

一类 Krylov 子空间方法在求解 Sylvester 方程的应用

张晓东,黄光鑫

(成都理工大学管理科学学院,成都 610059)

摘要:提出了一种求解 Sylvester 方程 $AX + XB = EF^T$ 的块 Krylov 子空间方法。当矩阵 A 和 B 非常大,并且右侧的秩很小时,给出如何求解精确低秩近似解。理论结果和数值实例证明了方法的有效性。

关键词:Krylov 子空间;Sylvester 方程;Arnoldi 算法

中图分类号:O151.21

文献标识码:A

引言

Sylvester 方程出现在很多的应用领域,如控制和通信理论,模型降解问题^[1-2]。矩阵方程

$$AX + XB = EF^T \quad (1)$$

也出现在矩阵微分 Riccati 数值解,普通或偏微分方程的去耦技术,图像修复等技术^[3-6]。

本文将研究大规模时间连续 Sylvester 方程(1)的数值解,其中未知矩阵,系数矩阵,是满秩矩阵, $r \ll n, s$ 。

假设 A 和 B 都是稳定的矩阵(即, A 和 B 的所有特征值都在开左半复平面)。在这个条件下,矩阵方程(1)有唯一的解 X ,可以表示为积分形式:

$$X = - \int_0^{+\infty} e^{At} EF^T e^{Bt} dt$$

结果对稳定矩阵 A 和 B 是有效的。如果这个条件不成立,那么应该进行其它的研究,在这个方面还不成熟。Hessenberg - Schur 方法^[7]是 Bartels - Stewart 算法^[8]的一种有效的变形,并且广泛应用于直接求解中小型 Sylvester 方程。这些方法需要将 A 和 B 中最大的简化成 Hessenberg 的形式,小的简化成实 Schur 的形式。因此, A 或 B 中的任意一个是中型或大型稀疏矩阵时,直接法将不适用求解。

对于系数矩阵 A 和 B 是中大型矩阵时,将会应用迭

代方法求解。如果矩阵稀疏并且 A 和 B 的谱信息未知,诸如基于 Arnoldi 过程^[9-10]的 Krylov 子空间方法^[11]是非常有吸引力的。当系数矩阵 A 和 B 是稠密时,如基于矩阵的符号函数的迭代方法非常有效。如果给定 A 和 B 的谱信息,可以应用 Smith 方法^[12]和交替方向隐式(ADI)迭代法^[3-4]。如果可以有效地计算出 A 和 B 的次优转换,ADI 迭代法比 Krylov 法的收敛速度快,并且可以以很小的代价有效地求出线性系统的转换系数矩阵。本文中提出一种应用到低秩时间连续 Sylvester 方程的块 Krylov 方法。

本文的符号标记:对在 $\mathbb{R}^{n \times s}$ 上的矩阵 X 和 Y ,定义下面的内积 $\langle X, Y \rangle_F = \text{trace}(X^T Y)$,其中 X^T 是 X 的转置矩阵。Frobenius 范数定义 $\|\cdot\|_F$,对于 $V \in \mathbb{R}^{n \times s}$,块 Krylov 子空间 $K_m(A, V)$ 是由矩阵 $A, AV, \dots, A^{m-1}V$ 的列生成的子空间。

1 块 Arnoldi 算法

Arnoldi 算法^[13-14]:设 V 是一个 $n \times s$ 的矩形矩阵,考虑块 Krylov 子空间

$$K_m(A, V) = \text{span}\{V, AV, \dots, A^{m-1}V\}$$

是由矩阵 $V, AV, \dots, A^{m-1}V$ 的列产生。

块 Arnoldi 算法构造了 Krylov 子空间 $K_m(A, V_1)$ 的一组正交基,算法描述如下:

算法1 块 Arnoldi 算法:

(1) 选择一个 $n \times s$ 的单位矩阵 V_1

(2) For $j = 1, \dots, m$ $W_j = AV_j$

For $i = 1, \dots, j$

$$H_{i,j} = V_i^T W_j$$

$$W_j = W_j - V_j H_{i,j}$$

$Q_j R_j = W_j$ (QR 分解, Q_j 和 R_j 的维数分别是 $n \times s, s \times s$)

$$V_{j+1} = Q_j; H_{j+1,j} = R_j$$

块 V_1, \dots, V_m 由该算法构建, 并且它们相互正交, 本文假设上三角矩阵 $H_{j+1,j}$ 有最大的秩。如果 $H_{j+1,j} = O_{s \times s}$, 那么 K_j 在 A 下不变。

矩阵 $v_m = [V_1, \dots, V_m]$ 的列形成块 Krylov 子空间 $K_m = span\{V_1, AV_1, \dots, A^{m-1}V_1\}$ 。

设 \tilde{H}_m 表示 $(ms + s) \times ms$ 的块上三角 Hessenberg 矩阵, 它的非零块 H 在算法1中定义。子矩阵 $H_{i,j}, 1 \leq i \leq j$ 是稠密的并且 $H_{i+1,i}$ 是上三角的。 H_m 是 $ms \times ms$ 矩阵, 通过 \tilde{H}_m 去掉最后 s 行得到。

从块 Arnoldi 算法, 可以推断出:

$$AV_m = V_m H_m + V_{m+1} H_{m+1} E_m^T$$

$$AV_m = V_{m+1} \tilde{H}_m, H_m = V_m^T AV_m, V_m^T V_m = I_{ms}$$

E_m 是 $ms \times ms$ 单位矩阵 I_{ms} 的最后 s 列。

用 C 替代 EFT, 方程(1)可表示为:

$$AX + XB = C \tag{2}$$

方程(2)有唯一解, 当且仅当 $Sp(A) \cap Sp(B) \neq 0$, 其中 $Sp(A)$ 是矩阵 A 的特征值集合, 在整篇文章中, 假设满足这个条件。

设 A 是线性算子, 定义如下:

$$A: \mathbb{R}^{n \times s} \rightarrow \mathbb{R}^{n \times s}$$

$$X \rightarrow A(X) = AX + XB$$

Sylvester 方程(1)被写成

$$A(X) = C \tag{3}$$

求解方程(3)等价于求解原始方程(1)。

2 块 GMRES - Sylvester 法

提出求解(2)近似解的方法是基于应用块 GMRES 法到方程(3)中。块 GMRES-Sylvester (BGS) 法从初始假设 X_0 开始计算, 到矩阵 Sylvester 方程(2)的近似解 X_m^G , 满足 $X_m^G - X_0 = V_m Y_m^G \in K_m(A, R_0)$ ($Y_m^G \in \mathbb{R}^{ms \times s}$), 这里的 R_0 表示初始残差, 对初始假设 $X_0, R_0 = C - AX_0 - X_0 B$, 并且残差 $R_m^G = R_0 - AX_m^G - X_m^G B$ 的 F 范数最小。可

以得到性质 $X_m^G - X_0 = Z_m \in K_m(A, R_0)$ 和正交关系 $R_m^G = C - A(X_m^G) \perp K_m(A, R_0)$ 。

命题1 假设 A 是之前定义的算子, 并且假定 R_0 是满秩的。那么, $K_m(A, R_0) = K_m(A, R_0)$, 所以, $X_m^G - X_0 = Z_m \in K_m(A, R_0), R_m^G = C - A(X_m^G) \perp K_m(A, R_0)$ 。

假设 R_0 的秩为 s , 并设 $R_0 = V_1 U_1$ 是 R_0 的 QR 分解, $n \times s$ 的矩阵 V_1 是正交的, U_1 是 $s \times s$ 的上三角矩阵。

由正交关系可以得出:

$$V_m^T (R_0 - AV_m Y_m^G + V_m Y_m^G B) = 0$$

所以, 低维的 Sylvester 方程:

$$H_m Y_m + Y_m B = \tilde{C}$$

有 $\tilde{C} = V_m^T R_0 = \tilde{E}_1 U_1$, \tilde{E}_1 是 $ms \times s$ 矩阵, 它的上 $s \times s$ 主要块是单位矩阵。

使用满足算法1的关系式, 残差 R_m^G 可以写成:

$$R_m^G = V_{m+1} \begin{pmatrix} H_m Y_m^G + Y_m^G B - \tilde{C} \\ H_{m+1,m} E_m^T Y_m^G \end{pmatrix}$$

但是, 由于矩阵 v_{m+1} 的列是正交的, 有:

$$\min_{Y_m^G} \|R_m^G\|_F = \min_{Y_m^G} \left\| \begin{pmatrix} H_m Y_m^G + Y_m^G B - \tilde{C} \\ H_{m+1,m} E_m^T Y_m^G \end{pmatrix} \right\|_F \tag{4}$$

命题2 最小值问题(4)等价于新的问题:

$$H_m^T (H_m Y_m^G + Y_m^G B - \tilde{C}) - (H_m Y_m^G + Y_m^G B - \tilde{C}) B^T + E_m H_{m+1,m}^T H_{m+1,m} E_m^T Y_m^G = 0 \tag{5}$$

残差范数满足:

$$\|R_m^G\|_F = [\text{trace}(\tilde{C}^T (-H_m Y_m^G - Y_m^G B + \tilde{C}))]^{1/2} \tag{6}$$

证明 从关系(4)式, 残差的 Frobenius 范数可以转换成

$$\|R_m^G\|_F^2 = \|H_m Y_m^G + Y_m^G B - \tilde{C}\|_F^2 + \|H_{m+1,m} E_m^T Y_m^G\|_F^2 \tag{7}$$

定义线性操作符 Φ_m 和 Ψ_m 如下: 对 $Y \in \mathbb{R}^{ms \times s}$, $\Phi_m(Y) = H_m Y + Y B, \Psi_m(Y) = H_{m+1,m} E_m^T Y$, (7)式可以表示为:

$$\|R_m^G\|_F^2 = \|\Phi_m(Y_m^G) - \tilde{C}\|_F^2 + \|\Psi_m(Y_m^G)\|_F^2 \tag{8}$$

因此 Y_m^G 是最小的残差 F 范数当且仅当他满足下面的线性矩阵方程:

$$\Phi_m^T(\Phi_m(Y_m^G) - \tilde{C}) + \Psi_m^T(\Psi_m(Y_m^G)) = 0 \tag{9}$$

其中线性操作符 Φ_m^T 和 Ψ_m^T 分别是操作符 Φ_m 和 Ψ_m 的转置。因此方程(9)可以写成:

$$H_m^T (H_m Y_m^G + Y_m^G B - \tilde{C}) - (H_m Y_m^G + Y_m^G B - \tilde{C}) B^T +$$

$$E_m H_{m+1,m}^T H_{m+1,m} H_m^T Y_m^G = 0$$

之后,给定残差范数

$$R_{mF}^G = - \langle \tilde{C}, \Phi_m(Y_m^G) - \tilde{C} \rangle_F = [trace(\tilde{C}^T(-H_m Y_m^G - Y_m^G B + \tilde{C}))]^{1/2}$$

证明结束。(6)式中的残差范数为 BGS 算法的实现给出了一个终止准则。它可以在不用精确计算残差的情况下测试其收敛性。

证明广义 Sylvester 矩阵方程的解 Y_m^G 存在,并且假若 H_m 和 B 没有相同特征值时唯一。

命题 3 假定在块 Arnoldi 算法执行到第 m 步,并且 $Sp(H_m) \cap Sp(B) = \emptyset$,那么

(1) 问题(5)式的解存在并且唯一。

(2) 对应的残差 R_m^G 满足 Galerkin 条件 $v_m^T(A^T R_m^G - R_m^G B^T) = 0$ 。

证明 设 $A_{m,1} = I_s \otimes H_m - B^T \otimes I_{ms}$ 和 $A_{m,2} = I_s \otimes (H_{m+1,m} E_m^T)$ 分别表示操作符 Φ_m 和 Ψ_m 的相关矩阵,假定

$$A_m = \begin{pmatrix} A_{m,1} \\ A_{m,2} \end{pmatrix} \tag{10}$$

问题(4)可以被写成

$$\min \|A_m \text{vec}(Y_m^G) - \text{vec}(\tilde{C})\|_2 \tag{11}$$

但是,因为矩阵 H_m 和 B 没有相同的特征值,矩阵 $A_{m,1}$ 是满秩,因此 A_m 也是满秩,并且先前的最小值问题有唯一解。命题的第二部分通过使用块 Arnoldi 算法生成的块的性质很容易证明。重启块 GMRES - Sylvester (BGS(m)) 总结如下:

算法 2 (块 GMRES - Sylvester (BGS)(m))

(1) 选择原始 $n \times s$ 矩阵 X_0 。

(2) 计算 $R_0 = C - AX_0 - X_0 B$ 和 QR 分解 $R_0 = V_1 U_1$, 其中 U_1 是 $s \times s$ 上三角矩阵, V_1 是 $n \times s$ 正交矩阵。

(3) 对 $k = 1, \dots, k_{max}$

(i) 利用算法 1 生成块 Arnoldi 块 V_1, \dots, V_m 和矩阵 $H_m = [H_{i,j}]_{i,j=1,\dots,m}$ 。

(ii) 求解问题(5)并计算近似解 $X_m^G = X_0 + v_m Y_m^G$ 。

(iii) 测试收敛性;如果 $\|R_m^G\|_F < tol$ 那么停止。

(iv) 设 $X_0 = X_m^G$ 回到第 2 步。

结束。

3 数值实例

给出 Sylvester 矩阵方程: $AX + XB = C$, 这里 A 和 B 是三角矩阵。

$$A = \text{tridiag}(-1 - vh, 2, -1 + vh)$$

$$B = \text{tridiag}(-1 - vk, 2, -1 + vk)$$

其中, $h = 1/(n + 1), k = 1/(s + 1)$ 。权值矩阵 C 是被随机值均匀分布在 $[0, 1]$ 上, 这里选取 $X_0 = 0$ 。

表 1 是选取 $n = 3000, s = 10, v = 10$ 。块 GMRES-Sylvester 算法得到的结果。这里选择重启步数为 2。

表 1 块 GMRES-Sylvester (BGS) 法

Method	BGS(2)
Iterations	16
Flops	2.0×10^8
Residual norms	$4e - 09$

表 2 是选取 $n = 4000, s = 10, m = 3, v = 1$, 块 GMRES-Sylvester 算法得到的结果。这里选择重启步数为 3。

表 2 块 GMRES-Sylvester (BGS) 法

Method	BGS(3)
Iterations	26
Flops	7.2×10^8
Residual norms	$6.7e - 09$

4 结束语

本文提出了一种块 Krylov 子空间算法,求解 Sylvester 矩阵方程。由于最小化性质,GMRES Sylvester 法看起来更加强健,但是对相当大的值 s 需要更多的计算。数值实例证明了方法的有效性。

参考文献:

[1] Datta B N, Datta K. Theoretical and computational aspects of some linear algebra problems in control theory[M]. Amsterdam: Elsevier, 1986.

[2] Van Dooren P. Gramian based model reduction of large scale-dynamical systems[M]. London: Chapman, Hall (eds). CRC press, 2000.

[3] Calvetti D, Reichel L. Application of ADI iterative methods to the restoration of noisy images[J]. SIAM Journal on Matrix Analysis and Applications, 1996, 17: 165-186.

[4] Baur U, Benner P. Factorized Solution of Lyapunov Equations Based on Hierarchical Matrix Arithmetic[J]. Computing, 2006, 78(3): 211-234.

[5] Datta B N. Numerical Methods for Linear Control Systems Design and Analysis[M]. Amsterdam: Elsevier Academic Press, 2003.

[6] 叶超, 胡劲松. Riccati 方程的几个可积类型[J]. 四川

- 理工学院学报:自然科学版,2008,21(4):18-20.
- [7] Golub G H, Nash S, Van Loan C. A Hessenberg-Schur method for the problem $AX + XB = C$ [J]. IEEE Transaction on Automatic Control, 1979, 24: 909-913.
- [8] Bartels R H, Stewart G W. Algorithm 432: Solution of the matrix equation $AX + XB = C$ [J]. Communications of the ACM, 1972, 15: 820-826.
- [9] El Guennouni A, Jbilou K, Riquet A J. Block Krylov subspace methods for solving large Sylvester equations [J]. Numerical Algorithms, 2002, 29: 75-96.
- [10] Jbilou K. Low rank approximate solutions to large Sylvester matrix equations [J]. Applied Mathematics and Computation, 2006, 177: 365-376.
- [11] Guennouni E A, Jbilou K, Riquet A J. Block Krylov subspace methods for solving large Sylvester equations [J]. Numerical Algorithms, 2002, 29: 75-96.
- [12] Penzl T. A cyclic low-rank Smith method for large sparse Lyapunov equations [J]. SIAM Journal on Scientific Computing, 2000, 21: 1401-1418.
- [13] Saad Y. Iterative Methods for Sparse Linear Systems [M]. New York: PWS, 1995.
- [14] Bouhamidi A A, Hached M, Heyouni M, et al. A preconditioned block Arnoldi method for large Sylvester matrix equation. Numer [J]. Numer. Lin. Alg, 2011, 236(6): 1531-1542.

Application of a Kind of Krylov Subspace Methods in Solving the Sylvester Equation

ZHANG Xiao-dong, HUANG Guang-xin

(College of Management Sciences, Chengdu University of Technology, Chengdu 610059, China)

Abstract: Block Krylov subspace methods for solving the Sylvester matrix equation $AX + XB = EF^T$ is proposed. When both matrices A and B are large and the right-hand side matrix is of small rank, it is shown that how to extract low-rank approximations. Some theoretical results are given and numerical experiments show the effectiveness of these block methods.

Key words: Krylov subspace; Sylvester equations; Arnoldi algorithm