Available online at www.sciencedirect.com
SCIENCE DIRECT®
Applied Mathematics Letters 19 (2006) 859–864
Applied Mathematics Letters
www.elsevier.com/locate/aml

The polynomial solution to the Sylvester matrix equation☆

Qingxi Hu, Daizhan Cheng*

Institute of Systems Science, Chinese Academy of Sciences, Beijing 100080, PR China

Received 14 April 2005; received in revised form 6 September 2005; accepted 27 September 2005

Abstract

For when the Sylvester matrix equation has a unique solution, this work provides a closed form solution, which is expressed as a polynomial of known matrices. In the case of non-uniqueness, the solution set of the Sylvester matrix equation is a subset of that of a deduced equation, which is a system of linear algebraic equations.

Keywords: Sylvester equation; Characteristic polynomial; Spectrum

1. Introduction

Let $A \in \mathbb{R}^{m\times m}$ and $B \in \mathbb{R}^{n\times n}$; the following matrix equation is called Sylvester equation:

$$AX - XB = C. \tag{1}$$

The Sylvester theorem tells us that[10,12]: for every matrix $C \in \mathbb{R}^{m\times n}$, the Sylvester equation (1) has a unique solution X if and only if $\sigma(A) \bigcap \sigma(B) = \varnothing$, where $\sigma(Z)$ denotes the spectrum of the matrix Z.

The Sylvester equation, containing the Lyapunov matrix equation as a special case, has numerous applications in control theory, signal processing, filtering, model reduction, image restoration, decoupling techniques for ordinary and partial differential equations, implementation of implicit numerical methods for ordinary differential equations, and block-diagonalization of matrices; see, for example [1,3–6,9,11] as a few references.

The problem was first discussed in a seminal book [7], where the corresponding homogeneous equation of (1) is defined as

$$AX - XB = 0. \tag{2}$$

Then the general solution X of Eq. (1) has the form

$$X = X_0 + X_1 \tag{3}$$

where X_0 is a fixed particular solution of (1), and X_1 is the general solution of Eq. (2).

☆ Supported partly by NNSF 60274010, 60221301, 60334040, 60228003 of China.
* Corresponding author. Tel.: +86 10 62651445; fax: +86 10 62587343.
E-mail addresses: huqx@mail.hainnu.edu.cn (Q. Hu), dcheng@iss03.iss.ac.cn (D. Cheng).

0893-9659/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aml.2005.09.005

Next, in order to solve Eq. (2), the author reduced the matrices A and B to their Jordan normal forms $\tilde{A}$ and $\tilde{B}$ respectively via similar transformations

$$A = U^{-1}\tilde{A}U, \qquad B = P^{-1}\tilde{B}P, \tag{4}$$

and convert then to a set of simple matrix equations.

This method depends on solving eigenvalues and converting the matrices to Jordan canonical form, which is in general very difficult. In addition, the book does not provide a method for obtaining one particular solution X_0 of Eq. (1).

Then some standard solving methods for the Sylvester equation (1) have been developed. Two widely used methods are the Stewart method [2] and the Hessenberg–Schur method [5,8]. These methods are based on transforming the coefficient matrices into Schur or Hessenberg form and then solving the corresponding linear equations directly by a back-substitution process. So these methods are called *direct methods*.

The main shortcoming of the aforementioned methods is that they do not provide an explicit formula for the solutions.

An alternative method [10] is to express (1) as

$$[I_n \otimes A - B^{\mathrm{T}} \otimes I_m]x = c \tag{5}$$

where $\otimes$ is the Kronecker product of matrices. $x = V_c(X)$ and $c = V_c(C)$ are the column stacking forms of the matrices, i.e.,

$$V_c(X) = (x_{11}, x_{21}, \dots, x_{m1}, \dots, x_{1n}, x_{2n}, \dots, x_{mn})^{\mathrm{T}}.$$

From (5) one sees easily that the coefficient matrix is non-singular, iff A and B have no common eigenvalue, which is the Sylvester theorem.

When $\sigma(A) \bigcap \sigma(B) = \varnothing$ Eq. (5) does provide a precise solution. But it is in a vector form, which means it is basically a numerical solution. In some applications it is not convenient.

In this work, we try to find the matrix form solution of the Sylvester equation when it has a unique solution. Moreover, the unique solution is a polynomial of the coefficient matrices A, B and C. When the uniqueness fails, we convert it to a common equation of the form $GX = H$. And the relationship between the solution sets of (1) and $GX = H$ is discussed. In our approach what do we need is the characteristic polynomials of the coefficient matrices A and B, which can be obtained via a routine computation.

2. Solving the Sylvester equation

In this section we provide a polynomial matrix form solution of the Sylvester equation.

Consider Eq. (1). Let

$$p(s) = \sum_{i=0}^{m} \alpha_i s^i = s^m + \alpha_{m-1}s^{m-1} + \alpha_{m-2}s^{m-2} + \cdots + \alpha_1 s + \alpha_0$$

and

$$q(s) = \sum_{i=0}^{n} \beta_i s^i = s^n + \beta_{n-1}s^{n-1} + \beta_{n-2}s^{n-2} + \cdots + \beta_1 s + \beta_0$$

(where $\alpha_m = 1$ and $\beta_n = 1$) be the characteristic polynomials of A and B, respectively.

We first prove a lemma.

Lemma 2.1. *Assume X is the solution of (1). Then for any $k \geq 1$*

$$A^k X - XB^k = \sum_{i=0}^{k-1} A^{k-1-i} C B^i. \tag{6}$$

Proof. We prove this by mathematical induction. When $k = 1$, it is exactly Eq. (1). Assume (6) holds for $k \leq N$, i.e.

$$A^N X - XB^N = \sum_{i=0}^{N-1} A^{N-1-i} C B^i. \tag{7}$$

After left-multiplying by A and right-multiplying by B on both sides of (7), we get

$$A^{N+1} X - AXB^N = A \sum_{i=0}^{N-1} A^{N-1-i} C B^i \tag{8}$$

and

$$A^N XB - XB^{N+1} = \sum_{i=0}^{N-1} A^{N-1-i} C B^i B. \tag{9}$$

Adding (8) to (9), we achieve

$$A^{N+1} X - XB^{N+1} + A^N XB - AXB^N = A \sum_{i=0}^{N-1} A^{N-1-i} C B^i + \sum_{i=0}^{N-1} A^{N-1-i} C B^i B. \tag{10}$$

Therefore,

$$\begin{aligned} &A^{N+1} X - XB^{N+1} \\ &= \sum_{i=0}^{N-1} A^{N-i} C B^i + \sum_{i=0}^{N-1} A^{N-1-i} C B^{i+1} + AXB^N - A^N XB \\ &= \sum_{i=0}^{N-1} A^{N-i} C B^i + \sum_{i=0}^{N-1} A^{N-1-i} C B^{i+1} - A(A^{N-1} X - XB^{N-1})B \\ &= \sum_{i=0}^{N-1} A^{N-i} C B^i + \sum_{i=0}^{N-1} A^{N-1-i} C B^{i+1} - A\left(\sum_{i=0}^{N-2} A^{N-2-i} C B^i\right) B \\ &= \sum_{i=0}^{N-1} A^{N-i} C B^i + \sum_{i=0}^{N-1} A^{N-1-i} C B^{i+1} - \sum_{i=0}^{N-2} A^{N-1-i} C B^{i+1} \\ &= \sum_{i=0}^{N-1} A^{N-i} C B^i + CB^N \\ &= \sum_{i=0}^{N} A^{N-i} C B^i. \quad \square \end{aligned} \tag{11}$$

Define $\eta(k, A, C, B) \triangleq \sum_{i=0}^{k} A^{k-i} C B^i$; then the equality (6) can be written in a compact form as

$$A^k X - XB^k = \eta(k-1, A, C, B). \tag{12}$$

So, we have

$$\begin{aligned} \sum_{k=1}^{n} \beta_k (A^k X - XB^k) &= \sum_{k=1}^{n} \beta_k (A^k X - XB^k) + \beta_0 (X - X) \\ &= \left(\sum_{k=1}^{n} \beta_k A^k X + \beta_0 X\right) - \left(\sum_{k=1}^{n} \beta_k XB^k + \beta_0 X\right) \\ &= \left(\sum_{k=1}^{n} \beta_k A^k + \beta_0 I_m\right) X - X\left(\sum_{k=1}^{n} \beta_k B^k + \beta_0 I_n\right) \\ &= q(A)X - Xq(B) = q(A)X. \end{aligned} \tag{13}$$

On the other hand,

$$\sum_{k=1}^{n}\beta_k(A^kX - XB^k) = \sum_{k=1}^{n}\beta_k\eta(k-1, A, C, B). \tag{14}$$

Define $\eta(A, C, B) \triangleq \sum_{k=1}^{n}\beta_k\eta(k-1, A, C, B)$. It is obvious that $\eta(A, C, B)$ is a polynomial of the matrices A, B and C. And this polynomial is determined by the coefficient matrices and the characteristic polynomial of B, which means that for each Sylvester equation of the form (1) there is a uniquely determined polynomial $\eta(A, C, B)$ of its coefficient matrices.

Consequently, we get the following equation:

$$q(A)X = \eta(A, C, B). \tag{15}$$

Theorem 2.2. *If A and B have no common eigenvalue, then (15) is equivalent to (1).*

Proof. ($\Rightarrow$): It was proved in the aforementioned argument.

($\Leftarrow$): Suppose X is a solution of (15); then

$$\begin{aligned} A(q(A)X) - (q(A)X)B &= q(A)(AX - XB) \\ &= A\eta(A, C, B) - \eta(A, C, B)B. \end{aligned}$$

Moreover,

$$\begin{aligned} &A\eta(A, C, B) - \eta(A, C, B)B \\ &\quad = \left(A\sum_{i=0}^{n}\beta_i\sum_{j=0}^{i-1}A^{i-1-j}CB^j - \sum_{i=0}^{n}\beta_i\sum_{j=0}^{i-1}A^{i-1-j}CB^jB\right) \\ &\quad = \left(\sum_{i=0}^{n}\beta_i\sum_{j=0}^{i-1}A^{i-j}CB^j - \sum_{i=0}^{n}\beta_i\sum_{j=0}^{i-1}A^{i-1-j}CB^{j+1}\right) \\ &\quad = \sum_{i=0}^{n}\beta_i\left(\sum_{j=0}^{i-1}A^{i-j}CB^j - \sum_{j=0}^{i-1}A^{i-1-j}CB^{j+1}\right) \\ &\quad = \sum_{i=0}^{n}\beta_i(A^iC - CB^i) = \sum_{i=0}^{n}\beta_iA^iC - C\sum_{i=0}^{n}\beta_iB^i \\ &\quad = q(A)C - Cq(B) = q(A)C. \end{aligned} \tag{16}$$

That is,

$$q(A)(AX - XB) = q(A)C.$$

Since $q(s)$ is the characteristic polynomial of B and A and B have no common eigenvalue, $q(A)$ is nonsingular. (1) follows. □

Proposition 2.3. *In the case of $\sigma(A)\bigcap\sigma(B) = \varnothing$ the solution of the Sylvester equation (1) is*

$$X = q(A)^{-1}\eta(A, C, B), \tag{17}$$

which is a polynomial of the matrices A, B, and C.

Proof. We have only to show that $q(A)^{-1}\eta(A, C, B)$ is a polynomial of A, B and C. It follows that what need to show is that $q(A)^{-1}$ is a polynomial of A. Let the characteristic polynomial of $q(A)$ be $f(s) = \sum_{k=0}^{m}\gamma_ks^k$ where $\gamma_m = 1$. Since $q(A)$ is invertible, we claim that $\gamma_0 \neq 0$. The Cayley–Hamilton theorem tells us that

$$f(q(A)) = \sum_{k=1}^{m}\gamma_k[q(A)]^k + \gamma_0I_m = 0. \tag{18}$$

And from the above equality we deduce that

$$q(A) - \frac{1}{\gamma_0} \sum_{k=1}^{m} \gamma_k [q(A)]^{k-1} = I_m. \tag{19}$$

So, $q(A)^{-1} = -\frac{1}{\gamma_0} \sum_{k=1}^{m} \gamma_k [q(A)]^{k-1}$, which is a polynomial of $q(A)$. But $q(A)$ is a polynomial of A. Therefore, the composition is also a polynomial of A. $\square$

For the case where A and B have common eigenvalues, Eqs. (1) and (15) are not equivalent, that is, their solution sets are not identical. We illustrate this through the following example.

Example 2.4. For

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}, \qquad C = \begin{pmatrix} 0 & 8 \\ 1 & 7 \\ 2 & 6 \end{pmatrix} \tag{20}$$

the general solution of (1) is

$$\begin{pmatrix} x_1 & -1 \\ 1 & -1 \\ 1 & -1 \end{pmatrix} \tag{21}$$

while the solution of (15) is

$$\begin{pmatrix} x_1 & x_2 \\ 1 & -1 \\ 1 & -1 \end{pmatrix} \tag{22}$$

where x_1 and x_2 are free arguments.

However, the following lemma is obvious.

Lemma 2.5. *When $\sigma(A) \bigcap \sigma(B) \neq \varnothing$, the solution set of (1) is contained in that of (15).*

The proof is obvious since every solution X of (1) satisfies (15). Although this lemma does not provide complete "equivalence" between the solution sets of (1) and (15), it is still meaningful in that we can verify the solutions of (15) to pick out the solutions Eq. (1).

3. Conclusions

In this work we proposed a quite different method for solving the well known Sylvester equation. When the solution is unique, a solution of closed form is obtained and expressed as a polynomial of known matrices. Otherwise, using the newly proposed approach, the Sylvester equation is transformed into the traditional form $GX = H$, where G, H are the coefficient matrices and X an unknown. Equations of this form can be solved in a standard way, say by the Gaussian method.

References

[1] F.A. Aliev, V.B. Larin, Optimization of Linear Control Systems: Analytical Methods and Computational Algorithms, in: Stability and Control: Theory, Methods and Applications, vol. 8, Gordon and Breach, 1998.

[2] R.H. Bartels, G.W. Stewart, Solution of the matrix equation $AX + XB = C$: Algorithm 432, Commun. ACM 15 (1972) 820–826.

[3] D. Calvetti, L. Reichel, Application of ADI iterative methods to the restoration of noisy images, SIAM J. Matrix Anal. Appl. 17 (1996) 165–186.

[4] L. Dieci, M.R. Osborne, R.D. Russel, A Riccati transformation method for solving linear bvps. I: Theoretical aspects, SIAM J. Numer. Anal. 25 (5) (1988) 1055–1073.

[5] W.H. Enright, Improving the efficiency of matrix operations in the numerical solution of stiff ordinary differential equations, ACM Trans. Math. Softw. 4 (1978) 127–136.

[6] M.A. Epton, Methods for the solution of $AXD - BXC = E$ and its applications in the numerical solution of implicit ordinary differential equations, BIT 20 (1980) 341–345.

[7] F.R. Gantmacher, The Theory of Matrices, vol. 1, Chelsea Pub. Co., 1959.
[8] G.H. Golub, S. Nash, C.F. Van Loan, A Hessenberg–Schur method for the problem $AX + XB = C$, IEEE Trans. Automat. Control AC-24 (1979) 909–913.
[9] G.H. Golub, C.F. Van Loan, Matrix computations, 3rd ed., Johns Hopkins University Press, Baltimore, 1996.
[10] P. Lancaster, M. Tismenetsky, The Theory of Matrices, 2nd ed., Academic Press, Orando, 1985.
[11] V. Sima, Algorithms for Linear–Quadratic Optimization, in: Pure and Applied Mathematics., vol. 200, Marcel Dekker, Inc., New York, NY, 1996.
[12] H.L. Trentelman, A.A. Stoorvogel, M. Hautus, Control Theory for Linear Systems, Springer, London, 2001.