World Applied Sciences Journal 12 (6): 830-834, 2011
ISSN 1818-4952

New Refinement Process for Solving Large Sparse Sylvester Equations

A.H. Refahi Sheikhani, S. Kordrostami and S. Poorjafar

Department of Mathematics, Faculty of Mathematical Sciences,
Islamic Azad University, Lahijan Branch, Lahijan, Iran

Abstract: In this paper we propose two iterative methods for solving large Sylvester equations. These methods reduce the given Sylvester equation to a Sylvester equation of smaller size by applying the weighted Arnoldi and block Arnoldi process. The numerical tests report the effectiveness of these methods.

MSC(2010)NO: 65F30 · 65F50
Key words: Phrases · Krylov subspace · Sylvester Equation · Schur · Hessenberg · Weighted Arnoldi · Refinement

INTRODUCTION

The Sylvester equation arise in a wide variety of applications, for example in many areas of control theory; see [1-4]. The existing numerical methods for solving the Sylvester equation, such as Hessenberg method is of theoretical interest only and solving the Sylvester equation cannot be recommended for use in practice. Also another group of these methods such as Schur-Hessenberg method are not suitable for large equation; see[5]. Like most large practical problem, their matrices are very sparse. The standard methods, are well known to destroy the sparsity of the problems and are attractive if the matrices are of small size; see [6-8]. A class of classical methods known as the Krylov subspace methods, that include the block Arnoldi and weighted Arnoldi, etc. have been found to be suitable for sparse matrix computations.

The Existence and Uniqueness of Solutions: In most numerical methods for solving matrix equation, it is implicity assumed that the equation to be solved has a unique solution and the methods then construct the unique solution. Thus the results on the existence and uniqueness of solutions of the Sylvester equations are of importance. We present some of these results in this section.

Theorem 2.1: Let $\lambda_1,\lambda_2,\ldots,\lambda_n$ be the eigenvalues of A and $\mu_1,\mu_2,\ldots,\mu_m$, be the eigenvalues of B. Then the Sylvester equation $XA + BX = C$ has a unique solution X if and only if $\lambda_i + \mu_j \neq 0$ for all $i = 1,\ldots,n$ and $j = 1,\ldots,m$. In other words, the Sylvester equation has a unique solution if and only if A and $-B$ do not have a common eigenvalue.

Proof: The Sylvester equation $XA + BX = C$ is equivalent to the $nm \times nm$ linear system

$$Px = c \tag{2.1}$$

Where $P = (I_n \otimes B) + (A^T \otimes I_m)$,

$$x = vec(X) = (x_{11},\ldots,x_{m1},x_{12},x_{22},\ldots,x_{m2},\ldots x_{1n},x_{2n},\ldots x_{mn})^{T},$$
$$c = vec(C) = (x_{11},\ldots,c_{m1},c_{12},c_{22},\ldots,c_{m2},\ldots c_{1n},c_{2n},\ldots x_{mn})^{T}.$$

Thus the Sylvester equation has a unique solution if and only if P is nonsingular.

Here $W \otimes Z$ is the Kronecker product of W and Z. We known that if $W = (w_{ij})$ and $Z = (z_{ij})$ are two matrices of orders $p \times p$ and $r \times r$, respectively, then their Kronecker product $W \otimes Z$ is defined by

$$W \otimes Z = \begin{pmatrix} w_{11}Z & w_{12}Z & \cdots & w_{1p}Z \\ w_{21}Z & w_{22}Z & \cdots & w_{2p}Z \\ \vdots & \vdots & & \vdots \\ w_{p1}Z & w_{p2}Z & \cdots & w_{pp}Z \end{pmatrix}$$

Thus, the Sylvester equation $XA + BX = C$ has a unique solution if and only if the matrix P of the system (2.1) is nonsingular.

Corresponding Author: H.S. Najafi,Department of Mathematics, University of Guilan,
P.O.Box 41335-1914, Rasht, Iran, E-mail: hnajafi@guilan.ac.ir.

Now, the eigenvalue of the matrix P are the mn numbers $\lambda_i + \mu_j$, where $i = 1,\ldots,n$ and $j = 1,\ldots,m$. Since the determinant of a matrix is equal to the products of its eigenvalues, this means that P is nonsingular if and only if $\lambda_i + \mu_j \neq 0$ for all $i = 1,\ldots,n$ and $j = 1,\ldots m$.

Weighted Krylov Method for Sylvester Equations: As we already mentioned, so far many numerical methods have been developed by different authors. For example the Hessenberg-Schur method is now widely used as an effective computational method for the Sylvester equation. But numerical stability of this method has not been investigated. As the iterative methods are very efficient for the solution of computational problems, therefore we thought it will be good idea to create an iterative method for solving the Sylvester equation $XA + BX = C$ where $A, B, C \in R^{n\times n}$. The following method is based on reduction of A and B to Hessenberg matrix with use of Weighted Arnoldi method.

The basis $U_m = [u_1,\ldots,u_m]$ and $V_m = [v_1,\ldots,v_m]$ constructed by the Weighted Arnoldi process are respectively D–orthonormal and $\hat{D}$– orthonormal, thus it holds

$$U_m^T D U_m = I_m \qquad V_m^T \hat{D} V_m = I_m \tag{3.1}$$

Where $U_m, V_m \in R^{n\times m}$ $(m < n)$ and $D, \hat{D} \in R^{n\times n}$ are two diagonal matrices.

The square Hessenberg matrices H_m and $\hat{H}_m$ whose nonzero entries are the scalars h_{ij} and $\hat{h}_{ij}$, constructed by the Weighted Arnoldi process can be expressed under the form

$$H_m = U_m^T D A U_m \qquad \hat{H}_m = V_m^T \hat{D} B V_m \tag{3.2}$$

Let X_0 be an initial approximate solution of the Sylvester equation and introduce the residual matrix

$$R_0 = C - (X_0 A + BX_0),$$

We wish to determine a correction F_0 and obtain a new approximate $X_1 = X_0 + F_0$. The correction F_0 can be written as

$$F_0 = \hat{D} V_m Y_m U_m^T D$$

Where $Y_m \in R^{m\times m}$ is the solution of the Sylvester equation

$$Y_m H_m + \hat{H}_m^T Y_m = V_m^T R_0 U_m. \tag{3.3}$$

Thus, the new residual matrix becomes

$$R_1 = C - (X_1 A + BX_1)$$

$$= C - ((X_0 + F_0)A + B(X_0 + F_0))$$

$$= R_0 - (F_0 A + BF_0)$$

$$= R_0 - (\hat{D} V_m Y_m U_m^T D A + B \hat{D} V_m Y_m U_m^T D)$$

Multiplying above relation on the left by V^T_M and on the right by U_m, we have

$$V_m^T R_1 U_m = V_m^T R_0 U_m - (V_m^T \hat{D} V_m Y_m U_m^T D A U_m + V_m^T B \hat{D} V_m Y_m U_m^T D U_m)$$

Now by using (3.1), (3.2) and (3.3) we get

$$V_m^T R_1 U_m = V_m^T R_0 U_m - (Y_m H_m + \hat{H}_m^T Y_m) = 0$$

In order to get $Y_m \in R^{m\times m}$, we need to solve the smaller Sylvester equation (3.3). According to the results we can develop an iterative method for the solving of the Sylvester equation. The algorithm is as follows:

Algorithm 1(Weighted Krylov method):

(1) Start:choose an initial solution X_0, new dimension m lesser than n and a tolerance ε.
(2) Compute $R_0 = C - (X_0 A + BX_0)$.
(3) Construct the D–orthonormal basis $U_m \in R^{n\times m}$ and $\hat{D}$– orthonormal basis $V_m \in R^{n\times m}$ by the Weighted Arnoldi process, such that

$$H_m = U_m^T D A U_m \qquad \hat{H}_m = V_m^T \hat{D} B V_m$$

(4) Solve the reduced Sylvester equation $Y_m H_m + \hat{H}_m^T Y_m = V_m^T R_0 U_m$.
(5) Set $X_1 = X_0 + \hat{D} V_m Y_m U_m^T D$.
(6) Compute residual matrix $R_1 = C - (X_1 A + BX_1)$.
(7) Restart: if $\|R_1\| \leq \varepsilon$ Stop
else set $X_0 = X_1$, $R_0 = R_1$ and goto 2.

Example 1: *Consider the Sylvester equation* $XA + BX = C$ *with* $n = 100$. *We apply Weighted Krylov method for solving this matrix equation and take* $\varepsilon = 10^{-6}$ *In the Table* 1, *we report the results for different value of* m.

$$A=\begin{pmatrix} 10 & 1.2 & .42 & .8 & 2.3 & .8 & 0 & \dots & \dots & 0 \\ 1.8 & 10 & 1.2 & .42 & .8 & 2.3 & .8 & 0 & \ddots & 0 \\ 1.6 & 1.8 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & \vdots \\ 1.64 & 1.6 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & .8 & 0 \\ 1.3 & 1.64 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 2.3 & .8 \\ 1.61 & 1.3 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & .8 & 2.3 \\ 0 & 1.61 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & .42 & .8 \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 1.2 & .42 \\ \vdots & \vdots & 0 & 1.61 & 1.3 & 1.64 & 1.6 & 1.8 & 10 & 1.2 \\ 0 & 0 & \dots & 0 & 1.61 & 1.3 & 1.64 & 1.6 & 1.8 & 10 \end{pmatrix}_{n\times n}$$

$$B=\begin{pmatrix} 10 & 2.1 & .38 & .7 & 1.5 & .4 & 0 & \dots & \dots & 0 \\ 1.21 & 10 & 2.1 & .38 & .7 & 1.5 & .4 & 0 & \ddots & 0 \\ 1.9 & 1.21 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & \vdots \\ .64 & 1.9 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & .4 & 0 \\ 1.9 & .64 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 1.5 & .4 \\ .87 & 1.9 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & .7 & 1.5 \\ 0 & .87 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & .38 & .7 \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 2.1 & .38 \\ \vdots & \vdots & 0 & .87 & 1.9 & .64 & 1.9 & 1.21 & 10 & 2.1 \\ 0 & 0 & \dots & 0 & .87 & 1.9 & .64 & 1.9 & 1.21 & 10 \end{pmatrix}_{n\times n}$$

$$C=\begin{pmatrix} .1 & 2.21 & 1.4 & 1.5 & .13 & 2.62 & 0 & \dots & \dots & 0 \\ 1.3 & .1 & 2.21 & 1.4 & 1.5 & .13 & 2.62 & 0 & \ddots & 0 \\ 2.6 & 1.3 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & \vdots \\ 1.7 & 2.6 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 2.62 & 0 \\ 2.3 & 1.7 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & .13 & 2.62 \\ 2.6 & 2.3 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 1.5 & .13 \\ 0 & 2.6 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 1.4 & 1.5 \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 2.21 & 1.4 \\ \vdots & \vdots & 0 & 2.6 & 2.3 & 1.7 & 2.6 & 1.3 & .1 & 2.21 \\ 0 & 0 & \dots & 0 & 2.6 & 2.3 & 1.7 & 2.6 & 1.3 & .1 \end{pmatrix}_{n\times n}$$

In Table 1, the results show that by increasing the values of m and l, the number of iterations decreases. The last column of Table 1 also shows the decreasing of time consumption. Note that the forth and fifth columns of this table are the errors of the orthogonalization method. The desired accuracy has been chosen as 10^{-6}, but the model works well with any choice of 10^{-t}.

Block Krylov Method for Large Sylvester Equations: In this section we propose to show that the obtained approximate solution of the Sylvester equation by any method can be improved, in other words the accuracy can be increased. If this idea is applicable then we have found an iterative method for solving of the Sylvester equation. Therefore let the basis $Vm = [v_1,...,v_m]$ and $Wm = [w_1,...,w_m]$ constructed by the Block Arnoldi process, thus we have

$$V_m^T V_m = I_m \qquad W_m^T W_m = I_m$$

The square block Hessenberg matrices H_m and $\hat{H}_m$ ($m = r*l$ where r and l are the dimensions of blocks) whose nonzero entries are the scalars h_{ij} and $\hat{h}_{ij}$, constructed by the Block Arnoldi process can be expressed as

$$H_m = V_m^T A^T V_m \qquad \hat{H}_m = W_m^T B W_m$$

Let X_0 be an initial approximate solution of the Sylvester equation $XA + BX = C$ where $A, B, C, X_0, \in R^{n\times n}$.

Also introduce the residual matrix

$$R_0 = C-(X_0A + BX_0)$$

And let $Y_m \in R^{m\times m}$ be the solution of the Sylvester equation:

$$Y_m H_m^T + \hat{H}_m Y_m = W_m^T R_0 V_m \tag{4.1}$$

Table 1: Implementation of Iterative Weighted Krylov method to solve the Sylvester equation with different values of m

m	r	l	$\left\Vert U_m^T DAU_m - H_m \right\Vert$	$\left\Vert V_m^T \hat{D}BV_m - \hat{H}_m \right\Vert$	*Iteration*	*Time*
4	2	2	5.33E-015	4.1E-016	397	8.91
8	2	4	2.51E-014	6.61E-016	321	7.35
10	2	5	2.98E-014	3.09E-015	261	5.89
20	2	10	3.33E-014	7.29E-015	183	4.11
30	2	15	4.11E-014	9.22E-015	138	2.86
40	2	20	7.09E-014	9.94E-015	68	1.25
50	2	25	1.21E-013	3.45E-014	18	0.2284

Table 2: Implementation of Iterative Block Krylov method to solve the Sylvester equation with different values of *m*

m	r	l	$\|U_m^T DAU_m - H_m\|$	$\|V_m^T \hat{D}BV_m - \hat{H}_m\|$	*Iteration*	*Time*
4	2	2	8.63E-014	4.05E-014	278	5.83
8	2	4	1.57E-0.13	5.49E-014	156	4.68
10	2	5	2.81E-0.13	1.98E-014	95	3.56
20	2	10	2.84E-0.13	1.77E-013	58	2.68
30	2	15	3.05E-0.13	4.46E-014	36	1.74
40	2	20	9.77E-0.13	8.36E-014	21	0.7811
50	2	25	3.17E-0.12	4.68E-013	2	0.0918

Table 3: Implementation of new Iterative methods and Hessenberg-Schur method for solving the Sylvester equation

	Hessenberg-Schur method		Weighted Krylov method		Block Krylov method		
n	Error	Time	Error	Time	Error	Time	Cond(B)
200	1.16E-010	0.6881	2.50E-012	0.4753	2.38E-014	0.2612	8.53E+033
400	3.89E-007	6.312	4.22E-008	4.642	6.15E-014	3.134	3.17E+004
600	0.0011	28.89	0.0033	65.76	6.95E-014	21.39	7.63E+005
800	8.931	85.75	13.01	174.32	8.37E-014	58.26	2.13E+007
1000	27.35	201.14	48.19	322.11	1.84E-015	121.53	1.50E+008

If set
$$X_1 = X_0 + W_m Y_m V_m^T \tag{4.2}$$

then the corresponding residual $R_1 = C-(X_1A + BX_1)$ satisfies:

$$\begin{aligned} R_1 &= C - ((X_0 + W_m Y_m V_m^T)A + B(X_0 + W_m Y_m V_m^T)) \\ &= R_0 - W_m Y_m V_m^T A - BW_m Y_m V_m^T \\ &= R_0 - W_m Y_m H_m^T V_m^T - W_m \hat{H}_m Y_m V_m^T \\ &= R_0 - W_m (Y_m H_m^T + \hat{H}_m Y_m) V_m^T \end{aligned}$$

Since Y_m is the solution of (4.1) we have:

$$R_1 = R_0 - W_m W_m^T R_0 V_m V_m^T = 0$$

According to the above results we can develop an iterative method for the solving of the Sylvester equation when the matrices A,B and C are large and sparse. For doing this idea if we choose $m < n$, then instead of solving $XB + BX = C$ we can solve (4.1). In other words in this method, first we transform the initial Sylvester equation to another Sylvester equation with less dimensions, then in each iteration step solve this matrix equation and extend the obtained solution to the solution of initial equation by (4.2). The algorithm is as follows:

Algorithm 2 (Block Krylov Method):

(1) Start: choose an initial solution X_0 and a tolerance ε.

(2) Select two numbers r and l for dimensions of block and set $m = r * l$ $(m < n)$.

(3) Compute $R_0 = C-(X_0A + BX_0)$.

(4) Construct the orthonormal basis V_m and $W_m \in R^{n\times m}$ by the Block Arnoldi process, such that

$$H_m = V_m^T A^T V_m \qquad \hat{H}_m = W_m^T BW_m$$

(5) Solve the reduced Sylvester equation $Y_m H_m^T + \hat{H}_m Y_m = W_m^T R_0 V_m$.

(6) Set $X_1 = X_0 + W_m Y_m V_m^T$.

(7) Compute $R_1 = C-(X_1A + BX_1)$.

(8) Restart: if $\|R_1\| \le \varepsilon$ Stop
else set $X_0 = X_1$, $R_0 = R_1$ and goto 3.

Example 2: *Consider the matrices A and B and C of Example 1 with n = 100. We apply the Iterative Block Krylov method for solving the XA + BX = C and take $\varepsilon = 10^{-6}$. In the Table 2, we report the results for different values of m.*

Example 3: *According to the results in Table 1 and Table 2, we see that the Block Krylov method in comparison with Weighted Krylov method works better. Now consider A and B and C are the same matrices that used in Example 3.1. We apply our two Iterative methods and Hessenberg-Schur method to solve the Sylvester equation when the dimension of the matrices are large. Results are shown in Table 3.*

COMMENTS AND CONCLUSION

1. Two new iterative methods presented in this paper can solve the Sylvester equations with any desired accuracy. Mean while, other methods like the Hessenberg-Schur method do not have this capability.
2. Results in Table 3 show that:
   (i) Weighted Krylov method works better than the Hessenberg-Schur method when the dimensions are not very large. In the case of large dimensions both methods are not working well.
   (ii) Block Krylov method works very well, when the dimension is large. Note also that in spite of having large condition numbers our new method works with high accuracy.
3. The Refinement methods presented in section 3 and 4 has the capability of improving the results obtained by any other method. Therefore we suggest using the Refinement method whenever a higher accuracy is required.
4. As we already mentioned in the Iterative method, the original Sylvester equation is first transformed to an equivalent equation with less dimension. Therefore solving the Sylvester equation with a large dimension would work very well.

REFERENCES

1. Datta, B.N., 2004. Numerical Methods for Linear Control Systems: Design and Analysis, Elsevier Academic Press, Amsterdam.
2. Datta, B.N. and K. Datta, 1986. Theoretical and computational aspects of some linear algebra problems in control theory, in: C.I. Byrnes, A. Lindquist (Eds.), Computational and Combinatorial Methods in Systems Theory, Elsevier, Amsterdam, pp: 201-212.
3. Higham, N.J., 2002. Accuracy and Stability of Numerical Algorithms, second ed., SIAM, Philadelphia, PA.
4. Hyland, C. and D. Bernstein, 1984. The optimal projection equations for fixed-order dynamic compensation, IEEE Trans. Automat. Control., 29, 1034-1037.
5. Bartels, R.H. and G.W. Stewart, 1972. Solution of the matrix equation AX + XB = C, Comm. ACM15, 820-826.
6. Golub, G.H., S. Nash and C.F. Van Loan, 1979. A Hessenberg–Shur method for the problem AX + XB = C, IEEE Trans. Automat. Control, 24: 909-913.
7. Golub, G.H. and C.F. Van Loan, 1996. Matrix Computations, third ed., Johns Hoplins U.P., Baltimore.
8. Guennouni, A. El., K. Jbilou and A.J. Riquet, 2002. Block Krylov subspace methods for solving large Sylvester equations, Numer. Alg., 29: 75-96.