Calcolo Numerico A - a.a. 2008/09

30 marzo 2009

Stima degli autovalori

Definizione (Cerchi di Gerschgorin)

Sia $A = (a_{ij}) \in \mathbb{C}^{n\times n}$*. I cerchi del piano complesso*

$$\mathcal{R}_i = \left\{ z \in \mathbb{C} \,\middle|\, |z - a_{ii}| \leq \sum_{j=1,\, j\neq i}^{n} |a_{ij}| \right\}, \quad j = 1, \ldots, n$$

di centri a_{ii} *e raggi* $r_i = \sum_{j=1,\, j\neq i}^{n} |a_{ij}|$ *sono detti cerchi di Gerschgorin.*

Teorema (Gerschgorin)

Gli autovalori di una matrice A sono contenuti nell'unione dei suoi cerchi di Gerschgorin:

$$\sigma(A) \subseteq \mathcal{R} := \bigcup_{i=1,\ldots,n} \mathcal{R}_i \tag{0.1}$$

dove $\sigma(A) := \{\lambda_1, \ldots, \lambda_n\}$ *è lo spettro di A.*

Sia λ un autovalore di A e x un autovettore corrispondente. Essendo $AX = \lambda x$, abbiamo:

$$\sum_{j=1}^{n} a_{ij} x_j = \lambda x_i, \quad i = 1, \ldots, n,$$

da cui segue

$$(\lambda - a_{ii}) x_i = \sum_{j=1, j \neq i}^{n} a_{ij} x_j, \quad i = 1, \ldots, n.$$

Sia x_k la componente di x di massimo modulo. A meno di moltiplicare x per uno scalare, possiamo supporre che essa sia uno, cioè $x = (x_1, \ldots, x_{k-1}, 1, x_{k+1}, \ldots, x_n)^T$ con $|x_i| \leq 1$. Allora, prendendo $i = k$ nella formula sopra e passando alle norme, otteniamo

$$|\lambda - a_{kk}| \leq \sum_{j=1, j \neq i}^{n} |a_{kj}||x_j| \leq \sum_{j=1, j \neq i}^{n} |a_{kj}|.$$

Si noti che non sappiamo a priori quale sia la componente di x di massimo modulo e dunque a quale cerchio appartenga λ.

Osservazione

Ricordo che una matrice e la sua trasposta hanno stesso polinomio caratteristico, infatti

$$p_A(x) := \det(A - x\, I_n) = \det(A - x\, I_n)^T = \det(A^T - x\, I_n) := p_{A^T}(x).$$

Ne segue che una stima simile a (0.1) vale usando le colonne anziché le righe, cioè:

$$\sigma(A) \subseteq \mathcal{C} := \bigcup_{i=1,\dots,n} \mathcal{C}_i$$

dove

$$\mathcal{C}_i = \left\{ z \in \mathbb{C} \,\middle|\, |z - a_{ii}| \leq \sum_{i=1,\ i \neq j}^{n} |a_{ij}| \right\}, \quad j = 1, \dots, n.$$

In conclusione:

$$\sigma(A) \subseteq \mathcal{R} \cap \mathcal{C} = \left(\bigcup_{i=1,\dots,n} \mathcal{R}_i \right) \cap \left(\bigcup_{i=1,\dots,n} \mathcal{C}_i \right).$$

Esempio

$$A = \begin{pmatrix} 15 & -2 & 2 \\ 1 & 10 & -3 \\ -2 & 1 & 0 \end{pmatrix}$$

$$\begin{aligned} \mathcal{R}_1 &= \{z \in \mathbb{C} \,:\, |z - 15| \leq 4\} \\ \mathcal{R}_2 &= \{z \in \mathbb{C} \,:\, |z - 10| \leq 4\} \\ \mathcal{R}_3 &= \{z \in \mathbb{C} \,:\, |z| \leq 3\} \end{aligned}$$

$$\begin{aligned} \mathcal{C}_1 &= \{z \in \mathbb{C} \,:\, |z - 15| \leq 3\} \\ \mathcal{C}_2 &= \{z \in \mathbb{C} \,:\, |z - 10| \leq 3\} \\ \mathcal{C}_3 &= \{z \in \mathbb{C} \,:\, |z| \leq 5\} \end{aligned}$$

$$\sigma \subseteq \{z \in \mathbb{C} : |z - 15| \leq 3\} \cap \{z \in \mathbb{C} : |z - 10| \leq 3\} \cap \{z \in \mathbb{C} : |z| \leq 3\}.$$

Matrici a diagonale dominante

Definizione

Una matrice $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ si dice

- *a dominanza diagonale per righe se*

$$|a_{kk}| \geq \sum_{j=1,\, j \neq k}^{n} |a_{kj}|, \quad \forall k = 1, \ldots, n,$$

- *a dominanza diagonale per colonne se*

$$|a_{kk}| \geq \sum_{i=1,\, i \neq k}^{n} |a_{ik}|, \quad \forall k = 1, \ldots, n.$$

Se le disuguaglianze precedenti sono verificate in senso stretto, cioè non valgono le uguaglianze, allora A si dice a dominanza diagonale stretta (per righe o per colonne, rispettivamente).

Corollario

Se A è a dominanza diagonale stretta (per righe o per colonne), allora i suoi autovalori sono diversi da zero. Quindi le matrici a dominanza diagonale stretta sono invertibili.

Segue dal fatto che i cerchi di Gerschgorin non intersecano lo zero.

Matrici definite positive

Definizione

Una matrice $A \in \mathbb{R}^{n \times n}$ si dice

- *definita positiva se*

$$(x, Ax) = x^T Ax > 0, \quad \forall x \neq 0,$$

- *semidefinita positiva se*

$$(x, Ax) = x^T Ax \geq 0, \quad \forall x \in \mathbb{R}^n.$$

Osservazione

Tutti gli elementi diagonali di una matrice A definita positiva sono positivi (infatti deve essere $e_i^T A e_i = a_{ii} > 0$).

Osservazione

Ricordo che ogni matrice quadrata reale $B \in \mathbb{R}^{n\times n}$ si scrive in modo unico come somma di una matrice simmetrica e una antisimmetrica:

$$B = B' + B''$$

dove

$$B' = \frac{1}{2}(B + B^T), \quad B'' = \frac{1}{2}(B - B^T).$$

Osservazione

Se A è una matrice antisimmetrica, allora si ha:

$$(x, Ax) = x^T A x = (x^T A x)^T = x^T A^T x = -x^T A x = -(x, Ax) = 0, \quad \forall x \in \mathbb{R}^n.$$

Quinidi una matrice simmetrica è sempre semidefinita positiva, ma non definita positiva, secondo la definizione data sopra.

Proposizione

Sia $A \in \mathbb{R}^{n \times n}$ una matrice simmetrica. Allora sono fatti equivalenti:

1. *A è definita positiva;*
2. *ogni minore principale di A è una matrice definita positiva;*
3. *gli autovalori di A sono positivi;*
4. *esiste una matrice non singolare S simmetrica tale che $A = S^2$;*
5. *esiste una matrice non singolare B tale che $A = B^T B$;*
6. *i minori principali dominanti hanno determinante positivo (criterio di Sylvester).*

Corollario

Una matrice simmetrica, a dominanza diagonale stretta, i cui elementi diagonali siano positivi, è definita positiva.

Infatti essa ha autovalori positivi per il teorema di Gerschgorin e quindi è definita positiva per la caratterizzazione sopra.

SISTEMI LINEARI

Condizionamento di una matrice

(Cfr. P.G. Ciarlet)

Esempio (R. S. Wilson)

Consideriamo il sistema lineare

$$\begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix}, \text{ *di soluzione* } \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Consideriamo il sistema perturbato in cui i secondi membri sono modificati leggermente, mentre la matrice A dei coefficienti è rimasta invariata:

$$\begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix} \begin{pmatrix} x_1 + \delta x_1 \\ x_2 + \delta x_2 \\ x_3 + \delta x_3 \\ x_4 + \delta x_4 \end{pmatrix} = \begin{pmatrix} 32,1 \\ 23,9 \\ 33,1 \\ 31,9 \end{pmatrix}, \text{ *di soluzione* } \begin{pmatrix} 9,2 \\ -12,6 \\ 4,5 \\ -1,1 \end{pmatrix}.$$

Un errore relativo dell'ordine di $1/200$ *sui dati (i termini noti) porta ad un errore di ordine* 10 *sui risutati (la soluzione del sistema), cioè si ha un rapporto di amplificazione degli errori relativi dell'ordine di 2000.*

Consideriamo analogamente il sistema perturbato in cui questa volta sono gli elementi della matrice che sono modificati leggermente:

$$\begin{pmatrix} 10 & 7 & 8,1 & 7,2 \\ 7,08 & 5,04 & 6 & 5 \\ 8 & 5,98 & 9,89 & 9 \\ 6,99 & 4,99 & 9 & 9,98 \end{pmatrix} \begin{pmatrix} x_1 + \Delta x_1 \\ x_2 + \Delta x_2 \\ x_3 + \Delta x_3 \\ x_4 + \Delta x_4 \end{pmatrix} = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix}, \text{ di soluzione } \begin{pmatrix} -81 \\ 137 \\ -34 \\ 22 \end{pmatrix}.$$

Anche in questo caso piccole variazioni dei dati (la matrice dei coefficienti) modificano completamente il risultato (la soluzione del sistema lineare). Tuttavia la matrice A dei coefficienti si presenta bene: è simmetrica e ha determinante uno; lo stesso si può dire dell'inversa

$$A^{-1} = \begin{pmatrix} 25 & -41 & 10 & -6 \\ -41 & 68 & -17 & 10 \\ 10 & -17 & 5 & -3 \\ -6 & 10 & -3 & 2 \end{pmatrix}.$$

Definizione

Un sistema lineare per cui a "piccoli" errori nei dati corrispondono "grandi" errori nella soluzione viene detto mal condizionato o mal posto.
Viceversa, un sistema lineare per cui a piccoli errori nei dati corrispondono piccoli errori nella soluzione si dice ben condizionato o ben posto.

Consideriamo un sistema lineare di n equazioni in n incognite

$$Ax = b \tag{0.2}$$

con coefficenti reali o complessi. Supporremo sempre che il sistema abbia almeno una soluzione. Diremo anche che il sistema è consistente.

Definizione

Data una norma indotta $\|\cdot\|_*$ su $\mathbb{C}^{n\times n}$, si dice condizionamento della matrice $A \in \mathbb{C}^{n\times n}$ il numero reale

$$\mu_*(A) = \|A\|_* \|A^{-1}\|_*$$

Teorema

Siano $\delta A \in \mathbb{C}^{n\times n}$ *e* $\delta b \in \mathbb{C}^n$ *pertubazioni della matrice A e del vettore b, rispettivamente. Consideriamo su* $\mathbb{C}^{\times n}$*una qualsiasi norma matriciale indotta* $\|\cdot\|$*.*
Supponiamo A non singolare. Se $\|A^{-1}\|\|\delta A\| < 1$*, allora* $A + \delta A$ *è non singolare.*
Supponiamo anche che il sistema sia non omogeneo (i.e. $b \neq 0$*). Indicata con* $x + \delta x$ *una soluzione del sistema perturbato*

$$(A + \delta A)(x + \delta x) = b + \delta b, \tag{0.3}$$

risulta

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\mu(A)}{1 - \mu(A)\|\delta A\|/\|A\|}\left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|}\right),$$

dove $\mu(A)$ *indica il condizionamento rispetto ad una qualsiasi norma indotta* $\|\cdot\|$*.*

Dal fatto che $\|A^{-1} \cdot \delta A\| \leq \|A^{-1}\|\|\delta A\| < 1$, segue che $I + A^{-1} \cdot \delta A$ è non singolare, e quindi anche $A + \delta A = A(I + A^{-1} \cdot \delta A)$ è non singolare (essendo il prodotto di due matrici non singolari). Inoltre:

$$\|(I + A^{-1} \cdot \delta A)^{-1}\| \leq \frac{1}{1 - \|A^{-1} \cdot \delta A\|} \leq \frac{1}{1 - \|A^{-1}\|\|\delta A\|}.$$

D'altra parte, ricavando δx dalla (0.3) e tenendo conto che $Ax = b$, si ha

$$\begin{aligned} \delta x &= (A + \delta A)^{-1}(-\delta A\, x + \delta b) \\ &= (I + A^{-1} \cdot \delta A)^{-1} A^{-1}(-\delta A\, x + \delta b), \end{aligned}$$

da cui passando alle norme e utilizzando la stima sopra si ricava

$$\|\delta x\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\|\delta A\|}(\|\delta A\|\|x\| + \|\delta b\|).$$

Essendo $b \neq 0$, si ha $x \neq 0$, quindi si possono dividere entrambi i membri $|x|$ e, utilizzando che $|b| \leq |A||x|$, si trova:

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\|\delta A\|}(\|\delta A\| + \|\delta b\|\|A\|/\|b\|) \leq \frac{\mu(A)}{1 - \mu(A)\|\delta A\|/\|A\|}\left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|}\right).$$

$\square$

Indicando con

$$\varepsilon_A = \frac{\|\delta A\|}{\|A\|}, \quad \varepsilon_b = \frac{\|\delta b\|}{\|b\|}, \quad \varepsilon_x = \frac{\|\delta x\|}{\|x\|},$$

le perturbazioni relative di A, b e x, rispettivamente, allora abbiamo la maggiorazione:

$$\varepsilon_x \leq \mu(A) \frac{\varepsilon_A + \varepsilon_B}{1 - \mu(A)\varepsilon_A}. \tag{0.4}$$

Nota

$$\mu(A) := \|A\| \|A^{-1}\| \geq \|AA^{-1}\| = \|I_n\| = 1$$

Osservazione

Se il condizionamento $\mu(A)$ assume valori piccoli, allora a piccole perturbazioni sui dati del sistema corrispondono piccole perturbazioni della soluzione x.
Questo significa che il sistema $Ax = b$ è ben condizionato se il condizionamento della matrice A è piccolo. In tal caso si dice anche che la matrice A è ben condizionata.

Proposizione

Per ogni matrice $A \in \mathbb{C}^{n \times n}$ si ha:

1. $\mu(A) \geq 1$
2. $\mu(A^{-1}) = \mu(A)$
3. $\mu(\alpha A) = \mu(A), \quad \forall \alpha \in \mathbb{C}$

Esempio (Matrice di Hilbert)

Sia $H^{(n)} \in \mathbb{R}^{n\times n}$ *la matrice definita da*

$$H_{ij}^{(n)} := \frac{1}{i+j-1}, \quad \textit{per } i,j = 1,\ldots,n.$$

Questo è un esempio classico di matrice mal condizionata, infatti si ha

Table: Condizionamento delle matrici di Hilbert

n	$\mu_2\left(H^{(n)}\right)$	$\mu_\infty\left(H^{(n)}\right)$
1	*1*	*1*
2	*1,505*	*27*
3	$5,241 \cdot 10^2$	$7,480 \cdot 10^2$
4	$1,551 \cdot 10^4$	$2,837 \cdot 10^4$
5	$4,766 \cdot 10^5$	$9,436 \cdot 10^5$
...	...	...
10	$1,603 \cdot 10^{13}$	$3,535 \cdot 10^{13}$

Si noti che la maggiorazione (0.4) può fornire una stima eccessiva della perturbazione ε_x. Si veda il seguente

Esempio

Prendiamo:

$$A = \begin{pmatrix} 1 & 1 \\ 0,99 & 1 \end{pmatrix}.$$

Allora:

$$A^{-1} = 10^2 \begin{pmatrix} 1 & -1 \\ -0,99 & 1 \end{pmatrix}.$$

Dunque:

$$\mu_\infty(A) = \|A\|_\infty \|A^{-1}\|_\infty \simeq 400$$

Prendendo $b = (0 - 0,01)^T$, il sistema $Ax = b$ a soluzione $x = (1, -1)^T$. Perturbiamo poi il sistema con una matrice

$$\delta A = 0,002 \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}, \quad \text{per cui } \varepsilon_A = \frac{\|\delta A\|_\infty}{\|A\|_\infty} = \frac{0,004}{2} = 0,002.$$

Abbiamo che il sistema lineare $(A + \delta A)x = b$, ha ancora soluzione $x = (1, -1)^T$. Quindi troviamo che la perturbazione relativa è nulla, cioè:

$$0 = \varepsilon_x \leq \mu(A) \frac{\varepsilon_A + \varepsilon_B}{1 - \mu(A)\varepsilon_A} = 4.$$

Esercizio

Calcolare il condizionamento in norma 2 e in norma ∞ delle seguenti matrici:

$$A = \begin{pmatrix} 1 & 2 \\ 1.001 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 39 & 16 \\ 71 & 29 \end{pmatrix}, \quad C = \begin{pmatrix} 100 & 99 \\ 99 & 98 \end{pmatrix},$$

$$D = \begin{pmatrix} 1 & -1 & 1 \\ -1 & \eta & \eta \\ 1 & \eta & \eta \end{pmatrix}, \quad 0 < \eta < \frac{1}{2}.$$

$\left[\mu_2(A) = 5001,\ \mu_\infty(A) = 6002, \ldots, \mu_2(D) = \frac{1}{\varepsilon},\ \mu_\infty(D) = \frac{3}{2}\left(1 + \frac{1}{\varepsilon}\right)\right].$

Esercizio

Data $A \in \mathbb{R}^{n \times n}$, si provi che:

$$\begin{aligned} \frac{1}{n}\,\mu_2(A) \ &\le \mu_1(A) \le n\,\mu_2(A) \\ \frac{1}{n}\,\mu_\infty(A) &\le \mu_2(A) \le n\,\mu_\infty(A) \\ \frac{1}{n^2}\mu_1(A) &\le \mu_\infty(A) \le n^2\mu_1(A). \end{aligned}$$

Ricerca delle soluzioni di un sistema lineare

1. Metodi diretti: forniscono la soluzione (esatta in assenza di errori di arrotodamento) in un numero finito di passi;
2. Metodi iterativi: forniscono la soluzione come limite di una successione di vettori (ad ogni iterazione viene richiesto il calcolo di un prodotto matrice-vettore).

Sistemi lineari con matrice triangolare

La soluzione di un sistema lineare $Ax = b$ è particolarmente semplice quando A è triangolare (superiore).

$$\begin{cases} a_{ii}x_i + \sum_{j=i+1}^{n} a_{ij}x_j = b_i, \quad i = 1, \ldots, n-1, \\ a_{nn}x_n = b_n. \end{cases} \tag{0.5}$$

Se A è non singolare, cioè $a_{ii} \neq 0$, per ogni $i = 1, \ldots, n$, abbiamo

$$\begin{cases} x_n = \frac{b_n}{a_{nn}}, \\ x_i = \frac{1}{a_{ii}}\left(b_i - \sum_{j=i+1}^{n} a_{ij}x_j\right), \quad i = n-1, \ldots, 1. \end{cases} \tag{0.6}$$

Quindi la soluzione si calcola con una "sostituzione all'indietro". Se la matrice è triangolare inferiore, la risoluzione avviene in modo analogo ("sostituzione in avanti").

Il costo computazionale di un sistema con matrice triangolare è determinato tenendo conto che la componente x_i viene calcolata con $n - i$ moltiplicazioni e una divisione, per cui risulta:

$$\sum_{i=1}^{n}(n - i + 1) = \sum_{k=1}^{n} i = \frac{n^2 + n}{2} \simeq \frac{n^2}{2}.$$

Fattorizzazioni

Supponiamo che la matrice A del sistema lineare

$$Ax = b,$$

si scomponga nel prodotto di due matrici

$$A = BC, \tag{0.7}$$

allora il sistema lineare risulta equivalente a

$$\begin{cases} By = b, \\ Cx = y. \end{cases} \tag{0.8}$$

Fattorizzazioni classiche associate a diversi metodi di risoluzione sono le sequenti:

1. Fattorizzazione LU: L matrice triangolare inferiore (lower) con elementi diagonali uguali ad uno ($l_{ii} = 1$), U matrice triangolare superiore (upper).
   Questa fattorizzazione è associata al metodo di Gauss.
2. Fattorizzazione LL^T: L matrice triangolare inferiore con elementi diagonali positivi ($l_{ii} > 0$).
   Questa fattorizzazione è associata al metodo di Cholesky.
3. Fattorizzazione QR: Q matrice unitaria (ortogonale nel caso reale), R matrice triangolare superiore.
   Questa fattorizzazione è associata al metodo di Householder.

Osservazione

- *Se la matrice A è reale, allora anche le matrici delle fattorizzazioni lo sono.*
- *Il costo computazionale delle fattorizzazioni si vedrà che sarà dato da un numero di operazioni dell'ordine di n^3, mentre il costo computazionale della risoluzione dei sistemi (0.8) sarà dato da un numero di operazioni dell'ordine di n^2.*
- *La fattorizzazione QR esiste sempre, mentre non sempre si possono avere le fattorizzazione LU e LL^T.*

Teorema

Siano $A_k = (a_{ij})_{1 \leq i,j \leq k}$ i minori principali dominanti (detti anche "di testa") ottenuti dalla matrice A.
Esiste ed è unica la fattorizzazione LU di A, se e solo se le matrici A_k sono non singolari (per ogni $k = 1, \ldots, n$).

Osservazione

La fattorizzazione LU è sempre possibile per i seguenti tipi di matrici:

- *Matrici a dominanza diagonale stretta (infatti anche i loro minori principali sono di questo tipo e quindi invertibili per il corollario del teorema di Gerschgorin).*
- *Matrici simmetriche definite positive (segue dal criterio di Sylvester).*

Teorema

Data A matrice quadrata reale o complessa, esiste una matrice di permutazione Π per cui la matrice ΠA ammette la fattorizzazione LU, cioè:

$$\Pi A = LU.$$

Teorema

Se la matrice A è hermitiana (simmetrica nel caso reale) definita positiva, allora esiste ed è unica la fattorizzazione LL^T di A.

Osservazione

La fattorizzazione QR di una matrice A non è unica, infatti, per ogni matrice unitaria (ortogonale nel caso reale) e diagonale, detta matrice di fase,

$$S = \begin{pmatrix} \theta_1 & & \\ & \ddots & \\ & & \theta_n \end{pmatrix}, \quad \text{con } |\theta_i| = 1,\ i = 1, \ldots, n,$$

risulta

$$QR = QS\bar{S}R = (QS)(\bar{S}R) = Q'R'.$$

Se però A è non singolare, allora la matrice triangolare R deve essere invertibile e quindi con elementi diagonali non nulli. Allora si può richiedere che essi siano reali positivi. Tale scomposizione è unica.