La forma normale di Schur

Dario A. Bini, Università di Pisa

24 ottobre 2019

Sommario

Questo modulo didattico contiene risultati relativi alla forma normale di Schur, alle sue proprietà e alle sue applicazioni.

1 Introduzione

Tra le diverse forme canoniche di una matrice disponibili sul mercato la forma di Schur è particolarmente utile poiché si ottiene con una trasformazione per similitudine data da una matrice unitaria. Ricordiamo che una matrice $U \in \mathbb{C}^{n\times n}$ è detta *unitaria* se $U^H U = UU^H = I$, dove I indica la matrice identica $n \times n$ e A^H indica la matrice trasposta hermitiana di A, cioè la matrice che si ottiene trasponendo A e coniugando gli elementi complessi. Una matrice reale e unitaria è detta *ortogonale* .

Enunciamo questo risultato e ne diamo una dimostrazione costruttiva che non utilizza la forma normale di Jordan. La dimostrazione che vedremo può essere opportunamente trasformata in un algoritmo di calcolo.

È importante osservare che il calcolo numerico della forma normale di Jordan è un compito numericamente intrattabile. Infatti tale forma non è stabile sotto perturbazioni della matrice pur arbitrariamente piccole purché non nulle. A questo riguardo si consideri un singolo blocco di Jordan di dimensione n, ad esempio con autovalore nullo

$$J = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}.$$

È evidente che la matrice J ha $\lambda = 0$ come unico autovalore di molteplicità algebrica n e di molteplicità geometrica 1. Si modifichi ora J alterando l'elemento di posto $(n, 1)$ con una perturbazione $\epsilon > 0$ ottenendo

$$J_\epsilon = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ \epsilon & & & 0 \end{bmatrix}.$$

Non è difficile verificare che questa matrice ha polinomio caratteristico $\lambda^n - \epsilon$ e quindi ha n autovalori distinti dati da $\lambda_i = \epsilon^{1/n}\omega_n^i$, $i = 1, \ldots, n$, dove $\omega_n = \cos(2\pi/n) + \mathtt{i}\sin(2\pi/n)$, dove $\mathtt{i}$ è l'unità immaginaria tale che $\mathtt{i}^2 = -1$. Quindi la sua forma normale di Jordan è data da una matrice *diagonale* .

La forma normale di Schur non presenta questo inconveniente.

2 Forma normale di Schur

Vale il seguente

Teorema 1 (Forma normale di Schur) *Per ogni matrice $A \in \mathbb{C}^{n\times n}$ esistono una matrice triangolare superiore T e una matrice unitaria U tali che*

$$U^H AU = T.$$

Dim. Si procede per induzione sulla dimensione n della matrice A. Se $n = 1$ la matrice A è un numero complesso e la decomposizione di Schur vale con $T = A$ e $U = 1$. In generale, supponiamo che la fattorizzazione valga per dimensione $n-1$. Consideriamo $A \in \mathbb{C}^{n\times n}$ e denotiamo con λ e x rispettivamente un autovalore e un autovettore di A, cioè $Ax = \lambda x$. Supponiamo che x sia normalizzato in modo che $x^H x = 1$ e consideriamo una base ortonormale formata dai vettori $y_2, \ldots, y_n$ dello spazio ortogonale a x. Costruiamo la matrice Q le cui colonne sono $x, y_2, \ldots, y_n$. Questa matrice è unitaria, cioè $Q^H Q = I$ e risulta $Qe^{(1)} = x$, $e^{(1)} = Q^H x$ dove $e^{(1)} = (1, 0, \ldots, 0)^T$. Inoltre vale

$$Q^H AQ = \begin{bmatrix} \lambda & u^T \\ 0 & A_{n-1} \end{bmatrix},$$

dove $A_{n-1} \in \mathbb{C}^{(n-1)\times(n-1)}$. Per verificare quest'ultima relazione basta considerare $Q^H AQe^{(1)}$ che è la prima colonna di $Q^H AQ$. Vale

$$Q^H AQe^{(1)} = Q^H Ax = Q^H \lambda x = \lambda Q^H x = \lambda e^{(1)}$$

come richiesto. Inoltre per l'ipotesi induttiva si ha $A_{n-1} = U_{n-1}T_{n-1}U_{n-1}^H$, da cui posto

$$U_n = Q \begin{bmatrix} 1 & 0 \\ & U_{n-1} \end{bmatrix},$$

si ottiene

$$U_n^H AU_n = \begin{bmatrix} \lambda & u^T U_{n-1}^H \\ 0 & T_{n-1} \end{bmatrix}$$

che è una matrice triangolare. $\square$

Si osservi che per trasformare la dimostrazione del teorema precedente in algoritmo di calcolo è sufficiente disporre di due "scatole nere": la prima che data in input una matrice A ci fornisce in output un suo autovalore λ e il corrispondente autovettore x; la seconda che dato in input un vettore x ci fornisce in output una base ortonormale del sottospazio ortogonale a x, o, equivalentemente, fornisce una matrice Q tale che $Qx = e^{(1)}$. Vedremo in un altro articolo come la seconda scatola nera sia di facile costruzione usando le matrici elementari di Householder.

Un'altra osservazione utile è che la forma di Schur non è unica. Infatti scegliendo diversi ordinamenti degli autovalori arriviamo a forme normali in generale diverse tra loro.

Seguendo l'impostazione della dimostrazione del teorema precedente è possibile dimostrare un risultato specifico per le matrici reali. Per questo abbiamo bisogno della seguente definizione

Definizione 1 *Una matrice $T \in \mathbb{R}^{n\times n}$ si dice* quasi triangolare *se si può scrivere nella forma*

$$T = \begin{bmatrix} T_{1,1} & \dots & T_{1,n} \\ & \ddots & \vdots \\ & & T_{m,m} \end{bmatrix}$$

dove $T_{i,i}$ per $i = 1, \dots, m$ possono essere matrici 2×2 con coppie di autovalori complessi coniugati, oppure matrici 1×1, cioè numeri reali.

Gli autovalori di una matrice quasitriangolare sono gli autovalori delle sottomatrici $T_{i,i}$ per $i = 1, \dots, m$.

Ad esempio, la matrice

$$\begin{bmatrix} 2 & 1 & 1 & 1 & 1 \\ -1 & 2 & 1 & 1 & 1 \\ 0 & 0 & 5 & 2 & 3 \\ 0 & 0 & 0 & 3 & -1 \\ 0 & 0 & 0 & 1 & 3 \end{bmatrix}$$

è quasi triangolare. I suoi autovalori sono dati dagli autovalori di $\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$ che sono $2 + \mathtt{i}$ e $2 - \mathtt{i}$, dagli autovalori di $\begin{bmatrix} 3 & -1 \\ 1 & 3 \end{bmatrix}$ che sono $3 + \mathtt{i}$, $3 - \mathtt{i}$, e da 5.

Teorema 2 (Forma reale di Schur) *Per ogni matrice $A \in \mathbb{R}^{n\times n}$ esistono una matrice quasi triangolare superiore $T \in \mathbb{R}^{n\times n}$ e una matrice ortogonale $Q \in \mathbb{R}^{n\times n}$ tali che*

$$Q^T AQ = T.$$

3 Applicazioni

La forma di Schur ha delle conseguenze interessanti, alcune di facile dimostrazione. Una prima conseguenza riguarda le matrici hermitiane, cioè quelle matrici A tali che $A = A^H$.

Infatti, se $U^H AU = T$ è la forma di Schur della matrice hermitiana A, allora la proprietà $A = A^H$ implica

$$T^H = U^H A^H U = U^H AU = T.$$

Cioè T è hermitiana e quindi, essendo triangolare, è diagonale. Inoltre gli elementi diagonali di T sono tali che $t_{i,i} = \bar{t}_{i,i}$ e quindi sono reali. Cioè si ottiene che le matrici hermitiane sono diagonalizzabili con una trasformazione per similitudine unitaria.

Se A è anti-hermitiana, cioè se $A = -A^H$ allora procedendo come sopra si ottiene che anche T è anti-hermitiana. Ne segue che T è una matrice diagonale tale che $t_{i,i} = -\bar{t}_{i,i}$. Cioè una matrice anti-hermitiana è diagonalizzabile da una trasformazione ortogonale e i suoi autovalori sono numeri immaginari.

Un risultato più generale è espresso dal seguente

Teorema 3 *Una matrice $A \in \mathbb{C}^{n\times n}$ ha forma normale di Schur diagonale se e solo se $A^H A = AA^H$.*

Dim. Se T è la forma di Schur di A, allora $A^H A = AA^H$ se e solo se $T^H T = TT^H$. Ciò segue dal fatto che

$$A^H A = U^H T^H UU^H TU = U^H T^H TU, \quad AA^H = U^H TUU^H T^H U = U^H TT^H U.$$

Se T è matrice diagonale allora $T^H T = TT^H$ e quindi $A^H A = AA^H$. Viceversa, se $A^H A = AA^H$ allora $T^H T = TT^H$. Dimostriamo che la condizione $T^H T = TT^H$ implica che T è matrice diagonale. Per questo procediamo per induzione su n. Se $n = 1$ non c'è nulla da dimostrare. Assumiamo vera la tesi per matrici di dimensione $n-1$ e consideriamo il caso di dimensione n. Leggendo la relazione $T^H T = TT^H$ sull'elemento di posto $(1, 1)$ otteniamo

$$\bar{t}_{1,1} t_{1,1} = \sum_{j=1}^{n} t_{1,j} \bar{t}_{1,j}$$

cioè $|t_{1,1}|^2 = |t_{1,1}|^2 + |t_{1,2}|^2 + \cdots + |t_{1,n}|^2$ da cui $|t_{1,2}|^2 + \cdots + |t_{1,n}|^2 = 0$. Data la non negatività degli addendi questo implica che $t_{1,j} = 0$ per $j = 2, \ldots, n$. Quindi la matrice T ha la forma

$$T = \begin{bmatrix} t_{1,1} & 0 \\ 0 & T_{n-1} \end{bmatrix}.$$

La condizione $T^H T = TT^H$ implica allora $T_{n-1}^H T_{n-1} = T_{n-1} T_{n-1}^H$, e, per l'ipotesi induttiva segue che T_{n-1} è matrice diagonale. $\square$

La classe di matrici che verificano la condizione $A^H A = AA^H$ è detta classe delle matrici *normali*

Una matrice unitaria è in particolare una matrice normale e quindi ha una forma di Schur diagonale. Inoltre da $A^H A = I$ segue che $|t_{i,i}|^2 = 1$. Cioè le matrici unitarie sono diagonalizzabili da una trasformazione ortogonale e hanno autovalori di modulo 1.

4 Esercizi

1. Determinare una forma di Schur della matrice $A = uv^T$ dove $u, v \in \mathbb{R}^n$.

2. Usando la forma normale di Schur determinare l'insieme dei possibili autovalori delle matrici A ad elementi complessi $n \times n$ che risolvono l'equazione $A^2 - 5A^H + 6I = 0$.

3. Sia α un numero reale e si consideri la classe $\mathcal{A}_n(\alpha)$ delle matrici $n \times n$ a elementi complessi tali che $\alpha A + A^H A + A^H = I$. Utilizzando la forma normale di Schur si descriva l'insieme $\Lambda(\alpha) = \{\lambda \in \mathbb{C} : \quad \exists A \in \mathcal{A}_n(\alpha),\ \det(A - \lambda I) = 0\}$.

4. Si ricavi la forma normale di Schur di una matrice A dalla forma normale di Jordan $A = SJS^{-1}$.

5. Siano $a, b \in \mathbb{R}$ e
$$C = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}. \tag{1}$$
Si determini una matrice Q unitaria tale che che $Q^H CQ = D$, dove D ̀diagonale con elementi diagonali $a \pm \mathbf{i}b$ dove $\mathbf{i}^2 = -1$. Si dimostri che tutte le matrici della forma (1) commutano.

   Soluzione. Calcolando gli zeri del polinomio caratteristico, troviamo che gli autovalori di C sono $\lambda_{1,2} = a \pm \mathbf{i}b$. Inoltre gli autovettori corrispondenti sono $x_1 = \begin{bmatrix} 1 \\ \mathbf{i} \end{bmatrix}$, $x_2 = \begin{bmatrix} 1 \\ -\mathbf{i} \end{bmatrix}$, che sono tra loro ortogonali. Dunque, se definiamo la matrice $Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ \mathbf{i} & -\mathbf{i} \end{bmatrix}$, questa è una matrice unitaria tale che
$$Q^H CQ = \begin{bmatrix} a + \mathbf{i}b & 0 \\ 0 & a - \mathbf{i}b \end{bmatrix}.$$
   In particolare, tutte le matrici della forma (1) sono diagonalizzabili mediante la stessa trasformazione, quindi commutano.

6. Dimostrare la forma normale reale di Schur del teorema 2.

   Soluzione. Si procede per induzione sulla dimensione n della matrice A con la stesso approccio usato per dimostrare il teorema di Schur nel caso generale. Se $n = 1$ non c'è nulla da dimostrare. Se $n = 2$ e gli autovalori sono reali, allora posto v un autovettore reale tale che $v^T v = 1$, la matrice Q le cui colonne sono v e u con u reale, ortogonale a v e tale che $u^T u = 1$, risulta $Q^T Q = I$ e $Q^T AQ = T$ triangolare superiore. Quindi T è in forma reale di Schur. Se invece con $n = 2$ la matrice ha una coppia di autovalori complessi coniugati allora è già nella forma reale di Schur. Per il passo induttivo, sia A matrice reale $n \times n$. Se λ è un autovalore reale si procede come nella dimostrazione del teorema di Schur nel caso generale e si riconduce il problema a dimensione $n - 1$. Se invece $\lambda = a + \mathtt{i}b$ e $\bar{\lambda} = a - \mathtt{i}b$, $a, b \in \mathbb{R}$, sono una coppia di autovalori complessi coniugati corrispondenti agli autovettori $v = u + \mathtt{i}w$, $\bar{v} = u - \mathtt{i}w$, con $u, w \in \mathbb{R}^n$ si osserva che

$$A \begin{bmatrix} u & w \end{bmatrix} = \begin{bmatrix} u & w \end{bmatrix} \begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

Cioè il sottospazio di $\mathbb{R}^n$ generato da u e w è invariante per A. I vettori u e w sono linearmente indipendenti perché, se non lo fossero, il vettore u sarebbe autovettore, dunque A avrebbe un autovettore reale corrispondente a un autovalore con parte immaginaria non nulla, che è un assurdo. Allora si costruisce una base ortonormale di questo sotto spazio, formata da due vettori $u^{(1)}$ e $u^{(2)}$ scegliendo combinazioni lineari di u e w, ad esempio

$$\begin{bmatrix} u^{(1)} & u^{(2)} \end{bmatrix} = \begin{bmatrix} u, w \end{bmatrix} S$$

con S opportuna matrice 2×2. In questo modo si ha

$$A \begin{bmatrix} u^{(1)} & u^{(2)} \end{bmatrix} = \begin{bmatrix} u^{(1)}, u^{(2)} \end{bmatrix} T_1$$

con $T_1 = S^{-1} \left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix} \right] S$ matrice 2×2. I vettori $u^{(1)}$ e $u^{(2)}$ si completano con una base ortonormale dello spazio ortogonale a $\mathrm{span}(u^{(1)}, u^{(2)})$. La matrice U che ha per colonne questi vettori è tale che $U^T AU$ è triangolare superiore a blocchi col primo blocco sulla diagonale principale che coincide con T_1. Mentre il blocco di posto $(2, 2)$ ha dimensione $n - 2$ quindi per l'ipotesi induttiva ha forma normale reale di Schur.

7. Usando la forma normale reale di Schur si dimostri che se A è una matrice $n \times n$ reale allora A è normale se e solo se esiste una matrice reale ortogonale Q tale che $Q^T AQ$ è una matrice diagonale a blocchi, reale, con blocchi di dimensione minore o uguale a 2 in cui i blocchi di dimensione 2 hanno la forma (1).

   Soluzione. Se la matrice $Q^T AQ$ è diagonale a blocchi con blocchi T_i di dimensione 1 o 2 e struttura definita in (1) allora poiché $A^T A = AA^T$ se e solo se $T_i^T T_i = T_i T_i^T$, basta dimostrare che $T_i^T T_i = T_i T_i^T$. Se la dimensione di T_i è 1 questo è banalmente vero. Se la dimensione è 2 e i blocchi hanno la struttura (1) allora la proprietà segue dall'esercizio 5. Per dimostrare l'implicazione opposta si procede per induzione come nella dimostrazione del teorema 3. Per $n = 1$ non c'è nulla da dimostrare. Per

$n = 2$ se la matrice ha autovalori reali allora il risultato segue dal teorema 3. Se invece $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ è reale con autovalori complessi coniugati, dalla relazione $A^T A - AA^T = 0$ segue $b^2 = c^2$, $(b-c)(d-a) = 0$ da cui $b = \pm c$. Se fosse $b = c$ la matrice sarebbe simmetrica e avrebbe autovalori reali. Allora deve essere $b = -c$ e conseguentemente $a = d$. Si ottiene quindi la struttura in (1). Per il passo induttivo si considera la forma normale di Schur reale T di A. Siano $T_{i,j}$ i blocchi di T. Leggendo l'uguaglianza $T^T T = TT^T$ nel blocco di posto $(1,1)$ si ha

$$T_{1,1}^T T_{1,1} = T_{1,1} T_{1,1}^T + \sum_{i \geq 2} T_{1,i} T_{1,i}^T$$

da cui

$$T_{1,1}^T T_{1,1} - T_{1,1} T_{1,1}^T = \sum_{i \geq 2} T_{1,i} T_{1,i}^T$$

La matrice a sinistra ha traccia nulla, e quindi anche la matrice a destra. Poiché ciascun addendo è semidefinito positivo deve avere traccia non negativa. Poiché la somma delle tracce è nulla allora ciascun addendo ha traccia nulla. Ma una matrice reale simmetrica semidefinita positiva con traccia nulla deve essere identicamente nulla. Ne segue che $T_{1,i} = 0$ per ogni i. Si osserva infatti che se $T_{1,i} \neq 0$ esisterebbe $x \neq 0$ tale che $T_{1,i}^T x \neq 0$, da cui $x^T T_{1,i} T_{1,i}^T x > 0$ per cui $T_{1,i} T_{1,i}^T$ non può esser nulla.

L'uguaglianza a zero della prima riga a blocchi di T, con eventuale esclusione di $T_{1,1}$ riconduce il problema a dimensione inferiore per cui si può applicare l'ipotesi induttiva.

Riferimenti bibliografici

[1] D. Bini, M. Capovani, O. Menchi. Metodi Numerici per l'Algebra Lineare. Zanichelli, Bologna 1988.