Sylvester equations

Sylvester equation

$$AX - XB = C$$

$A \in \mathbb{C}^{m\times m}$, $C, X \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{n\times n}$.

Assume $m \geq n$ for simplicity (otherwise: transpose everything).

Kronecker products

$$X \otimes Y = \begin{bmatrix} x_{11}Y & x_{12}Y & \dots & x_{1n}Y \\ x_{21}Y & x_{22}Y & \dots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ x_{m1}Y & x_{m2}Y & \dots & x_{mn}Y \end{bmatrix}.$$

Properties:

- $(A \otimes B)(C \otimes D) = (AC \otimes BD)$, when dimensions are compatible.
- $\operatorname{vec} AXB = (B^T \otimes A) \operatorname{vec} X$. (Warning: not B^H).

One can "factor" several decompositions, e.g.,

$$A \otimes B = (U_1 S_1 V_1^T) \otimes (U_2 S_2 V_2^T) = (U_1 \otimes U_2)(S_1 \otimes S_2)(V_1 \otimes V_2)^T.$$

Solvability criterion

The Sylvester equation is solvable for all C iff $\Lambda(A) \cap \Lambda(B) = \emptyset$.

$AX - XB = C \iff$

$$(I_n \otimes A - B^T \otimes I_m)\,\mathrm{vec}(X) = \mathrm{vec}(C).$$

Schur decompositions of A, B^T: if $\Lambda(A) = \{\lambda_1, \ldots, \lambda_m\}$, $\Lambda(B) = \{\mu_1, \ldots, \mu_n\}$, then $\Lambda(I_n \otimes A - B^T \otimes I_m) = \{\lambda_i - \mu_j : i, j\}$.

Solution algorithms

The naive algorithm costs $O((mn)^3)$. One can get down to $O(m^3 n^2)$ (full steps of GMRES, for instance.)
Bartels–Stewart algorithm (1972): $O(m^3 + n^3)$.

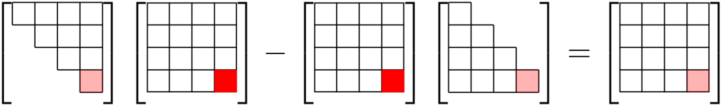
Step 1: Schur decompositions $A = Q_A T_A Q_A^*$, $B^* = Q_B T_B Q_B^*$.

$$Q_A T_A Q_A^* X - X Q_B T_B^* Q_B^* = C$$

$$T_A \widehat{X} - \widehat{X} T_B^* = \widehat{C}, \quad \widehat{X} = Q_A^* X Q_B, \widehat{C} = \widehat{Q_A^* C Q_B}.$$

Bartels–Stewart algorithm

Step 2: back-substitution.



Comments

- Works also with the real Schur form: back-sub yields block equations which are tiny 2×2 or 4×4 Sylvesters.
- Backward stable (as a system of *mn* linear equations): it's orthogonal transformations + back-sub.
- Not backward stable in the sense of $\widetilde{A}\widetilde{X} - \widetilde{X}\widetilde{B} = \widetilde{C}$ [Higham '93].
  (Sketch of proof: backward error given by a linear least squares system with matrix $\left[\widetilde{X}^T \otimes I \quad I \otimes \widetilde{X} \quad I\right]$). Its singular values depend on those of $\widetilde{X}$.)

Comments

Condition number: depends on

$$\mathrm{sep}(A, B) = \sigma_{\min}(I \otimes A - B^T \otimes I) = \min_Z \frac{\|AZ - ZB\|_F}{\|Z\|_F}.$$

(If A, B normal, simply the minimum difference of their eigenvalues.)

Decoupling eigenvalues

Solving a Sylvester equation means finding

$$\begin{bmatrix} I & -X \\ 0 & I \end{bmatrix} \begin{bmatrix} A & C \\ 0 & B \end{bmatrix} \begin{bmatrix} I & X \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}.$$

Idea Indicates how 'difficult' (ill-conditioned) it is to go from block-triangular to block-diagonal. (Compare also with the scalar case / Jordan form.)

Similar problem: reordering Schur forms (swapping blocks). One uses the Q factor from the QR of $\left[\begin{smallmatrix} I & X \\ 0 & I \end{smallmatrix}\right]\dots$

Invariant subspaces

Invariant subspace (for a matrix M): any subspace $\mathcal{U}$ such that $M\mathcal{U} \subseteq \mathcal{U}$. Completing a basis U_1 to one $U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$ of $\mathbb{C}^m$, we get

$$U^{-1}MU = \begin{bmatrix} A & C \\ 0 & B \end{bmatrix}.$$

$MU_1 = U_1 A$. $\Lambda(A) \subseteq \Lambda(M)$.
Idea: invariant subspaces are 'the span of some eigenvectors' (usually).

Sensitivity of invariant subspaces

If I perturb M to $M + \delta_M$, how much does U_1 change?

Proof (sketch:)

- Suppose $U = I$ for simplicity (just a change of basis).
- $M + \delta M = \begin{bmatrix} A+\delta_A & C+\delta_C \\ \delta_D & B+\delta_B \end{bmatrix}$
- Look for a transformation $V^{-1}(M + \delta M)V$ of the form $V = \begin{bmatrix} I & 0 \\ X & I \end{bmatrix}$ that zeroes out the $(2, 1)$ block.
- Formulate a Riccati equation
  $XA - BX = \delta_D - X(C + \delta C)X - X\delta_A + \delta_B X.$
- See as a fixed-point problem.
- Pass to norms to see when the map sends a $B(0, \rho)$ to itself: $\|X\|_F \leq \|T^{-1}\|(\dots)$. For a sufficiently small perturbation, it does.

Theorem [Stewart Sun book V.2.2]

Let $M = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$, $\delta_M = \begin{bmatrix} \delta_A & \delta_B \\ \delta_D & \delta_C, \end{bmatrix}$, $a = \|\delta_A\|$ and so on.
If $4(\operatorname{sep}(A, B) - a - b)^2 - d(\|C\| + c) \geq 0$, then there is a (unique) X with $\|X\| \leq 2\frac{d}{\operatorname{sep}(A,B)-a-b}$ such that $\begin{bmatrix} I \\ X \end{bmatrix}$ is an invariant subspace of $M + \delta_M$.

Speak about angles between subspaces $\begin{bmatrix} I \\ 0 \end{bmatrix}$ and $\begin{bmatrix} I \\ X \end{bmatrix}$.

Symmetric version ("Davis-Kahan $\sin\Theta$ theorem"):

$\|U_1^* \widetilde{U}_2\|_F \leq \frac{\|U_1^* \delta_M \widetilde{U}_2\|_F}{\delta}$, δ eigenvalue gap.