Analisi numerica di un'equazione matriciale quadratica (Calcolo Scientifico - proff. L. Gemignani e D. Bini)

Martino De Leo - 439650 - deleo@mail.dm.unipi.it

Indice

1	Equazioni matriciali, problema agli autovalori generalizzato	1
1.1	Fasci lineari	3
1.1.1	Forma normale di Schur generalizzata	3
2	Metodi di risoluzione	4
2.1	Relazione tra GEP e l'equazione matriciale quadratica associata	4
2.2	Metodo agli autovalori	5
2.2.1	Abbassare il grado	5
2.3	Metodo di Schur	6
2.3.1	Stabilità numerica: riscalatura	7
2.4	Metodo iterativo di Bernoulli	8
2.4.1	Iterazione di Bernoulli	8
2.4.2	Non-singularità della Vandermonde a blocchi	9
2.4.3	Calcolo delle soluzioni	10
3	Un'applicazione: problemi di vibrazioni	11
3.1	Risonanza	12
3.2	Dimensione uno	12
3.2.1	Caso non smorzato	13
3.2.2	Caso sovrasmorzato	13
3.2.3	Caso sottosmorzato	13
3.3	Dimensione arbitraria	14
3.3.1	Caso sovrasmorzato	14

1 Equazioni matriciali, problema agli autovalori generalizzato

Studieremo un'equazione matriciale quadratica

$$Q(X) = AX^2 + BX + C = 0, \quad (1)$$

dove sia i coefficienti A, B, C che l'indeterminata X sono matrici in $\mathbb{C}^{n \times n}$. È questo un caso particolare di un *polinomio* o *fascio* di matrici, generalmente di *grado* d anziché 2:

$$Q(X) = \sum_{i=0}^d A_i X^i = A_d X^d + A_{d-1} X^{d-1} + \dots + A_1 X + A_0, \quad (2)$$

dove $A_d \neq 0$.

Altre definizioni alternative nel caso quadratico sono, tra le altre, l'*equazione di Riccati algebrica* (proveniente da un problema di Controllo Ottimo)

$$XAX + BX + XC + D = 0,$$

l'equazione matriciale quadratica *sinistra*

$$X^2 A + XB + C = 0,$$

mentre un caso significativo di grado 1 è l'*equazione di Sylvester-Lyapounov* (sempre dalla Teoria del Controllo)

$$AX - XB - C = 0.$$

Corrispondentemente a un'equazione matriciale del tipo (2) si può considerare il *problema agli autovalori generalizzato* (GEP , Generalized Eigenvalue Problem) ottenuto sostituendo $X = \lambda I$:

$$Q(\lambda) = \lambda^d A_d + \lambda^{d-1} A_{d-1} + \cdots + \lambda A_1 + A_0$$

per il quale si definisce il *polinomio caratteristico* come $p_Q(\lambda) := \det(Q(\lambda))$; se $\tilde{\lambda}$ è una radice di p_Q si dice *autovalore* del fascio $Q(\lambda)$, e un corrispondente $x \in \mathbb{C}^n \setminus \{0\}$ è detto $\tilde{\lambda}$ -autovettore (destra) per $Q(\lambda)$ se $Q(\tilde{\lambda})x = 0 \iff x \in \ker Q(\tilde{\lambda})$.

Definizione 1. Il fascio $Q(\lambda)$ si dice *regolare* se il suo polinomio caratteristico non è identicamente nullo.

Osservazione 1. Se $Q(\lambda)$ è regolare allora $\bigcap_{i=0}^d \ker A_i = \{0\}$. Infatti se per assurdo esistesse un $x \in \mathbb{C}^n \setminus \{0\}$ tale che $A_i x = 0$ per ogni $i = 0, \dots, d$ si avrebbe $Q(\lambda)x = 0$ ($\forall \lambda \in \mathbb{C}$), implicando $p_Q \equiv 0$, ciò che è assurdo.

Un fascio regolare ammette al più dn autovalori; quindi rispetto al problema “classico” agli autovalori l'introduzione di una non linearità in λ determina una perdita di struttura: non è più garantito che il polinomio caratteristico abbia grado (massimo) fissato.

Esempio 1. Per $n = 2$ prendendo $A = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 3 \\ 0 & 1 \end{pmatrix}$ e $C = I_2$ si ha

$$p_Q(\lambda) = \det \begin{pmatrix} \lambda^2 + 1 & 3\lambda - \lambda^2 \\ 0 & \lambda + 1 \end{pmatrix} = \lambda^3 + \lambda^2 + \lambda + 1,$$

quindi il fascio è regolare ma ha soltanto i *tre* autovalori $\pm i, -1$.

Il numero massimo di autovalori è assicurato nel caso in cui il coefficiente di testa del fascio sia invertibile:

Lemma 1. *Se $Q(\lambda)$ ha grado d e il suo coefficiente di testa è non singolare allora ha esattamente nd autovalori.*

Dimostrazione. Per induzione sulla dimensione n .

Moltiplicando $Q(\lambda)$ a destra per A_d^{-1} si ottiene, con la regola di Binet

$$p_Q(\lambda) = \det \left(\sum_{i=0}^d \lambda^i A_i A_d^{-1} \right) \cdot \det A_d =: \hat{p}(\lambda) \cdot \det A_d,$$

dove $\hat{p}$ ha lo stesso grado (e le stesse radici) di p_Q perché sono associati. Sviluppando sulla prima riga il primo fattore si trova

$$\begin{aligned} \hat{p}(\lambda) &= \det \left(\lambda^d I_n + \sum_{i=0}^{d-1} \lambda^i A_i A_d^{-1} \right) = \\ &= \sum_{j=1}^n (-1)^{1+j} \left[\left(\lambda^d \delta_{j1} + \sum_{i=0}^{d-1} \lambda^i (A_i A_d^{-1})_{1j} \right) \cdot \det \left(\left(\lambda^d I_n + \sum_{i=0}^{d-1} \lambda^i A_i A_d^{-1} \right)_{\hat{1}j} \right) \right] = \\ &= \left(\lambda^d + \sum_{i=0}^{d-1} \lambda^i (A_i A_d^{-1})_{11} \right) \cdot \det \left(\lambda^d I_{n-1} + \sum_{i=0}^{d-1} \lambda^i (A_i A_d^{-1})_{\hat{1}1} \right) + \\ &+ \sum_{j=2}^n (-1)^{1+j} \left[\left(\sum_{i=0}^{d-1} \lambda^i (A_i A_d^{-1})_{1j} \right) \cdot \det \left(\left(\lambda^d I_n + \sum_{i=0}^{d-1} \lambda^i A_i A_d^{-1} \right)_{\hat{1}j} \right) \right] =: A(\lambda) + C(\lambda), \end{aligned}$$

dove si è indicato con $(B)_{ij}$ l'entrata b_{ij} di una matrice B e corrispondentemente con $B_{\hat{i}j}$ il complemento ottenuto da B cancellandone la riga i -esima e la colonna j -esima.

Il primo fattore di $A(\lambda)$ ha grado d , mentre il secondo è il polinomio caratteristico di un fascio di grado d e dimensione $n-1$, con coefficiente di testa I_{n-1} invertibile, che per ipotesi induttiva ha grado $(n-1)d$; in conclusione $\deg A(\lambda) = d + (n-1)d = dn$. Si può scrivere $B(\lambda) = \sum_{j=2}^n (-1)^{1+j} \left[q^{(j)}(\lambda) \cdot \det Q_{n-1}^{(j)}(\lambda) \right]$, dove $\deg q^{(j)} \leq d-1$ per ogni $j = 2, \dots, n$, mentre i fasci $Q_{n-1}^{(j)}$ hanno polinomio caratteristico di grado $(n-1)d$ (sempre per l'ipotesi induttiva); complessivamente quindi $\deg B(\lambda) \leq (d-1) + (n-1)d = nd - 1 < nd$. Questo conclude il passo induttivo.

Per la base d'induzione, quando $n = 1$ l'equazione scalare

$$\lambda^d a_d + \lambda^{d-1} a_{d-1} + \cdots + \lambda a_1 + a_0 = 0$$

con $a_d \neq 0$ è un polinomio di grado d , che ha d radici per il Teorema fondamentale dell'Algebra. $\square$

Un'altra differenza sostanziale rispetto al problema “classico” agli autovalori $(A - \lambda I)x = 0$ è la seguente: un fascio $Q(\lambda)$ di grado d regolare ammette certe autocopie $\{(\lambda_i, v_i)\}_{i=1}^p$, con $n \leq p \leq dn$. Non tutte le possibili scelte (in numero di $\binom{p}{n}$) di n autovettori $(v_{i_1}, \dots, v_{i_n})$ forniscono una base per $\mathbb{C}^n$; anzi non è neanche detto che una tale selezione esista, perché in generale non è vero che autovettori relativi ad autovalori diversi sono linearmente indipendenti.

1.1 Fasci lineari

Per $d = 1$ si può scrivere il fascio *lineare* $Q(\lambda) = A - \lambda B$. Il problema standard agli autovalori è un caso speciale di fascio lineare con $B = I$.

Se il fascio è regolare, si può estendere la definizione di autovalori di $Q(\lambda)$ in modo da averne sempre n anche quando $\deg p_Q < n$; questa situazione si verifica se e solo se B non è invertibile (coerentemente con il Lemma 1, vedasi anche il Lemma 2).

Definizione 2. Gli autovalori di $Q(\lambda) = A - \lambda B$, fascio regolare, sono:

1. gli zeri di p_Q ;
2. ∞ , con molteplicità $n - \deg p_Q$, se $\deg p_Q < n$. In tal caso un ∞ -autovettore (destra) per $Q(\lambda)$ è un $x \neq 0$, $x \in \ker B$.

Lemma 2. (*Caratterizzazione degli autovalori di un fascio lineare regolare*) Sia il fascio lineare $Q(\lambda) = A - \lambda B$ regolare.

1. Se B è non singolare, tutti gli autovalori di $Q(\lambda)$ sono finiti e coincidono con gli autovalori di AB^{-1} oppure di $B^{-1}A$; se B è singolare, $Q(\lambda)$ ha un autovalore ∞ di molteplicità $n - \text{rank } B$;
2. Se A è non singolare, gli autovalori di $Q(\lambda)$ coincidono con i reciproci degli autovalori di $A^{-1}B$ oppure di BA^{-1} , dove a un autovalore nullo di $A^{-1}B$ corrisponde un autovalore ∞ di $Q(\lambda)$.

Dimostrazione. (1) Se B è invertibile,

$$p_Q(\lambda) = \det(A - \lambda B) = \begin{cases} \det((A - \lambda B)B^{-1}B) = \det(AB^{-1} - \lambda I) \cdot \det B \\ \det(BB^{-1}(A - \lambda B)) = \det B \cdot \det(B^{-1}A - \lambda I). \end{cases}$$

Se invece $\text{rank } B = r < n$, scriviamo la SVD di $B = U\Sigma V^H$ da cui

$$p_Q(\lambda) = \det(A - \lambda U\Sigma V^H) = \det(U(U^H A V - \lambda \Sigma)V^H) = \pm \det(U^H A V - \lambda \Sigma),$$

e siccome Σ è diagonale con soltanto i primi r elementi non nulli, p_Q ha grado r .

(2) Se A è non singolare

$$\det(A - \lambda B) = 0 \iff \det(I - \lambda B A^{-1}) = 0 \iff \det(I - \lambda A^{-1} B) = 0$$

e questa uguaglianza è verificata se e soltanto se $\lambda \neq 0$ e $1/\lambda$ è un autovalore di $A^{-1}B$. □

1.1.1 Forma normale di Schur generalizzata

Anche per un fascio lineare si può effettuare un cambio di base che ovviamente preserva gli autovalori dal momento che $\det(A - \lambda B) = 0 \iff \det(V(A - \lambda B)S) = \det(VAS - \lambda VBS) = 0$ per ogni V, S invertibili.

Quindi per il nostro scopo di calcolare gli autovalori di $Q(\lambda)$ converrà cercare una forma più “semplice” possibile tramite una trasformazione numericamente più stabile possibile. Si può ottenere una forma triangolare con due trasformazioni unitarie:

Teorema 1. (*Forma normale di Schur generalizzata, caso regolare*) Se $Q(\lambda) = A - \lambda B$ è un fascio regolare, esistono $Q, Z \in \mathbb{C}^{n \times n}$ unitarie tali che

$$Q(A - \lambda B)Z^H = T - \lambda S$$

con T, S triangolari superiori. Gli autovalori di $Q(\lambda)$ sono i rapporti tra gli elementi diagonali di T ed S : $\lambda_i = t_{ii}/s_{ii}$ (se $s_{ii} = 0$ allora $\lambda_i = \infty$).

Esiste una forma normale di Schur generalizzata anche nel caso di un fascio non regolare.

Dimostrazione. Per induzione su n .

Nel caso base l'equazione scalare $a - \lambda b = 0$ è già nella forma richiesta (prendasi dunque $q = z = 1$).

Per il passo induttivo sia λ un autovalore e x un corrispondente autovettore con $\|x\|_2 = 1$: da $Ax - \lambda Bx = 0$ segue che Ax e Bx sono paralleli fra loro, quindi sono anche paralleli a un vettore y con $\|y\|_2 = 1$. L'ipotesi di regolarità del fascio si usa in questo punto per dire che y è ben definito, a meno del segno, anche nel caso in cui uno tra Ax e Bx è il vettore nullo: infatti per l'osservazione 1 questo non si verifica per entrambi.

Completando a basi ortonormali si ottengono due matrici unitarie

$$X = (x | \tilde{X}), \quad Y = (y | \tilde{Y}),$$

da cui

$$Y^H AX = \begin{pmatrix} y^H \\ \tilde{Y}^H \end{pmatrix} (Ax | A\tilde{X}) = \begin{pmatrix} \tilde{a}_{11} | & \tilde{a}_{12} \\ 0 | & \\ \vdots | & \tilde{A}_{22} \\ 0 | & \end{pmatrix} \quad \tilde{a}_{11} \in \mathbb{C}, \tilde{a}_{21} \in \mathbb{C}^{1 \times (n-1)}, \tilde{A}_{22} \in \mathbb{C}^{(n-1) \times (n-1)},$$

dove gli zeri sottodiagonali nella prima colonna appaiono perché le colonne di $\tilde{Y}$ sono ortogonali ad Ax che è parallelo a y .

Con lo stesso argomento si vede che

$$Y^H BX = \begin{pmatrix} \tilde{b}_{11} | & \tilde{b}_{12} \\ 0 | & \\ \vdots | & \tilde{B}_{22} \\ 0 | & \end{pmatrix},$$

con le dimensioni coerenti con le precedenti; in particolare anche $\tilde{B}_{22}$ è quadrata di ordine $(n-1)$, quindi si conclude facilmente con l'ipotesi induttiva.

Per quanto riguarda l'asserzione sugli autovalori, se $Bx \neq 0$ (altrimenti è vera per definizione con $\lambda = \infty$)

$$\frac{\tilde{a}_{11}}{\tilde{b}_{11}} = \frac{y^H Ax}{y^H Bx} = \frac{y^H (\lambda Bx)}{y^H Bx} = \lambda,$$

e y non è ortogonale a Bx per costruzione; ciò conclude la dimostrazione. $\square$

Osservazione 2. (Algoritmo QZ) La dimostrazione è costruttiva ed esibisce una versione preliminare dell'*algoritmo QZ*, standard *de facto* per il calcolo degli autovalori di un fascio lineare. Si noti che QZ calcola *implicitamente* gli autovalori di $B^{-1}A$ senza effettivamente calcolare l'inversa di B .

2 Metodi di risoluzione

2.1 Relazione tra GEP e l'equazione matriciale quadratica associata

Tornando alla (1), per risolvere $Q(X) = 0$ si può procedere con un attacco diretto oppure considerando il problema agli autovalori generalizzato ad essa associato $Q(\lambda)x = 0$.

Si noti anzitutto che $Q(X) = 0$ può avere un numero finito di soluzioni, un numero infinito, oppure nessuna:

Esempio 2. In dimensione $n = 2$, per $A = -C = I_2$ e $B = 0$, $Q(X) = X^2 - I_2 = 0$ ha infinite soluzioni; infatti I_2 ha infinite radici:

$$\frac{1}{t} \begin{pmatrix} s & r \\ r & -s \end{pmatrix}, \quad s^2 + r^2 = t^2.$$

La matrice $C = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}$ invece non ha alcuna radice quadrata: se fosse

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 = \begin{pmatrix} a^2 + bc & b(a+d) \\ c(a+d) & d^2 + bc \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \Rightarrow c = 0 \Rightarrow a = d = 0,$$

ciò che è assurdo. Quindi $X^2 - \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = 0$ non ha soluzioni.

Se si deve risolvere il problema $Q(\lambda)x = 0$, ci si chiede come una soluzione della quadratica $Q(X) = 0$ possa semplificare il lavoro.

Ogni soluzione S della $Q(X) = 0$ permette di abbassare il grado dell'istanza quadratica di GEP, riducendola a un fascio lineare: difatti se $Q(S) = 0$ allora $S - \lambda I$ è un divisore destro di $Q(\lambda)$:

$$\lambda^2 A + \lambda B + C = -(B + AS + \lambda A)(S - \lambda I),$$

come è immediato verificare. Quindi ogni autocoppia per S è anche una autocoppia per $Q(\lambda)$.

Quindi per risolvere il GEP $Q(\lambda)$ si può procedere trovando tutte le soluzioni S per $Q(X)$, e corrispondentemente a ciascuna di esse

1. risolvere il problema standard agli autovalori per S , trovando $(\lambda_i^{(S)}, v_i^{(S)})$, $i = 1, \dots, n$;
2. risolvere l'istanza lineare di GEP $(B + AS + \lambda A)x = 0$ in $(\lambda, x) =: (\lambda_i^{(*)}, v_i^{(*)})$ trovando $p_S \leq n$ autovalori finiti.

Questa era la fase “divide”. Per ricomporre le soluzioni (“impera”), sempre per ogni S , se qualche $\lambda_i^{(*)}$ è anche autovalore di S lo si scarta (già considerato). Altrimenti $S - \lambda_i^{(*)}I$ è non singolare e si prende $u_i^{(*)} := (S - \lambda_i^{(*)}I)^{-1} v_i^{(*)}$ come autovettore.

Se A è invertibile l'istanza lineare di GEP si trasforma nel problema standard agli autovalori $-(A^{-1}B + S) - \lambda I)x = 0$.

Viceversa, dalle soluzioni di GEP si possono ricavare delle soluzioni *diagonalizzabili* di $Q(X) = 0$, ammesso che esistano. Supponiamo infatti che $Q(\lambda_i)v_i = 0$ per $i = 1, \dots, p \geq n$; prendendo

$$V = (v_1 | \dots | v_n), \quad \Lambda = \text{diag}(\lambda_i),$$

si ha allora $AV\Lambda^2 + BV\Lambda + CV = 0$. Se V è invertibile allora $S = V\Lambda V^{-1}$ è una soluzione di $Q(X) = 0$.

Però in generale V non è invertibile, perché come abbiamo detto non è garantita l'esistenza di n autovettori linearmente indipendenti, anche se sono relativi ad autovalori distinti.

2.2 Metodo agli autovalori

Supponiamo di aver risolto l'istanza di GEP associata a $Q(X)$, ottenendo delle coppie autovalore, autovettore (λ_i, v_i) per $i = 1, \dots, p$ con $n \leq p \leq 2n$:

$$(\lambda_i^2 A + \lambda_i B + C)v_i = 0 \quad i = 1, \dots, p.$$

Per identificare i gruppi di n autovettori linearmente indipendenti si può applicare l'algoritmo QR con pivotaggio delle colonne alla matrice che ha per colonne gli autovettori:

$$\mathbb{C}^{n \times p} \ni V = (v_1 | \dots | v_p)$$

ottenendo

$$V\Pi = Q(R_1 | R_2)$$

dove Q è unitaria, $R_1 \in \mathbb{C}^{n \times n}$ è triangolare superiore e Π è una matrice di permutazione delle colonne di V : se $\sigma \in S_p$ è la permutazione rappresentata da Π allora

$$V\Pi = ((v_{\sigma(1)} | \dots | v_{\sigma(i)} | \dots | v_{\sigma(p)})).$$

Se a questo punto R_1 è non singolare (i.e. non ha elementi diagonali nulli) allora

$$W = (v_{\sigma(1)} | \dots | v_{\sigma(n)})$$

è non singolare e $S = W \text{diag}(\lambda_{\sigma(1)}, \dots, \lambda_{\sigma(n)}) W^{-1}$ è una soluzione di $Q(X) = 0$.

Se invece R_1 è singolare possiamo solo concludere che non esistono soluzioni *diagonalizzabili* per l'equazione quadratica (1).

2.2.1 Abbassare il grado

Si utilizza un “trucco” ben noto per le equazioni differenziali ordinarie omogenee, che consiste nell'abbassare il grado al prezzo di aumentare la dimensione. Si definiscano le due matrici

$$F = \begin{pmatrix} 0 & I_n \\ -C & -B \end{pmatrix}, \quad G = \begin{pmatrix} I_n & 0 \\ 0 & A \end{pmatrix} \in \mathbb{C}^{2n \times 2n}. \quad (3)$$

Lemma 3. $S \in \mathbb{C}^n$ è una soluzione di $Q(X) = 0$ se e soltanto se

$$F \begin{pmatrix} I \\ S \end{pmatrix} = G \begin{pmatrix} I \\ S \end{pmatrix} S = G \begin{pmatrix} S \\ S^2 \end{pmatrix}.$$

Dimostrazione. La verifica è un calcolo:

$$F \begin{pmatrix} I \\ S \end{pmatrix} = \begin{pmatrix} 0 & I \\ -C & -B \end{pmatrix} \begin{pmatrix} I \\ S \end{pmatrix} = \begin{pmatrix} S \\ -C - BS \end{pmatrix}, \quad G \begin{pmatrix} I \\ S \end{pmatrix} S = \begin{pmatrix} I & 0 \\ 0 & A \end{pmatrix} \begin{pmatrix} S \\ S^2 \end{pmatrix} = \begin{pmatrix} S \\ AS^2 \end{pmatrix},$$

e l'equazione significativa è quella della seconda riga a blocchi. □

La (3) è soltanto una tra le tante possibili riduzioni.

Per vedere che il fascio $F - \lambda G$ è equivalente (per GEP) a $Q(\lambda) = \lambda^2 A + \lambda B + C$, si consideri la trasformazione

$$x \mapsto y = \begin{pmatrix} x \\ \lambda x \end{pmatrix},$$

da cui $(F - \lambda G)y = 0_{2n} \iff Q(\lambda)x = 0_n$. Se A è invertibile allora lo è anche G (e vale ovviamente anche il viceversa); in tal caso il problema si riduce ulteriormente al problema standard agli autovalori $(H - \lambda I)y = 0$, dove

$$H = G^{-1}F = \begin{pmatrix} 0_n & I_n \\ -A^{-1}C & -A^{-1}B \end{pmatrix}. \quad (4)$$

Non sempre la risoluzione di un'istanza lineare di GEP di dimensione doppia è conveniente rispetto alla corrispondente quadratica.

2.3 Metodo di Schur

Teorema 2. (Caratterizzazione delle soluzioni delle equazioni quadratiche matriciali) Tutte e sole le soluzioni di $Q(X) = AX^2 + BX + C = 0$ sono della forma $X = Z_{21}Z_{11}^{-1} = Q_{11}T_{11}S_{11}^{-1}Q_{11}^{-1}$, dove

$$Q^H F Z = T \quad Q^H G Z = S$$

è una forma normale di Schur generalizzata del fascio $F - \lambda G$ e F, G sono come definite in (3). La precedente si intende 2×2 a blocchi $n \times n$.

Dimostrazione. Facciamo vedere che $X = Z_{21}Z_{11}^{-1}$ risolve.

$$\begin{aligned} FZ = QT &\iff \begin{pmatrix} 0 & I_n \\ -C & -B \end{pmatrix} \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix} \iff \\ &\iff \begin{pmatrix} Z_{21} & Z_{22} \\ -CZ_{11} - BZ_{21} & -CZ_{12} - BZ_{22} \end{pmatrix} = \begin{pmatrix} Q_{11}T_{11} & Q_{11}T_{12} + Q_{12}T_{22} \\ Q_{21}T_{11} & Q_{21}T_{12} + Q_{22}T_{22} \end{pmatrix} \iff \\ &\iff \begin{cases} Z_{21} = Q_{11}T_{11} \\ Z_{22} = Q_{11}T_{12} + Q_{12}T_{22} \\ -CZ_{11} - BZ_{21} = Q_{21}T_{11} \\ -CZ_{12} - BZ_{22} = Q_{21}T_{12} + Q_{22}T_{22}. \end{cases} \end{aligned}$$

In maniera analoga, prendendo solo la prima colonna a blocchi di $GZ = QS$ si trova in particolare

$$\begin{cases} Z_{11} = Q_{11}S_{11} \\ AZ_{21} = Q_{21}S_{11}, \end{cases} \quad (5)$$

da cui segue che anche Q_{11} e S_{11} sono invertibili. Mettendo insieme con le equazioni della prima colonna a blocchi di $FZ = QT$ si trova

$$-C - BZ_{21}Z_{11}^{-1} = Q_{21}T_{11}Z_{11}^{-1} = Q_{21}Q_{11}^{-1}Z_{21}Z_{11}^{-1} = Q_{21}S_{11}^{-1}Z_{21}Z_{11}^{-1} = A(Z_{21}Z_{11}^{-1})^2,$$

e ciò equivale a dire che $X = Z_{21}Z_{11}^{-1}$ risolve.

Vediamo il viceversa; sia X una soluzione. Per il lemma 3 allora $F \begin{pmatrix} I_n \\ X \end{pmatrix} = G \begin{pmatrix} X \\ X^2 \end{pmatrix}$. Prendo una fattorizzazione QR di $\begin{pmatrix} I_n \\ X \end{pmatrix}$:

$$\begin{pmatrix} I \\ X \end{pmatrix} = ZR' = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix} \begin{pmatrix} R \\ 0_n \end{pmatrix}$$

con $Z \in \mathbb{C}^{2n \times 2n}$ unitaria e $R' \in \mathbb{C}^{2n \times n}$, $R \in \mathbb{C}^{n \times n}$ triangolare superiore. Da $Z_{11}R = I_n$ segue che entrambi i blocchi sono non singolari e anzi l'uno l'inverso dell'altro. Usando questo fatto nella seconda riga a blocchi troviamo

$$X = Z_{21}R = Z_{21}Z_{11}^{-1}.$$

Prendiamo adesso una fattorizzazione QR di GZ :

$$GZ = \begin{pmatrix} I_n & 0 \\ 0 & A \end{pmatrix} Z = QS = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix}$$

e premoltiplichiamo la $F \begin{pmatrix} I_n \\ X \end{pmatrix} = G \begin{pmatrix} X \\ X^2 \end{pmatrix}$ per Q^H :

$$\underbrace{Q^H F Z}_{T:=} \begin{pmatrix} R \\ 0 \end{pmatrix} = \underbrace{Q^H G Z}_{=S} \begin{pmatrix} R \\ 0 \end{pmatrix} X \iff \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} R \\ 0 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix} \begin{pmatrix} R \\ 0 \end{pmatrix} X \iff \begin{pmatrix} T_{11}R \\ T_{21}R \end{pmatrix} = \begin{pmatrix} S_{11}RX \\ 0 \end{pmatrix},$$

e siccome R è invertibile questo implica $T_{21} = 0$.

Abbiamo ottenuto $Q^H F Z = T$ e $Q^H G Z = S$, con Q, Z unitarie e S, T triangolari soltanto a blocchi: bisognerebbe che i loro blocchi diagonali fossero triangolari. Dobbiamo cioè completare la $Q^H(F, G)Z = (T, S)$ a una forma normale di Schur del fascio (F, G) mantenendo però $X = Z_{21}Z_{11}^{-1}$.

Un modo per far ciò è calcolare due forme di Schur per i blocchi diagonali corrispondenti (T_{ii}, S_{ii}) per $i = 1, 2$:

$$U_1^H (T_{11}, S_{11}) V_1 = (T'_{11}, S'_{11}), \quad U_2^H (T_{22}, S_{22}) V_2 = (T'_{22}, S'_{22}),$$

dove tutti i secondi membri sono triangolari superiori; $U = \text{diag}(U_1, U_2)$ e $V = \text{diag}(V_1, V_2)$ sono unitarie quindi per chiusura della classe lo sono anche $Q' = QU$ e $Z' = ZV$ e si verifica facilmente che questa è una forma normale di Schur per (F, G) :

$$Q'^H F Z' = U^H Q^H F Z V = U^H T V = \begin{pmatrix} T'_{11} & * \\ 0 & T'_{22} \end{pmatrix},$$

e analogamente che anche $Q'^H G Z'$ è triangolare superiore.

In queste nuove coordinate $X' = Z'_{21}Z'_{11}{}^{-1} = Z_{21}V_1(Z_{11}V_1)^{-1} = Z_{21}V_1V_1^{-1}Z_{11}^{-1} = Z_{21}Z_{11}^{-1} = X$. $\square$

Osservazione 3. Dalla (5) si vede che se A è singolare allora almeno una tra Q_{21} e S_{11} è singolare; in quest'ultimo caso, (sempre dalla (5)) anche Z_{11} è singolare. La parte “viceversa” della dimostrazione dice però che in questo caso, se esiste una soluzione di $Q(X) = 0$ allora è possibile trovare una decomposizione di Schur per cui Z_{11} è invertibile.

2.3.1 Stabilità numerica: riscalatura

L'algoritmo standard per calcolare la forma normale di Schur del fascio (3) è il QZ (vedasi il Teorema 1): alcune varianti di QZ ottengono un blocco Z_{11} invertibile, quando possibile.

Questo algoritmo è molto stabile. L'unico problema sorge con l'inversione di Z_{11} , che a priori può essere mal condizionata.

Lemma 4. *Sia $X = Z_{21}Z_{11}^{-1}$ una soluzione di $Q(X)$ calcolata a partire dalla forma normale di Schur generalizzata di (F, G) come nel teorema 2. Allora*

$$\kappa_2(Z_{11}) \leq 1 + \|X\|_2^2$$

Dimostrazione. Siccome Z è unitaria: $Z^H Z = I_{2n}$, il primo blocco diagonale dà

$$Z_{11}^H Z_{11} + Z_{21}^H Z_{21} = I_n, \tag{6}$$

da cui

$$1 = \|Z_{11}^H Z_{11} + Z_{21}^H Z_{21}\|_2 \geq \| \|Z_{21}^H Z_{21}\|_2 - \|Z_{11}^H Z_{11}\|_2 \| \geq \| \|Z_{11}^H Z_{11}\|_2 \| = \|Z_{11}^H Z_{11}\|_2 \tag{7}$$

Sempre dalla (6) si ha:

$$\begin{aligned} I_n = Z_{11}^H Z_{11} + Z_{21}^H Z_{21} &= Z_{11}^H Z_{11} + Z_{11}^H Z_{11}^{-H} (Z_{21}^H Z_{21}) Z_{11}^{-1} Z_{11} = Z_{11}^H Z_{11} + Z_{11}^H \underbrace{(Z_{21} Z_{11}^{-1})^H}_{X^H} \underbrace{Z_{21} Z_{11}^{-1}}_X Z_{11} = \\ &= Z_{11}^H (I_n + X^H X) Z_{11} \Rightarrow Z_{11}^{-1} = Z_{11}^H (I_n + X^H X), \end{aligned}$$

da cui, passando alle norme e usando (7),

$$\|Z_{11}^{-1}\|_2 = \|Z_{11}^H (I_n + X^H X)\|_2 \leq \underbrace{\|Z_{11}^H\|_2}_{=\|Z_{11}\|_2 \leq 1} \underbrace{\|(I_n + X^H X)\|_2}_{\leq 1 + \|X\|_2^2},$$

e infine, sempre per la (7),

$$\kappa_2(Z_{11}) = \|Z_{11}\|_2 \|Z_{11}^{-1}\|_2 \leq 1 \cdot (1 + \|X\|_2^2) = 1 + \|X\|_2^2. \quad \square$$

Siccome non c'è un metodo generale per stabilire in anticipo la norma di una soluzione X , una via è la seguente: si calcola X e se si scopre che $\|X\|_F$ è molto grande, si ripete il calcolo sul problema riscaldato

$$AX^2 + BX + C \longrightarrow \tilde{A}\left(\frac{1}{\rho}X\right)^2 + \tilde{B}\left(\frac{1}{\rho}X\right) + \tilde{C}$$

con $\tilde{A} = \rho^2 A$, $\tilde{B} = \rho B$, $\tilde{C} = C$, $\rho = \|X\|_2$. Gli autovalori del fascio riscaldato sono ρ^{-1} volte quelli del fascio originario $F - \lambda G$ quindi l'ordinamento è mantenuto. Dopo aver risolto in $\frac{1}{\rho}X$ l'equazione riscaldata, basta moltiplicarla per ρ ottenendo una soluzione meglio condizionata rispetto alla precedente: ρ scarica la difficoltà del calcolo dell'inversione di Z_{11} su una omotetia.

2.4 Metodo iterativo di Bernoulli

Per un'equazione quadratica scalare esistono metodi di punto fisso che convergono alla radice di massimo e minimo modulo, supposto che questi siano diversi. Generalizziamo questa definizione al caso matriciale usando lo spettro di una soluzione.

Supporremo che A sia invertibile, cosicché $Q(\lambda)$ ammetta esattamente $2n$ autovalori, come affermato nel Lemma 1. Ordiniamo gli autovalori per modulo decrescente:

$$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_{2n}| \quad (8)$$

Definizione 3. (Soluzione dominante, soluzione minimale) Una soluzione S di $Q(X) = 0$ si dice *dominante* se $\text{sp}(S) = \{\lambda_1, \dots, \lambda_n\}$ e $|\lambda_n| > |\lambda_{n+1}|$; si dice *minimale* se $\text{sp}(S) = \{\lambda_{n+1}, \dots, \lambda_{2n}\}$ e $|\lambda_n| > |\lambda_{n+1}|$.

In generale $Q(X)$ non ammette una soluzione minimale o dominante, anche se ha soluzioni. Inoltre può esistere una soluzione dominante senza che esista una soluzione minimale, e viceversa.

Però se esiste una soluzione dominante, questa è unica, e lo stesso dicasi per la minimale.

Osservazione 4. Se $Q(X)$ ammette una soluzione dominante S_1 , questa è invertibile.

Osservazione 5. Se $Q(\lambda)$ ha $2n$ autovalori, ordinati come in (8), che soddisfano $|\lambda_n| > |\lambda_{n+1}|$ ed *entrambi* i due gruppi di autovettori corrispondenti $\{v_1, \dots, v_n\}$ e $\{v_{n+1}, \dots, v_{2n}\}$ sono linearmente indipendenti, allora $Q(X)$ ammette una soluzione dominante e una soluzione minimale.

Infatti basta prendere $S_1 = V_1 \text{diag}(\lambda_1, \dots, \lambda_n) V_1^{-1}$ e $S_2 = V_2 \text{diag}(\lambda_{n+1}, \dots, \lambda_{2n}) V_2^{-1}$, dove

$$V_1 = (v_1 | \dots | v_n), \quad V_2 = (v_{n+1} | \dots | v_{2n}).$$

Un sistema vibrante sovrasmorzato (Definizione 4) verifica queste ipotesi.

2.4.1 Iterazione di Bernoulli

Consideriamo la ricorsione a due termini (con Y_0, Y_1 condizioni iniziali)

$$AY_{i+1} + BY_i + CY_{i-1} = 0 \quad (9)$$

e la matrice "di Vandermonde a blocchi"

$$V(S_1, S_2) = \begin{pmatrix} I_n & I_n \\ S_1 & S_2 \end{pmatrix} \in \mathbb{C}^{2n \times 2n}.$$

Teorema 3. Se S_1, S_2 sono soluzioni di $Q(X)$ tali che $V(S_1, S_2)$ è non singolare allora la soluzione generale della relazione ricorsiva (9) è

$$Y_i = S_1^i \Omega_1 + S_2^i \Omega_2, \quad (10)$$

dove le $\Omega_1, \Omega_2 \in \mathbb{C}^{n \times n}$ sono determinate dalle condizioni iniziali, e per la precisione

$$\begin{pmatrix} \Omega_1 \\ \Omega_2 \end{pmatrix} = V(S_1, S_2)^{-1} \begin{pmatrix} Y_0 \\ Y_1 \end{pmatrix}.$$

Dimostrazione. Per vedere che la (10) definisce effettivamente una soluzione basta sostituirla nella (9):

$$\begin{aligned} AY_{i+1} + BY_i + CY_{i-1} &= A(S_1^{i+1} \Omega_1 + S_2^{i+1} \Omega_2) + B(S_1^i \Omega_1 + S_2^i \Omega_2) + C(S_1^{i-1} \Omega_1 + S_2^{i-1} \Omega_2) = \\ &= (AS_1^2 + BS_1 + C) S_1^{i-1} \Omega_1 + (AS_2^2 + BS_2 + C) S_2^{i-1} \Omega_2 = 0_n S_1^{i-1} \Omega_1 + 0_n S_2^{i-1} \Omega_2 = 0. \end{aligned}$$

Per $i = 0, 1$ la (10) fornisce rispettivamente $\Omega_1 + \Omega_2 = Y_0$ e $S_1\Omega_1 + S_2\Omega_2 = Y_1$, che si può scrivere in forma matriciale (a blocchi)

$$\begin{pmatrix} I & I \\ S_1 & S_2 \end{pmatrix} \begin{pmatrix} \Omega_1 \\ \Omega_2 \end{pmatrix} = \begin{pmatrix} Y_0 \\ Y_1 \end{pmatrix},$$

che ammette un'unica soluzione per l'ipotesi di non singolarità di $V(S_1, S_2)$.

Viceversa, se Z_i è una soluzione della (9), prendasi Y_i come nella (10) con condizioni iniziali $Y_0 = Z_0$ e $Y_1 = Z_1$. Allora

$$AY_2 = -BY_1 - CY_0 = -BZ_1 - CZ_0 = AZ_2$$

da cui $Z_2 = Y_2$ per la non singolarità di A . Per induzione vale allora $Z_i = Y_i$ per ogni $i \in \mathbb{N}$. $\square$

2.4.2 Non-singolarità della Vandermonde a blocchi

Per la matrice di Vandermonde scalare

$$V(x_1, \dots, x_p) = (x_j^{i-1})_{i,j=1}^p$$

condizione sufficiente¹ per la non-singolarità è che gli x_i siano tutti diversi.

Questo criterio non è estendibile in generale al caso della Vandermonde a blocchi, a meno che non si tratti di soluzioni di un'equazione matriciale quadratica.

Esempio 3. Prendansi $X_1 = \begin{pmatrix} 2 & 0 \\ -2 & 1 \end{pmatrix}$ e $X_2 = \begin{pmatrix} 4 & 2 \\ 0 & 3 \end{pmatrix}$, cosicché $\text{sp}(X_1) \cap \text{sp}(X_2) = \{1, 2\} \cap \{3, 4\} = \emptyset$. Ma la Vandermonde a blocchi

$$V(X_1, X_2) = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 2 & 0 & 4 & 2 \\ -2 & 1 & 0 & 3 \end{pmatrix}$$

è singolare.

Teorema 4. *Siano S_1 e S_2 soluzioni di $Q(X) = 0$. Se $\text{sp}(S_1) \cap \text{sp}(S_2) = \emptyset$ allora $V(S_1, S_2)$ è invertibile.*

Dimostrazione. Per assurdo.

Supponiamo $V(S_1, S_2)$ singolare e sia $0 \neq v \in \ker(V(S_1, S_2))$. In particolare la seconda riga a blocchi di $V(S_1, S_2)v = 0_{2n}$ implica che $\begin{pmatrix} S_1 & S_2 \end{pmatrix} v = 0_n$.

Per l'invertibilità di A si possono scrivere le $Q(S_i) = 0$ anche come

$$(A^{-1}C \quad A^{-1}B) \begin{pmatrix} I_n \\ S_i \end{pmatrix} = -S_i^2,$$

che messe insieme in forma matriciale danno

$$(A^{-1}C \quad A^{-1}B) \begin{pmatrix} I_n & I_n \\ S_1 & S_2 \end{pmatrix} = -\begin{pmatrix} S_1^2 & S_2^2 \end{pmatrix},$$

il che implica in particolare $\begin{pmatrix} S_1^2 & S_2^2 \end{pmatrix} v = 0_n$.

Mettendole insieme entrambe

$$0_{2n} = \begin{pmatrix} S_1 & S_2 \\ S_1^2 & S_2^2 \end{pmatrix} v = \begin{pmatrix} I_n & I_n \\ S_1 & S_2 \end{pmatrix} \begin{pmatrix} S_1 & 0_n \\ 0_n & S_2 \end{pmatrix} v,$$

ossia $\text{diag}(S_1, S_2)v \in \ker(V(S_1, S_2))$, e questo dimostra che $\ker(V(S_1, S_2))$ è un sottospazio invariante per $\text{diag}(S_1, S_2)$.

Quindi $\ker(V(S_1, S_2))$ contiene un autovettore w di $\text{diag}(S_1, S_2)$; siccome per ipotesi gli spettri di S_1 e S_2 sono disgiunti, w deve avere per forza la forma $w = \begin{pmatrix} w_1^T & 0_n^T \end{pmatrix}^T$ oppure $w = \begin{pmatrix} 0_n^T & w_2^T \end{pmatrix}^T$.

In entrambi i casi

$$0_{2n} = V(S_1, S_2)w = \begin{pmatrix} I_n & I_n \\ S_1 & S_2 \end{pmatrix} w$$

implica $w = 0_{2n}$, ciò che è assurdo. $\square$

¹e necessaria.

Sappiamo che in generale per due matrici quadrate Z_1 e Z_2 , se

$$\min_{\lambda \in \text{sp}(Z_1)} |\lambda| > \max_{\lambda \in \text{sp}(Z_2)} |\lambda| \quad (11)$$

allora

$$\exists \lim_{i \rightarrow +\infty} \|Z_2^i\| \|Z_1^{-i}\| = 0$$

Siamo ora pronti per enunciare il risultato principale di convergenza:

Teorema 5. (Convergenza alla soluzione dominante del metodo di Bernoulli) Sia $Q(X) = 0$ un'equazione matriciale quadratica che ammetta una soluzione dominante S_1 e una minimale S_2 . Sia Y_i la soluzione dell'equazione ricorsiva (9) con condizioni iniziali $Y_0 = 0_{n \times n}$ e $Y_1 = I_n$.

Allora Y_i è definitivamente non singolare e

$$\lim_{i \rightarrow +\infty} Y_i Y_{i-1}^{-1} = S_1.$$

Dimostrazione. Gli spettri delle due soluzioni dominante e massimale sono disgiunti², quindi per il Teorema 4 la $V(S_1, S_2)$ è non singolare e allora per il Teorema 3 la soluzione generale della (9) è data dalla (10): $Y_i = S_1^i \Omega_1 + S_2^i \Omega_2$.

Le condizioni iniziali danno

$$\Omega_1 + \Omega_2 = Y_0 = 0, \quad S_1 \Omega_1 + S_2 \Omega_2 = Y_1 = I,$$

da cui segue $(S_1 - S_2) \Omega_1 = I$, quindi in particolare Ω_1 è invertibile.

Dimostriamo anzitutto che Y_i è definitivamente invertibile. Dalla (10), siccome S_1 è invertibile, si ha

$$Y_i = S_1^i (\Omega_1 + S_1^{-i} S_2^i \Omega_2).$$

Come conseguenza delle definizioni di soluzione dominante e minimale vale la (11), da cui segue che $S_1^{-i} S_2^i \rightarrow 0_{n \times n}$ e quindi, per la continuità di $\det$ e il Teorema di permanenza del segno, che Y_i è non singolare per i abbastanza grande.

Usando di nuovo la definizione di Y_i data dalla (10) si trova

$$\begin{aligned} Y_i Y_{i-1}^{-1} &= (S_1^i \Omega_1 + S_2^i \Omega_2) (S_1^{i-1} \Omega_1 + S_2^{i-1} \Omega_2)^{-1} = \\ &= \left(S_1 + S_2^i \Omega_2 \Omega_1^{-1} S_1^{-(i-1)} \right) (S_1^{i-1} \Omega_1) \left(\left(I + S_2^{i-1} \Omega_2 \Omega_1^{-1} S_1^{-(i-1)} \right) (S_1^{i-1} \Omega_1) \right)^{-1} = \\ &= \left(S_1 + S_2^i \Omega_2 \Omega_1^{-1} S_1^{-(i-1)} \right) \left(I + S_2^{i-1} \Omega_2 \Omega_1^{-1} S_1^{-(i-1)} \right)^{-1} =: \Gamma_1 \Gamma_2^{-1}. \end{aligned}$$

Sempre dal fatto che S_1 e S_2 verificano la (11), seguono le

$$\lim_{i \rightarrow +\infty} S_2^i S_1^{-(i-1)} = 0_{n \times n} \implies \lim_{i \rightarrow +\infty} S_2^i \Omega_2 \Omega_1^{-1} S_1^{-(i-1)} = 0_{n \times n} \implies \lim_{i \rightarrow +\infty} \Gamma_1 = S_1,$$

e la

$$\lim_{i \rightarrow +\infty} S_2^{i-1} S_1^{-(i-1)} = 0_{n \times n} \implies \lim_{i \rightarrow +\infty} S_2^{i-1} \Omega_2 \Omega_1^{-1} S_1^{-(i-1)} = 0_{n \times n} \implies \lim_{i \rightarrow +\infty} \Gamma_2^{-1} = I_n^{-1} = I_n,$$

che insieme danno la tesi. $\square$

È importante rimarcare il fatto che l'ipotesi del Teorema richiede che *entrambe* le soluzioni minimale e dominante esistano; ci sono controesempi in cui $Q(X)$ ammette una soluzione dominante ma non una soluzione minimale (o viceversa) per cui l'iterazione definita sopra non converge.

2.4.3 Calcolo delle soluzioni

Modifichiamo la (9) in modo da calcolare direttamente la soluzione dominante di $Q(X) = 0$. Postmoltiplicandola per Y_{i-1}^{-1} si ottiene

$$A Y_{i+1} Y_{i-1}^{-1} + B Y_i Y_{i-1}^{-1} + C = 0,$$

che si può riscrivere anche

$$(A Y_{i+1} Y_i^{-1} + B) Y_i Y_{i-1}^{-1} + C = 0,$$

²sono anzi separati nel senso dei moduli: $|\lambda_1| > |\lambda_2|$ ($\forall \lambda_1 \in \text{sp}(S_1), \lambda_2 \in \text{sp}(S_2)$).

3.1 Risonanza

Supponiamo che le masse siano sollecitate da una forza esterna di tipo oscillatorio semplice, dove le frequenze di oscillazione sono uguali per tutte le masse mentre l'intensità può variare; si ottiene

$$M\ddot{x} + B\dot{x} + Kx = F, \quad F(t) = e^{i\omega_0 t} f, \quad f \in \mathbb{R}^n.$$

Sempre supponendo l'esistenza di $2n$ autovalori λ_i , autovettori destri e sinistri v_i e u_i , si ottiene la soluzione particolare, purché $i\omega_0$ **non** sia un autovalore,

$$x_p(t) = e^{i\omega_0 t} \sum_{j=1}^{2n} \frac{u_j^H f}{i\omega_0 - \lambda_j} v_j.$$

Da ciò si vede che, fintanto che gli autovettori sinistri u_i non sono tutti ortogonali al vettore delle intensità della forza esterna f , all'avvicinarsi della frequenza di sollecitazione ω_0 a una delle frequenze naturali del sistema⁴, indipendentemente dall'intensità della forza il sistema si avvicina a uno stato di risonanza diventando estremamente instabile.

Con il “trucco standard” si può abbassare l'ordine da 2 a 1, al prezzo di raddoppiare la dimensione: si pone

$$y(t) = \begin{pmatrix} \dot{x}(t) \\ x(t) \end{pmatrix} \Rightarrow \dot{y}(t) = \begin{pmatrix} \ddot{x}(t) \\ \dot{x}(t) \end{pmatrix}$$

da cui

$$\dot{y}(t) = Ay(t), \quad A = \begin{pmatrix} -M^{-1}B & -M^{-1}K \\ I_n & 0_n \end{pmatrix}, \quad (13)$$

che è un sistema dinamico lineare di dimensione $2n$ la cui soluzione generale $y(t) = e^{tA}y_0$, $y_0 := y(0)$ è calcolabile a partire dagli autovalori e autovettori di A .

Consideriamo quindi il problema agli autovalori (standard) $Ay = \lambda y$.

Osservazione 6. A è la matrice “companion a blocchi” del fascio $Q(\lambda) = \lambda^2 M + \lambda B + K$.

Osservazione 7. Se invece utilizziamo la (3) per trasformare il GEP $Q(\lambda) = \lambda^2 M + \lambda B + K$ in un fascio lineare si ottiene

$$F = \begin{pmatrix} 0_n & I_n \\ -K & -B \end{pmatrix}, \quad G = \begin{pmatrix} I_n & 0_n \\ 0_n & M \end{pmatrix}$$

da cui il fascio $F - \lambda G$ che, essendo G invertibile perché lo è M , è equivalente (cfr. (4)) al problema agli autovalori standard $\det(G^{-1}F - \lambda I_{2n}) = 0$, dove

$$G^{-1}F = \begin{pmatrix} 0_n & I_n \\ -M^{-1}K & -M^{-1}B \end{pmatrix}$$

si ottiene da A tramite una trasposizione delle righe e delle colonne (a blocchi): $G^{-1}F = \begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix} A \begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix}$.

Quindi le due riduzioni sono sostanzialmente equivalenti.

M è sempre invertibile, ma è tanto peggio condizionata quanto più è grande la massima differenza tra le masse: infatti $\kappa_2(M) = \|M\|_2 \|M^{-1}\|_2 = \rho(M)\rho(M^{-1}) = \frac{\max_i m_i}{\min_i m_i}$. In tal caso il calcolo dell'inversa M^{-1} è numericamente instabile, si può affrontare una forma equivalente del problema $Ay = \lambda y$ moltiplicandone la prima riga a blocchi per M , ottenendo il GEP lineare

$$\begin{pmatrix} -B & -K \\ I & 0 \end{pmatrix} y = \lambda \begin{pmatrix} M & 0 \\ 0 & I \end{pmatrix} y$$

3.2 Dimensione uno

Consideriamo dapprima il caso $n = 1$, cosicché c'è una sola massa (m), una sola molla (di costante k), un solo ammortizzatore (di costante b):

$$m\ddot{x}(t) + b\dot{x}(t) + kx(t) = 0. \quad (14)$$

⁴gli autovalori immaginari puri.

3.2.1 Caso non smorzato

Eliminando anche l'ammortizzatore si ottiene un oscillatore armonico semplice

$$\ddot{x} + \omega^2 x = 0, \quad \omega = \sqrt{\frac{k}{m}}$$

la cui soluzione generale è

$$x(t) = c_1 \cos(\omega t) + c_2 \sin(\omega t),$$

dove le costanti c_1, c_2 che parametrizzano la famiglia si possono ricavare dalle condizioni iniziali: infatti

$$x(0) = c_1, \quad \dot{x}(0) = [-c_1 \omega \sin(\omega t) + c_2 \omega \cos(\omega t)]_{t=0} = c_2 \omega \Rightarrow c_2 = \frac{\dot{x}(0)}{\omega}.$$

Il periodo è $T = \frac{2\pi}{\omega}$.

In generale $m, b, k \neq 0$ e la (14) si trasforma in un sistema dinamico lineare con la trasformazione standard

$$\dot{y} = Ly, \quad y = \begin{pmatrix} \dot{x} \\ x \end{pmatrix}, \quad L = \begin{pmatrix} -\frac{b}{m} & -\frac{k}{m} \\ 1 & 0 \end{pmatrix}$$

(cfr. (13)). Le parti reali degli autovalori sono concordi perché $\det(L) = k/m > 0$ e negative perché $\text{tr}(L) = -b/m < 0$. Il polinomio caratteristico è $p_L(\lambda) = \lambda^2 + \frac{b}{m}\lambda + \frac{k}{m}$, con discriminante

$$\Delta = \frac{1}{m^2} (b^2 - 4km).$$

Distinguiamo tre casi qualitativamente differenti a seconda del segno di Δ .

3.2.2 Caso sovrasmorzato

Gli autovalori sono reali e distinti se e solo se $\Delta > 0$, cioè quando

$$b^2 > 4km. \tag{15}$$

In questo caso la famiglia a due parametri delle soluzioni è

$$x(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t},$$

quindi tutte le soluzioni tendono esponenzialmente all'equilibrio per ogni condizione iniziale; una soluzione passa *trasversalmente* per il punto d'equilibrio al più una volta.

Infatti una soluzione $x(t)$ passa trasversalmente dall'equilibrio all'istante $t = \tau$ se $x(\tau) = 0$ e $\dot{x}(\tau) \neq 0$. Ma $\dot{x}(t) = c_1 \lambda_1 e^{\lambda_1 t} + c_2 \lambda_2 e^{\lambda_2 t}$ cambia segno al più una volta.

Smorzamento critico. Se $\Delta = 0$ ovvero $b^2 = 4km$ c'è un unico autovalore $\lambda = -\frac{b}{2m} < 0$ doppio con molteplicità geometrica 1. Quindi la forma di Jordan di L è

$$\begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix},$$

e la soluzione generale ha la forma

$$x(t) = (c_1 + c_2 t) e^{\lambda t},$$

Questo è un caso limite del caso sovrasmorzato, infatti il sistema continua a tendere all'equilibrio in maniera asintoticamente esponenziale e la massa m attraversa la posizione di equilibrio al più una volta. Però in questo caso lo smorzamento è appena sufficiente per evitare oscillazioni.

3.2.3 Caso sottosmorzato

Come è ragionevole aspettarsi, in questo caso lo smorzamento non è sufficiente ad evitare il moto oscillatorio periodico. Infatti $\Delta < 0 \iff b^2 < 4km$, ci sono due autovalori complessi coniugati con parte reale negativa

$$\overline{\lambda_2} = \lambda_1 = r + i\omega, \quad r = -\frac{b}{2m}, \quad \omega = \frac{\sqrt{4km - b^2}}{2m},$$

e la soluzione generale

$$x(t) = c_1 e^{rt} \cos(\omega t) + c_2 e^{rt} \sin(\omega t)$$

decade sempre esponenzialmente verso l'equilibrio ma accompagnata da oscillazioni di frequenza costante ω , e la posizione di equilibrio viene attraversata trasversalmente infinite volte. In questo senso, un sistema sottosmorzato esibisce un comportamento caratteristico di *vibrazione*.

3.3 Dimensione arbitraria

3.3.1 Caso sovrasmorzato

Si estende la definizione (15) al caso generale:

Definizione 4. (Sovrasmorzamento) Il problema agli autovalori generalizzato $Q(\lambda)x = (\lambda^2 M + \lambda B + K)x = 0$ si dice *sovrasmorzato* se

1. $M, B, K \in \mathbb{R}^{n \times n}$ sono simmetriche;
2. M, B sono definite positive e K è semidefinita positiva;
3. $(x^T B x)^2 > 4 (x^T M x) (x^T K x) \quad (\forall x \in \mathbb{C}^n \setminus \{0\})$.

Nel caso sovrasmorzato sia M che B sono non singolari, quindi per il Lemma 1 $Q(\lambda)$ ammette $2n$ autovalori. Lancaster nel 1966 ha dimostrato che in un problema sovrasmorzato

- gli autovalori sono tutti reali e non positivi;
- gli autovalori dominanti e minimali sono separati: cioè secondo l'ordinamento (8) $\lambda_n > \lambda_{n+1}$.
- corrispondentemente agli autovalori dominanti e minimali i due gruppi di autovettori $v_1, \dots, v_n$ e $v_{n+1}, \dots, v_{2n}$ (rispettivamente) sono linearmente indipendenti.

Da ciò segue che un problema sovrasmorzato ammette una soluzione dominante e una soluzione minimale (Definizione 3).

La condizione di sovrasmorzamento non è di semplice verifica in generale, ma è ovviamente implicata dalla

$$\min^2 \text{sp}(B) > 4 \cdot \max \text{sp}(M) \cdot \max \text{sp}(K),$$

per verificare quest'ultima si può procedere ad esempio con il metodo delle potenze (per i fattori a membro destro) e delle potenze inverse (per il fattore a membro sinistro).

Un'istanza sovrasmorzata di GEP può essere attaccata con il metodo di Bernoulli, che è generalmente più appetibile rispetto al calcolo degli autovalori del fascio di dimensione doppia.

Per il risultato di Lancaster esposto sopra, si può applicare il metodo di Bernoulli usando l'iterazione (12) per calcolare la soluzione dominante S_1 di $Q(X) = 0$ e la sua analoga per calcolare la soluzione minimale S_2 (perché anch'essa è invertibile).

A questo punto le coppie autovalore/autovettore corrispondenti si trovano risolvendo i due problemi standard agli autovalori $(S_i - \lambda I)x = 0$; altri autovalori/autovettori si possono trovare risolvendo il GEP lineare associato $((B + MS_i) + \lambda M)x = 0$ (cfr. p. 4).

Riferimenti bibliografici

- [1] James W. Demmel. *Applied Numerical Linear Algebra*. SIAM, 1997.
- [2] Nicholas J. Higham e Hyun-Min Kim. «Numerical analysis of a quadratic matrix equation». In: *IMA Journal of Numerical Analysis* 20 (2000), pp. 499–519.
- [3] Ilse C. F. Ipsen. «Accurate Eigenvalues for Fast Trains». In: *SIAM News* 37.9 (nov. 2004).
- [4] Françoise Tisseur e Karl Meerbergen. «The Quadratic Eigenvalue Problem». In: *SIAM Review* 43.2 (2001), pp. 235–286.