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

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:

$$AXB - X + EF^T = 0 \tag{1}$$

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:

$$(B^T \otimes A - I_s \otimes I_n)vec(X) = -vec(EF^T) \tag{2}$$

where $vec(X)$ 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 $vec(AXB) = (B^T \otimes A)vec(X)$. 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:

$$\mathbb{K}_m(A, V) = Range(\{V, AV, A^2V, \cdots, A^{m-1}V\})$$

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

$$\begin{aligned}
\mathcal{K}_m^e(A, V) &= Range\{V, A^{-1}V, AV, A^{-2}V, A^2V, \cdots, A^{m-1}V, A^{-m}V\}) \\
&= \mathbb{K}_m(A, V) + \mathbb{K}_m(A^{-1}, A^{-1}V)
\end{aligned}$$

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 $\mathbb{V}_0 = [\,]$
(4) for $j = 1, 2, ..., m$

    (a) Set $V_j^{(1)}$: first r columns of V_j; $V_j^{(2)}$: second r columns of V_j
    (b) $\mathbb{V}_j = \left[\mathbb{V}_{j-1}, V_j\right]$; $\hat{V}_{j+1} = \left[A\, V_j^{(1)}, A^{-1}\, V_j^{(2)}\right]$
    (c) Orthogonalize $\hat{V}_{j+1}$ w.r. to $\mathbb{V}_j$ to get V_{j+1}, i.e.

        ∗ for $i = 1, 2, \dots, j$
        ∗ $H_{i,j} = V_i^T\, \hat{V}_{j+1}$;
        ∗ $\hat{V}_{j+1} = \hat{V}_{j+1} - V_i\, H_{i,j}$
        ∗ End for
    (d) Compute the QR decomposition of $\hat{V}_{j+1}$, i.e., $\hat{V}_{j+1} = V_{j+1}\, H_{j+1,j}$

(5) End For

This algorithm allows us to construct an orthonormal matrix $\mathbb{V}_m = [V_1, V_2, \dots, V_m]$ 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 $\mathbb{T}_m = \mathbb{V}_m^T\, A\, \mathbb{V}_m$.

Let $\bar{\mathbb{T}}_m = \mathbb{V}_{m+1}^T A \mathbb{V}_m$. Then, we have the following relations [25]:

$$
\begin{aligned}
A\mathbb{V}_m &= \mathbb{V}_{m+1}\bar{\mathbb{T}}_m \\
&= \mathbb{V}_m\mathbb{T}_m + V_{m+1}T_{m+1,m}\mathbb{E}_m^T
\end{aligned}
$$

where $\mathbb{E}_m = [0_{2r \times 2(m-1)r}, I_{2r}]^T$ 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 $\mathcal{K}_m^e(A, E)$ and $\mathcal{K}_m^e(B^T, F)$ associated with the pairs (A, E) and (B^T, F), respectively, and get orthonormal bases $\{V_1, V_2, ..., V_m\}$ and $\{W_1, W_2, ..., W_m\}$. We then consider approximate solutions of the Stein matrix Equation (1) that have the low-rank form:

$$
X_m^{GA} = \mathbb{V}_m Y_m^{GA} \mathbb{W}_m^T \tag{3}
$$

where $\mathbb{V}_m = [V_1, V_2, ..., V_m]$ and $\mathbb{W}_m = [W_1, W_2, ..., W_m]$.

The matrix Y_m^{GA} is determined from the following Galerkin orthogonality condition:

$$
\mathbb{V}_m^T \mathcal{R}_m^{GA} \mathbb{W}_m = \mathbb{V}_m^T (AX_m^{GA}B - X_m^{GA} + EF^T)\mathbb{W}_m = 0 \tag{4}
$$

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

$$
\mathbb{T}_A Y_m^{GA}(\mathbb{T}_B)^T - Y_m^{GA} + \widetilde{E}\widetilde{F}^T = 0 \tag{5}
$$

where $\widetilde{E} = \mathbb{V}_m^T E$, $\widetilde{F} = \mathbb{W}_m^T F$, $\mathbb{T}_A = \mathbb{V}_m^T A\mathbb{V}_m$, and $\mathbb{T}_B = \mathbb{W}_m^T B^T \mathbb{W}_m$.

Assuming that $\lambda_i(\mathbb{T}_A)\lambda_j(\mathbb{T}_B) \neq 1$ for any $i = 1, 2, ..., 2mr$ and $j = 1, 2, ..., 2mr$, the solution Y_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 X_m^{GA}.

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

$$
\|\mathcal{R}_m^G\|_F = \sqrt{\alpha_m^2 + \beta_m^2 + \gamma_m^2} \tag{6}
$$

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

$$
\gamma_m = \left\|T_{m+1,m}^A \mathbb{E}_m^T Y_m \mathbb{E}_m (T_{m+1,m}^B)^T\right\|_F
$$

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 $X - X_m^{GA}$.

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:*

$$\|X - X_m^{GA}\|_2 \leq \frac{\|A\|_2\|T_{m+1,m}^B\|_2 + \|B\|_2\|T_{m+1,m}^A\|_2 + \|T_{m+1,m}^A\|_2\|T_{m+1,m}^B\|_2}{1 - \|A\|_2\|B\|_2}\|Y_m\|_2 \qquad (7)$$

Proof. The proof is similar to the one given at Theorem 2 in [27]. □

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:

$$Y_m^{GA} = \widetilde{Y}_1 \Sigma \widetilde{Y}_2^T$$

where Σ 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 $\widetilde{Y}_1$ and $\widetilde{Y}_2$, 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:

$$X_m^{GA} \approx Z_{1,m} \, Z_{2,m}^T \qquad (8)$$

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 $tol > 0$, a maximum number of *itermax* iterations.
(3) For $m = 1, 2, 3, \ldots, itermax$
(4) Compute $\mathbb{V}_m$, $\mathbb{T}_m^A$, by Algorithm 1 applied to (A, E).
(5) Compute $\mathbb{W}_m$, $\mathbb{T}_m^B$, by Algorithm 1 applied to (B^T, F).
(6) Solve the low order Stein Equation (5) and compute $\|\mathcal{R}_m\|_F$ given by Equation (6)
(7) if $\|\mathcal{R}_m\|_F \leq tol$, 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:

$$AXB - X + E = 0 \qquad (9)$$

where E is a matrix of size $n \times s$ with $s << n$.

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

$$X_m = \mathbb{V}_m Y_m$$

where $\mathbb{V}_m$ is the orthonormal basis obtained by applying the extended block Krylov subspace $\mathcal{K}^e_m(A, E)$. The orthogonality Galerkin condition gives:

$$\mathbb{V}^T_m \mathcal{R}_m = 0 \tag{10}$$

where $\mathcal{R}_m$ is the m-th residual given by $\mathcal{R}_m = AX_mB - X_m + E$. Therefore, we obtain the projected Stein matrix equation:

$$\mathbb{T}_A Y^{GA}_m B - Y^{GA}_m + \widetilde{E} = 0 \tag{11}$$

where $\mathbb{T}_A = \mathbb{V}^T_m A \mathbb{V}_m$ and $\widetilde{E} = \mathbb{V}^T_m E$.

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

Theorem 3. *Let Y^{GA}_m the exact solution of the reduced Stein matrix Equation ([11](#)) and let $X^{GA}_m = \mathbb{V}_m Y^{GA}_m$ be the approximate solution of Equation ([9](#)) with $\mathcal{R}_m = \mathcal{R}(X^{GA}_m)$ the corresponding residual. Then:*

$$\|\mathcal{R}_m\|_F = \left\| T^A_{m+1,m} \mathbb{E}^T_m Y^{GA}_m B \right\|_F \tag{12}$$

Proof. The residual is given by $\mathcal{R}_m = AX^{GA}_m B - X^{GA}_m + E$. Since E is belonging to $\mathcal{K}^e_m(A, E)$, then $\mathbb{V}_m \mathbb{V}^T_m E = E$. Using the relation $A\mathbb{V}_m = \mathbb{V}_{m+1}\bar{\mathbb{T}}^A_m$, we have:

$$
\begin{aligned}
\|\mathcal{R}_m\|_F &= \left\| A\mathbb{V}_m Y^{GA}_m B - \mathbb{V}_m Y^{GA}_m + E \right\|_F \\
&= \left\| \mathbb{V}_{m+1}\bar{\mathbb{T}}^A_m Y^{GA}_m B - \mathbb{V}_m Y^{GA}_m + \mathbb{V}_m \mathbb{V}^T_m E \right\|_F \\
&= \left\| \mathbb{V}_{m+1}\bar{\mathbb{T}}^A_m Y^{GA}_m B - \mathbb{V}_{m+1}\begin{pmatrix} I \\ 0 \end{pmatrix} Y^{GA}_m + \mathbb{V}_{m+1}\begin{pmatrix} I \\ 0 \end{pmatrix}\widetilde{E} \right\|_F \\
&= \left\| \mathbb{V}_{m+1}\left[\bar{\mathbb{T}}^A_m Y^{GA}_m B - \begin{pmatrix} I \\ 0 \end{pmatrix} Y^{GA}_m + \begin{pmatrix} \widetilde{E} \\ 0 \end{pmatrix}\right] \right\|_F
\end{aligned}
$$

As the matrix $\mathbb{V}_{m+1}$ is orthogonal and $\overline{\mathbb{T}}^A_m = \begin{bmatrix} \mathbb{T}^A_m \\ T^A_{m+1,m}\mathbb{E}^T_m \end{bmatrix}$, we have:

$$
\begin{aligned}
\|\mathcal{R}_m\|_F &= \left\| \left[\begin{bmatrix} \mathbb{T}^A_m \\ T^A_{m+1,m}\mathbb{E}^T_m \end{bmatrix} Y^{GA}_m B - \begin{pmatrix} Y^{GA}_m \\ 0 \end{pmatrix} + \begin{pmatrix} \widetilde{E} \\ 0 \end{pmatrix} \right] \right\|_F \\
&= \left\| \left[\begin{pmatrix} \mathbb{T}^A_m Y^{GA}_m B - Y^{GA}_m + \widetilde{E} \\ T^A_{m+1,m}\mathbb{E}^T_m Y^{GA}_m B \end{pmatrix} \right] \right\|_F
\end{aligned}
$$

Therefore:

$$\|\mathcal{R}_m\|_F = \left\| T^A_{m+1,m}\mathbb{E}^T_m Y^{GA}_m B \right\|_F$$

□

This result is very important because it allows us to calculate the Frobenius norm of $\mathcal{R}_m(X^{GA}_m)$ 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:*

$$(A - F_m)X_m B - X_m = E \tag{13}$$

where $F_m = V_{m+1} T^A_{m+1,m} V^T_m$.

Proof. Multiplying the Equation (11) from the left by $\mathbb{V}_m$, we obtain:

$$[A\mathbb{V}_m - V_{m+1} T^A_{m+1,m} \mathbb{E}^T_m]Y_m B - \mathbb{V}_m Y_m = \mathbb{V}_m \tilde{E} \tag{14}$$

As $\mathbb{V}_m \tilde{E} = E$, we get:

$$(A - F_m)X_m B - X_m = E \tag{15}$$

where:

$$F_m = V_{m+1} T^A_{m+1,m} V^T_m$$

□

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:*

$$\|X - X_m\|_2 \leq \frac{\| \ T^A_{m+1,m} \mathbb{E}_m Y_m B \ \|_2}{1 - \|A\|_2 \|B\|_2} \tag{16}$$

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

$$A(X - X_m)B - (X - X_m) = -F_m X_m B \tag{17}$$

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

$$X_m - X = \sum_{i=0}^{+\infty} A^i [F_m X_m B] B^i \tag{18}$$

$$\begin{aligned}
\|X_m - X\|_2 &\leq \sum_{i=0}^{+\infty} \|A^i\|_2 \|F_m X_m B\|_2 \|B^i\|_2 & (19) \\
&\leq \|F_m X_m B\|_2 \sum_{i=0}^{+\infty} (\|A\|_2 \|B\|_2)^i & (20) \\
&\leq \frac{\|F_m X_m B\|_2}{1 - \|A\|_2 \|B\|_2} & (21) \\
&\leq \frac{\| \ T^A_{m+1,m} E_m Y_m B\|_2}{1 - \|A\|_2 \|B\|_2} & (22)
\end{aligned}$$

□

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 $X^{MR}_m = \mathbb{V}_m Y^{MR}_m \mathbb{W}^T_m$ satisfying the following minimization property:

$$X^{MR} = \arg \min_{X_m = \mathbb{V}_m Y_m \mathbb{W}^T_m} \left\| AX_m B - X_m + EF^T \right\|_F \tag{23}$$

We have the following result.

Theorem 6. *The solution X_m^{MR} of the the minimization problem:*

$$X_m^{MR} = \arg \min_{X_m = \mathbb{V}_m Y_m \mathbb{W}_m^T} \left\| A X_m B - X_m + E F^T \right\|_F$$

is given by:

$$X_m^{MR} = \mathbb{V}_m Y_m^{MR} \mathbb{W}_m^T$$

where Y_m^{MR} solves the following low dimentional minimization problem:

$$Y_m^{MR} = \arg \min \left\| \bar{\mathbb{T}}_m^A Y_m \left(\bar{\mathbb{T}}_m^B \right)^T - \begin{pmatrix} I \\ 0 \end{pmatrix} Y_m \begin{pmatrix} I & 0 \end{pmatrix} + \begin{pmatrix} R_E R_F^T & 0 \\ 0 & 0 \end{pmatrix} \right\|_F \tag{24}$$

with $E = V_1 R_E$ and $F = W_1 R_F$, the QR factorization of E and F, respectively.

Proof. We have:

$$\begin{aligned}
&\min_{X = \mathbb{V}_m Y_m \mathbb{W}_m^T} \left\| A X B - X + E F^T \right\|_F \\
&= \min_{Y_m} \left\| A \mathbb{V}_m Y_m \mathbb{W}_m^T B - \mathbb{V}_m Y_m \mathbb{W}_m^T - V_1 R_E R_F^T W_1^T \right\|_F \\
&= \min_{Y_m} \left\| \mathbb{V}_{m+1} \left[\bar{\mathbb{T}}_m^A Y_m \left(\bar{\mathbb{T}}_m^B \right)^T - \begin{pmatrix} I \\ 0 \end{pmatrix} Y_m \begin{pmatrix} I & 0 \end{pmatrix} + \begin{pmatrix} R_E R_F^T & 0 \\ 0 & 0 \end{pmatrix} \right] \mathbb{W}_{m+1}^T \right\|_F \\
&= \min_{Y_m} \left\| \left[\bar{\mathbb{T}}_m^A Y_m \left(\bar{\mathbb{T}}_m^B \right)^T - \begin{pmatrix} I \\ 0 \end{pmatrix} Y_m \begin{pmatrix} I & 0 \end{pmatrix} + \begin{pmatrix} R_E R_F^T & 0 \\ 0 & 0 \end{pmatrix} \right] \right\|_F
\end{aligned}$$

$\square$

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) [28,29] to solve the reduced minimization problem (24). The normal equation associated with (24) is given by:

$$\mathcal{L}_m^*(\mathcal{L}_m(Y)) = \mathcal{L}_m^*(\mathcal{C}) \tag{25}$$

where:

$$\mathcal{L}_m(Y) = \bar{\mathbb{T}}_m^A Y \left(\bar{\mathbb{T}}_m^B \right)^T - \begin{pmatrix} I \\ 0 \end{pmatrix} Y \begin{pmatrix} I & 0 \end{pmatrix}$$

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:

$$\mathcal{L}_m^*(Z) = (\bar{\mathbb{T}}_m^A)^T Z \bar{\mathbb{T}}_m^B - \begin{pmatrix} I & 0 \end{pmatrix} Z \begin{pmatrix} I \\ 0 \end{pmatrix}$$

and:

$$\mathcal{C} = \begin{pmatrix} R_E R_F^T & 0 \\ 0 & 0 \end{pmatrix}$$

We can decompose the matrices $\overline{\mathbb{T}}_m^A$ and $\overline{\mathbb{T}}_m^B$ as follows:

$$\overline{\mathbb{T}}_m^A = \begin{pmatrix} \mathbb{T}_m^A \\ h_m^A \end{pmatrix} \text{ and } \overline{\mathbb{T}}_m^B = \begin{pmatrix} \mathbb{T}_m^B \\ h_m^B \end{pmatrix}$$

where h_m^A and h_m^B represent the last $2r$ rows of the matrices $\overline{\mathbb{T}}_m^A$ and $\overline{\mathbb{T}}_m^B$, respectively. Therefore, the normal Equation (25) can be written as:

$$\overline{\mathbb{T}}_m^{A^T} \overline{\mathbb{T}}_m^A Y \overline{\mathbb{T}}_m^{B^T} \overline{\mathbb{T}}_m^B + Y - \mathbb{T}_m^{A^T} Y \mathbb{T}_m^B - \mathbb{T}_m^A Y (\mathbb{T}_m^B)^T - \mathcal{L}_m^*(\mathcal{C}) = 0 \tag{26}$$

Considering the singular value decomposition (SVD) of the matrices $\overline{\mathbb{T}}_m^A$ and $\overline{\mathbb{T}}_m^B$:

$$\overline{\mathbb{T}}_m^A = \overline{U}_A \overline{\Sigma}_A \overline{V}_A^T; \qquad \overline{\mathbb{T}}_m^B = \overline{U}_B \overline{\Sigma}_B \overline{V}_B^T \tag{27}$$

we get the eigendecomposition:

$$\overline{\mathbb{T}}_m^{A^T} \overline{\mathbb{T}}_m^A = Q_A D_A Q_A^T, \qquad \overline{\mathbb{T}}_m^{B^T} \overline{\mathbb{T}}_m^B = Q_B D_B Q_B^T \tag{28}$$

where $Q_A = \overline{V}_A$, $Q_B = \overline{V}_B$ and $D_A = \overline{\Sigma}_A^T \overline{\Sigma}_A$.

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

$$D_A \widetilde{Y} D_B + \widetilde{Y} - \widetilde{\mathbb{T}}_m^A \widetilde{Y} (\widetilde{\mathbb{T}}_m^B)^T - (\widetilde{\mathbb{T}}_m^A)^T \widetilde{Y} (\widetilde{\mathbb{T}}_m^B) - \widetilde{\mathcal{C}} = 0 \tag{29}$$

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

$$\mathcal{P}(\widetilde{Y}) = D_A \widetilde{Y} D_B + \widetilde{Y} \tag{30}$$

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

$$\widetilde{\mathcal{L}}_m(\widetilde{Y}) = \widetilde{\mathcal{T}}_m^A \widetilde{Y} (\widetilde{\mathcal{T}}_m^B)^T - \begin{pmatrix} Q_A \\ 0 \end{pmatrix} \widetilde{Y} \begin{pmatrix} Q_B^T & 0 \end{pmatrix} \tag{31}$$

where $\widetilde{\mathcal{T}}_m^A = \overline{\mathbb{T}}_m^A Q_A$ and $\widetilde{\mathcal{T}}_m^B = \overline{\mathbb{T}}_m^B Q_B$. 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 $\widetilde{Y}_0 = 0$
    Compute $\widetilde{R}_0 = \mathcal{C} - \widetilde{\mathcal{L}}_m(\widetilde{Y}_0)$; $S_0 = \widetilde{\mathcal{L}}_m^*(\widetilde{R}_0)$, $Z_0 = \mathcal{P}^{-1}(S_0)$; $P_0 = S_0$
(2) **For** $j = 0, 1, 2, ..., j_{max}$

    (a) $W_j = \widetilde{\mathcal{L}}_m(P_j)$
    (b) $\alpha_j = \langle S_j, Z_j \rangle_F / |W_j|_F^2$
    (c) $\widetilde{Y}_{j+1} = \widetilde{Y}_j + \alpha_j P_j$
    (d) $\widetilde{R}_{j+1} = \widetilde{R}_j - \alpha_j W_j$
    (e) **If** $\|\widetilde{R}_{j+1}\|_F$ is small enough, then **stop**
    Else
    (f) $S_{j+1} = \widetilde{\mathcal{L}}_m^*(\widetilde{R}_{j+1})$
    (g) $Z_{j+1} = \mathcal{P}^{-1}(S_{j+1})$

(h) $\beta_j = \langle S_{j+1}, Z_{j+1} \rangle_F \,/\, \langle S_j, Z_j \rangle_F$
(i) $P_{j+1} = Z_{j+1} + \beta_j P_j$

(3) **End For**

Notice that the use of the preconditioner $\mathcal{P}$ 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:

Algorithm 4. The Minimal Residual (MR) Method for Nonsymmetric Stein Matrix Equations

(1) Choose a tolerance $tol > 0$, a maximum number of *itermax* iterations
(2) For $m = 1, 2, 3, ..., itermax$
(3) Update $\mathbb{V}_m, \overline{\mathbb{T}}_m^A$, by algorithm 1 (EBA) applied to $(A,\ E)$
(4) Update $\mathbb{W}_m, \overline{\overline{\mathbb{T}}}_m^B$, by algorithm 1 (EBA) applied to $(B^T,\ F)$
(5) Solve the low order problem (24)
(6) if $\|\mathcal{R}_m\|_F \leq tol$, 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:

$$X_m^{MR} = \mathbb{V}_m Y_m^{MR}$$

with:

$$X^{MR} = \arg \min_{X_m = \mathbb{V}_m Y_m} \|A X_m B - X_m + E\|_F \tag{32}$$

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

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

$$X_m^{MR} = \arg \min_{X_m = \mathbb{V}_m Y_m} \|A X_m B - X_m + E\|_F$$

is given by:

$$X_m^{MR} = \mathbb{V}_m Y_m^{MR}$$

where:

$$Y_m^{MR} = \arg \min_{Y_m} \left\| \left[\overline{\mathbb{T}}_m^A Y_m B - \begin{bmatrix} Y_m \\ 0 \end{bmatrix} + \begin{bmatrix} R_E \\ 0 \end{bmatrix} \right] \right\|_F \tag{33}$$

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 $tol_l = 10^{-12}$ or when a maximum of $kmax = 200$ iterations was achieved. The algorithms were coded in Matlab 8.0 (2014). The stopping criterion used for EBA-MR and GA was $\|\mathcal{R}(X_m)\|_F < 10^{-7}$ 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 $A = fdm(n0,'f\_1(x,y)','f\_2(x,y)','f(x,y)')$. In this example, $n = 10,000$ and $s = 4900$, respectively, and are named as $A = \mathrm{fdm}(n0, f_1(x, y), f_2(x, y), f(x, y))$ and $B = \mathrm{fdm}(s0, g_1(x, y), g_2(x, y), g(x, y))$ 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.
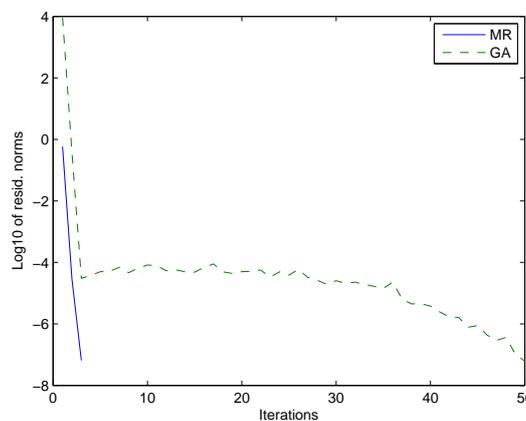
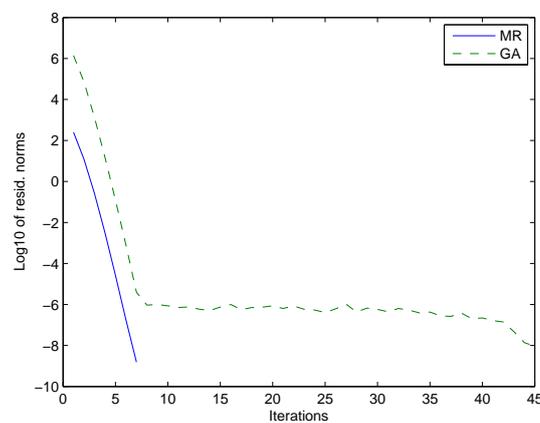**Figure 1.** Galerkin approach (GA): dashed line, minimal residual (MR): solid line.

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.

| Test Problem	Method	Iterations	Residual Norm	Times (s)
$n = 8100, s = 3600, r = 2$	GA	43	7.56×10^{-8}	4.80
	MR	**3**	1.46×10^{-8}	**1.87**
$n = 10{,}000, s = 4900, r = 4$	GA	45	4.99×10^{-8}	26.52
	MR	**3**	6.28×10^{-8}	**3.75**
$n = 12{,}100, s = 7900, r = 3$	GA	49	8.93×10^{-8}	12.96
	MR	**3**	4.98×10^{-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 $A = \texttt{pde2961}$ and $B = \texttt{fdm}(s0, 100e^x, 12xy, \sqrt{x^2 + y^2})$ with dimensions $n = 2961$ and $s = 3600$, respectively, and $r = 3$.

Figure 2. GA: dashed line, MR: solid line.

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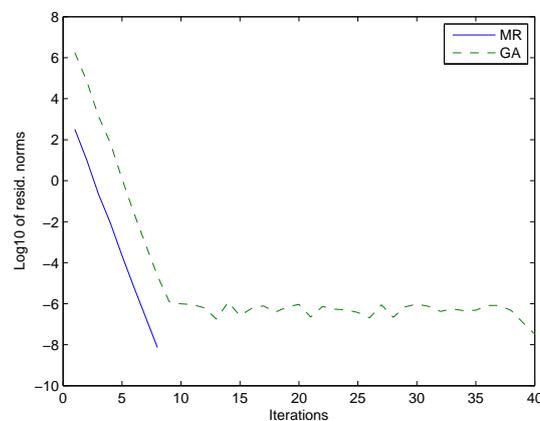
Figure 3. GA: dashed line, MR: solid line.

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 2. Results for Example 5.2.

Test Problem	Method	Iterations	Residual Norm	Time (s)
$n = 2961, s = 3600, r = 2$, A = `pde2961`	GA	45	9.10×10^{-9}	3.7440
and B = `fdm`$(s_0, 100e^x, 12xy, \sqrt{x^2 + y^2})$	MR	**7**	1.54×10^{-9}	**1.0296**
$n = 3456, s = 8100, r = 3$, A = `Thermal`	GA	40	3.27×10^{-8}	10.1245
and B = `fdm`$(e^x, \sin(xy), x^2 - y^2)$	MR	**8**	7.29×10^{-9}	**7.3008**

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

1. Bouhamidi, A.; Heyouni, M.; Jbilou, K. Block Arnoldi-based methods for large scale discrete-time algebraic Riccati equations. *J. Comput. Appl. Math.* **2011**, *236*, 1531–1542.
2. Bouhamidi, A.; Jbilou, K. Sylvester Tikhonov-regularization methods in image restoration. *J. Comput. Appl. Math.* **2007**, *206*, 86–98.
3. Zhou, B.; Lam, J.; Duan, G.-R. On Smith-type iterative algorithms for the Stein matrix equation. *Appl. Math. Lett.* **2009**, *22*, 1038–1044.
4. Zhou, B.; Lam, J.; Duan, G.-R. Toward solution of matrix equation $X = Af(X)B + C$. *Linear Algebra Appl.* **2011**, *435*, 1370–1398.
5. Zhou, B.; Duan, G.-R.; Lam, J. Positive definite solutions of the nonlinear matrix equation. *Appl. Math. Comput.* **2013**, *219*, 7377–7391.
6. Li, Z.-Y.; Zhou, B.; Lam, J. Towards positive definite solutions of a class of nonlinear matrix equations. *Appl. Math. Comput.* **2014**, *237*, 546–559.
7. Van Dooren, P. Gramian Based Model Reduction of Large-Scale Dynamical Systems. In *Numerical Analysis*; Chapman and Hall/CRC Press: London, UK, 2000; pp. 231–247.
8. Datta, B.N. *Numerical Methods for Linear Control Systems*; Academic Press: New York, NY, USA, 2003.
9. Datta, B.N.; Datta, K. Theoretical and computational aspects of some linear algebra problems in control theory. In *Computational and Combinatorial Methods in Systems Theory*; Byrnes, C.I., Lindquist, A., Eds.; Elsevier: Amsterdam, The Netherlands, 1986; pp. 201–212.
10. Calvetti, D.; Levenberg, N.; Reichel, L. Iterative methods for $X - AXB = C$. *J. Comput. Appl. Math.* **1997**, *86*, 73–101.
11. Bartels, R.H.; Stewart, G.W. Solution of the matrix equation A X + X B = C, Algorithm 432. *Commun. ACM* **1972**, *15*, 820–826.
12. Golub, G.H.; Nash, S.; van Loan, C. A Hessenberg Schur method for the problem $AX + XB = C$. *IEEE Trans. Autom. Control* **1979**, *24*, 909–913.
13. Simoncini, V. Computational methods for linear matrix equations. *SIAM Rev.* **2016**, *58*, 377–441.
14. Agoujil, S.; Bentbib, A.H.; Jbilou, K.; Sadek, E.L.M. A minimal residual norm method for large-scale Sylvester matrix equations. *Elect. Trans. Numer. Anal.* **2014**, *43*, 45–59.

15. Bentbib, A.H.; Jbilou, K.; Sadek, E.M. On some Krylov subspace based methods for large-scale nonsymmetric algebraic Riccati problems. *Comput. Math. Appl.* **2015**, 2555–2565.

16. Druskin, V.; Knizhnerman, L. Extended Krylov subspaces: Approximation of the matrix square root and related functions. *SIAM J. Matrix Anal. Appl.* **1998**, *19*, 755–771.

17. El Guennouni, A.; Jbilou, K.; Riquet, A.J. Block Krylov subspace methods for solving large Sylvester equations. *Numer. Algorithms* **2002**, *29*, 75–96.

18. Heyouni, M. Extended Arnoldi methods for large low-rank Sylvester matrix equations. *Appl. Numer. Math.* **2010**, *60*, 1171–1182.

19. Jaimoukha, I.M.; Kasenally, E.M. Krylov subspace methods for solving large Lyapunov equations. *SIAM J. Numer. Anal.* **1994**, *31*, 227–251.

20. Jbilou, K. Low-rank approximate solution to large Sylvester matrix equations. *Appl. Math. Comput.* **2006**, *177*, 365–376.

21. Jbilou, K.; Riquet, A.J. Projection methods for large Lyapunov matrix equations. *Linear Algebra Appl.* **2006**, *415*, 344–358.

22. Lin, Y.; Simoncini, V. Minimal residual methods for large scale Lyapunov equations. *Appl. Numer. Math.* **2013**, *72*, 52–71.

23. Simoncini, V. A new iterative method for solving large-scale Lyapunov matrix equations. *SIAM J. Sci. Comput.* **2007**, *29*, 1268–1288.

24. Jagels, C.; Reichel, L. Recursion relations for the extended Krylov subspace method. *Linear Algebra Appl.* **2011**, *434*, 1716–1732.

25. Heyouni, M.; Jbilou, K. An extended Block Arnoldi algorithm for large-scale solutions of the continuous-time algebraic Riccati equation. *Elect. Trans. Numer. Anal.* **2009**, *33*, 53–62.

26. Saad, Y. Numerical solution of large Lyapunov equations. In *Signal Processing, Scattering, Operator Theory and Numerical Methods*; Kaashoek, M.A., van Shuppen, J.H., Ran, A.C., Eds.; Birkhaser: Boston, MA, USA, 1990; pp. 503–511.

27. Bouhamidi, A.; Hached, M.; Heyouni, M.; Jbilou, K. A preconditioned block Arnoldi method for large Sylvester matrix equations. *Numer. Linear Algebra Appl.* **2011**, *20*, 208–219.

28. Saad, Y. *Iterative Methods for Sparse Linear Systems*, 2nd ed.; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2003.

29. Saad, Y.; Yeung, M.; Erhel, J.; Guyomarc'h, F. A deflated version of the conjugate gradient algorithm. *SIAM J. Sci. Comput.* **2000**, *21*, 1909–1926.

30. Penzl, T. LYAPACK A MATLAB Toolbox for Large Lyapunov and Riccati Equations, Model Reduction Problems, and Linear-Quadratic Optimal Control Problems. Available online: http://www.tu-chemintz.de/sfb393/lyapack (Accessed on 10 June 2016).

31. Davis, T. The University of Florida Sparse Matrix Collection, NA Digest, Volume 97, No. 23, 7 June 1997. Available online: http://www.cise.ufl.edu/research/sparse/matrices (Accessed on 10 June 2016).