International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:6, No:10, 2012

Approximate solutions to large Stein matrix equations

Khalide Jbilou

Abstract—In the present paper, we propose numerical methods for solving the Stein equation $AXC - X - D = 0$ where the matrix A is large and sparse. Such problems appear in discrete-time control problems, filtering and image restoration. We consider the case where the matrix D is of full rank and the case where D is factored as a product of two matrices. The proposed methods are Krylov subspace methods based on the block Arnoldi algorithm. We give theoretical results and we report some numerical experiments.

Keywords—IEEEtran, journal, LaTeX, paper, template.

I. Introduction

We consider the Stein matrix equation

$$AXC - X - D = 0 \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{p\times p}$, $D \in \mathbb{R}^{n\times p}$ and $X \in \mathbb{R}^{n\times p}$.
The matrix equation (1) plays an important role in linear control and filtering theory for discrete-time large-scale dynamical systems and other problems; see [5], [6], [8], [17] and the references therein. They also appear in image restoration techniques [4] and in each step of Newton's method for discrete-time algebraic Riccati equations [11]. Equation (1) is also referred to as discrete Sylvester equation.
Direct methods for solving the matrix equation (1) such as those proposed in [2], [3], [9] are attractive if the matrices are of small size. The matrix equation (1) can be formulated as an $np \times np$ large linear system using the Kronecker formulation

$$(A \otimes C^T - I_{np})\, vec(X) = vec(D) \tag{2}$$

where $\otimes$ denotes the Kronecker product; ($F \otimes G = [f_{i,j}\, G]$), $vec(X)$ is the vector of $\mathbb{R}^{np}$ obtained by stacking the columns of the matrix X and I_{np} is the $np \times np$ identity matrix. Krylov subspace methods such as the GMRES algorithm [13] could be used to solve the linear system (2). However, for large problems this approach cannot be applied directly.

The matrix Equation (1) has a unique solution if and only if $\lambda_i(A)\lambda_j(C) \neq 1$ for all $i = 1\ldots,n$; $j = 1,\ldots,p$ where $\lambda_i(A)$ is the i-th eigenvalue of the matrix A. This will be assumed through this paper. In particular, if $\rho(A)\rho(C) < 1$ where $\rho(A)$ denotes the spectral radius of the matrix A, equation (1) has a unique solution.

In this work, we present Galerkin projection methods based on the block Arnoldi algorithm [14], [15]. We first consider the case where the $n \times p$ matrix D is of full rank and $p \ll n$. The second part of this paper is devoted to the case where both matrices A and C are large and D is factored as $D = EF^T$ with a low rank.

LMPA Université du Littoral, 50 rue F. Buisson BP 699, F-62228 Calais Cedex, France E-mail: jbilou@lmpa.univ-littoral.fr.

II. The block Arnoldi Algorihm

In this section, we recall the block Arnoldi process applied to the matrix A and starting with the $n \times p$ orthonormal matrix V_1.
The block Krylov subspace $\mathcal{K}_k(A, V_1) = span\{V_1, A\,V_1, \ldots, A^{k-1}\,V_1\}$, is the subspace generated by the columns of the matrices $V_1, AV_1, \ldots, A^{k-1}V_1$.
The block Arnoldi algorithm constructs the blocks $V_1, \ldots, V_k$ whose columns form an orthonormal basis of the block Krylov subspace $\mathcal{K}_k(A, V_1)$. The algorithm is described as follows

Algorithm **1** **The block Arnoldi algorithm**

1) Choose a unitary $n \times p$ matrix V_1.
2) For $j = 1, \ldots, k$
   - $W_j = AV_j$,
   - for $i = 1, 2, \ldots, j$
     - $H_{i,j} = V_i^T\, W_j$,
     - $W_j = W_j - V_j\, H_{i,j}$,
   - end for i
   - $Q_jR_j = W_j$ (QR decomposition)
   - Set $V_{j+1} = Q_j$ and $H_{j+1,j} = R_j$.
3) End

The blocks $V_1, \ldots, V_k$ constructed by Algorithm 1 have their columns mutually orthogonal provided that the upper triangular matrices $H_{j+1,j}$ are of maximum rank. If $H_{j+1,j} = 0$ then $\mathcal{K}_j$ is invariant under A.

Let $\tilde{\mathcal{H}}_k$ denotes the $(k+1)p \times kp$ upper band-Hessenberg matrix whose nonzero entries $h_{i,j}$; $i = 1, \ldots, (k+1)p$ and $j = 1, \ldots, kp$ are defined by Algorithm 1. $\tilde{\mathcal{H}}_k$ is the $kp \times kp$ matrix obtained from $\tilde{\mathcal{H}}_k$ by deleting the last p-rows and $H_{k+1,k}$ is the $p \times p$ submatrix of the last p-rows and the last p-columns of $\tilde{\mathcal{H}}_k$.
The matrix $\mathcal{V}_k$ is defined by $\mathcal{V}_k = [V_1, \ldots, V_k]$ where V_i, $i = 1, \ldots, k$ is the i-th block constructed by the block Arnoldi algorithm. From the block Arnoldi algorithm we can deduce the following relations

$$A\,\mathcal{V}_k = \mathcal{V}_k\,\mathcal{H}_k + V_{k+1}H_{k+1,k}E_k^T; \quad A\mathcal{V}_k = \mathcal{V}_{k+1}\tilde{\mathcal{H}}_k, \tag{3}$$

and

$$\mathcal{H}_k = \mathcal{V}_k^T A \mathcal{V}_k \qquad \text{and} \qquad \mathcal{V}_k^T\,\mathcal{V}_k = I_k, \tag{4}$$

where E_k is the matrix of the last p columns of the $kp \times kp$ identity matrix I_{kp}.

III. The case where D is full rank

In this section, we consider the case where the $n \times p$ right-hand side matrix D of (1) is of full rank, C nonsingular and we assume that $p \ll n$.

Let $\mathcal{A}$ be the linear operator from $\mathbb{R}^{n\times p}$ onto $\mathbb{R}^{n\times p}$ defined as follows

$$\mathcal{A}: \ X \longrightarrow \mathcal{A}(X) = AXC - X. \tag{5}$$

Then the Stein equation (1) can be written as

$$\mathcal{A}(X) = D. \tag{6}$$

We will solve the problem (6) which is equivalent to the initial problem (1).
Starting from an initial guess X_0 and the corresponding residual $R_0 = D - A\,X_0C + X_0$, the block Arnoldi Stein method constructs, at step k, the new approximation X_k such that

$$X_k^{(i)} - X_0^{(i)} = Z_k^{(i)} \in \mathcal{K}_k(\mathcal{A}, R_0);\ i = 1,\dots,p \tag{7}$$

with the orthogonality relation

$$R_k^{(i)} \perp \mathcal{K}_k(\mathcal{A}, R_0);\ i = 1,\dots,p \tag{8}$$

where $R_k^{(i)}$ is the ith component of the residual $R_k = D - \mathcal{A}(X_k)$ and $X_k^{(i)}$ is the ith of component X_k. We give the following result which is easy to prove [7].

Theorem 1: Let $\mathcal{A}$ be the operator defined by (5) and assume that R_0 is of full rank. Then

$$\mathcal{K}_k(\mathcal{A}, R_0) = \mathcal{K}_k(A, R_0).$$

Using this last property, the relations (7) and (8) are written as

$$X_k^{(i)} - X_0^{(i)} = Z_k^{(i)} \in \mathcal{K}_k(A, R_0), \tag{9}$$

and

$$R_k^{(i)} \perp \mathcal{K}_k(A, R_0);\ i = 1,\dots,p. \tag{10}$$

Assume that R_0 is of rank p and let $R_0 = V_1U_1$ be the QR decomposition of R_0 where the $n\times p$ matrix V_1 is orthogonal and U_1 is $p\times p$ upper triangular.
Now as the columns of the matrix $\mathcal{V}_k$ (constructed by the block Arnoldi algorithm) form a basis of the block Krylov subspace $\mathcal{K}_k(A, R_0)$, the relation (9) implies that $X_k = X_0 + \mathcal{V}_k\,Y_k$ where Y_k is a $kp\times p$ matrix. Using the orthogonality relation (10), it follows that

$$\mathcal{V}_k^T\,(R_0 - A\,\mathcal{V}_k\,Y_kC + \mathcal{V}_k\,Y_k) = 0.$$

We finally obtain the low-dimensional Stein equation

$$\mathcal{H}_k\,Y_kC - Y_k = \tilde{D} \tag{11}$$

with $\tilde{D} = \tilde{E}_1U_1$ where $\tilde{E}_1$ is the $kp\times p$ matrix whose upper $p\times p$ principal block is the identity matrix.
The matrix equation (11) will be solved by using a direct method such as the Hessenberg-Schur method [5]. We assume that during the iterations $\lambda_i(\mathcal{H}_k)\,\lambda_j(C) < 1$ and this implies that the equation (11) has a unique solution.

Let us give now an expression of the residual norm that can be used to stop the iterations in the block-Arnoldi Stein algorithm without having to compute an extra product with the matrix A.

Theorem 2: At step k, the norm of the residual R_k is given by

$$\begin{aligned}\| R_k \|_F &= \| H_{k+1,k}E_k^TY_kC \|_F \\ &= \| H_{k+1,k}\tilde{Y}_kC \|_F,\end{aligned}$$

where $\tilde{Y}_k$ is the $p\times p$ matrix corresponding to the last p rows of the matrix Y_k.

Proof: At step k, the residual $R_k = D - AX_kC + X_k$, with $X_k = X_0 + \mathcal{V}_kY_k$, is expressed as

$$R_k = R_0 - A\mathcal{V}_kY_kC + \mathcal{V}_kY_k$$

and from the relation $A\,\mathcal{V}_k = \mathcal{V}_k\,\mathcal{H}_k + V_{k+1}H_{k+1,k}E_k^T$, it follows that

$$R_k = \mathcal{V}_k[\tilde{D} - \mathcal{H}_kY_kC + Y_k] - V_{k+1}H_{k+1,k}E_k^TY_kC.$$

Therefore using (11) and the fact that the matrix V_{k+1} is orthogonal the result follows. ∎
The next result shows that the approximate solution X_k is an exact solution of a perturbed Stein matrix equation.

Theorem 3: Assume that k steps of the block Arnoldi Stein method have been run and let $X_k = X_0 + \mathcal{V}_k\,Y_k$, be the obtained approximate solution to (1) where Y_k satisfies (11). Then X_k is a solution of the perturbed problem

$$(A - F_k)XC - X = D - F_kX_0C,$$

with $F_k = V_{k+1}H_{k+1,k}V_k^T$ and $\| F_k \|_F = \| H_{k+1,k} \|_F$.

Proof: Multiplying on the left the equation (11) by the matrix $\mathcal{V}_k$ we get

$$\mathcal{V}_k\mathcal{H}_kY_kC - \mathcal{V}_kY_k = \mathcal{V}_k\tilde{D}.$$

Using the relation $A\mathcal{V}_k = \mathcal{V}_k\,\mathcal{H}_k + V_{k+1}H_{k+1,k}E_k^T$ and the fact that $\mathcal{V}_k$ is orthogonal it follows that

$$A\mathcal{V}_kY_kC - V_{k+1}H_{k+1,k}E_k^T\mathcal{V}_k^T\,\mathcal{V}_kY_kC - \mathcal{V}_kY_k = \mathcal{V}_k\tilde{D}.$$

Then as $X_k = X_0 + \mathcal{V}_kY_k$, $\mathcal{V}_kE_k = V_k$ and $\mathcal{V}_k\tilde{D} = R_0$, we get

$$(A - F_k)\,X_kC - X_k = D - F_kX_0C$$

where $F_k = V_{k+1}H_{k+1,k}V_k^T$ and then $\| F_k \|_F = \| H_{k+1,k} \|_F$. ∎

Note that when $H_{k+1,k} = 0$, $F_k = 0$ and hence X_k is the exact solution of the Stein matrix equation (1). In practice, the computational requirements growth with the iteration and then the block Arnoldi algorithm will be computed in a restarted mode. The block-Arnoldi algorithm for solving (1) is summarized as follows

Algorithm **2** **The block Arnoldi algorithm for Stein equations**

1) Choose a tolerance tol, an initial guess X_0 and an integer $kmax$.
2) Compute $R_0 = D + X_0 - AX_0C$ and $R_0 = V_1U_1$: (QR decomposition.)
3) For $k = 1,\dots,kmax$,
   - Apply Algorithm 1 to the pair (A, V_1) to generate $V_1,\dots,V_{k+1}$ and the block Hessenberg matri $\mathcal{H}_k$.
   - Solve by a direct method the low-order Stein equation $\quad \mathcal{H}_kY_kC - Y_k = \tilde{D}$.

- If $\| R_k \|_F < tol$, stop.

4) End

IV. LOW-RANK APPROXIMATE SOLUTIONS TO LARGE STEIN EQUATIONS

In this section, we consider large Stein matrix equations with low-rank right-hand sides

$$AXC - X = EF^T \tag{12}$$

where $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{p\times p}$, $E \in \mathbb{R}^{n\times r}$ and $F \in \mathbb{R}^{p\times r}$. We assume that n and p are large; $r \ll n$ and $r \ll p$. From now on, we suppose that $\rho(A)\rho(C) < 1$ which ensures that (12) has a unique solution.

Equations of the form (12) arise in many application such as control theory and model reduction in large scale discrete-time dynamical sytems [17]. This is the case for example when one has to compute the controllability X_c and observability X_o Gramians by solving two symmetric Stein equations

$$AX_cA^T - X_c + EE^T = 0 \text{ and } A^TX_oA - X_o + FF^T = 0.$$

The Gramians of linear time-invariant systems play a fundamental role in many analysis and design problems such as computing the Hankel singular values, the H_2 norm of dynamical systems and model reduction techniques [8], [17].

Next, we will show how to extract low-rank approximate solutions to (12) via the block Arnoldi algorithm. At step k, let $\mathcal{K}_k(A, E)$ and $\mathcal{K}_k(C^T, F)$ be the block Krylov subspaces associated with (A, E) and (C^T, F), respectively. Consider the QR decompositions $E = V_{1,A}U_1$, $F = V_{1,C}U_2$ and apply the block Arnoldi process to the pairs (A, E) and (C^T, F) starting with $V_{1,A}$ and $V_{1,C}$ respectively. We obtain two block orthonormal bases $\{V_{1,A}, \dots, V_{k,A}\}$ and $\{V_{1,C}, \dots, V_{k,C}\}$ of the Krylov subspaces $\mathcal{K}_k(A, E)$ and $\mathcal{K}_k(C^T, F)$ respectively. We denote by $\mathcal{H}_{k,A}$ and $\mathcal{H}_{k,C}$ the block upper Hessenberg matrices given by

$$\mathcal{H}_{k,A} = \mathcal{V}_{k,A}^T\, A\, \mathcal{V}_{k,A} \qquad \text{and} \qquad \mathcal{H}_{k,C} = \mathcal{V}_{k,C}^T\, C^T\, \mathcal{V}_{k,C}$$

where $\mathcal{V}_{k,A} = [V_{1,A}, \dots, V_{k,A}]$, $\mathcal{V}_{k,C} = [V_{1,C}, \dots, V_{k,C}]$ and $\mathcal{H}_{k,A} = [H_{i,j}^A]_{i,j=1,\dots,p}$. We also have the following relations

$$A\, \mathcal{V}_{k,A} = \mathcal{V}_{k,A}\, \mathcal{H}_{k,A} + V_{k+1,A} H_{k+1,k}^A E_k^T. \tag{13}$$

and

$$C^T\, \mathcal{V}_{k,C} = \mathcal{V}_{k,C}\, \mathcal{H}_{k,C} + V_{k+1,C} H_{k+1,k}^C E_k^T. \tag{14}$$

where E_k is the matrix of the last r columns of the $kr \times kr$ identity matrix I_{kr}.

The following result gives the exact solution of (12) in terms of the two block Arnoldi bases.

Theorem 4: Let q and l be the degrees of the minimal polynomials of A for E and C^T for F respectively. Then the exact solution of the Stein equation (12) is given by

$$X = \mathcal{V}_{q,A} Z \mathcal{V}_{l,C}^T \tag{15}$$

where Z solves the problem

$$\mathcal{H}_{q,A} Z \mathcal{H}_{l,C}^T - Z = \tilde{E}\tilde{F}^T \tag{16}$$

with $\tilde{E} = \tilde{E}_1 U_1$, $\tilde{F} = \tilde{E}_1 U_2$ and $\tilde{E}_1$ is the $kp \times p$ matrix whose upper $p \times p$ principal block is the identity matrix I_p.

Proof: Since q and l are the degrees of the minimal polynomials of A for E and C^T for F, respectively, it follows that

$$A\mathcal{V}_{q,A} = \mathcal{V}_{q,A}\mathcal{H}_{q,A} \quad \text{and} \quad C^T\mathcal{V}_{l,C} = \mathcal{V}_{l,C}\mathcal{H}_{l,C}. \tag{17}$$

Multiplying on the left the two sides of (16) by $\mathcal{V}_{q,A}$ and $\mathcal{V}_{l,C}^T$ respectively, we get

$$\mathcal{V}_{q,A}\mathcal{H}_{q,A} Z \mathcal{H}_{l,C}^T \mathcal{V}_{l,C}^T - \mathcal{V}_{q,A} Z \mathcal{V}_{l,C}^T = \mathcal{V}_{q,A}\tilde{E}\tilde{F}^T\mathcal{V}_{l,C}^T. \tag{18}$$

Using (17) and the fact that $\mathcal{V}_{q,A}\tilde{E} = E$ and $\mathcal{V}_{l,C}\tilde{F} = F$, equation (18) is written as

$$A\mathcal{V}_{q,A} Z \mathcal{V}_{l,C}^T C - \mathcal{V}_{q,A} Z \mathcal{V}_{l,C}^T = EF^T.$$

This shows that $X = \mathcal{V}_{q,A} Z \mathcal{V}_{l,C}^T$ is the solution (unique) of (12). ■

Following the result of Theorem 4.1, we consider low-rank approximations of the form

$$X_k = \mathcal{V}_{k,A} Z_k \mathcal{V}_{k,C}^T \tag{19}$$

where $Z_k \in \mathbb{R}^{kp\times kp}$ is solution of the low order Stein equation

$$\mathcal{H}_{k,A} Z_k \mathcal{H}_{k,C}^T - Z_k = \tilde{E}\tilde{F}^T \tag{20}$$

The low-dimensional discrete Stein equation (20) will be solved by a direct method such as the Hessenberg-Schur method [2]. We assume that during the iterations, $\lambda_i(\mathcal{H}_{k,C})\, \lambda_i(\mathcal{H}_{k,A}) < 1$ which ensures that (20) has a unique solution.

In the following, we give some theoretical results. The next theorem shows that the low-order approximate solution X_k is a solution of a perturbed Stein equation.

Theorem 5: At step k, let X_k be the low-rank approximate solution given by (19) and (20. Then

$$(A - A_k)\, X_k\, (C - C_k) - X_k = EF^T \tag{21}$$

where $A_k = V_{k+1,A}\, H_{k+1,k}^A\, V_{k,A}^T$ and $C_k = (V_{k+1,C}\, H_{k+1,k}^C\, V_{k,V}^T)^T$

Proof: Multiplying the low order Stein equation (20) on the left by $\mathcal{V}_{k,A}$ and on the right by $\mathcal{V}_{k,C}^T$, using the relations (13) and (14) and the fact that the two matrices $\mathcal{V}_{k,A}$, $\mathcal{V}_{k,C}$ are orthogonal, we get

$$AX_kC - X_k - AX_kC_k - A_kX_kC + A_kX_kC_k = EF^T \tag{22}$$

with $A_k = V_{k+1,A}\, H_{k+1,k}^A\, V_{k,A}^T$ and $C_k = (V_{k+1,C}\, H_{k+1,k}^C\, V_{k,V}^T)^T$. This shows the result. ■

The computation of the approximation X_k given by (19) needs the product of three matrices and this becomes very expensive as k increases. In the next theorem, we show how to compute the residual norms used to stop the iterations without computing the approximation X_k. When convergence is achieved, X_k is given in a factored form and not formed explicitly.

Theorem 6: Let $X_k = \mathcal{V}_{k,A} Z_k \mathcal{V}_{k,C}^T$ be the approximate solution obtained, at step k, by the block Arnoldi Stein method where Z_k is the solution of (20) and let R_k be the corresponding residual. Then

$$\| R_k \|_F^2 = \|R_{k,1}\|^2 + \|R_{k,2}\|^2 + \|R_{k,3}\|^2 \qquad (23)$$

with $R_{k,1} = \mathcal{H}_{k,A} Z_k E_k (H_{k+1,k}^C)^T$, $R_{k,2} = H_{k+1,k}^A E_k^T Z_k \mathcal{H}_{k,C}^T$ and $R_{k,3} = H_{k+1,k}^A E_k^T Z_k E_k {H_{h+1,k}^C}^T$ where E_k is the $kp \times p$ matrix of the last p columns of the identity matrix I_{kp}.

Proof: The residual is given by $R_k = EF^T - A\mathcal{V}_{k,A} Z_k \mathcal{V}_{k,C}^T C + \mathcal{V}_{k,A} Z_k \mathcal{V}_k^T$.
Then using the relations $A\mathcal{V}_{k,A} = \mathcal{V}_{k+1,A}\,\tilde{\mathcal{H}}_{k,A}$, $C^T \mathcal{V}_{k,C} = \mathcal{V}_{k+1,C}\,\tilde{\mathcal{H}}_{k,C}$ and the expressions $\tilde{\mathcal{H}}_{k,A} = \begin{pmatrix} \mathcal{H}_{k,A} \\ H_{k+1,k}^A E_k^T \end{pmatrix}$ and $\tilde{\mathcal{H}}_{k,C} = \begin{pmatrix} \mathcal{H}_{k,C} \\ H_{k+1,k}^C E_k^T \end{pmatrix}$, the residual R_k can be expressed in a matrix form

$$R_k = \mathcal{V}_{k+1,A} \mathcal{Z}_k \mathcal{V}_{k+1,C}^T \qquad (24)$$

with

$$\mathcal{Z}_k = \begin{pmatrix} 0 & \mathcal{H}_{k,A} Z_k E_k (H_{k+1,k}^C)^T \\ H_{k+1,k}^A E_k^T Z_k \mathcal{H}_{k,C}^T & H_{k+1,k}^A E_k^T Z_k E_k (H_{k+1,k}^C)^T \end{pmatrix}$$

where Z_k solves (20). Therefore taking the norm of (24) and using the fact that $\mathcal{V}_{k+1,A} = [\mathcal{V}_{k,A}, V_{k+1,A}]$ and $\mathcal{V}_{k+1,C} = [\mathcal{V}_{k,C}, V_{k+1,C}]$ are orthonormal matrices, the result (23) follows. ■

The block-Arnoldi algorithm for solving (12) is summarized as follows

Algorithm **3** **The block Arnoldi algorithm for Stein equations**

1) Choose a tolerance tol and an integer $kmax$.
2) Compute $E = V_{1,A} U_1$ and $F = V_{1,C} U_2$: (QR)
3) For $k = 1, \ldots, kmax$
   - Apply Algorithm 1 to (A, V_1) and (C^T, V_1) to generate $V_{1,A}, \ldots, V_{k+1,A}$; $V_{1,C}, \ldots, V_{k+1,C}$ and the block Hessenberg matrices $\mathcal{H}_{k,A}$ and $\mathcal{H}_{k,C}$.
   - Solve by a direct method the low-order Stein equation: $\mathcal{H}_{k,A} Z_k \mathcal{H}_{k,C} - Z_k = \tilde{E}\tilde{F}^T$.
   - If $\| R_k \|_F < tol$, stop
4) End.

V. The symmetric Stein equation

In this section, we consider symmetric Stein equations

$$AXA^T - X + BB^T = 0 \qquad (25)$$

where $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times p}$ with $p \ll n$.
If $\rho(A) < 1$ where $\rho(A)$ denotes the spectral radius of A, the symmetric Stein equation (25) (called also Schur stable) has a unique solution given by (see [11])

$$X = \sum_{i=0}^{\infty} A^i \, BB^T \, {A^i}^T . \qquad (26)$$

As in [10], we apply the block Arnoldi algorithm to the pair (A, B) and get the matrices $\mathcal{V}_k$ and $\mathcal{H}_k$. We then consider approximations of the form $X_k = \mathcal{V}_k Z_k \mathcal{V}_k^T$ where Z_k solves the low-order symmetric Stein equation

$$\mathcal{H}_k Z_k \mathcal{H}_k^T - Z_k = \tilde{B}\tilde{B}^T. \qquad (27)$$

with $\tilde{B} = \mathcal{V}_k^T B$.
Using Theorem 6 and Theorem 7, we get the following results

Theorem 7: Let X_k be the low-rank approximate solution obtained at step k and let X be the exact solution of the symmetric Stein equation (25). Then

$$(A - A_k)\, X_k \,(A - A_k)^T - X_k + BB^T = 0 \qquad (28)$$

and

$$\| R_k \|_F^2 = 2 \,\| H_k Y_k E_k H_{k+1,k}^T \|_F^2 + \| H_{k+1,k} E_k^T Y_k E_k H_{k+1,k}^T \|_F^2 \qquad (29)$$

where E_k is the $kp \times p$ matrix of the last p columns of the identity matrix $I_{kp\times kp}$ and $A_k = V_{k+1}\, H_{k+1,k}\, V_k^T$.
In the following theorem, we give an upper bound of the norm of the error $X - X_k$ where X is the exact solution of the problem (25) and X_k is the low-rank approximate solution of (25) obtained at step k by applying the block Arnoldi algorithm.

Theorem 8: Assume that k steps of the block Arnoldi Stein algorithm have been run and let X_k be the obtained low-rank approximation. Then if $\| A \|_2 < 1$, we have

$$\| X - X_k \|_F \leq 2\sqrt{p}\, \| A \|_2 \, \frac{\| H_{k+1,k} \|_F \, \| Y_k \|_F}{1 - \| A \|_2^2}.$$

Proof: Subtracting (28) from (25), it follows that the error $X - X_k$ is the unique solution of the symmetric Stein equation

$$A(X - X_k)A^T - (X - X_k) = -AX_k A_k^T - A_k X_k A^T. \qquad (30)$$

Now since $\rho(A) < 1$, the unique solution of (30) is written as

$$X - X_k = \sum_{i=0}^{\infty} A^i \left[AX_k A_k^T + A_k X_k A^T\right] {A^i}^T.$$

Therefore

$$\| X - X_k \|_2 \leq 2 \,\| AX_k A_k^T \|_2 \sum_{i=0}^{\infty} \| A \|_2^{2i} . \qquad (31)$$

On the other hand if $G \in \mathbb{R}^{n\times p}$ we have $\| G \|_2 \leq \| G \|_F \leq \sqrt{p}\, \| G \|_2$.
Invoking the expression of A_k used in Theorem 5.1 and the fact that $X_k = \mathcal{V}_k Z_k \mathcal{V}_k^T$, we obtain

$$\| X_k A_k^T \|_F \leq \| H_{k+1,k} \|_F \, \| Y_k \|_F . \qquad (32)$$

Then using (31) and (32), we obtain the desired result. ■

VI. Numerical examples

The tests reported in this section were run on SUN Microsystems workstations using Matlab. In all our experiments, we divided the matrices A and C by $\| A \|_1$ and $\| C \|_1$ respectively.
We considered the Stein equation

$$AXC - X = EF^T$$

where the matrices A, C, E and F are of dimension $n \times n$, $p \times p$, $n \times r$ and $p \times r$ respectively with $r << n, p$.
For all our experiments, the tests where stopped when $\| R_k \|_F \leq 10^{-8}$.

TABLE I
MATRICES FROM HARWELL BOEING COLLECTION

Matrices A, B	CPU-time	iter.	res. norms
A=Sherman5 C=Serman4 $n = 3312$, $p = 1104$	0.58	7	5.9×10^{-9}
A=Pde2961 C=Fidap009 $n = 2961$, $p = 3363$	2.11	13	2.5×10^{-8}

Example 1 For this experiment, we used matrices from Harwell-Boeing Collection: Sherman4 ($n = 1104$ and $nnz(A) = 3786$), PDE2961 ($n = 2961$ and $nnz(A) =$), Sherman5 ($n = 3312$ and $nnz(A) = 20793$) and Fidap009 ($n = 3363$ and $nnz(A) = 99397$) where $nnz(A)$ denotes the number of nonzero coefficients in A.
The entries of the matrices E and F were random values uniformly distributed on $[0, 1]$ and we used $r = 4$.

In Table I, we listed the results obtained with different matrices. A maximum number of $itemax = 50$ iterations was allowed to the block-Arnoldi Stein algorithm (Algorithm 3). The expression given in Theorem 7 was used to compute the norm of the residual R_k without computing the approximation X_k which is given in a factored form when convergence is achieved.

Example 2 In this experiment, we applied the block-Arnoldi Stein algorithm (Algorithm 3) with matrices A and B defined as follows. The matrix A is generated from the 5-point discretization of the operator

$$L_1(u) = \Delta u - f_1(x, y)\,\frac{\partial u}{\partial x} - f_2(x, y)\,\frac{\partial u}{\partial y} - f_3(x, y)\,u$$

on the unit square $[0, 1] \times [0, 1]$ with homogeneous Dirichlet boundary conditions. We set $f_1(x, y) = e^{x^2+y}$, $f_2(x, y) = 2xy$ and $f_3(x, y) = cos(xy)$.
The matrix C is also generated from the 5-point discretization of the operator

$$L_2(u) = -\Delta u + g_1(x, y)\,\frac{\partial u}{\partial x} + g_2(x, y)\,\frac{\partial u}{\partial y} + g_3(x, y)\,u$$

on the unit square $[0, 1] \times [0, 1]$ with homogeneous Dirichlet boundary conditions. We set $g_1(x, y) = sin(x + 2y)$, $g_2(x, y) = e^{xy}$ and $g_3(x, y) = xy$.
The entries of the matrices E and F were random values uniformly distributed on $[0, 1]$. The dimensions of the matrices A and C are $n = n_0^2$ and $p = p_0^2$ respectively, where n_0 and p_0 are the number of inner grid points in each direction. For this experiment we used $n = 40.000$, $p = 10.000$, which corresponds to a very large linear system of dimension $4\,10^8 \times 4\,10^8$. We used different values of r ($r = 5$, $r = 10$, $r = 20$ and $r = 30$). The obtained results are reported in Table II.

TABLE II
RESULTS WITH $n = 40.000$ AND $p = 10.000$

Values of r	5	10	20	30
iteration	14	14	13	12
res. norms	2.6×10^{-8}	3.3×10^{-8}	2.2×10^{-8}	1.6×10^{-8}
cpu-time	9.9	22.2	60.3	125.1

VII. CONCLUSION

We proposed in this paper block Krylov subspace methods for solving large and sparse Stein matrix equations. We first considered the case when the right hand side is of full rank. In the second part of the paper, we showed how to apply the block Arnoldi algorithm to derive low-rank approximate solutions to Stein matrix equations with factored right-hand sides. In the two cases, we gave some theoretical results. The numerical examples show that the proposed methods are attractive and could be used for large problems.

REFERENCES

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.
[2] A. Y. BARRAUD, *A numerical algorithm to solve $A^TXA - X = Q$*, IEEE Trans. Autom. Contr., AC-22(1977), pp. 883–885.
[3] R. H. BARTELS AND G. W. STEWART, *Algorithm 432: Solution of the matrix equation AX+XB=C*, Circ. Syst. and Signal Proc., 13(1994), pp. 820–826.
[4] A. BOUHAMIDI AND K. JBILOU, *Sylvester Tikhonov-regularization methods in image restoration*, J. Comput. Appl. Math., to appear.
[5] B.N. DATTA, *Numerical Methods for Linear Control Systems*, Elsevier Academic press, 2004.
[6] B.N. DATTA, *Krylov-subspace methods for large scale matrix problems in control*, Generation of Computer Systems, 19(2003), pp. 125–126.
[7] A. EL GUENNOUNI, K. JBILOU AND A.J. RIQUET , *Block Krylov subspace methods for solving large Sylvester equations*, Numer. Alg., 29(2002), pp. 75–96.
[8] K. GLOVER, D.J.N. LIMEBEER, J.C. DOYLE, E.M. KASENALLY AND M.G. SAFONOV, *A characterisation of all solutions to the four block general distance problem*, SIAM J. Control Optim., 29(1991), pp. 283–324.
[9] G. H. GOLUB, S. NASH AND C. VAN LOAN, *A Hessenberg-Schur method for the problem AX+XB=C*, IEEC Trans. Autom. Contr., AC-24(1979), pp. 909–913.
[10] I. M. JAIMOUKHA AND E. M. KASENALLY, *Krylov subspace methods for solving large Lyapunov equations*, SIAM J. Numer. Anal., 31(1994), pp. 227–251.
[11] P. LANCASTER AND L. RODMAN, *The Algebraic Riccati Equations*, Clarendon Press, Oxford 1995.
[12] A. RUHE, *Rational Krylov sequence methods for eigenvalue computations*, Linear Algebra Appl., 58(1984), pp. 391–405.
[13] Y. SAAD AND M. H. SCHULTZ, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statis. Comput., 7(1986), pp. 856–869.
[14] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, PWS Press, New York, 1995.
[15] M. SADKANE, *Block Arnoldi and Davidson methods for unsymmetric large eigenvalue problems*, Numer. Math., 64(1993), pp. 687–706.
[16] R. SMITH, *Matrix equation XA+BX=C*, SIAM J. Appl. Math., 16(1)(1968), pp. 198–201.
[17] P. VAN DOOREN, *Gramian based model reduction of large-scale dynamical systems* , in Numerical Analysis, Chapman and Hall, pp. 231-247, CRC Press London, 2000.