Sylvester equations

Sylvester equation

$$\boxed{AX - XB = C}$$

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

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 $O(m^3n^2)$ (full steps of GMRES, for instance.)
Bartels–Stewart algorithm (1972): $O(m^3 + n^3)$.

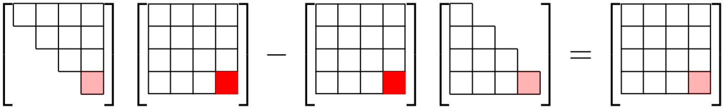
Step 1: Schur decompositions $A = Q_A T_A Q_A^*$, $B^* = Q_B T_B Q_B^*$.

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

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

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 $\begin{bmatrix} \widetilde{X}^T \otimes I & I \otimes \widetilde{X} & I \end{bmatrix}$). Its singular values depend on those of $\widetilde{X}$.)

Comments

Condition number: depends on

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

(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 $\left[\begin{smallmatrix} I & X \\ 0 & I \end{smallmatrix}\right]$...

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 $U = [\, U_1 \; U_2 \,]$ of $\mathbb{C}^m$, we get

$$U^{-1}MU = \left[\begin{array}{c|c} A & C \\ \hline 0 & B \end{array}\right] \cdot \begin{bmatrix} x \\ 0 \end{bmatrix} = \begin{bmatrix} Ax \\ 0 \end{bmatrix}$$

$\boxed{MU_1 = U_1 A.}$ $\Lambda(A) \subseteq \Lambda(M)$.

Idea: invariant subspaces are 'the span of some eigenvectors' (usually).

Example: stable invariant subspace: x s.t. $\{\lim_{k\to\infty} A^k x = 0\} =: S$

= span of all eigenvectors / Jordan chains of A with eigenvalues $|\lambda| < 1$

$\lambda_1 \neq \lambda_2 \neq \dots \lambda_n \qquad Mv_i = \lambda v_i$

$V = \operatorname{span}(v_i)$ is invariant: $M(\alpha v_i) = \alpha M v_i =$
$= \alpha \lambda_i v_i \in V$

$V = \operatorname{span}(v_1, v_2, \dots v_k)$ is invariant:

$$M(\alpha_1 v_1 + \dots + \alpha_k v_k) = \alpha_1 \lambda_1 v_1 + \dots \alpha_k \lambda_k v_k \subseteq V$$

$\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$ has 3 inv. subspaces:
$\{0\}, \left\langle \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\rangle, \mathbb{C}^2$

$\begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}$ has inv. subspaces $\underset{\{0\}}{}\left\langle \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\rangle, \left\langle \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix} \right\rangle, \left\langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & 0 \\ 0 & 0 & 0 \end{bmatrix} \right\rangle, \dots$

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 $M = \begin{bmatrix} A & B \\ \boxed{0} & C \end{bmatrix}$, $\delta_M = \begin{bmatrix} \delta_A & \delta_B \\ \delta_D & \delta_C, \end{bmatrix}$, $a = \|\delta_A\|$ and so on.

If $4(\mathrm{sep}(A, B) - a - b)^2 - 4d(\|C\| + c) \geq 0$, then there is a (unique) X with $\|X\| \leq 2\frac{\boxed{d}}{\underline{\mathrm{sep}(A,B) - a - b}}$ such that $\begin{bmatrix} I \\ X \end{bmatrix}$ is an invariant subspace of $M + \delta_M$.

(Not exactly what we obtain directly from the above argument — handles a and b in a slightly different way.)

evil case:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1+\varepsilon \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 \\ \varepsilon & 1+\varepsilon \end{bmatrix}$$

Let $M = \begin{bmatrix} A & C \\ 0 & B \end{bmatrix}$ (with inv. subspace $\begin{bmatrix} I \\ 0 \end{bmatrix}$), and

let $M + \delta_M = \begin{bmatrix} A + \delta A & C + \delta C \\ \delta D & B + \delta B \end{bmatrix}$. if

$a = \|\delta A\|_F$
$b = \|\delta B\|_F$
c ⋮
d

$$\boxed{\left(\operatorname{sep}(A,B) - a - b\right)^2 - 4d\left(\|C\|_F + c\right) > 0}$$ then

there exists X with $\|X\|_F \leq \dfrac{2d}{\operatorname{sep}(A,B) - a - b}$

such that $\begin{bmatrix} I \\ X \end{bmatrix}$ is an inv. subspace of $M + \delta M$

$$\operatorname{sep}(A,B) = \sigma_{\min}\left(I \otimes A - B^T \otimes I\right)$$

$$\begin{bmatrix} A & & & 0 \\ & A & & \\ & & A & \\ 0 & & & \ddots \quad A \end{bmatrix} - \begin{bmatrix} b_{11} I & b_{21} I & \cdots & \\ \vdots & & & \\ b_{1m} I & \cdots & & b_{mm} I \end{bmatrix} =: T$$

$$\begin{bmatrix} I & 0 \\ -X & I \end{bmatrix} \begin{bmatrix} A+\delta A & C+\delta C \\ \delta D & B+\delta B \end{bmatrix} \begin{bmatrix} I & 0 \\ X & I \end{bmatrix} = \begin{bmatrix} * & * \\ 0 & * \end{bmatrix}$$

$$= \begin{bmatrix} * & * \\ \delta D - X(A+\delta A) & B+\delta B - X(C+\delta C) \end{bmatrix} \begin{bmatrix} I & 0 \\ X & I \end{bmatrix} =$$

$\begin{bmatrix} * & * \\ \square & * \end{bmatrix}$ we want this to be zero:

$$\hookrightarrow \delta D - X(A+\delta A) + (B+\delta B)X - X(C+\delta C)X \overset{!}{=} 0$$

$$X(A+\delta A) - (B+\delta B)X = \delta D - X(C+\delta C)X$$

$$\tilde{T} \cdot \operatorname{vec} X = \operatorname{vec}\left(\delta D - X(C+\delta C)X\right)$$

$$\tilde{T} = (A+\delta A)^T \otimes I - I \otimes (B+\delta B)$$

$$\tilde{T} = A^T \otimes I - I \otimes B + \delta A^T \otimes I - I \otimes \delta B$$

$$\|\tilde{T} z\| \geq \|(A^T \otimes I - I \otimes B)z\| - \|(\delta A^T \otimes I)z\| - \|(I \otimes \delta B)z\|$$

$$\geq \left[\sigma_{min}\left(A^T \otimes I + I \otimes B\right) - \sigma_{max}\left(\delta A^T \otimes I\right) - \sigma_{max}\left(I \otimes \delta B\right)\right] \|x\|$$

$$\sigma_{min}(\tilde{T}) \geq \sigma_{min}\left(A^T \otimes I + I \otimes B\right) - a - b$$

$$\operatorname{vec} X = \tilde{T}^{-1} \operatorname{vec}\left(\delta D + X\left(C + \delta C\right) X\right)$$

DEF: $\varphi(x) := \tilde{T}^{-1} \operatorname{vec}\left(\delta D + (\operatorname{vec}^{-1} x) \cdot (C + \delta C) \cdot \operatorname{vec}^{-1} x\right)$

$x = \varphi(x)$ We wish to prove that φ sends $B(0, r)$ into itself

---

Lemma: $\|M \otimes N\|_2 = \|M\|_2 \cdot \|N\|_2$

(facendo SVD fattore per fattore)

$$\|\varphi(x)\| \leq \frac{1}{\sigma_{min}(\tilde{T})} \cdot \|vec[\delta D + X(C+\delta C)X]\|$$

$$\|vec[\delta D + X(C+\delta C)X]\|_2 = \|\delta D + X(C+\delta C)X\|_F$$

$$\leq \|\delta D\|_F + \|X\|_F(\|C\|_F + \|\delta C\|_F)\|X\|_F$$

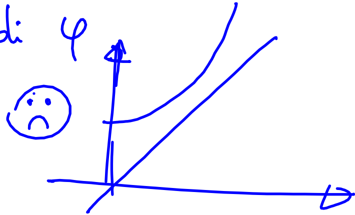
$$\|\varphi(x)\| \leq \frac{1}{sep(B,A)-a-b} \cdot \left(d + (\|C\|+c)\cdot\|x\|^2\right)$$

ci piacerebbe che

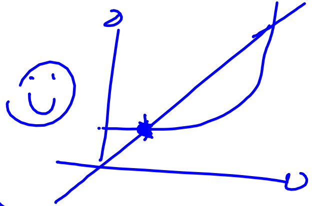
$$\frac{1}{sep(B,A)-a-b}\left(d + (\|C\|+c)r^2\right) \leq r$$

se è soddisfatta, $\varphi(B(0,r)) \subseteq B(0,r)$

e quindi esiste un p.f. di φ
$\|X\|_F \leq r$.



Discriminante di quella equazione: condizione in rosso nel teorema.



Se è soddisfatta, esiste un pto fisso di φ con $\|X\|_F \leq r \leq \dfrac{2a}{sep(B,A)-a-b}$ e quindi

$$\begin{bmatrix} 1 & 0 \\ -X & 1 \end{bmatrix} \begin{bmatrix} M+\delta M \end{bmatrix} \begin{bmatrix} 1 & 0 \\ X & 1 \end{bmatrix} = \begin{bmatrix} \tilde{A} & \tilde{C} \\ 0 & \tilde{B} \end{bmatrix}$$

$$\begin{bmatrix} M+\delta M \end{bmatrix} \begin{bmatrix} 1 & 0 \\ X & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ X & 1 \end{bmatrix} \begin{bmatrix} \tilde{A} & \tilde{C} \\ 0 & \tilde{B} \end{bmatrix}$$

$$\begin{bmatrix} M+\delta M \end{bmatrix} \begin{bmatrix} 1 \\ X \end{bmatrix} = \begin{bmatrix} 1 \\ X \end{bmatrix} \tilde{A}$$

Applications of Sylvester equations

Apart from the ones we have seen (more 'theoretical'):

- Computing matrix functions.
- Stability of linear dynamical systems.
  Lyapunov equations $AX + XA^T = B$, B symmetric.
- As a step to solve more complicated matrix equations (Newton's method $\to$ linearization).

Will see them later in the course (time permitting)

open problem: solve

$$AXB + CXD + EXF = G$$

in $O(n^3)$