

# I Problemi dell'Algebra Lineare Numerica: Aspetti Computazionali e Condizionamento

Luca Gemignani  
luca.gemignani@unipi.it

8 marzo 2018

## Indice

<b>Lezione 1: Norme Matriciali e Norme Vettoriali.</b>	<b>1</b>
<b>Lezione 2: Il Problema della Risoluzione di un Sistema Lineare ed il suo Condizionamento.</b>	<b>4</b>
<b>Lezione 3: Il Problema del Calcolo degli Autovalori di una Matrice ed il suo Condizionamento.</b>	<b>6</b>
<b>Lezione 4: Teoremi di Localizzazione per Autovalori.</b>	<b>9</b>

## Lezione 1: Norme Matriciali e Norme Vettoriali.

I principali problemi dell'algebra lineare numerica concernono la risoluzione di sistemi lineari ed il calcolo di autovalori e/o autovettori di matrici. Studiarne il condizionamento significa misurare la sensibilità del problema considerato rispetto a perturbazioni dei dati (vettori e matrici) forniti in ingresso. Risulta pertanto essenziale disporre di strumenti per valutare la distanza tra vettori e matrici.

**Definizione 1.1.** Si dice *norma vettoriale* su  $\mathbb{R}^n$  una funzione  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  che soddisfa le seguenti proprietà:

1.  $\forall \mathbf{v} \in \mathbb{R}^n, f(\mathbf{v}) \geq 0$  ed inoltre  $f(\mathbf{v}) = 0 \iff \mathbf{v} = \mathbf{0}$ ;
2.  $\forall \mathbf{v} \in \mathbb{R}^n, \forall \alpha \in \mathbb{R}, f(\alpha \mathbf{v}) = |\alpha|f(\mathbf{v})$ ;
3.  $\forall \mathbf{v}, \mathbf{z} \in \mathbb{R}^n, f(\mathbf{v} + \mathbf{z}) \leq f(\mathbf{v}) + f(\mathbf{z})$ .

Se  $f$  è una norma vettoriale su  $\mathbb{R}^n$  indicheremo per comodità di notazione  $f(\mathbf{v}) = \|\mathbf{v}\|$ .

Si osservi che:

- Una norma vettoriale su  $\mathbb{R}^n$  induce una *distanza*  $d: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  tra elementi (punti) di  $\mathbb{R}^n$  definita come

$$\forall \mathbf{v}, \mathbf{z} \in \mathbb{R}^n, \quad d(\mathbf{v}, \mathbf{z}) = \|\mathbf{v} - \mathbf{z}\|.$$

Le proprietà della norma si traducono in analoghe proprietà della distanza indotta:

1. (*non negatività*)  $\forall \mathbf{v}, \mathbf{z} \in \mathbb{R}^n, \quad d(\mathbf{v}, \mathbf{z}) \geq 0$  e  $d(\mathbf{v}, \mathbf{z}) = 0 \iff \mathbf{v} = \mathbf{z}$ ;
  2. (*simmetria*)  $\forall \mathbf{v}, \mathbf{z} \in \mathbb{R}^n, \quad d(\mathbf{v}, \mathbf{z}) = d(\mathbf{z}, \mathbf{v})$ ;
  3. (*diseguaglianza triangolare*)  $\forall \mathbf{v}, \mathbf{z}, \mathbf{w} \in \mathbb{R}^n, \quad d(\mathbf{v}, \mathbf{z}) \leq d(\mathbf{v}, \mathbf{w}) + d(\mathbf{w}, \mathbf{z})$ .
- Le seguenti funzioni  $f(\mathbf{v}), \mathbf{v} = [v_1, \dots, v_n]^T \in \mathbb{R}^n$  sono norme su  $\mathbb{R}^n$  (rispettivamente dette *norma euclidea*, *norma 1* e *norma infinito*):

$$\begin{aligned} f(\mathbf{v}) = \|\mathbf{v}\|_2 &= \sqrt{\sum_{i=1}^n v_i^2} = \sqrt{\mathbf{v}^T \mathbf{v}}; \\ f(\mathbf{v}) = \|\mathbf{v}\|_1 &= \sum_{i=1}^n |v_i|; \\ f(\mathbf{v}) = \|\mathbf{v}\|_\infty &= \max_{1 \leq i \leq n} |v_i| \end{aligned}$$

- Sebbene l'utilizzo di differenti norme conduca a risultati quantitativamente differenti le proprietà qualitative degli oggetti e dei fenomeni analizzati risultano preservate. Vale infatti il seguente *Principio di equivalenza topologica (metrica) delle norme*:

**Teorema 1.1.** Siano  $\|\cdot\|$  e  $\|\cdot\|'$  due norme su  $\mathbb{R}^n$ . Allora esistono costanti  $\alpha, \beta > 0$  tali che:

$$\alpha \|\mathbf{v}\|' \leq \|\mathbf{v}\| \leq \beta \|\mathbf{v}\|', \quad \forall \mathbf{v} \in \mathbb{R}^n.$$

In particolare questo risultato implica che le proprietà (topologiche) di convergenza/divergenza di successioni e di continuità delle funzioni sono invarianti rispetto alla norma considerata.

♠ FAC Per una dimostrazione del risultato dalla disequaglianza triangolare segue che

$$|\|\mathbf{x}\| - \|\mathbf{y}\|| \leq \|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Ciò implica la continuità della funzione  $\mathbf{x} \rightarrow \|\mathbf{x}\|, \mathbf{x} \in \mathbb{R}^n$ . Dal teorema di Weierstrass si ricava quindi che la funzione valutata sull'insieme  $\mathcal{S} = \{\mathbf{v} \in \mathbb{R}^n : \|\mathbf{v}\|' = 1\}$  ammette massimo detto  $\beta$  e minimo detto  $\alpha$ .

Le norme vettoriali possono essere estese alle matrici.

**Definizione 1.2.** Si dice *norma matriciale* su  $\mathbb{R}^{n \times n}$  una funzione  $f: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$  che soddisfa le seguenti proprietà:

1.  $\forall A \in \mathbb{R}^{n \times n}, f(A) \geq 0$  ed inoltre  $f(A) = 0 \iff A = 0$ ;
2.  $\forall A \in \mathbb{R}^{n \times n}, \forall \alpha \in \mathbb{R}, f(A) = |\alpha|f(A)$ ;
3.  $\forall A, B \in \mathbb{R}^{n \times n}, f(A + B) \leq f(A) + f(B)$ ;
4.  $\forall A, B \in \mathbb{R}^{n \times n}, f(A \cdot B) \leq f(A) \cdot f(B)$ .

Se  $f$  è una norma vettoriale su  $\mathbb{R}^{n \times n}$  indicheremo per comodità di notazione  $f(A) = \|A\|$ .

Si osserva che:

- Analogamente a sopra una norma matriciale su  $\mathbb{R}^{n \times n}$  induce una *distanza*  $d: \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$  tra elementi di  $\mathbb{R}^{n \times n}$  definita come

$$\forall A, B \in \mathbb{R}^{n \times n}, d(A, B) = \|A - B\|.$$

- Dalle proprietà 1) e 4) si ricava che  $\|I_n\| \geq 1$ .

La seguente definizione descrive l'estensione di una norma vettoriale ad una norma matriciale detta *norma matriciale indotta o compatibile* con la norma vettoriale.

**Definizione 1.3.** Data  $\|\cdot\|$  una norma vettoriale su  $\mathbb{R}^n$  si dice *norma matriciale indotta o compatibile* con la norma vettoriale la funzione  $f: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$  definita da

$$\forall A \in \mathbb{R}^{n \times n}, f(A) = \max_{\{v \in \mathbb{R}^n: \|v\|=1\}} \|Av\|.$$

Per comodità di notazione scriveremo  $f(A) = \|A\|$  utilizzando lo stesso simbolo della norma vettoriale che induce la norma matriciale.

Si osserva che:

- La definizione è ben posta.

♠ FAC Si può infatti dimostrare come sopra che data  $A \in \mathbb{R}^{n \times n}$  la funzione  $v \rightarrow \|Av\|$  è continua per cui dal teorema di Weierstrass esiste il massimo. Inoltre la funzione  $f$  così definita verifica le proprietà 1–4 delle norme matriciali.

- Per una norma matriciale indotta da una norma vettoriale vale  $\|I_n\| = 1$ .
- Esistono norme matriciali che non sono indotte da una norma vettoriale. Ad esempio si consideri la funzione *norma di Frobenius* definita come

$$\forall A = (a_{i,j}) \in \mathbb{R}^{n \times n}, \|A\| = \sqrt{\sum_{1 \leq i,j \leq n} a_{i,j}^2} = \sqrt{\text{traccia}(A^T A)}.$$

Il seguente risultato esprime un'ulteriore proprietà di fondamentale importanza per l'analisi dei problemi e dei metodi computazionali.

**Teorema 1.2.** Sia  $\| \cdot \|$  una norma vettoriale su  $\mathbb{R}^n$  e sia  $\| \cdot \|$  la norma matriciale indotta. Vale allora che

$$\forall A \in \mathbb{R}^{n \times n}, \forall \mathbf{v} \in \mathbb{R}^n, \quad \| A \mathbf{v} \| \leq \| A \| \| \mathbf{v} \|.$$

*Dimostrazione.* Se  $\mathbf{v} = \mathbf{0}$  allora la relazione vale. Assumiamo pertanto che  $\mathbf{v} \neq \mathbf{0}$ . Si ha

$$\| A \frac{\mathbf{v}}{\| \mathbf{v} \|} \| \leq \| A \| = \max_{\{\mathbf{z} \in \mathbb{R}^n : \| \mathbf{z} \| = 1\}} \| A \mathbf{z} \|,$$

e quindi la tesi segue per la proprietà 2) delle norme vettoriali.  $\square$

La valutazione delle norme matriciali indotte dalla norma euclidea, dalla norma 1 e dalla norma infinito in accordo alla definizione risulta computazionalmente non praticabile. Vengono pertanto fornite le seguenti caratterizzazioni che seguono dalla definizione e mediante l'individuazione del punto di massimo forniscono lo strumento per il calcolo effettivo delle norme.

**Teorema 1.3.** Sia  $A = (a_{i,j}) \in \mathbb{R}^{n \times n}$ . Si ha

$$\begin{aligned} \| A \|_{\infty} &= \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{i,j}|; \\ \| A \|_1 &= \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{i,j}|. \end{aligned}$$

Detto inoltre  $\rho(B)$  il *raggio spettrale* di una matrice  $B \in \mathbb{R}^{n \times n}$  definito come il modulo dell'autovalore di modulo massimo di  $B$ , i.e.,

$$\forall B \in \mathbb{R}^{n \times n}, \quad \rho(B) = \max_{1 \leq i \leq n} |\lambda_i|, \quad \lambda_i, 1 \leq i \leq n, \text{ autovalore di } B,$$

allora vale

$$\| A \|_2 = \sqrt{\rho(A^T A)}.$$

Si osservi che mentre per la norma infinito e la norma 1 il calcolo si realizza facilmente con una sequenza finita di operazioni aritmetiche e confronti a partire dagli elementi della matrice, per la norma euclidea si richiede la risoluzione di un problema ausiliario (il calcolo del modulo dell'autovalore dominante di una matrice associata) di assai maggiore difficoltà (vedi Lezione 4 successiva).

## Lezione 2: Il Problema della Risoluzione di un Sistema Lineare ed il suo Condizionamento.

Il problema della risoluzione di un sistema lineare viene formulato come segue.

**Problema 2.1.** Data  $A = (a_{i,j}) \in \mathbb{R}^{n \times n}$  matrice invertibile e dato  $\mathbf{b} \in \mathbb{R}^n$  si cerca un vettore  $\mathbf{x} \in \mathbb{R}^n$  tale che

$$A \mathbf{x} = \mathbf{b}. \tag{1}$$

La matrice  $A$  è detta *matrice dei coefficienti* del sistema lineare. Il vettore  $\mathbf{b}$  è detto *vettore dei termini noti*. Il vettore  $\mathbf{x}$  è detto *vettore delle incognite*. La scrittura (1) è formalmente equivalente al sistema di equazioni lineari

$$\begin{cases} \sum_{j=1}^n a_{1,j}x_j = b_1 \\ \sum_{j=1}^n a_{2,j}x_j = b_2 \\ \vdots \\ \sum_{j=1}^n a_{n,j}x_j = b_n \end{cases}$$

Dall'invertibilità della matrice dei coefficienti segue l'esistenza di un tale  $\mathbf{x} \in \mathbb{R}^n$  (ad esempio  $\mathbf{x} = A^{-1}\mathbf{b}$ ) e la sua unicità (altrimenti si avrebbe  $\dim(\ker(A)) > 0$ ). Inoltre la *regola di Cramer* fornisce una descrizione delle componenti del vettore  $\mathbf{x}$  come funzioni razionali negli elementi della matrice  $A$  e del vettore  $\mathbf{b}$ . Quindi in linea di principio è possibile studiare formalmente il condizionamento del problema della risoluzione di un sistema lineare con gli strumenti (differenziali e sviluppi al primo ordine) visti in precedenza. Un tale studio tuttavia risulta poco informativo non evidenziando una misura generale "facilmente" calcolabile del condizionamento del problema in oggetto. Differentemente l'approccio descritto di seguito fornisce una maggiorazione dell'errore imputabile nel risultato alla perturbazione dei dati iniziali  $(A, \mathbf{b})$  in termini di un parametro detto *numero di condizione o di condizionamento* del sistema lineare che funge in questo contesto da analogo dei coefficienti di amplificazione.

La memorizzazione in macchina dei dati in ingresso  $(A, \mathbf{b})$  conduce a dati perturbati  $(\hat{A}, \hat{\mathbf{b}})$  con

$$\begin{aligned} \hat{A} &= (\hat{a}_{i,j}), \quad \hat{a}_{i,j} = a_{i,j}(1 + \epsilon_{i,j}), \quad |\epsilon_{i,j}| \leq u, \\ \hat{\mathbf{b}} &= [\hat{b}_1, \dots, \hat{b}_n]^T, \quad \hat{b}_j = b_j(1 + \delta_j), \quad |\delta_j| \leq u. \end{aligned}$$

Si può dunque scrivere

$$\hat{A} = A + F, \quad F = (a_{i,j}\epsilon_{i,j}); \quad \hat{\mathbf{b}} = \mathbf{b} + \mathbf{e}, \quad \mathbf{e} = [b_1\delta_1, \dots, b_n\delta_n]^T.$$

Per semplicità di analisi assumiamo di perturbare il solo termine noto  $\mathbf{b} \neq \mathbf{0}$ . Le stime così ottenute si estendono facilmente al caso più generale in cui si assumono perturbazioni sia su  $A$  che su  $\mathbf{b}$ . Sia  $\hat{\mathbf{x}}$  la soluzione del sistema perturbato, i.e.,

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}}.$$

Si ha

$$\hat{\mathbf{x}} - \mathbf{x} = A^{-1}\hat{\mathbf{b}} - A^{-1}\mathbf{b} = A^{-1}\mathbf{e},$$

e dunque passando ad una valutazione in norma

$$\|\hat{\mathbf{x}} - \mathbf{x}\| = \|A^{-1}\mathbf{e}\| \leq \|A^{-1}\| \|\mathbf{e}\|,$$

ove la norma matriciale è quella indotta dalla norma vettoriale iniziale. D'altra parte abbiamo che

$$\|\mathbf{b}\| = \|A\mathbf{x}\| \leq \|A\| \|\mathbf{x}\|,$$

da cui

$$\| \mathbf{x} \| \geq \| \mathbf{b} \| / \| A \| .$$

Combinando insieme queste relazioni si ottiene

$$\frac{\| \hat{\mathbf{x}} - \mathbf{x} \|}{\| \mathbf{x} \|} \leq \mathcal{K}(A) \frac{\| \mathbf{e} \|}{\| \mathbf{b} \|}, \quad \mathcal{K}(A) = \| A \| \| A^{-1} \| .$$

Questa relazione esprime il fatto che l'errore relativo in norma imputato alla perturbazione dei dati iniziali si magiora con l'errore relativo sui dati moltiplicato per il parametro  $\mathcal{K}(A)$  detto *numero di condizione o di condizionamento* del sistema lineare che assume il ruolo di coefficiente di amplificazione. Per una norma matriciale indotta da una norma vettoriale vale

$$1 = \| I_n \| = \| A A^{-1} \| \leq \| A \| \| A^{-1} \| = \mathcal{K}(A).$$

Se  $\mathcal{K}(A)$  è qualitativamente elevato il sistema lineare è detto *mal condizionato*. Se altrimenti  $\mathcal{K}(A)$  è qualitativamente modesto il sistema lineare è detto *ben condizionato*.

In norma infinito si ha

$$\| \mathbf{e} \|_{\infty} \leq u \| \mathbf{b} \|_{\infty},$$

da cui

$$\frac{\| \hat{\mathbf{x}} - \mathbf{x} \|_{\infty}}{\| \mathbf{x} \|_{\infty}} \leq u \mathcal{K}_{\infty}(A), \quad \mathcal{K}_{\infty}(A) = \| A \|_{\infty} \| A^{-1} \|_{\infty} .$$

Ne ricaviamo che una stima di  $\mathcal{K}_{\infty}(A)$  fornisce un'indicazione sul numero di cifre al più corrette attese sul risultato.

### Lezione 3: Il Problema del Calcolo degli Autovalori di una Matrice ed il suo Condizionamento.

Il problema del calcolo di un autovalore di una matrice viene formulato come segue.

**Problema 3.1.** Data  $A = (a_{i,j}) \in \mathbb{C}^{n \times n}$  si cerca  $\lambda \in \mathbb{C}$  tale per cui

$$\exists \mathbf{x} \in \mathbb{C}^n, \mathbf{x} \neq \mathbf{0} : A \mathbf{x} = \lambda \mathbf{x}, \quad (2)$$

o, equivalentemente,

$$\exists \mathbf{y} \in \mathbb{C}^n, \mathbf{y} \neq \mathbf{0} : \mathbf{y}^H A = \lambda \mathbf{y}^H. \quad (3)$$

Il vettore  $\mathbf{x}$  è detto *autovettore (destro)* relativo all'*autovalore*  $\lambda$  di  $A$ . Il vettore  $\mathbf{y}$  è detto *autovettore (sinistro)* relativo all'*autovalore*  $\lambda$  di  $A$ . Dalla relazione (2) o (3) segue che  $\lambda$  è autovalore di  $A$  se e soltanto se

$$\dim(\ker(A - \lambda I_n)) > 0$$

o, equivalentemente, se e soltanto se

$$\det(A - \lambda I_n) = 0.$$

Introdotta pertanto il polinomio

$$p(x) = \det(A - xI_n)$$

detto *polinomio caratteristico* della matrice  $A$  segue che  $\lambda \in \mathbb{C}$  è autovalore di  $A$  se e soltanto se

$$p(\lambda) = \det(A - \lambda I_n) = 0,$$

ovvero  $\lambda$  è una radice dell'equazione  $p(x) = 0$ .

La caratterizzazione degli autovalori come zeri del polinomio caratteristico ha importanti implicazioni teoriche e computazionali. Dal *teorema fondamentale dell'algebra* sappiamo che l'equazione algebrica  $p(x) = 0$  ammette esattamente  $n$  radici contate con la loro molteplicità. Dette  $\lambda_1, \lambda_2, \dots, \lambda_n$  queste radici (e quindi gli autovalori di  $A$ ) possiamo scrivere la fattorizzazione ottenuta raggruppando le radici uguali nella forma

$$p(x) = (-1)^n \prod_{i=1}^k (x - \lambda_{j_i})^{\sigma_i},$$

con  $1 \leq j_i \leq n$ ,  $\lambda_{j_i} \neq \lambda_{j_\ell}$  se  $i \neq \ell$ ,  $\sum_{i=1}^k \sigma_i = n$ . Il numero  $\sigma_i$  è detto *molteplicità algebrica* dell'autovalore  $\lambda_{j_i}$  e indica il numero di volte che  $\lambda_{j_i}$  compare nell'insieme  $\mathcal{S} = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$  detto *spettro* di  $A$ . Il numero

$$\tau_i = \dim(\ker(A - \lambda_{j_i} I_n)), \quad 1 \leq i \leq k,$$

è detto *molteplicità geometrica* dell'autovalore  $\lambda_{j_i}$  e indica il numero di autovettori destri o sinistri linearmente indipendenti relativi all'autovalore  $\lambda_{j_i}$ . È noto che

$$\sigma_i \geq \tau_i, \quad 1 \leq i \leq k,$$

e la situazione in cui vale l'uguaglianza tra i due indici per ogni  $i$  risulta di particolare interesse.

**Definizione 3.1.** Sia  $A \in \mathbb{C}^{n \times n}$  e  $V \in \mathbb{C}^{n \times n}$  invertibile allora la trasformazione  $A \rightarrow B = V^{-1}AV$  è detta *trasformazione per similitudine* di  $A$  mediante  $V$ .

Nel contesto del calcolo degli autovalori di una matrice le trasformazioni per similitudine sono rilevanti in quanto preservano gli autovalori insieme agli indici associati. Vale infatti che

$$\begin{aligned} \det(B - xI_n) &= \det(V^{-1}AV - xI_n) = \det(V^{-1}AV - xV^{-1}V) = \\ \det[V^{-1}(A - xI_n)V] &= \det(V^{-1})\det(A - xI_n)\det(V) = \det((A - xI_n)), \end{aligned}$$

e quindi  $A$  e  $B$  condividono lo stesso polinomio caratteristico. Analogamente si mostra che

$$\dim(\ker(A - \lambda_{j_i} I_n)) = \dim(\ker(B - \lambda_{j_i} I_n)), \quad 1 \leq i \leq k.$$

**Definizione 3.2.** La matrice  $A \in \mathbb{C}^{n \times n}$  è detta *diagonalizzabile* se esiste una trasformazione per similitudine che rende  $A$  diagonale, ovvero,

$$\exists V, \det(V) \neq 0: V^{-1}AV = D \text{ diagonale.}$$

Segue facilmente che se  $A$  è diagonalizzabile allora  $D = \text{diag}[\lambda_1, \dots, \lambda_n]$  ed inoltre  $AV = VD$  implica che le colonne di  $V$  definiscono gli autovettori (destri) corrispondenti. Il seguente teorema fornisce una condizione necessaria e sufficiente per la diagonalizzabilità di una matrice.

**Teorema 3.1.**  $A \in \mathbb{C}^{n \times n}$  è diagonalizzabile se e soltanto se  $\sigma_i = \tau_i, 1 \leq i \leq n$ .

Le seguenti classi di matrici diagonalizzabili risultano di interesse:

1. *Matrici simmetriche.* Una matrice  $A \in \mathbb{R}^{n \times n}$  si dice simmetrica se  $A = A^T$ . Le matrici simmetriche sono diagonalizzabili e la matrice  $V$  può essere scelta ortonormale, i.e.,  $V^T V = V V^T = I_n$ . Gli autovalori di una matrice simmetrica sono reali.
2. *Matrici hermitiane.* Una matrice  $A \in \mathbb{C}^{n \times n}$  si dice hermitiana se  $A = A^H$ . Le matrici hermitiane sono diagonalizzabili e la matrice  $V$  può essere scelta ortonormale, i.e.,  $V^H V = V V^H = I_n$ . Gli autovalori di una matrice hermitiana sono reali.
3. *Matrici con autovalori distinti.* Qualora  $\lambda_i \neq \lambda_j$  se  $i \neq j$  allora  $k = n$ ,  $\sigma_i = 1, 1 \leq i \leq n$  e dunque  $\tau_i = \sigma_i, 1 \leq i \leq n$ . Un autovalore  $\lambda$  per cui  $\sigma = \tau = 1$  è detto *semplice*.

Lo studio del condizionamento del calcolo degli autovalori risulta semplificato nell'ipotesi di autovalori semplici. In particolare si assuma che  $\lambda$  è autovalore di  $A$  con  $\sigma = \tau = 1$  e autovettore destro e sinistro denotati rispettivamente con  $\mathbf{x}$  e  $\mathbf{y}$ . Si consideri una perturbazione  $\hat{A}$  di  $A$  della forma  $\hat{A} = A + \epsilon F$ . Per  $\epsilon$  sufficientemente piccolo si può dimostrare che  $\hat{A}$  ammette un autovalore  $\lambda(\epsilon) \doteq \lambda + \epsilon \eta$  con corrispondente autovettore  $\mathbf{x}(\epsilon) \doteq \mathbf{x} + \epsilon \mathbf{z}$ . Dalla relazione

$$(A + \epsilon F)(\mathbf{x} + \epsilon \mathbf{z}) \doteq (\lambda + \epsilon \eta)(\mathbf{x} + \epsilon \mathbf{z}),$$

segue

$$A\mathbf{z} + F\mathbf{x} \doteq \lambda\mathbf{z} + \eta\mathbf{x},$$

che moltiplicando ambo i membri per  $\mathbf{y}^H$  implica

$$\eta \doteq \frac{\mathbf{y}^H F \mathbf{x}}{\mathbf{y}^H \mathbf{x}}$$

da cui la stima

$$|\lambda(\epsilon) - \lambda| \doteq |\epsilon| \left| \frac{\mathbf{y}^H F \mathbf{x}}{\mathbf{y}^H \mathbf{x}} \right|.$$

Dalla disuguaglianza di Cauchy-Schwarz si ha che

$$|\mathbf{y}^H F \mathbf{x}| \leq \|\mathbf{y}\|_2 \|F \mathbf{x}\|_2 \leq \|\mathbf{y}\|_2 \|\mathbf{x}\|_2 \|F\|_2$$



e quindi considerando autovettori normalizzati si perviene alla relazione

$$|\lambda(\epsilon) - \lambda| \doteq |\epsilon| \left| \frac{\mathbf{y}^H F \mathbf{x}}{\mathbf{y}^H \mathbf{x}} \right| \leq \| \epsilon F \|_2 \frac{1}{|\mathbf{y}^H \mathbf{x}|}.$$

Questa relazione indica che il condizionamento (rispetto all'errore assoluto) del calcolo di un autovalore semplice di una matrice è misurato dal reciproco del valore assoluto del prodotto scalare tra i rispettivi autovettori destro e sinistro normalizzati. Il lettore ne derivi che il problema del calcolo di un autovalore semplice di una matrice simmetrica o hermitiana è ben condizionato.

## Lezione 4: Teoremi di Localizzazione per Autovalori.

Gli aspetti computazionali della caratterizzazione degli autovalori di una matrice come soluzioni dell'equazione caratteristica associata sono pure rilevanti. Il teorema di Abel-Ruffini afferma che non esiste nessuna formula per le radici di una generica equazione algebrica di grado  $\geq 5$  in funzione dei coefficienti del polinomio, usando solo le operazioni aritmetiche e l'applicazione di radicali (radici quadrate, radici cubiche, ecc.). Ne consegue che, a differenza della soluzione del sistema lineare, gli autovalori di una matrice non sono generalmente esprimibili come funzioni razionali negli elementi della matrice. Pertanto la loro approssimazione si baserà sulla costruzione iterativa di successioni  $\{a_k\}_{k \in \mathbb{N}}$  di approssimanti che sotto ipotesi convenienti convergeranno ad un autovalore. Per la convergenza risulta essenziale disporre di un'approssimazione iniziale "sufficientemente buona". I teoremi di localizzazione individuano regioni del piano complesso ove gli autovalori sono confinati. Il più classico è il seguente detto *teorema di Gershgorin*.

**Teorema 4.1.** Sia  $A = (a_{i,j}) \in \mathbb{C}^{n \times n}$ . Definiamo i *cerchi di Gershgorin*  $K_i$ ,  $1 \leq i \leq n$ , come

$$K_i = \{z \in \mathbb{C} : |z - a_{i,i}| \leq \sum_{j=1, j \neq i}^n |a_{i,j}|\}, \quad 1 \leq i \leq n.$$

Allora

$$\lambda \text{ autovalore di } A \Rightarrow \lambda \in \cup_{i=1}^n K_i.$$

*Dimostrazione.* Sia  $\lambda$  autovalore di  $A$  con corrispondente autovettore destro  $\mathbf{x}$ . La relazione  $A\mathbf{x} = \lambda\mathbf{x}$  implica che

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

da cui

$$(\lambda - a_{i,i})x_i = \sum_{j=1, j \neq i}^n a_{i,j} x_j, \quad 1 \leq i \leq n. \quad (4)$$

Sia  $p$  l'indice di una componente di modulo massimo di  $\mathbf{x}$ , i.e.,  $|x_p| = \|\mathbf{x}\|_\infty$ . Poichè  $\mathbf{x} \neq 0$  si ha  $|x_p| > 0$ . La relazione per  $i = p$  in (4) porge

$$(\lambda - a_{p,p})x_p = \sum_{j=1, j \neq p}^n a_{p,j}x_j$$

da cui passando ai valori assoluti

$$|(\lambda - a_{p,p})x_p| = |\lambda - a_{p,p}||x_p| = \left| \sum_{j=1, j \neq p}^n a_{p,j}x_j \right| \leq \sum_{j=1, j \neq p}^n |a_{p,j}||x_j|,$$

e quindi dividendo ambo i membri per  $|x_p|$

$$|\lambda - a_{p,p}| \leq \sum_{j=1, j \neq p}^n |a_{p,j}| \frac{|x_j|}{|x_p|} \leq \sum_{j=1, j \neq p}^n |a_{p,j}|.$$

Questa relazione implica che  $\lambda \in K_p$  e dunque la tesi.  $\square$

Un risultato di inclusione generalmente più debole è fornito dal seguente detto *teorema di Hirsch*.

**Teorema 4.2.** Sia  $A = (a_{i,j}) \in \mathbb{C}^{n \times n}$  e sia  $\|\cdot\|$  una norma matriciale indotta da una norma vettoriale su  $\mathbb{C}^n$ . Allora

$$\lambda \text{ autovalore di } A \Rightarrow |\lambda| \leq \|A\|.$$

*Dimostrazione.* La relazione  $A\mathbf{x} = \lambda\mathbf{x}$  implica che

$$\|\lambda\mathbf{x}\| = |\lambda| \|\mathbf{x}\| = \|A\mathbf{x}\| \leq \|A\| \|\mathbf{x}\|$$

da cui la tesi.  $\square$