1}V)\]

The EBA Algorithm 1 is defined as follows [15,16,18,23]:

Algorithm 1. The Extended Block Arnoldi (EBA) Algorithm

1. Inputs: \(A\) an \(n \times n\) matrix, \(V\) an \(n \times r\) matrix and \(m\) an integer
2. Compute the QR decomposition of \([V, A^{-1}V]\), i.e., \([V, A^{-1}V] = V_1 \Lambda\)
3. Set \(V_0 = [\ ]\)
4. for \(j = 1, 2, \ldots, m\)
   
   a. Set \(V_j^{(1)}\), first \(r\) columns of \(V_j\); \(V_j^{(2)}\), second \(r\) columns of \(V_j\)
   
   b. \(V_j = [V_{j-1}, V_j]; \tilde{V}_j = [V_j^{(1)}, A^{-1} V_j^{(2)}]\)
   
   c. Orthogonalize \(\tilde{V}_j\) w.r. to \(V_j\) to get \(V_{j+1}\), i.e.
      
      * for \(i = 1, 2, \ldots, j\)
      * \(H_{ij} = V_j^T \tilde{V}_{j+1}^i\)
      * \(\tilde{V}_{j+1} = \tilde{V}_{j+1} - V_i H_{ij}\)
      * End for
   
   d. Compute the QR decomposition of \(\tilde{V}_{j+1}\), i.e., \(\tilde{V}_{j+1} = V_{j+1} H_{j+1}\)
5. End for

This algorithm allows us to construct an orthonormal matrix \(V_m = [V_1, V_2, \ldots, V_m]\) that is a basis of the block extended Krylov subspace \(\mathcal{K}_m^e(A, V)\). The restriction of the matrix \(A\) to the block extended Krylov subspace \(\mathcal{K}_m^e(A, V)\) is given by \(T_m = V_m^T A V_m\).
Let $\tilde{\mathbf{r}}_m = \mathbf{V}^T_{m+1} A \mathbf{V}_m$. Then, we have the following relations [25]:

$$A \mathbf{V}_m = \mathbf{V}_{m+1} \tilde{\mathbf{r}}_m = \mathbf{V}_m \mathbf{T}_m + \mathbf{V}_{m+1} \mathbf{E}_m \mathbf{T}_m^T$$

where $\mathbf{E}_m = [0_{2r \times 2(m-r) \times 2r}]^T$ is the matrix of the last $2r$ columns of the identity matrix $I_{2mr}$ [23,25]. In the next section, we will define the GA for solving Stein matrix equations.

3. Galerkin-Based Methods

In this section, we will apply the Galerkin projection method to obtain low-rank approximate solutions of the nonsymmetric Stein matrix Equation (1). This approach has been applied for Lyapunov, Sylvester or Riccati matrix equations [1,14,15,19–21,23,25,26].

3.1. The Case: Both A and B Are Large Matrices

We consider here a nonsymmetric Stein matrix equation, where $A$ and $B$ are large and sparse matrices with $r \ll n$ and $r \ll s$. We project the initial problem by using the extended block Krylov subspaces $\mathcal{K}_m^a(A, E)$ and $\mathcal{K}_m^b(B^T, F^T)$ associated with the pairs $(A, E)$ and $(B^T, F^T)$, respectively, and get orthonormal bases $\{\mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_m\}$ and $\{\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_m\}$. We then consider approximate solutions of the Stein matrix Equation (1) that have the low-rank form:

$$\mathbf{X}_m^{GA} = \mathbf{V}_m \gamma_m \mathbf{W}_m^T$$

where $\mathbf{V}_m = [\mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_m]$ and $\mathbf{W}_m = [\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_m]$.

The matrix $\gamma_m^{GA}$ is determined from the following Galerkin orthogonality condition:

$$\mathbf{V}_m^T \gamma_m^{GA} \mathbf{W}_m = \mathbf{V}_m^T (A \mathbf{X}_m^{GA} - \mathbf{X}_m^{GA} + E \mathbf{F}^T) \mathbf{W}_m = 0$$

Now, replacing $\mathbf{X}_m^{GA} = \mathbf{V}_m \gamma_m^{GA} \mathbf{W}_m^T$ in Equation (4), we obtain the reduced Stein matrix equation:

$$T_A \gamma_m^{GA} (T_B)^T - \gamma_m^{GA} + \tilde{\mathbf{E}} \tilde{\mathbf{F}}^T = 0$$

where $\tilde{\mathbf{E}} = \mathbf{V}_m^T \mathbf{E}$, $\tilde{\mathbf{F}} = \mathbf{W}_m^T \mathbf{F}$, $T_A = \mathbf{V}_m^T A \mathbf{V}_m$, and $T_B = \mathbf{W}_m^T B \mathbf{W}_m$.

Assuming that $\lambda_i(T_A) \lambda_j(T_B) \neq 1$ for any $i = 1, 2, ..., 2mr$ and $j = 1, 2, ..., 2mr$, the solution $\gamma_m$ of the low-order Stein Equation (5) can be obtained by a direct method such as those described in [11]. The following result on the norm of the residual $\mathcal{R}_m$ allows us to stop the iterations without having to compute the approximation $\mathbf{X}_m^{GA}$.

**Theorem 1.** Let $\mathbf{X}_m^{GA}$ be the approximation obtained at step $m$ by the EBA algorithm. Then, the Frobenius norm of the residual $\mathcal{R}_m^{GA}$ associated to the approximation $\mathbf{X}_m^{GA}$ is given by:

$$\|\mathcal{R}_m^{GA}\|_F = \sqrt{\alpha_m^2 + \beta_m^2 + \gamma_m^2}$$

where $\alpha_m = \left\|T_A \mathbf{Y}_m \mathbf{E}_m (T_{m+1,m}^B)^T\right\|_F$, $\beta_m = \left\|T_A \mathbf{E}_m (T_{m+1,m}^B)^T\right\|_F$, and:

$$\gamma_m = \left\|T_A \mathbf{E}_m \mathbf{Y}_m \mathbf{E}_m (T_{m+1,m}^B)^T\right\|_F$$

**Proof.** The proof is similar to the one given at proposition 6 in [17].

In the following result, we give an upper bound for the norm of the error $\mathbf{X} - \mathbf{X}_m^{GA}$. 

Theorem 2. Assume that \( \|A\|_2 < 1 \) and \( \|B\|_2 < 1 \), and let \( Y_m^{GA} \) be the exact solution of projected Stein matrix Equation (5) and \( X_m^{GA} \) be the approximate solution given by running \( m \) steps of the EBA algorithm. Then:

\[
\|X - X_m^{GA}\|_2 \leq \frac{\|A\|_2\|T_{B_m}^{A} + 1, m\|_2 + \|B\|_2\|T_{A_m}^{A} + 1, m\|_2}{1 - \|A\|_2\|B\|_2} \|Y_m\|_2
\]

(7)

Proof. The proof is similar to the one given at Theorem 2 in [27]. \( \square \)

The approximate solution \( X_m^{GA} \) can be given as a product of two matrices of low rank. Consider the singular value decomposition of the \( 2mr \times 2mr \) matrix:

\[
Y_m^{GA} = \tilde{Y}_1 \Sigma \tilde{Y}_2^T
\]

where \( \Sigma \) is the diagonal matrix of the singular values of \( Y_m^{MR} \) sorted in decreasing order. Let \( Y_{1, l} \) and \( Y_{2, l} \) be the \( 2mr \times l \) matrices of the first \( l \) columns of \( \tilde{Y}_1 \) and \( \tilde{Y}_2 \), respectively, corresponding to the \( l \) singular values of magnitude greater than some tolerance. We obtain the truncated singular value decomposition:

\[
Y_m^{GA} \approx U_{1, l} \Sigma_l U_{2, l}^T
\]

where \( \Sigma_l = \text{diag}[\sigma_1, \ldots, \sigma_l] \). Setting \( Z_{1,m} = \mathbb{V}_m U_{1, l} \Sigma_l^{1/2} \) and \( Z_{2,m} = \mathbb{W}_m U_{2, l} \Sigma_l^{1/2} \), it follows that:

\[
X_m^{GA} \approx Z_{1,m} Z_{2,m}^T
\]

(8)

This is very important for large problems when one doesn’t need to compute and store the approximation \( X_m \) at each iteration.

The GA is given in Algorithm 2:

Algorithm 2. Galerkin Approach (GA) for the Stein Matrix Equations

1. Inputs: \( A \) an \( n \times n \) matrix, \( B \) an \( s \times s \) matrix, \( E \) an \( n \times r \) matrix and \( F \) an \( s \times r \) matrix.
2. Choose a tolerance \( \text{tol} > 0 \), a maximum number of \( \text{itermax} \) iterations.
3. For \( m = 1, 2, 3, \ldots, \text{itermax} \)
4. Compute \( \mathbb{V}_m, T_{A_m}^A \), by Algorithm 1 applied to \((A, E)\).
5. Compute \( \mathbb{W}_m, T_{B_m}^B \), by Algorithm 1 applied to \((B^T, F)\).
6. Solve the low order Stein Equation (5) and compute \( \|R_m\|_F \) given by Equation (6)
7. if \( \|R_m\|_F \leq \text{tol} \), stop,
8. Using Equation (8), the approximate solution \( X_m^{GA} \) is given by \( X_m^{GA} \approx Z_{1,m} Z_{2,m}^T \).

In the next section, we consider the case where the matrix \( A \) is large while \( B \) has a moderate or a small size.

3.2. The Case: \( A \) Large and \( B \) Small

In this section, we consider the Stein matrix equation:

\[
AXB - X = E
\]

(9)

where \( E \) is a matrix of size \( n \times s \) with \( s \ll n \).

In this case, we will consider approximations of the exact solution \( X \) as:

\[
X_m = \mathbb{V}_m Y_m
\]
where $V_m$ is the orthonormal basis obtained by applying the extended block Krylov subspace $K_m^e(A, E)$. The orthogonality Galerkin condition gives:

$$V_m^T R_m = 0 \quad (10)$$

where $R_m$ is the $m$-th residual given by $R_m = AX_mB - X_m + E$. Therefore, we obtain the projected Stein matrix equation:

$$T_A Y_m^{GA} B - Y_m^{GA} + \tilde{E} = 0 \quad (11)$$

where $T_A = V_m^T A V_m$ and $\tilde{E} = V_m^T E$.

The next result gives a useful expression of the norm of the residual.

**Theorem 3.** Let $Y_m^{GA}$ the exact solution of the reduced Stein matrix Equation (11) and let $X_m^{GA} = V_m Y_m^{GA}$ be the approximate solution of Equation (9) with $R_m = \mathcal{R}(X_m^{GA})$ the corresponding residual. Then:

$$\|R_m\|_F = \left\| T_{m+1}^A E_m^T Y_m^{GA} B \right\|_F \quad (12)$$

**Proof.** The residual is given by $R_m = AX_m^{GA}B - X_m^{GA} + E$. Since $E$ is belonging to $K_m^e(A, E)$, then $V_m Y_m^{GA} B = E$. Using the relation $A V_m = V_m + T_A \quad (I_0)$, we have:

$$\|R_m\|_F = \left\| V_m + T_A Y_m^{GA} B - V_m Y_m^{GA} + Y_m^{GA} + T_{m+1} \left( \begin{array}{c} 0 \\ I \end{array} \right) + \left( \begin{array}{c} \tilde{E} \\ 0 \end{array} \right) \right\|_F$$

As the matrix $V_{m+1}$ is orthogonal and $T_m^A = \left[ \begin{array}{c} T_m^A \\ T_{m+1,1,m} \end{array} \right]$, we have:

$$\|R_m\|_F = \left\| \left[ \begin{array}{c} T_m^A \\ T_{m+1,1,m} \end{array} \right] Y_m^{GA} B - Y_m^{GA} + \left( \begin{array}{c} \tilde{E} \\ 0 \end{array} \right) \right\|_F$$

Therefore:

$$\|R_m\|_F = \left\| T_{m+1,1,m} E_m^T Y_m^{GA} B \right\|_F$$

This result is very important because it allows us to calculate the Frobenius norm of $R_m(X_m^{GA})$ without having to compute the approximate solution.

Next, we give a result showing that the error $X - X_m$ is an exact solution of a perturbed Stein matrix equation.

**Theorem 4.** Let $X_m$ be the approximate solution of Equation (9) obtained after $m$ iterations of the EBA algorithm. Then:

$$(A - F_m)X_mB - X_m = E \quad (13)$$
where \( F_m = V_{m+1}T^A_{m+1,m}V_m^T \).

**Proof.** Multiplying the Equation (11) from the left by \( V_m \), we obtain:

\[
[AV_m - V_{m+1}T^A_{m+1,m}E_m]Y_mB - V_mY_m = \tilde{Y}_m
\]

(14)

As \( \tilde{V}_m \tilde{E} = E \), we get:

\[
(A - F_m)X_mB - X_m = E
\]

(15)

where:

\[
F_m = V_{m+1}T^A_{m+1,m}V_m^T
\]

\( \square \)

We can now state the following result, which gives an upper bound for the norm of the error.

**Theorem 5.** If \( \|A\|_2 < 1 \) and \( \|B\|_2 < 1 \), then we have:

\[
\|X - X_m\|_2 \leq \frac{\|T^A_{m+1,m}E_mY_mB\|_2}{1 - \|A\|_2\|B\|_2}
\]

(16)

**Proof.** By subtracting Equation (13) from Equation (9), we get:

\[
A(X - X_m)B - (X - X_m) = -F_mX_mB
\]

(17)

The error \( X_m - X \) is the solution of the Stein matrix Equation (17) and can be expressed as:

\[
X_m - X = \sum_{i=0}^{+\infty} A^i [F_mX_mB]B^i
\]

(18)

\[
\|X_m - X\|_2 \leq \sum_{i=0}^{+\infty} \|A^i\|_2 \|F_mX_mB\|_2 \|B^i\|_2
\]

(19)

\[
\leq \|F_mX_mB\|_2 \sum_{i=0}^{+\infty} (\|A\|_2\|B\|_2)^i
\]

(20)

\[
\leq \frac{\|F_mX_mB\|_2}{1 - \|A\|_2\|B\|_2}
\]

(21)

\[
\leq \frac{\|T^A_{m+1,m}E_mY_mB\|_2}{1 - \|A\|_2\|B\|_2}
\]

(22)

\( \square \)

In the next section, we present projection methods based on extended block Krylov subspaces and MR property.

4. Minimal Residual Method for Large Scale Stein Matrix Equations

In this section, we present a MR method for solving large scale Stein matrix equations. A MR method for solving large scale Lyapunov matrix equation is given in [22].

4.1. The Case: Both \( A \) and \( B \) Are Large

Instead of using a Galerkin condition as we explained in the preceding section, we consider here approximate solutions \( X_m^{MR} = \tilde{V}_mY_m^{MR}W_m \) satisfying the following minimization property:

\[
X^{MR} = \arg \min_{X_m=\tilde{V}_mY_mW_m} \|AX_mB - X_m + EF^T\|_F
\]

(23)
We have the following result.

**Theorem 6.** The solution $X_{m}^{MR}$ of the minimization problem:

$$X_{m}^{MR} = \arg \min_{X_{m}=V_{m}Y_{m}W_{m}^{T}} \|AX_{m}B - X_{m} + EF^{T}\|_{F}$$

is given by:

$$X_{m}^{MR} = V_{m}Y_{m}^{MR}W_{m}^{T}$$

where $Y_{m}^{MR}$ solves the following low dimensional minimization problem:

$$Y_{m}^{MR} = \arg \min \| \mathbb{T}^{A}Y_{m}(\mathbb{T}^{B}_{m})^{T} - \left( \begin{array}{c} 1 \\ 0 \end{array} \right) Y_{m}( \begin{array}{c} 1 \\ 0 \end{array} ) + \left( \begin{array}{c} R_{E}R_{F}^{T} \\ 0 \\ 0 \\ 0 \end{array} \right) \mathbb{W}_{m+1}^{T} \|_{F}$$

(24)

with $E = V_{1}R_{E}$ and $F = W_{1}R_{F}$, the QR factorization of $E$ and $F$, respectively.

**Proof.** We have:

$$\min_{X=V_{m}Y_{m}W_{m}^{T}} \|AXB - X + EF^{T}\|_{F}$$

$$= \min_{Y_{m}} A V_{m}Y_{m}W_{m}^{T}B - V_{m}Y_{m}W_{m}^{T} - V_{1}R_{E}R_{F}^{T}W_{1}^{T}$$

$$= \min_{Y_{m}} \| V_{m+1} \left[ \mathbb{T}^{A}Y_{m}(\mathbb{T}^{B}_{m})^{T} - \left( \begin{array}{c} 1 \\ 0 \end{array} \right) Y_{m}( \begin{array}{c} 1 \\ 0 \end{array} ) + \left( \begin{array}{c} R_{E}R_{F}^{T} \\ 0 \\ 0 \\ 0 \end{array} \right) \mathbb{W}_{m+1}^{T} \right] \|_{F}$$

One advantage of using the minimization approach is the fact that the projected problem (24) always has a solution that is not the case when one uses a GA.

The main problem is now how to solve the reduced order minimization problem (24). One possibility is the use of the preconditioned global conjugate gradient (PGCG) method.

### 4.2. The Preconditioned Global CG Method for Solving the Reduced Minimization Problem

In this section, we adopt the preconditioned conjugate gradient method (PCG) \cite{28,29} to solve the reduced minimization problem (24). The normal equation associated with (24) is given by:

$$\mathcal{L}^{*}_{m}(\mathcal{L}_{m}(Y)) = \mathcal{L}^{*}_{m}(C)$$

(25)

where:

$$\mathcal{L}_{m}(Y) = \mathbb{T}_{m}^{A}Y(\mathbb{T}_{m}^{B})^{T} - \left( \begin{array}{c} 1 \\ 0 \end{array} \right) Y( \begin{array}{c} 1 \\ 0 \end{array} )$$

Notice that $\mathcal{L}^{*}_{m}$ is the adjoint of the linear operator $\mathcal{L}_{m}$ with respect to the Frobenius inner product is given by:

$$\mathcal{L}^{*}_{m}(Z) = (\mathbb{T}_{m}^{A})^{T}Z\mathbb{T}_{m}^{B} - \left( \begin{array}{c} 1 \\ 0 \end{array} \right) Z \left( \begin{array}{c} 1 \\ 0 \end{array} \right)$$

and:

$$C = \left( \begin{array}{c} R_{E}R_{F}^{T} \\ 0 \\ 0 \\ 0 \end{array} \right)$$
We can decompose the matrices \( T_m^A \) and \( T_m^B \) as follows:

\[
T_m^A = \begin{pmatrix} T_m^A & h_m^A \end{pmatrix} \quad \text{and} \quad T_m^B = \begin{pmatrix} T_m^B & h_m^B \end{pmatrix}
\]

where \( h_m^A \) and \( h_m^B \) represent the last 2\( r \) rows of the matrices \( T_m^A \) and \( T_m^B \), respectively. Therefore, the normal Equation (25) can be written as:

\[
T_m^A T_m^A T_m^B + Y - T_m^A Y T_m^B = T_m^A Y (T_m^B)^T - \mathcal{L}_m^*(C) = 0 \tag{26}
\]

Considering the singular value decomposition (SVD) of the matrices \( T_m^A \) and \( T_m^B \):

\[
T_m^A = U_A \Sigma_A V_A^T, \quad T_m^B = U_B \Sigma_B V_B^T
\]

we get the eigendecomposition:

\[
T_m^A T_m^A = Q_A D_A Q_A^T, \quad T_m^B T_m^B = Q_B D_B Q_B^T
\]

where \( Q_A = V_A, Q_B = V_B \) and \( D_A = \Sigma_A \Sigma_A \).

Let \( \tilde{Y} = Q_A^T Y Q_B \) and \( \tilde{C} = Q_A^T \mathcal{L}_m^*(C) Q_B \), and then the normal Equation (26) is now expressed as:

\[
D_A \tilde{Y} D_B + \tilde{Y} - \tilde{T}_m^A \tilde{Y} (\tilde{T}_m^B)^T - (\tilde{T}_m^A)^T \tilde{Y} (\tilde{T}_m^B) - \tilde{C} = 0 \tag{29}
\]

where \( \tilde{T}_m^A = Q_A^T T_m^A Q_A \) and \( \tilde{T}_m^B = Q_B^T T_m^B Q_B \). This expression suggests that one can use the first part as a preconditioner, that is, the matrix operator:

\[
\mathcal{P}(\tilde{Y}) = D_A \tilde{Y} D_B + \tilde{Y} \tag{30}
\]

It can be seen that the expression (29) corresponds to the normal equation of the following matrix operator:

\[
\mathcal{E}_m(\tilde{Y}) = \tilde{T}_m^A \tilde{Y} (\tilde{T}_m^B)^T - \begin{pmatrix} Q_A^T & 0 \end{pmatrix} \tilde{Y} \begin{pmatrix} Q_B^T & 0 \end{pmatrix} \tag{31}
\]

where \( \tilde{T}_m^A = \tilde{T}_m^A Q_A \) and \( \tilde{T}_m^B = \tilde{T}_m^B Q_B \). Then, the preconditioned global CG algorithm is obtained by applying the preconditioner (30) to the normal equation associated with the matrix linear operator defined by Equation (31). This is summarized in Algorithm 3.

**Algorithm 3. The Preconditioned Global Conjugate Gradient (PGCG) Algorithm.**

1. Set \( \tilde{Y}_0 = 0 \)
   
   Compute \( \tilde{R}_0 = C - \tilde{L}_m(\tilde{Y}_0); S_0 = \tilde{L}_m(\tilde{R}_0), Z_0 = \mathcal{P}^{-1}(S_0); P_0 = S_0 \)

2. **For** \( j = 0, 1, 2, ..., j_{\text{max}} \)
   
   (a) \( W_j = \tilde{L}_m(P_j) \)
   
   (b) \( \alpha_j = \langle S_j, Z_j \rangle_j / \| W_j \|_F^2 \)
   
   (c) \( \tilde{Y}_{j+1} = \tilde{Y}_j + \alpha_j P_j \)
   
   (d) \( \tilde{R}_{j+1} = \tilde{R}_j - \alpha_j W_j \)
   
   (e) **If** \( \| \tilde{R}_{j+1} \|_F \) is small enough, then **stop**
   
   **Else**
   
   (f) \( S_{j+1} = \tilde{L}_m(\tilde{R}_{j+1}) \)
   
   (g) \( Z_{j+1} = \mathcal{P}^{-1}(S_{j+1}) \)
(h) \( \beta_j = \frac{\langle S_{j+1}, Z_{j+1} \rangle_F}{\langle S_j, Z_j \rangle_F} \)
(i) \( P_{j+1} = Z_{j+1} + \beta_j P_j \)

(3) End For

Notice that the use of the preconditioner \( P \) requires the solution, at each iteration, of a Stein equation. As the matrices \( D_A \) and \( D_B \) of these Stein matrix equations are diagonal matrices, this reduces the costs.

The MR Algorithm 4 for the Stein matrix equations is summarized as follows:


1. Choose a tolerance \( tol > 0 \), a maximum number of \( \text{itermax} \) iterations
2. For \( m = 1, 2, 3, \ldots, \text{itermax} \)
3. Update \( V_m, \hat{\Pi}_m^A \) by algorithm 1 (EBA) applied to \( (A, E) \)
4. Update \( W_m, \hat{\Pi}_m^B \) by algorithm 1 (EBA) applied to \( (B^T, F) \)
5. Solve the low order problem (24)
6. if \( \| R_m \|_F \leq tol \), stop
7. Using Equation (8), the approximate solution \( X_{m}^{MR} \) is given by \( X_m \approx Z_{1,m} Z_{2,m}^T \)

4.3. The Case: A Large and B Small

In this subsection, we apply the MR norm method to the nonsymmetric Stein Equation (9) in the case \( A \) large and \( B \) small. The approximate solution is given by:

\[
X_{m}^{MR} = V_m Y_m^{MR}
\]

with:

\[
X^{MR} = \min_{X_m = V_m Y_m} \| A X_m B - X_m + E \|_F
\]

We have the following result, which is not difficult to prove.

**Theorem 7.** The solution of the minimization problem:

\[
X_{m}^{MR} = \min_{X_m = V_m Y_m} \| A X_m B - X_m + E \|_F
\]

is given by:

\[
X_{m}^{MR} = V_m Y_m^{MR}
\]

where:

\[
Y_{m}^{MR} = \min_{Y_m} \left\| \hat{\Pi}_m^A Y_m B - \begin{bmatrix} Y_m \\ 0 \end{bmatrix} + \begin{bmatrix} R_E \\ 0 \end{bmatrix} \right\|_F
\]

with \( E = V_1 R_E \) being the QR decomposition of \( E \).

The reduced minimization problem (33) can also be solved by using the preconditioned global CG method (PGCG), as we did for the problem (24).

5. Numerical Experiments

In this section, we present some numerical experiments of large and sparse Stein matrix equations. We compared EBA-MR and EBA-GA methods. For the GA and at each iteration \( m \), we solved the projected Stein matrix equations by using the Bartels–Stewart algorithm [11]. When solving the
minimization reduced problem by the PGCG, we stopped the iterations when the relative norm of the residual was less than $t_{ol} = 10^{-12}$ or when a maximum of $k_{max} = 200$ iterations was achieved. The algorithms were coded in Matlab 8.0 (2014). The stopping criterion used for EBA-MR and GA was $\|R(X_m)\|_F < 10^{-7}$ or a maximum of $m_{max} = 100$ iterations was achieved.

In all of the examples, the coefficients of the matrices $E$ and $F$ were random values uniformly distributed on $[0, 1]$.

**Example 1.** In this first example, the matrices $A$ and $B$ are obtained from the centered finite difference discretization of the operators:

$$L_A(u) = \Delta u + f_1(x,y)\frac{\partial u}{\partial x} + f_2(x,y)\frac{\partial u}{\partial y} + f(x,y)u$$

$$L_B(u) = \Delta u + g_1(x,y)\frac{\partial u}{\partial x} + g_2(x,y)\frac{\partial u}{\partial y} + g(x,y)u$$

on the unit square $[0, 1] \times [0, 1]$ with homogeneous Dirichlet boundary conditions. The number of inner grid points in each direction was $n_0$ and $s_0$ for the operators $L_A$ and $L_B$, respectively. The matrices $A$ and $B$ were obtained from the discretization of the operator $L_A$ and $L_B$ with the dimensions $n = n_0^2$ and $s = s_0^2$, respectively. The discretization of the operator $L_A(u)$ and $L_B(u)$ yields matrices extracted from the Lyapack package [30] using the command `fdm_2d_matrix` and denoted as $A = \text{fdm}(n_0,f_1(x,y),f_2(x,y),f(x,y))$ and $B = \text{fdm}(s_0,g_1(x,y),g_2(x,y),g(x,y))$ with $f_1(x,y) = -e^{xy}$, $f_2(x,y) = -\sin(xy)$, $f(x,y) = y^2$, $g_1(x,y) = -100e^x$, $g_2(x,y) = -12xy$ and $g(x,y) = \sqrt{x^2+y^2}$. For this experiment, we used $r = 3$.

In Figure 1, we plotted the Frobenius norms of the residuals versus the number of iterations for the MR and the GAs.

![Figure 1](image-url)  
**Figure 1.** Galerkin approach (GA): dashed line, minimal residual (MR): solid line.

In Table 1, we compared the performances of the MR method and the GA. For both methods, we listed the residual norms, the maximum number of iteration and the corresponding execution time.
Table 1. Results for Example 1.

<table>
<thead>
<tr>
<th>Test Problem</th>
<th>Method</th>
<th>Iterations</th>
<th>Residual Norm</th>
<th>Times (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n=8100$, $s=3600$, $r=2$</td>
<td>GA</td>
<td>43</td>
<td>$7.56 \times 10^{-8}$</td>
<td>4.80</td>
</tr>
<tr>
<td></td>
<td>MR</td>
<td>3</td>
<td>$1.46 \times 10^{-8}$</td>
<td>1.87</td>
</tr>
</tbody>
</table>

| $n=10,000$, $s=4900$, $r=4$ | GA     | 45         | $4.99 \times 10^{-8}$ | 26.52     |
|                            | MR     | 3          | $6.28 \times 10^{-8}$ | 3.75      |

| $n=12,100$, $s=7900$, $r=3$ | GA     | 49         | $8.93 \times 10^{-8}$ | 12.96     |
|                            | MR     | 3          | $4.98 \times 10^{-8}$ | 3.63      |

**Example 2.** For the second set of experiments, we considered matrices from the University of Florida Sparse Matrix Collection [31] and from the Harwell Boeing Collection (http://math.nist.gov/MatrixMarket).

In Figure 2, we used the matrices $A = \text{pde2961}$ and $B = \text{fdm}(s_0, 100e^x, 12xy, \sqrt{x^2+y^2})$ with dimensions $n = 2961$ and $s = 3600$, respectively, and $r = 3$.

![Figure 2. GA: dashed line, MR: solid line.](image)

In Figure 3, we used the matrices $A = \text{Themal}$ and $B = \text{fdm}(s_0, e^{xy}, \sin(xy), x^2 - y^2)$ with dimensions $n = 3456$ and $s = 6400$, respectively, and $r = 3$.

![Figure 3. GA: dashed line, MR: solid line.](image)
In Table 2, we compared the performances of the MR method and the GA. For both methods, we listed the residual norms, the maximum number of iterations and the corresponding execution time.

<table>
<thead>
<tr>
<th>Test Problem</th>
<th>Method</th>
<th>Iterations</th>
<th>Residual Norm</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 2961, s = 3600, r = 2, A = \text{pde2961} ) and ( B = \text{fdm}(80, 100, 12, xy, \sqrt{x^2 + y^2}) )</td>
<td>GA</td>
<td>45</td>
<td>( 9.10 \times 10^{-9} )</td>
<td>3.7440</td>
</tr>
<tr>
<td>( n = 3456, s = 8100, r = 3, A = \text{Thermal} ) and ( B = \text{fdm}(e^x, \sin(xy), x^2 - y^2) )</td>
<td>MR</td>
<td>7</td>
<td>( 1.54 \times 10^{-9} )</td>
<td>1.0296</td>
</tr>
</tbody>
</table>

6. Conclusions

We presented in this paper two iterative methods for computing numerical solutions for large scale Stein matrix equations with low rank right-hand sides. The proposed methods are based on projection onto extended block Krylov subspaces with a Galerkin or a minimal residual approach. The approximate solutions are given as products of two low rank matrices and allow for saving memory for large problems. The numerical experiments show that the proposed Krylov-based methods are effective for large and sparse matrices.

Author Contributions: Authors have contributed equally in the mathematical part, the editorial as well as the experimental part.

Conflicts of Interest: The authors declare no conflict of interest.

References