Norme di vettori e matrici

Dario A. Bini, Università di Pisa

7 luglio 2020

Sommario

Questo modulo didattico contiene risultati e proprietà relativi alle norme di vettori e di matrici.

1 Introduzione

Nello studio dei metodi di risoluzione di sistemi lineari è utile disporre del concetto di *norma* per valutare attraverso un numero reale non negativo la grandezza di un vettore o di una matrice.

1.1 Norme di vettori

Diamo la seguente

Definizione 1 *Una applicazione* $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$ *viene detta* norma vettoriale *se soddisfa alle seguenti proprietà*

- $\|x\| \geq 0$*, per ogni* $x \in \mathbb{R}$*;* $\|x\| = 0$ *se e solo se* $x = 0$*;*
- $\|\alpha x\| = |\alpha| \, \|x\|$*, per ogni* $\alpha \in \mathbb{C}$*,* $x \in \mathbb{C}^n$*;*
- $\|x + y\| \leq \|x\| + \|y\|$*, per ogni* $x, y \in \mathbb{C}^n$ *(diseguaglianza triangolare)*

Esempi importanti di norme sono:

- $\|x\|_1 = \sum_{i=1}^n |x_i|$ (norma 1)
- $\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{\frac{1}{2}}$ (norma euclidea, o norma 2)
- $\|x\|_\infty = \max_i |x_i|$ (norma infinito, o norma del massimo)

Queste norme sono casi speciali della *norma di Hölder*

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}, \quad p \geq 1,$$

dove $\|x\|_\infty = \lim_{p\to\infty} \|x\|_p$.

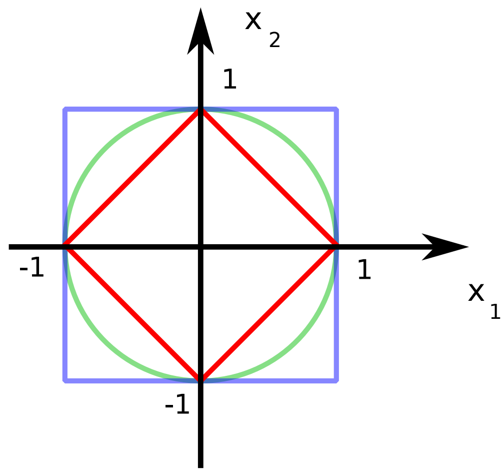

Figura 1: Palle unitarie in norma 1 (rosso), norma 2 (verde) e norma infinito (blu).

Nella figura 1 si riportano le sfere unitarie, cioè gli insiemi $S = \{x \in \mathbb{C}^n : \|x\| = 1\}$ nel caso delle norme 1,2 e ∞ per $n = 2$. In blu la palla per la norma infinito, in verde quella della norma 2 e in rosso quella della norma 1.

Non è difficile dimostrare che per una norma l'insieme $\{x \in \mathbb{C}^n : \quad \|x\| \leq 1\}$ è un insieme convesso.

Dim. Siano infatti x, y, tali che $\|x\|, \|y\| \leq 1$ e si consideri un generico punto z del segmento di estremi x e y. Vale allora $z = \alpha x + (1-\alpha)y$ con $0 \leq \alpha \leq 1$. Dalla diseguaglianza triangolare segue

$$\|z\| \leq \alpha\|x\| + (1-\alpha)\|y\| \leq \alpha + 1 - \alpha = 1$$

e quindi $\|z\| \leq 1$. $\square$

Questo fatto ci permette di dire che per $0 < p < 1$ l'espressione $\|x\|_p$ non può essere una norma essendo l'insieme $\{x \in \mathbb{C}^n : \quad \|x\| \leq 1\}$ non convesso. Nella figura 2 si riporta l'insieme S_p per valori di $p < 1$.

Si ricorda che un *prodotto scalare* su $\mathbb{C}^n$ è una applicazione da $\mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ che alla coppia (x, y) associa il numero $\langle x, y\rangle$ tale che

- $\langle x, y\rangle = \overline{\langle y, x\rangle}$.
- $\langle x, \alpha y\rangle = \alpha\langle x, y\rangle$, per ogni $\alpha \in \mathbb{C}$
- $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$
- $\langle x, x\rangle \geq 0$
- $\langle x, x\rangle = 0$ se e solo se $x = 0$

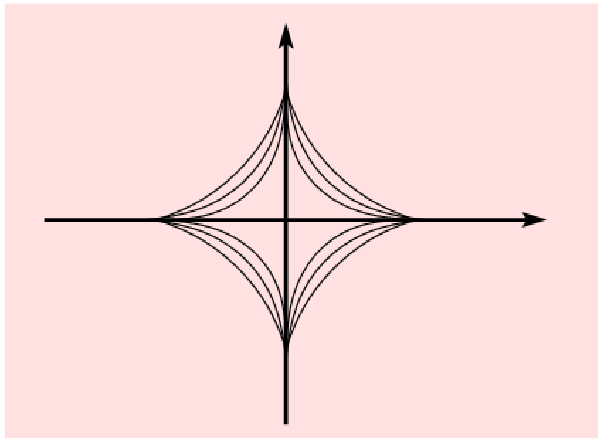

Figura 2: Insiemi S_p per valori di p minori di 1

Dalle proprietà del prodotto scalare discende

- $\langle \alpha x, y\rangle = \overline{\alpha}\langle x, y\rangle$, per ogni $\alpha \in \mathbb{C}$
- $\langle x, y+z\rangle = \langle x, y\rangle + \langle x, z\rangle$

Un esempio di prodotto scalare su $\mathbb{R}^n$ è dato da $\langle x, y\rangle = x^T y$; esso è detto *prodotto scalare euclideo*. Un esempio di prodotto scalare su $\mathbb{C}^n$ è dato da $\langle x, y\rangle = x^H y$; esso è detto *prodotto scalare hermitiano*.

Si osserva che, dato un prodotto scalare $\langle x, y\rangle$ su $\mathbb{C}^n$, allora l'applicazione $x \to \langle x, x\rangle^{1/2}$ è una norma. Questa norma viene detta *norma indotta dal prodotto scalare*. Ad esempio, la norma 2 è la norma indotta dal prodotto scalare hermitiano $\langle x, y\rangle = x^H y$.

Un prodotto scalare soddisfa la diseguaglianza di Cauchy-Schwarz nota anche come diseguaglianza di Cauchy-Bunyakowski-Schwarz

$$|\langle x, y\rangle|^2 \leq \langle x, x\rangle\langle y, y\rangle, \tag{1}$$

di cui si riporta la breve dimostrazione

dim. Se $y = 0$ la diseguaglianza è soddisfatta. Se $y \neq 0$ si pone $t = \langle y, x\rangle/\langle y, y\rangle$ per cui

$$0 \leq \langle x - ty, x - ty\rangle = \langle x, x\rangle - |\langle x, y\rangle|^2/\langle y, y\rangle$$

da cui, moltiplicando per $\langle y, y\rangle$, si ottiene la diseguaglianza di Cauchy-Schwarz $\square$

La diseguaglianza di Cauchy-Schwarz implica che $\langle x, y\rangle \leq 1$ per ogni coppia di vettori x e y tali che $\|x\| = \|y\| = 1$, dove $\|\cdot\|$ è la norma indotta dal prodotto scalare. Nel caso di $\mathbb{R}^n$ questo ci permette di definire l'angolo $\theta \in [0, \pi]$ formato da due vettori x e y mediante l'espressione $\cos\theta = \langle x, y\rangle/(\|x\|\,\|y\|)$. Nel caso di $\mathbb{C}^n$ si può definire l'angolo $\theta \in [0, \pi/2]$ mediante la relazione $\cos\theta = |\langle x, y\rangle|/(\|x\|\,\|y\|)$. Questo ci permette di definire l'ortogonalità di due vettori x, y quando $\langle x, y\rangle = 0$.

Ci si può chiedere se la norma 1 o la norma 2 siano o meno indotte da un prodotto scalare. Per questo vale il seguente risultato.

Teorema 1 *Una norma $\|\cdot\|$ è indotta da un prodotto scalare se e solo se vale la legge del parallelogramma*

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2). \tag{2}$$

Inoltre, su $\mathbb{R}^n$ il prodotto scalare che induce la norma $\|\cdot\|$ è dato da

$$\langle x, y\rangle = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2),$$

mentre su $\mathbb{C}^n$ il prodotto scalare è dato da

$$\langle x, y\rangle = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2 + \mathtt{i}(\|x + \mathtt{i}y\|^2 - \|x - \mathtt{i}y\|^2)).$$

La condizione (2) data nel teorema precedente, che abbiamo chiamato legge del parallelogramma, esprime il fatto che la somma dei quadrati delle diagonali di un parallelogramma è uguale alla somma dei quadrati dei quattro lati. Si può verificare che la norma infinito non soddisfa la legge del parallelogramma. Ad esempio, in $\mathbb{R}^2$, scegliendo $x = (1, 0)$ e $y = (0, 1)$ vale $\|x+y\|_\infty^2 = 1$, $\|x-y\|_\infty^2 = 1$, mentre $2(\|x\|_\infty^2 + \|y\|_\infty^2) = 4$. Gli stessi vettori forniscono un controesempio per dimostrare che anche la norma 1 non è indotta da un prodotto scalare.

La diseguaglianza triangolare si può esprimere in modo equivalente nella seguente forma

$$|\|x\| - \|y\|| \le \|x - y\|, \quad x, y \in \mathbb{C}^n. \tag{3}$$

Infatti, consideriamo $a, b \in \mathbb{C}^n$. Allora dalla diseguaglianza triangolare applicata a a e b, si ha $\|a + b\| \le \|a\| + \|b\|$ da cui $\|a + b\| - \|a\| \le \|b\|$. Quindi ponendo $x = a + b$ e $y = a$ si ha $\|x\| - \|y\| \le \|x - y\|$. Scambiando x con y si deduce che $\|y\| - \|x\| \le \|y - x\| = \|x - y\|$ da cui segue la diseguaglianza (3).

Una conseguenza immediata della diseguaglianza triangolare espressa nella forma (3) è che ogni norma è una funzione *uniformemente continua* cioè vale

$$\forall\ \epsilon > 0\ \exists\ \delta > 0 :\ \forall x, y \in \mathbb{C}^n\ |x_i - y_i| \le \delta\ \Rightarrow\ |\,\|x\| - \|y\|\,| \le \epsilon.$$

L'uniformità della continuità sta nel fatto che il valore di δ non dipende dalla scelta di x e y in $\mathbb{C}^n$. Questa utile proprietà si dimostra nel modo seguente

Dim. Si consideri la base canonica di $\mathbb{C}^n$ formata dai vettori $e^{(j)} = (e_i^{(j)})$, $j = 1, \ldots, n$ tali che $e_i^{(j)} = \delta_{i,j}$, dove $\delta_{i,j}$ è il delta di Kronecker. In questo modo $x = \sum_{i=1}^n x_i e^{(i)}$ e $y = \sum_{i=1}^n y_i e^{(i)}$. Vale allora

$$\|x - y\| = \|\sum_{i=1}^n (x_i - y_i) e^{(i)}\| \le \sum_{i=1}^n |x_i - y_i| \, \|e^{(i)}\|,$$

da cui, se $|x_i - y_i| \le \delta$ ne segue che

$$\|x - y\| \le \delta\gamma, \quad \gamma = \sum_{i=1}^n \|e^{(i)}\|,$$

dove γ è indipendente da x e da y. Quindi dalla (3) si deduce che

$$| \, \|x\| - \|y\| \, | \le \delta\gamma.$$

Allora per completare la dimostrazione basta scegliere $\delta < \epsilon/\gamma$. $\square$

La continuità delle norme vettoriali implica la seguente proprietà di equivalenza delle norme su $\mathbb{C}^n$.

Teorema 2 (Proprietà di equivalenza delle norme) *Per ogni coppia di norme $\| \cdot \|'$ e $\| \cdot \|''$ su $\mathbb{C}^n$ esistono costanti positive α, β tali che per ogni $x \in \mathbb{C}^n$ vale*

$$\alpha \|x\|' \le \|x\|'' \le \beta \|x\|'.$$

Dim. Se $x = 0$ la proprietà è verificata. Supponiamo allora $x \neq 0$. Possiamo scegliere una delle due norme a piacere ad esempio la norma infinito. Infatti se dimostriamo la doppia diseguaglianza per la coppia $\| \cdot \|', \| \cdot \|_\infty$ e per la coppia $\| \cdot \|'', \| \cdot \|_\infty$, componendo le due doppie diseguaglianze otteniamo le costanti che legano $\| \cdot \|'$ e $\| \cdot \|''$. Scegliamo allora $\| \cdot \|' = \| \cdot \|_\infty$ e osserviamo che, per la definizione di $\| \cdot \|_\infty$, l'insieme $S_\infty = \{x \in \mathbb{C}^n : \quad \|x\|_\infty = 1\}$ è limitato. Inoltre, essendo S_∞ l'immagine inversa dell'insieme chiuso $\{1\}$ tramite $\| \cdot \|_\infty$ ed essendo la norma infinito continua, ne segue che S_∞ è anch'esso chiuso. Sappiamo allora che essendo S_∞ un sottoinsieme chiuso e limitato di $\mathbb{C}^n$ esso è compatto. Quindi la funzione continua $\| \cdot \|''$ ha massimo e minimo su S_∞. Siano β e α i valori di questo massimo e minimo. Allora, poiché vale $y = \frac{1}{\|x\|_\infty} x \in S_\infty$, ne segue che

$$\alpha \le \|y\|'' \le \beta$$

da cui

$$\alpha \|x\|_\infty \le \|x\|'' \le \beta \|x\|_\infty.$$

$\square$

Se $x = (x_i) \in \mathbb{C}^n$, definiamo $y = (y_i)$ dove $y_i = |x_i|$. Se $\| \cdot \|$ è una delle tre norme 1,2 e infinito vale $\|x\| = \|y\|$. In generale questo non è vero. Si provi a trovare degli esempi di questo fatto. Una norma che verifica questa proprietà è detta *norma assoluta*.

Si possono agevolmente determinare le costanti α e β che legano le norme 1,2 e infinito. Vale infatti il seguente

Teorema 3 *Per ogni $x \in \mathbb{C}^n$ si ha*

$$\begin{array}{l} \|x\|_\infty \le \|x\|_1 \le n\|x\|_\infty \\ \|x\|_2 \le \|x\|_1 \le \sqrt{n}\|x\|_2 \\ \|x\|_\infty \le \|x\|_2 \le \sqrt{n}\|x\|_\infty \end{array} \tag{4}$$

Dim. Poiché $|x_i| \le \|x\|_\infty$ si ha $\|x\|_1 = \sum_{i=1}^n |x_i| \le n\|x\|_\infty$ e $\|x\|_1 \ge \|x\|_\infty$. Da cui segue la prima coppia di diseguaglianze. Dall'identità $\left(\sum_{i=1}^n |x_i|\right)^2 = \sum_{i=1}^n |x_i|^2 + 2\sum_{i>j} |x_i x_j|$ segue $\|x\|_1^2 \ge \|x\|_2^2$ e quindi $\|x\|_1 \ge \|x\|_2$. Per dimostrare la diseguaglianza $\|x\|_1 \le \sqrt{n}\|x\|_2$ si deve ricorrere alla diseguaglianza di Cauchy-Schwarz soddisfatta dal prodotto scalare hermitiano. Infatti scegliendo $y = (y_i)$ con $|y_i| = 1$ e $y_i \bar{x}_i = |x_i|$, dalla diseguaglianza di Cauchy-Schwarz (1) si ottiene

$$(\sum_{i=1}^n |x_i|)^2 = |x^H y|^2 \le \langle x, x\rangle\langle y, y\rangle = n\sum_{i=1}^n |x_i|^2$$

da cui $\|x\|_1 \le \sqrt{n}\|x\|_2$. Per quanto riguarda la terza coppia di diseguaglianze si osserva che la seconda si ottiene dalla formula $\|x\|_2^2 = \sum_{i=1}^n |x_i|^2$ maggiorando ciascun $|x_i|$ con $\|x\|_\infty$. Per la prima diseguaglianza basta osservare che $\sum_{i=1}^n |x_i|^2 \ge |x_j|^2$ per ogni j. $\square$

Una proprietà interessante della norma 2 è che è invariante sotto trasformazioni unitarie. Infatti se $y = Ux$ ed U è unitaria, cioè $U^H U = I$, allora

$$\|y\|_2^2 = y^H y = (Ux)^H (Ux) = x^H U^H U x = x^H x = \|x\|_2^2.$$

1.2 Alcuni Esercizi

1. Se $x \to \|x\|$ è norma allora $x \to \alpha\|x\|$ è norma per ogni $\alpha > 0$.

2. Se $x \to \|x\|$ è norma e S è matrice invertibile allora $x \to \|Sx\|$ è norma.

3. Se S è matrice $m \times n$ di rango massimo con $m \ge n$, e se $\|\cdot\|$ è una norma su $\mathbb{R}^m$, è vero che $x \to \|Sx\|$ è norma su $\mathbb{R}^n$?

4. Se $\|x\|'$ e $\|x\|''$ sono norme allora $\alpha\|x\|' + \beta\|x\|''$ è norma per $\alpha, \beta > 0$.

5. Dire se le seguenti applicazioni sono norme

   - $x \to \min_i |x_i|$
   - $x \to \max_i |x_i| + \min_i |x_i|$
   - $(x_1, x_2) \to |x_1 - x_2| + |x_1|$
   - $(x_1, x_2) \to |x_1| + \min\{|x_1|, |x_2|\}$

6. Sia $x \in \mathbb{R}^{2n}$ e si partizioni x in due sottovettori $x^{(1)} = (x_1, \ldots, x_n)$ e $x^{(2)} = (x_{n+1}, \ldots, x_{2n})$. Siano $\|\cdot\|'$ una norma su $\mathbb{R}^n$ e $\|\cdot\|''$ una norma su $\mathbb{R}^2$. dire se l'applicazione $x \to \|(\|x^{(1)}\|', \|x^{(2)}\|')\|''$ è una norma.

7. Sia A matrice hermitiana definita positiva. Dimostrare che $(x, y) \to y^H Ax$ è un prodotto scalare.

8. Dimostrare che una matrice A è unitaria se e solo se per ogni $x \in \mathbb{C}^n$ vale $\|x\|_2 = \|Ax\|_2$.

2 Norme di matrici

L'insieme delle matrici $n \times n$ è uno spazio vettoriale di dimensioni n^2. In altri termini una matrice A $n \times n$ può essere vista come un vettore di n^2 componenti. Per cui in linea teorica potremmo usare le definizioni e le proprietà delle norme di vettori per matrici in generale. Però è conveniente usare una definizione leggermente più forte che impone qualche utile proprietà.

Definizione 2 (Norma di matrice) *Si dice* norma di matrice *una applicazione* $\|\cdot\|$ *da* $\mathbb{C}^{n\times n}$ *in* $\mathbb{R}$ *tale che*

- $\|A\| \geq 0, \qquad \|A\| = 0$ *se e solo se* $A = 0$
- $\|\alpha A\| = |\alpha|\|A\|$ *per* $\alpha \in \mathbb{C}$
- $\|A + B\| \leq \|A\| + \|B\|$
- $\|AB\| \leq \|A\|\,\|B\|$

Le prime tre proprietà sono le analoghe di quelle date nel caso dei vettori. La quarta, più specifica al contesto delle matrici, è detta *proprietà submoltiplicativa.* Un esempio di norma di matrice è la *norma di Frobenius* definita da

$$\|A\|_F = \left(\sum_{i=1}^{n}\sum_{j=1}^{n} |a_{i,j}|^2\right)^{1/2}.$$

Si osserva che la norma di Frobenius non è altro che la norma euclidea applicata al vettore

$$vec(A) = (a_{1,1}, a_{2,1}, \ldots, a_{2,n}, a_{1,2}, \ldots, a_{n,n})^T$$

che si ottiene sovrapponendo una sull'altra le colonne di A. La norma di Frobenius può anche essere scritta come $\|A\|_F = \text{traccia}(A^H A)^{1/2}$ ed è la norma indotta dal prodotto scalare $\langle A, B\rangle = \text{traccia}(A^H B)$.

Tra le norme di matrici giocano un ruolo importante le *norme di matrici indotte* da una norma vettoriale, dette anche *norme operatore.*

Sia $\|\cdot\|$ una norma su $\mathbb{C}^n$ e A una matrice $n \times n$. Poiché la sfera unitaria $S = \{x \in \mathbb{C}^n : \quad \|x\| = 1\}$ è un chiuso e limitato quindi un compatto, ed essendo la norma una funzione continua, esiste il

$$\max_{x\in S} \|Ax\|.$$

È facile dimostrare che l'applicazione che associa ad A questo massimo è una norma che chiamiamo norma di matrice indotta dalla norma vettoriale $\|\cdot\|$ e denotiamo col simbolo $\|A\|$. Poniamo quindi

$$\|A\| := \max_{\|x\|=1} \|Ax\|.$$

La dimostrazione delle quattro proprietà è una semplice verifica e non viene riportata.

Se guardiamo ad una matrice come un operatore lineare tra due spazi vettoriali, la norma indotta di una matrice ci dice qual è il massimo allungamento che l'operatore produce nel trasformare i vettori. L'allungamento viene misurato nella norma vettoriale assegnata. Per questo motivo questa norma viene chiamata anche norma operatore.

Una proprietà che segue direttamente dalla definizione e che è molto utile nelle elaborazioni successive è la seguente

$$\|Ax\| \leq \|A\| \, \|x\|. \tag{5}$$

Un'altra conseguenza interessante della definizione di norma di matrice indotta è $\|I\| = 1$. Questa proprietà non è verificata dalla norma di Frobenius per cui $\|I\| = \sqrt{n}$. Quindi la norma di Frobenius non è una norma indotta.

Ci possiamo chiedere come sono fatte le norme di matrice indotte dalle norme vettoriali 1,2 e infinito. Vale per questo il seguente risultato che ci permette di valutare queste norme in modo agevole in almeno due casi su 3.

Teorema 4 *Per le norme di matrice indotte dalla norma 1,2 e infinito vale*

$$\begin{array}{l} \|A\|_1 = \max_j \sum_{i=1}^n |a_{i,j}| \\ \|A\|_2 = \left(\rho(A^H A)\right)^{1/2} \\ \|A\|_\infty = \max_i \sum_{j=1}^n |a_{i,j}| \end{array} \tag{6}$$

dove $\rho(A)$ denota il raggio spettrale di una matrice A, cioè il massimo dei moduli dei suoi autovalori.

Dim. La metodologia dimostrativa è la stessa nel caso delle tre norme. In un primo passo si dimostra che le relazioni (6) valgono con il segno $\leq$. In un secondo passo si dimostra che esiste un vettore x di norma unitaria per cui $\|Ax\|$ coincide con il membro destro in (6) nei tre casi distinti. Partiamo con la norma 1 e dimostriamo che $\|A\|_1 \leq \max_j \sum_{i=1}^n |a_{i,j}|$. Dato $x \in \mathbb{C}^n$ tale che $\|x\|_1 = 1$ vale

$$\|Ax\|_1 = \sum_{i=1}^n \left| \sum_{j=1}^n a_{i,j} x_j \right| \leq \sum_{i=1}^n \sum_{j=1}^n |a_{i,j}| \, |x_j| = \sum_{j=1}^n |x_j| \sum_{i=1}^n |a_{i,j}|.$$

Maggiorando $\sum_{i=1}^n |a_{i,j}|$ con $\max_j \sum_{i=1}^n |a_{i,j}|$ si ottiene

$$\|Ax\|_1 \leq \max_j \sum_{i=1}^n |a_{i,j}| \sum_{j=1}^n |x_j| = \max_j \sum_{i=1}^n |a_{i,j}|.$$

Inoltre se $\max_j \sum_{i=1}^n |a_{i,j}|$ è preso sulla colonna k-esima, il vettore $x = e^{(k)}$ che ha componenti nulle tranne la k-esima che vale 1, è tale che $\|x\|_1 = 1$ e

$$\|Ax\|_1 = \sum_{i=1}^n |a_{i,k}| = \max_j \sum_{i=1}^n |a_{i,j}|.$$

La dimostrazione procede in modo analogo con la norma infinito. Infatti, se x è un vettore tale che $\|x\|_\infty = 1$, vale

$$\|Ax\|_\infty = \max_i \left| \sum_{j=1}^n a_{i,j} x_j \right| \le \max_i \sum_{j=1}^n |a_{i,j}|\,|x_j|.$$

Maggiorando $|x_j|$ con 1, si ottiene

$$\|Ax\|_\infty \le \max_i \sum_{j=1}^n |a_{i,j}|.$$

Supponiamo che il massimo sia raggiunto sull'indice k. Allora scegliendo il vettore x in modo che $x_j = \bar{a}_{k,j}/|a_{k,j}|$ se $a_{k,j} \neq 0$ e $x_j = 1$ altrimenti, risulta $\|x\|_\infty = 1$ e la k-esima componente di Ax è uguale a $\sum_{j=1}^n |a_{k,j}| = \max_i \sum_{j=1}^n |a_{i,j}|$.

Per la norma 2 si procede in modo analogo. Se x è tale che $\|x\|_2 = 1$ allora $\|Ax\|_2^2 = x^H A^H A x$, e, poiché $A^H A$ è hermitiana, esiste una matrice unitaria U tale che $U^H A^H A U = D$, con D matrice diagonale e $d_{i,i} \ge 0$. Risulta allora, con $y = U^H x$,

$$\|Ax\|_2^2 = x^H (UDU^H) x = y^H D y = \sum_{i=1}^n |y_i|^2 \lambda_i \le \rho(A^H A) \sum_{i=1}^n |y_i|^2 = \rho(A^H A),$$

dove l'uguaglianza è raggiunta quando x è autovettore di $A^H A$ corrispondente a $\rho(A^H A)$. $\square$

Si può notare come la norma 1 e la norma infinito siano facilmente calcolabili. Mentre la norma 2, richiedendo il calcolo degli autovalori di una matrice è calcolabile con maggiori difficoltà computazionali.

Si è già osservato che dalla definizione di norma 2 e di norma di Frobenius segue $\|A\|_2 \le \|A\|_F$. Infatti $\|A\|_F^2 = \text{traccia}(A^H A)$ è la somma degli autovalori della matrice semidefinita positiva $A^H A$, mentre $\|A\|_2^2 = \rho(A^H A)$ è il massimo degli autovalori di $A^H A$.

Dalle relazioni (4) si possono ricavare agevolmente le costanti che legano attraverso diseguaglianze le norme di matrice indotte dalla norma 1,2 e infinito.

È interessante osservare che se U e V sono matrici unitarie e $B = UAV$ allora

$$B^H B = (UAV)^H (UAV) = V^H A^H U^H U A V = V^H A^H A V$$

cioè $A^H A$ e $B^H B$ sono (unitariamente) simili ed hanno quindi gli stessi autovalori per cui $\rho(A^H A) = \rho(B^H B)$ e $\text{traccia}(A^H A) = \text{traccia}(B^H B)$.

Una proprietà che lega la norma 2 e la norma di Frobenius è che $\|A\|_2 \leq \|A\|_F$. Ciò vale poiché

$$\|A\|_2^2 = \rho(A^H A) \leq \text{traccia}(A^H A) = \|A\|_F^2,$$

infatti la traccia è la somma degli autovalori, che per $A^H A$ sono tutti maggiori o uguali a zero essendo $A^H A$ semidefinita positiva. Questo ci permette di dimostrare agevolmente la proprietà submoltiplicativa della norma di Frobenius. Infatti vale

$$\|Ax\|_2 \leq \|A\|_2 \|x\|_2 \leq \|A\|_F \|x\|_2$$

Questo implica che

$$\|AB\|_F^2 = \sum_{j=1}^n \|ABe^{(j)}\|_2^2 \leq \sum_{j=1}^n \|A\|_2^2 \|Be^{(j)}\|_2^2 \leq \|A\|_F^2 \sum_{j=1}^n \|Be^{(j)}\|_2^2 = \|A\|_F^2 \|B\|_F^2.$$

2.1 Norme e raggio spettrale

Ci sono delle relazioni molto strette tra le norme di matrici indotte e il raggio spettrale. La prima osservazione interessante è che se λ è autovalore di A, cioè $Ax = \lambda x$ per $x \neq 0$, allora per la (5)

$$\|Ax\| \leq \|A\| \, \|x\|$$

da cui $|\lambda| \, \|x\| \leq \|A\| \, \|x\|$. Ne segue che $\|A\| \geq |\lambda|$ per ogni autovalore λ per cui

$$\|A\| \geq \rho(A). \tag{7}$$

Ci possiamo chiedere allora se $\rho(A)$ può essere una norma per ogni matrice A. Per questo vale il seguente risultato

Teorema 5 *Per ogni matrice A e per ogni $\epsilon > 0$ esiste una norma di matrice indotta $\| \cdot \|$ tale che*

$$\rho(A) \leq \|A\| \leq \rho(A) + \epsilon.$$

Inoltre, se gli autovalori di A di modulo uguale al raggio spettrale stanno in blocchi di Jordan di dimensione 1, allora esiste una norma indotta che applicata ad A coincide con $\rho(A)$.

Dim. Se S è una matrice invertibile allora si verifica facilmente che, data una norma $\| \cdot \|$ vettoriale, l'applicazione $x \to \|Sx\|$ è una norma. Definiamo allora $\|x\|_S := \|Sx\|$. Verifichiamo che la norma di matrice indotta dalla norma $\| \cdot \|_S$ è $\|A\|_S = \|SAS^{-1}\|$. Infatti

$$\|A\|_S := \max_{\|x\|_S=1} \|Ax\|_S = \max_{\|Sx\|=1} \|SAx\| = \max_{\|y\|=1} \|SAS^{-1}y\| = \|SAS^{-1}\|,$$

con $y = Sx$. Ora portiamo A in forma di Jordan, $J = W^{-1}AW$, e fissato $\epsilon > 0$, consideriamo la matrice diagonale D_ϵ i cui elementi diagonali sono $1, \epsilon, \epsilon^2, \ldots, \epsilon^{n-1}$. Osserviamo che la matrice $\widehat{J} = D_\epsilon^{-1} J D_\epsilon$ è diagonale a blocchi con blocchi che differiscono dai blocchi di Jordan per il fatto che gli elementi sopra diagonali sono uguali a ϵ. Per cui la norma infinito di $\widehat{J}$ è data da $\rho(A)+\epsilon$ se esiste un blocco di Jordan di dimensione maggiore di 1 con autovalore di modulo uguale al raggio spettrale. Mentre, se tutti gli autovalori di modulo uguale al raggio spettrale stanno in blocchi di Jordan di dimensione 1 e se ϵ è scelto abbastanza piccolo in modo che $|\lambda| + \epsilon \leq \rho(A)$ per ogni altro autovalore λ, allora $\|\widehat{J}\|_\infty = \rho(A)$. Ne segue che il teorema vale con $\|A\| = \|A\|_S$ dove $S = D_\epsilon^{-1} W^{-1}$. Inoltre $\|A\| = \rho(A)$ se tutti gli autovalori di modulo $\rho(A)$ stanno in blocchi di dimensione 1. □

Un altro risultato che lega il raggio spettrale con le norme di matrici è dato dal seguente.

Teorema 6 *Per ogni norma di matrice $\|\cdot\|$ e per ogni matrice A vale*

$$\lim_{k\to\infty} \|A^k\|^{\frac{1}{k}} = \rho(A)$$

Dim. La dimostrazione consiste di due parti. Una prima parte in cui si fa vedere che se il limite esiste allora non dipende dalla norma. Una seconda parte in cui si sceglie una norma speciale per cui il limite è proprio $\rho(A)$. Per la prima parte supponiamo di avere due norme $\|\cdot\|$ e $\|\cdot\|'$ e di sapere che esiste il $\lim_{k\to\infty} \|A^k\|^{1/k} = \ell$. Per il teorema di equivalenza delle norme si ha che esistono costanti α e β per cui vale la relazione

$$\alpha\|X\| \leq \|X\|' \leq \beta\|X\|$$

qualunque sia la matrice X. Scegliendo allora $X = A^k$ si ha

$$\alpha\|A^k\| \leq \|A^k\|' \leq \beta\|A^k\|$$

e, prendendo la radice k-esima delle tre espressioni si ottiene

$$\alpha^{1/k}\|A^k\|^{1/k} \leq (\|A^k\|')^{1/k} \leq \beta^{1/k}\|A^k\|^{1/k}$$

Prendendo il limite per $k \to \infty$, $\alpha^{1/k}$ e $\beta^{1/k}$ convergono ad 1 per cui la parte di sinistra e la parte di destra convergono allo stesso limite ℓ. Per il teorema dei carabinieri la parte centrale converge anch'essa ad ℓ.

Rimane ora da dimostrare che per una particolare norma vale $\ell = \rho(A)$. Per questo ricorriamo alla forma di Jordan $J = S^{-1}AS$ di A e, come abbiamo già fatto, consideriamo la norma infinito di J come norma particolare di A. Denotiamo quindi $\|A\| = \|J\|_\infty$. Valutiamo quindi $\|A^k\| = \|J^k\|_\infty$. Poiché la matrice J^k è diagonale a blocchi con le potenze k-esime dei blocchi di Jordan di A sulla diagonale principale, la norma infinito di J^k è la massima norma infinito delle potenze k-esime dei blocchi di Jordan di A. Sia $\widehat{J}$ un generico blocco di Jordan corrispondente all'autovalore λ. Se il blocco ha dimensione 1 la sua norma è $\|\widehat{J}^k\|_\infty = |\lambda|^k$, se il blocco ha dimensione m maggiore di 1 allora vale:

$$J^k = \begin{bmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \dots & \binom{k}{m-1}\lambda^{k-m+1} \\ & \lambda^k & \ddots & \vdots \\ & & \ddots & \binom{k}{1}\lambda^{k-1} \\ & & & \lambda^k \end{bmatrix}$$

Ciò si vede applicando la formula del binomio di Newton al blocco di Jordan $\widehat{J}$ scritto nella forma $\widehat{J} = \lambda I + H$ dove $H = (h_{i,j})$ è la matrice che ha tutti elementi nulli tranne $h_{i,i+1} = 1$ per $i = 1, \dots, m-1$, essendo $H^k = 0$ per $k \geq m$. Quindi, se $\lambda = 0$ vale $J^k = 0$ per $k \geq m$. Se $\lambda \neq 0$ allora prendendo la norma infinito di $\widehat{J}^k$ si ottiene

$$\|\widehat{J}^k\| = \sum_{i=0}^{m-1} |\lambda|^{k-i} \binom{k}{i} = |\lambda|^k \sum_{i=0}^{m-1} |\lambda|^{-i} \binom{k}{i}$$

Ora, prendendo la radice k-esima dell'espressione precedente e prendendo il limite per $k \to \infty$ si ottiene

$$\lim_{k\to\infty} \left(\|\widehat{J}^k\|_\infty\right)^{1/k} = |\lambda| \lim_{k\to\infty} (\sum_{i=0}^{m-1} |\lambda|^{-i} \binom{k}{i})^{1/k} = \lambda$$

infatti la parte sotto radice k-esima, come funzione di k, è un polinomio di grado $m-1$ per cui $\lim_{k\to\infty}(\sum_{i=0}^{m-1} |\lambda|^{-i}\binom{k}{i})^{1/k} = 1$. □

Se A è tale che $\|A\| < 1$ per qualche norma di matrice indotta, allora, per la (7), A ha autovalori in modulo più piccoli di 1 e conseguentemente $I - A$ è invertibile. Una diseguaglianza utile che fornisce una stima di $\|(I-A)^{-1}\|$ è la seguente:

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

Infatti, da $(I-A)^{-1}(I-A) = I$ si ricava che

$$(I-A)^{-1} = (I-A)^{-1}A + I.$$

Prendendo le norme e usando la diseguaglianza triangolare si ottiene

$$\|(I-A)^{-1}\| \leq 1 + \|A\| \, \|(I-A)^{-1}\|$$

da cui la tesi.

Si osservi che sostituendo $-A$ ad A si ottiene $\|(I+A)^{-1}\| \leq 1/(1 - \|A\|)$.

3 Numero di condizionamento

Lo studio del condizionamento di un sistema lineare $Ax = b$ ha come obiettivo capire in quali condizioni una piccola perturbazione introdotta nel vettore dei termini noti o negli elementi della matrice A, si ripercuota più o meno amplificato nella soluzione. Uno strumento per fare questo sono i coefficienti di amplificazione. Ricordiamo che nel calcolo di una funzione $f(t_1, \ldots, t_n)$ di n variabili il coefficiente di amplificazione rispetto a t_i è dato da $t_i \frac{\partial f}{\partial t_i}/f$. Nel nostro caso le funzioni di cui calcolare i coefficienti di amplificazione sono $x_i(b_1, \ldots, b_n, a_{1,1}, \ldots, a_{n,n})$ per $i = 1, \ldots, n$ di $n^2 + n$ variabili dove $x = A^{-1}b$. usare questo strumento nel caso dei sistemi lineari porta ad una complicazione formale molto pesante. È allora più conveniente, anziché valutare le perturbazioni componente a componente, dare una valutazione globale usando le norme. Anche se le stime che otteniamo in questo modo sono meno precise, esse hanno il vantaggio di essere di uso più facile e immediato.

Siano $\delta_b \in \mathbb{C}^n$ e $\delta_A \in \mathbb{C}^{n\times n}$ perturbazioni che introduciamo rispettivamente in b e in A.

Assumendo $b \neq 0$ e $A \neq 0$, definiamo ora

$$\epsilon_b = \|\delta_b\|/\|b\|, \qquad \epsilon_A = \|\delta_A\|/\|A\|$$

le perturbazioni relative espresse in norma. Denotiamo con $x + \delta_x$ la soluzione del sistema perturbato, cioè tale che

$$(A + \delta_A)(x + \delta_x) = b + \delta_b$$

e definiamo $\epsilon_x = \|\delta_x\|/\|x\|$ la variazione relativa nel risultato conseguente alle due perturbazioni introdotte. Si osserva che se $\delta_A = 0$, dalle relazioni $Ax = b$ e $A(x + \delta_x) = b + \delta_b$ si ricava

$$A\delta_x = \delta_b$$

e, assumendo $\det A \neq 0$, si ricava $\delta_x = A^{-1}\delta_b$ da cui $\|\delta_x\| \leq \|A^{-1}\|\,\|\delta_b\|$. D'altra parte, poiché $Ax = b$ ne segue $\|b\| \leq \|A\|\,\|x\|$ da cui

$$\frac{\|\delta_x\|}{\|x\|} \leq \|A\|\,\|A^{-1}\|\,\frac{\|\delta_b\|}{\|b\|}.$$

Cioè la perturbazione relativa in norma ϵ_b introdotta nel termine noto ha causato una variazione relativa in norma ϵ_x nella soluzione limitata da $\epsilon_x \leq \|A\|\,\|A^{-1}\|\epsilon_b$.

Il numero $\mu(A) = \|A\|\,\|A^{-1}\|$ è detto *numero di condizionamento* di A ed esprime la massima amplificazione che può subire l'errore introdotto nel termine noto.

Una maggiorazione analoga anche se un po' più complessa vale nel caso in cui $\delta_A \neq 0$.

Teorema 7 *Se* $\det A \neq 0$ *e* $\|A^{-1}\|\,\|\delta_A\| < 1$ *allora* $\det(A + \delta_A) \neq 0$*, inoltre*

$$\epsilon_x \leq \frac{\|A\|\,\|A^{-1}\|}{1 - \epsilon_A\|A\|\,\|A^{-1}\|}(\epsilon_b + \epsilon_A).$$

Dim. Dalle relazioni $Ax = b$ e $(A+\delta_A)(x+\delta_x) = b+\delta_b$, sottraendo membro a membro si ottiene $(A + \delta_A)\delta_x = -\delta_A x + \delta_b$. Poiché $A + \delta_A = A(I + A^{-1}\delta_A)$ e $\|A^{-1}\delta_A\| \leq \|A^{-1}\|\,\|\delta_A\| < 1$, la matrice $I + A^{-1}\delta_A$ risulta invertibile e vale

$$\delta_x = (I + A^{-1}\delta_A)^{-1}A^{-1}(-\delta_A x + \delta_b).$$

Poiché $\|(I + A^{-1}\delta_A)^{-1}\| \le 1/(1 - \|A^{-1}\delta_A\|) \le 1/(1 - \|A^{-1}\|\,\|\delta_A\|)$, risulta

$$\|\delta_x\| \le \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\,\|\delta_A\|}(\|\delta_A\|\,\|x\| + \|\delta_b\|)$$

che assieme alla relazione $\|b\| \le \|A\|\,\|x\|$ conduce al risultato $\square$

Si osservi che in questo caso il coefficiente di amplificazione ha una forma più complessa. Però nel caso in cui ϵ_A è molto più piccolo di $\|A\|\,\|A^{-1}\|$, allora il coefficiente di amplificazione è ben approssimato dal numero di condizionamento $\mu(A)$ di A.

Se la matrice A è hermitiana e definita positiva, ordinando i suoi autovalori come

$$0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n,$$

risulta $\|A\|_2 = \lambda_n$, $\|A^{-1}\|_2 = \lambda_1^{-1}$. Per cui

$$\mu_2(A) = \|A\|_2\|A^{-1}\|_2 = \frac{\lambda_n}{\lambda_1}.$$

Cioè il numero di condizionamento in norma 2 è dato dal rapporto tra il massimo e il minimo autovalore si A.

Se A è hermitiana ma non è definita positiva, ordinando i suoi autovalori in modo che $|\lambda_i| \le |\lambda_{i+1}|$ per $i = 1, \ldots, n-1$, vale $\|A\|_2\|A^{-1}\|_2 = \frac{|\lambda_n|}{|\lambda_1|}$, cioè il rapporto tra l'autovalore di massimo modulo e quello di minimo modulo di A.

Se A è normale allora dal fatto che la sua forma di Schur $D = Q^H A Q$ è diagonale segue che il numero condizionamento di A in norma 2 è ancora dato dal rapporto tra l'autovalore di massimo modulo e quello di minimo modulo.

Se A non è normale la situazione è più complicata. Prendiamo ad esempio la matrice

$$A = \begin{bmatrix} 1 & -a & & & \\ & 1 & -a & & \\ & & \ddots & \ddots & \\ & & & 1 & -a \\ & & & & 1 \end{bmatrix}$$

e valutiamo il numero di condizionamento in norma infinito. Vale

$$A^{-1} = \begin{bmatrix} 1 & a & a^2 & \ldots & a^{n-2} & a^{n-1} \\ & 1 & a & a^2 & \ddots & a^{n-2} \\ & & \ddots & \ddots & \ddots & \vdots \\ & & & 1 & a & a^2 \\ & & & & 1 & a \\ & & & & & 1 \end{bmatrix}$$

da cui $\|A\|_\infty = 1 + |a|$, mentre, assumendo per semplicità $|a| \ne 1$ si ha $\|A^{-1}\|_\infty = \sum_{i=0}^{n-1} |a|^i = \frac{|a|^n - 1}{|a| - 1}$. Si osserva allora che, se $|a| > 1$ il numero di condizionamento cresce esponenzialmente con n su base $|a|$. Ad esempio, se $a = 10$ e $n = 100$ il numero di condizionamento è dell'ordine di 10^{100}. Questa è la situazione tipica di un blocco di Jordan relativo ad un autovalore λ tale che $|\lambda| < 1$.

4 Esercizi

1. Sia $S \in \mathbb{R}^{m\times n}$, con $m \geq n$, e sia $\|\cdot\|$ una norma su $\mathbb{R}^m$. Si dica sotto quali ipotesi su S l'applicazione $x \longrightarrow \|Sx\|$ è una norma su $\mathbb{R}^n$.

2. Si dica, motivando la risposta, quali delle seguenti funzioni è una norma su $\mathbb{R}^n$:

   (a) $x \longrightarrow |x_1| + |x_n|$

   (b) $x \longrightarrow \max_i |x_i| + \min_i |x_i|$

   (c) $x \longrightarrow |x_1| + \max_i |x_i|$

   (d) su $\mathbb{R}^2$, $x \longrightarrow |x_1 - x_2| + |x_1|$

3. Dire se $\frac{1}{\sqrt{n}}\|\cdot\|_F$ è norma indotta.

4. Dimostrare che l'applicazione $A \to \max_{i,j} |a_{i,j}|$ soddisfa ai primi tre assiomi delle norme ma non è submultiplicativa.

5. Dimostrare che $\|A\|_F \leq \sqrt{r}\|A\|_2$ dove r è il rango di A.

6. Dimostrare che $\|A\|_2 \leq \sqrt{\|A\|_1\|A\|_\infty}$.

7. Dimostrare che per una matrice unitaria il numero di condizionamento in norma 2 è 1.

8. Stimare il numero di condizionamento in norma 2 e in norma infinito di un blocco di Jordan $n \times n$ di autovalore λ.

9. Stimare il numero di condizionamento in norma 2 e in norma infinito di una matrice elementare, cioè del tipo $I - uv^T$ dove $u, v \in \mathbb{C}^n$. (Si veda l'articolo sulle matrici elementari)

10. Se A è matrice triangolare non singolare si dimostri che $\|A\|_2\|A^{-1}\|_2 \geq \max_i |a_{i,i}|/\min_i |a_{i,i}|$.

11. Si dimostri che una matrice non singolare A ha numero di condizionamento 1 in norma 2 se e solo se $A^H A = \alpha I$.

12. Si determinino costanti α e β tali che $\alpha\mu_\infty(A) \leq \mu_2(A) \leq \beta\mu_\infty(A)$.

13. Si dimostri che se A è hermitiana allora $\mu_2(A) \leq \mu(A)$, dove $\mu(A)$ è il numero di condizionamento rispetto a una qualsiasi norma indotta.

14. Si dica qual è il massimo numero di condizionamento in norma infinito di una matrice triangolare inferiore con elementi diagonali uguali a 1 ed elementi non diagonali di modulo minore o uguale a 1.

15. Se x è la soluzione del sistema $Ax = b$ e y è una sua approssimazione, posto $r = Ay - b$ il *residuo* di y, si dimostri che vale

$$\frac{\|x-y\|}{\|x\|} \leq \mu(A)\frac{\|r\|}{\|b\|}.$$

16. Siano $\|\cdot\|'$ e $\|\cdot\|''$ due norme su $\mathbb{R}^n$. Si consideri l'applicazione che alla matrice $n \times n$ reale A associa il numero reale $f(A) = \max_{\|x\|'=1} \|Ax\|''$. Dire quali proprietà della norma di matrici soddisfa questa applicazione. Si dimostri che scegliendo $\|\cdot\|' = \|\cdot\|_1$ e $\|\cdot\|'' = \|\cdot\|_\infty$ si ha $f(A) = \max_{i,j} |a_{i,j}|$.

17. Sia $A = \begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix}$. Si dica se esiste una norma matriciale indotta $\|\cdot\|$ tale che:

    (a) $\|A\| = 4$

    (b) $\|A\| = 2 + 10^{-20}$

    (c) $\|A\| = 2$.

Riferimenti bibliografici

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