CAPITOLO I

SPAZI VETTORIALI NORMATI

In questo capitolo i concetti fondamentali dell'algebra lineare vengono ripresi e riformulati con una notazione più funzionale agli scopi del calcolo numerico. Vengono inoltre ripresi ed ampliati i concetti relativi alle norme dei vettori e delle matrici in vista della loro utilizzazione nell'analisi dell'errore di ogni procedimento numerico atto ad approssimare la soluzione dei vari problemi che verranno proposti nel corso.

1-NOZIONI PRELIMINARI.

Consideriamo lo spazio vettoriale R^n, (C^n) sul corpo R (C), cioè le n-uple ordinate di numeri reali (complessi) $(x_1, x_2, ..., x_n)$ con l'operazione di somma:

$$(x_1, x_2, ..., x_n) + (y_1, y_2, ..., y_n) = (x_1+y_1, x_2+y_2, ..., x_n+y_n)$$

e l'operazione esterna

$$\lambda(x_1, x_2, ..., x_n) = (\lambda x_1, \lambda x_2, ..., \lambda x_n) \qquad \lambda \in R \ (C).$$

Sia $\mathcal{L}(R^n, R^n)$ l'insieme delle applicazioni lineari di R^n in se. Esso è a sua volta spazio vettoriale su R ed ogni applicazione $A \in \mathcal{L}(R^n, R^n)$ è rappresentata da una matrice n x n con coefficienti a_{ij}, che indicheremo ancora con $A=(a_{ij})$. Analogamente, l'insieme $\mathcal{L}(C^n, C^n)$ è lo spazio delle matrici a coefficienti complessi. La somma ed il prodotto esterno rispetto al quale le matrici formano spazio vettoriale sono $A+B=(a_{ij}+b_{ij})$ ed $\alpha A=(\alpha a_{ij})$. Per le matrici si definisce anche una moltiplicazione AB che corrisponde alla applicazione composta $A \circ B$. Tale operazione non è commutativa.

In C^n si può definire una forma lineare, detta prodotto scalare,

$$<x,y>:=\bar{x}_1y_1+\bar{x}_2y_2+\cdots+\bar{x}_ny_n$$

che gode delle proprietà:

1- $<x,x> \geq 0$, $<x,x>=0 \Leftrightarrow x=0$

2- $<x,y> = \overline{<y,x>}$,

3- $<\alpha x,y> = \overline{\alpha}, <x,y>$ $\alpha \in C$

4- $<x+y,z>=<x,z>+<y,z>$

Se ora interpretiamo i vettori come "*vettori colonna*", cioè come matrici *n*x*1* appartenenti a $\mathcal{L}(C^1, C^n)$, si vede che $<x,y>$ altro non è che il prodotto matriciale di x^H per y ,dove

$$x^H := (\overline{x}_1, \overline{x}_2, ..., \overline{x}_n)$$

è detto "*trasposto coniugato*" di x. Naturalmente, con questa convenzione, il generico vettore x, che abbiamo deciso di interpretare come vettore colonna, andrà rappresentato con

$$x=(x_1, x_2, ..., x_n)^T.$$

Nella precedente notazione matriciale, anche il prodotto scalare puo essere rappresentato in maniera più compatta nella seguente forma:

$\forall x,y \in R^n$ $<x,y> = x^T y$

$\forall x,y \in C^n$ $<x,y> = x^H y$.

Di conseguenza, le proprietà del prodotto scalare possono essere riscritte nel seguente modo:

1- $x^H x \geq 0$ $x^H x = 0 \Leftrightarrow x \equiv 0$

2- $x^H y = \overline{y^H x}, = (y^H x)^H$

3- $(\alpha x)^H y = \overline{\alpha}\, x^H y$

4- $(x+y)^H z = x^H z + y^H z$

Ricordiamo ora alcune definizioni e notazioni:

Due vettori x ed y si dicono **ortogonali** ($x \perp y$) se $x^H y = 0$

Sia $A \in \mathcal{L}(R^n, R^n)$, $A = (a_{ij})$, $a_{i\,j} \in R$

La matrice $A^T = (a_{j\,i})$ è detta matrice **trasposta** di A;

A è detta **simmetrica** se $A = A^T$;

A è detta **ortogonale** se $A A^T = A^T A = I$, cioè se le sue colonne e le sue righe sono a due a due ortogonali in R^n.

La matrice A è detta **regolare**, o **non singolare**, se esiste una matrice, detta **inversa** di A ed indicata con A^{-1}, tale che $A A^{-1} = A^{-1} A = I$

A è detta **involutoria** se $A = A^{-1}$ e quindi $A^2 = I$.

Sia ora $A \in \mathcal{L}(C^n, C^n)$, $A = (a_{ij})$, $a_{ij} \in C$

La matrice $A^H = (\overline{a_{ji}})$ è detta matrice **trasposta coniugata** di A;

A è detta **hermitiana** se $A = A^H$;

A è detta **unitaria** se $A A^H = A^H A = I$, cioè se le sue colonne e le sue righe sono a due a due ortogonali in C^n.

Le nozioni di matrice regolare ed involutoria rimangono le stesse del caso reale.

Salvo esplicita dichiarazione contraria, d'ora in poi considereremo sempre matrici e vettori complessi e tutte le definizioni e proprietà che dimostreremo varranno in particolare nel caso reale.

La matrice A è detta **semidefinita positiva** se $x^H A x \geq 0 \quad \forall\, x \neq 0$,
ed è detta **definita positiva** se $x^H A x > 0 \quad \forall\, x \neq 0$.

Definizione. *Il numero* $\lambda \in C$ *è detto* **autovalore** *di* A *se* $\exists$ *un vettore* $x \neq 0$,*detto* **autovettore** *o* **autosoluzione** *associata a* λ, *tale che* $Ax = \lambda x$.

L'autovettore associato è definito a meno di una costante moltiplicativa. Ad autovalori distinti corrispondono autovettori indipendenti, mentre ad un autovalore possono essere associati più autovettori. Il numero di vettori linearmente indipendenti associati ad un autovalore è detto **molteplicità geometrica** dell'autovalore. Gli autovalori sono dati dalle radici della seguente **equazione caratteristica** associata

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

Il termine $\det(A-\lambda I)$ è un polinomio di grado *n* a coefficienti reali o complessi e quindi possiede *n* radici complesse, il cui prodotto è il termine noto del polinomio, e quindi è il det(A), e la cui somma è il coefficiente del termine di grado *n-1*, e quindi coincide con la somma degli elementi diagonali di A, che è detta **traccia** di A. In breve:

$$\prod \lambda_i = \det(A)$$
$$\sum \lambda_i = tr(A) := \sum a_{ii}$$

Qualcuna delle radici può essere multipla e la molteplicita della radice è detta **molteplicità algebrica** dell'autovalore. La molteplicità algebrica è maggiore od uguale alla molteplicità geometrica. L'insieme degli autovalori è detto brevemente **spettro** di A ed è indicato con S(A), mentre l'insieme degli autovettori è detto **autospazio** di A.

Valgono le seguenti proposizioni le cui dimostrazioni sono immediate.

PROPOSIZIONE 1. $(Ax)^H = x^H A^H$ e quindi, più in generale, $(AB)^H = B^H A^H$

PROPOSIZIONE 2. $(x^H A y)^H = (Ay)^H x = y^H A^H x$

PROPOSIZIONE 3. A hermitiana $\Rightarrow$ $x^H A x$ è reale.

PROPOSIZIONE 4. A def. pos. $\Rightarrow |A| \neq 0$.

Dim. Se A fosse singolare esisterebbe $x \neq 0$ tale che $Ax=0$ ed anche $x^H A x=0$.

PROPOSIZIONE 5. A def. pos. $\Rightarrow a_{ii} > 0$; A semi-def. pos. $\Rightarrow a_{ii} \geq 0 \ \forall i$.

Dim. Si prenda il vettore canonico $e_i = (0,0,...,1,...,0)$ e si osservi che $e_i^H A e_i = a_{ii}$

PROPOSIZIONE 6. A hermitiana $\Rightarrow$ autovalori reali.

Dim. Sia $Ax=\lambda x$. Allora $x^H A x = x^H(\lambda x) = \lambda x^H x$. Passando ai coniugati

$$(x^H A x)^H = (\lambda x^H x)^H$$
$$x^H A^H x = \bar{\lambda} x^H x$$
$$x^H A x = \bar{\lambda} x^H x \quad \text{da cui segue } \bar{\lambda} = \lambda$$

PROPOSIZIONE 7. Per ogni autovalore λ, A def. pos. $\Rightarrow \lambda > 0$ e A semidef. pos. $\Rightarrow \lambda \geq 0$

Dim. Sia $Ax=\lambda x$. Allora $x^HAx=\lambda x^Hx$ da cui segue la tesi

PROPOSIZIONE 8. $A^H=A^{-1} \Rightarrow x^Hy=(Ax)^HAy$

In altre parole: le matrici unitarie conservano il prodotto scalare. Infatti

$(Ax)^HAy=x^HA^HAy=x^HA^{-1}Ay=x^Hy$.

PROPOSIZIONE 9 (Disuguaglianza di Schwartz): $|x^Hy|^2 \le x^Hx \cdot y^Hy$

o equivalentemente:

$$|x^Hy| \le \sqrt{x^Hx}\sqrt{y^Hy}$$

$$\left|\sum_i \bar{x}_i y_i\right| \le \sqrt{\sum_i \bar{x}_i x_i}\sqrt{\sum_i \bar{y}_i y_i}$$

PROPOSIZIONE 10: Una matrice A che goda di 2 delle tre proprietà:

A è hermitiana

A è unitaria

A è involutoria,

gode anche della terza.

TEOREMA 1.1 (DI GERSCHGORIN). *Data la matrice* $A=(a_{ij})$, *ogni autovalore di* A *è contenuto nella riunione dei dischi* $D_i = \{ z\in C : |z-a_{ii}| \le \sum_{i\neq j} |a_{ij}| \}$ $i=1,..n$. *Inoltre se* k *dischi* D_i *sono disgiunti dagli altri, allora* k *autovalori appartengono all'insieme* U, *riunione di quei* k *dischi*.

Dim: Sia λ un autovalore arbitrario di A ed x il corrispondente autovettore. La generica riga k-esima della relazione $Ax=\lambda x$ è:

$$\sum_j a_{kj}x_j = \lambda x_k.$$

Da essa si ha:

$$\sum_{j\neq k} a_{kj}x_j = (\lambda-a_{kk})x_k$$

$$|\lambda-a_{kk}|\,|x_k| \le \sum_{j\neq k} |a_{kj}||x_j|.$$

Sia x_i la componente di modulo massimo di x, cioè tale che $|x_j| \leq |x_i|$ per ogni j. Per l'indice i si ha allora

$$|\lambda - a_{ii}|\,|x_i| \leq \Sigma_{j\neq i}\,|a_{ij}||x_j|$$

$$|\lambda - a_{ii}| \leq \Sigma_{j\neq i}\,|a_{ij}|\left|\frac{x_j}{x_i}\right| \leq \Sigma_{j\neq i}\,|a_{ij}|$$

cioè $\lambda \in D_i$. Non potendo sapere qual'è l'indice i, possiamo concludere che $\lambda \in$######$_i$ D_i. La prima parte del teorema è così dimostrata. Per quanto riguarda la seconda parte, sia:

$$A(t)=D + t(A-D) \qquad\qquad t \in [0,1]$$

e si osservi che:

$$A(0)=D \quad \text{ed} \quad A(1)=A.$$

Mentre t cresce da 0 ad 1 la matrice A(t) passa con continuità dalla matrice D alla matrice A, e così pure gli autovalori $\lambda_i(t)$ delle matrici A(t) percorrono con continuità delle traiettorie da $\lambda_i(0)$ a $\lambda_i(1)$ (la continuità di ciascuna traiettoria è dovuta al fatto che le radici di un polinomio sono funzioni continue dei suoi coefficienti). Così pure i dischi $D_i(t)$ della matrice A(t), che sono centrati negli elementi diagonali a_{ii} per ogni t, hanno raggio crescente al crescere di t in [0,1]. Da un lato, per t=0, essi si riducono all'elemento diagonale stesso a_{ii}, dall'altro, per t=1, essi sono i dischi D_i della matrice A.

Se D_p è uno dei dischi inclusi in U, anche $D_p(t)$ è incluso in U per ogni t. Per la prima parte del teorema la traiettoria $\lambda_p(t)$, uscente da $a_{pp}=\lambda_p(0)$ deve rimanere in U per ogni valore di t, altrimenti verrebbe a mancare la continuità. In particolare per t=1, l'autovalore $\lambda_p=\lambda_p(1)$ appartiene ad U. Ciò è vero per ciascuno dei k dischi che compongono U, quindi U include almeno k autovalori. D'altra parte, per quanto abbiamo appena visto, i rimanenti n-k autovalori devono stare nell'insieme V, costituito dalla riunione dei rimanenti dischi, che risulta a sua volta disgiunto da U.

Consideriamo ora alcune matrici di forma particolare che interverranno spesso in seguito.

Matrici di permutazione: Sono matrici che hanno un coefficiente uguale ad 1 su ciascuna riga e ciascuna colonna e tutti gli altri coefficienti uguali a 0, ad esempio la seguente matrice 5×5:

$$P = \begin{Vmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{Vmatrix}$$

Per una matrice di permutazione P che abbia il coefficiente $p_{i\,j}$ uguale ad 1, il prodotto PA provoca, sulla matrice A, il trasferimento della riga j-esima nella i-esima. Il prodotto AP provoca invece lo stesso trasferimento sulle colonne. Poichè ogni riga ed ogni colonna di P ha uno ed un solo coefficiente unitario, i prodotti PA ed AP rappresentano matrici ottenute con permutazioni delle righe e delle colonne di A.
In particolare, ,, chiamando $r_1 ... r_5$ le righe della matrice A ed usando la comoda notazione $A=(r_1, r_2, r_3, r_4, r_5)$, si ottiene $P\,A=(r_2, r_3, r_1, r_5, r_4)$.

Per ogni matrice P di permutazione si ha $\det(P)=(-1)^n$. Una particolare matrice di permutazione è la matrice

$$Q = \begin{Vmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{Vmatrix}$$

che, moltiplicata a sinistra o a destra di A, provoca una riflessione sugli assi di simmetria orizzontale e, rispettivamente, verticale. Q risulta simmetrica, involutoria ed ortogonale. Lo scambio di due righe o di due colonne si ottine con una matrice del tipo

$$
P = \begin{Vmatrix}
1 & & & & & & \\
 & 1 & & & & & \\
 & & 0 & \ldots & 1 & & \\
 & & \cdot & & \cdot & & \\
 & & \cdot & & \cdot & & \\
 & & 1 & \ldots & 0 & & \\
 & & & & & 1 & \\
 & & & & & & 1
\end{Vmatrix}
\begin{matrix}
\\
\\
\leftarrow \text{ i-esima riga} \\
\\
\\
\leftarrow \text{ j-esima riga} \\
\\
\\
\end{matrix}
$$

↑ i-esima colonna ↑ j-esima colonna

Il prodotto PA (o AP) realizza lo scambio tra la i-esima e la j-esima riga (o colonna) di A.

Matrici triangolari. Sono dette matrici triangolari inferiori (o superiori) quelle matrici i cui elementi sopradiagonali (o sottodiagonali) sono nulli.

$$
L = \begin{Vmatrix}
x & & & & \\
\cdot & x & & & \\
\cdot & & \cdot & & \\
\cdot & & & x & \\
x & \cdot & \cdot & \cdot & x
\end{Vmatrix}
\qquad
U = \begin{Vmatrix}
x & \cdot & \cdot & \cdot & x \\
 & x & & & \cdot \\
 & & \cdot & & \cdot \\
 & & & x & \cdot \\
 & & & & x
\end{Vmatrix}
$$

E' universalmente adottata la notazione L (Lower) per le matrici triangolari inferiori e U (Upper) per quelle triangolari superiori. Si osservi che, in riferimento alla matrice Q di riflessione, QLQ e QUQ sono, rispettivamente, triangolari superiore e inferiore. Si osservi infine che il prodotto di matrici triangolari L od U è ancora una matrice dello stesso tipo.

Matrici diagonali. Sono dette matrici diagonali le matrici con elementi non nulli solo sulla diagonale:

$$
A = \begin{Vmatrix}
\lambda_1 & & & & & & \\
 & \lambda_2 & & & & & \\
 & & \cdot & & & & \\
 & & & \cdot & & & \\
 & & & & \cdot & & \\
 & & & & & \cdot & \\
 & & & & & & \lambda_n
\end{Vmatrix}
$$

Esse saranno rappresentate, più semplicemente, nel seguente modo:

$$A= \mathrm{diag}(\lambda_1,\lambda_2,...,\lambda_n)$$

Il passaggio dalla matrice diagonale al vettore degli elementi diagonali della matrice si ottiene moltiplicando A per il **vettore unitario** $u=(1,1,...,1)$: $Au=(\lambda_1,\lambda_2,...,\lambda_n)^T$.

Gli autovalori di una matrice diagonale o triangolare sono gli elementi della diagonale principale.

Definizione: *Le matrici* A *e* B *si dicono simili* ($A\approx B$) *se esiste una matrice non singolare* P *tale che* $B=P^{-1}AP$.

E' immediato osservare che se λ è autovalore di A con autovettore x, allora λ è autovalore della matrice B con autovettore $P^{-1}x$. Infatti: $Ax=\lambda x \Leftrightarrow P^{-1}Ax=\lambda P^{-1}x \Leftrightarrow P^{-1}APP^{-1}x=\lambda P^{-1}x \Leftrightarrow BP^{-1}x=\lambda P^{-1}x$. Inoltre poichè P è non singolare, essa trasforma vettori indipendenti in vettori indipendenti e quindi la matrice B conserva la molteplicità geometrica degli autovalori. La molteplicità algebrica è pure conservata poiche A e B hanno la stessa equazione caratteristica associata. Infatti :

$\det(A-\lambda I) = \det(P^{-1}P) \det(A-\lambda I) = \det(P^{-1}) \det(A-\lambda I) \det(P) = \det(P^{-1}(A-\lambda I)P) = \det (P^{-1}AP-\lambda I)$.

Possiamo concludere con il seguente teorema:

TEOREMA 1.2. *Due matrici simili hanno gli stessi autovalori con la stessa molteplicità algebrica e geometrica. Gli autovettori sono i trasformati secondo la matrice di similitudine* P.

L'interesse per le traformazioni di similitudine stà nel fatto che le matrici possono essere trasformate in matrici simili di forma molto particolare, dette **forme canoniche**. Come avremo modo di vedere nel seguito, le trasformazioni di similitudine rappresentano uno strumento molto potente sul piano teorico ma sfortunatamente non sul piano numerico.

Teorema 1.3. (Forma canonica di Schur).*Per ogni matrice* A *esiste una matrice* P *unitaria tale che* P^HAP *risulta triangolare*.

Corollario 1.4. *Se* $A^H=A$ *allora* P^HAP è *diagonale.*

Dim. Poichè P^HAP è triangolare ed hermitiana, deve essere diagonale.

Corollario 1.5. *Se* $A^H=A$ *allora* A *possiede* *n autovettori linearmente indipendenti ed ortogonali (base ortogonale di* C^n*)*

Dim. $A \approx P^HAP = \mathrm{diag}(\lambda_1,\lambda_2,...,\lambda_n)=D$ con $P^{-1}=P^H$. Moltiplicando a sinistra per P si ottiene:

$$P(P^HAP)=PD$$

$$AP=PD.$$

Uguagliando le corrispondenti colonne , si ottine infine:

$$Ap_i=\lambda_i p_i$$

il che significa che le colonne p_i di P sono autovettori di A. Esse sono indipendenti perchè P è non singolare, e sono ortogonali perchè P è unitaria.

Corollario 1.6. *Se* $A^H=A$ *con autovalori (reali)* $\lambda_1 \le \lambda_2 \le ... \le \lambda_n$ *allora*

$$\lambda_1 x^Hx \le x^HAx \le \lambda_n x^Hx \qquad \forall x \in C^n.$$

Dim. Per il teorema di Schur:

$$A \approx P^HAP = \mathrm{diag}(\lambda_1,\lambda_2,...,\lambda_n)=D$$

Ponendo $y=P^Hx$ e quindi $x=Py$ si ottiene:

$$x^HAx = (Py)^HAPy = y^HP^HAPy = y^HDy = \Sigma\lambda_i|y_i|^2 ,$$

dove, per ipotesi,

$$\Sigma\lambda_i|y_i|^2 \le \Sigma\lambda_n |y_i|^2 = \lambda_n y^Hy$$

$$\Sigma\lambda_i|y_i|^2 \ge \Sigma \lambda_1|y_i|^2 = \lambda_1 y^Hy$$

Poiche P conserva il prodotto scalare, $y^Hy=x^Hx$ da cui la tesi.

Corollario 1.7. *Se* $\lambda_1,\lambda_2,...,\lambda_n$ *sono autovalori di* A *allora* $\lambda^{k,1},\lambda^{k,2},...,\lambda^{k,n}$ *sono tutti e soli gli autovalori di* A^k , k>1.

Dim. E' ovvio che se λ è autovalore di A, λ^k è autovalore di A^k. Dimostriamo che non ce ne sono altri. Poichè A è simile ad una matrice triangolare $T=P^{-1}AP$, i cui autovalori stanno sulla diagonale, consideriamo la matrice $R=P^{-1}A^kP$, simile ad A^k, che risulta ancora triangolare, infatti:

$$R=P^{-1}A^kP=P^{-1}AA...AP=P^{-1}APP^{-1}AP....P^{-1}AP=T^k.$$

Gli autovalori di R sono dunque gli elementi diagonali di T^k che sono del tipo $\lambda^{k,i}$.

Si osservi ancora che se $Ax=\lambda x$, anche $(A-\mu I)x=(\lambda-\mu)x$ e quindi, in base al corollario precedente, vale la seguente proposizione:

PROPOSIZIONE 10. Per ogni matrice A e per ogni polinomio p , detti λ ed x un autovalore e la corrispondente autosoluzione di A, si ha $p(A)x=p(\lambda)x$.

Definizione. *Data una matrice* A*, si chiama raggio spettrale di* A, *e si indica con* $\rho(A)$, *il numero* $\max_i |\lambda_i|$.

Teorema 1.8. $\rho^k(A) = \rho(A^k)\ \ \forall k>1$.

La dimostrazione discende dal corollario precedente.

Teorema 1.9. *Per ogni matrice non singolare, si ha:*

$$\rho(A^{-k}) = \rho^k(A^{-1}) =1/(\min_i |\lambda_i|)^k \ \ \forall k>1.$$

Dim. Si vede facilmente che se λ è autovalore di A, λ^{-1} è autovalore di A^{-1} da cui $\rho(A^{-1}) =1/(\min_i |\lambda_i|)$. Poichè $A^{-k}=(A^{-1})^k$ si ha, per il teorema 1.8, $\rho(A^{-k}) = \rho^k(A^{-1}) =1/(\min_i |\lambda_i|)^k$.

Teorema 1.10. (Forma canonica di Jordan). *Ogni matrice* $A\in \mathcal{L}(C^n, C^n)$ *è simile ad una matrice diagonale a blocchi del tipo:*

$$A \approx P^{-1}AP = J = \mathrm{diag}(J_1, J_2, ..., J_k)$$

dove i blocchi $J_i \in \mathcal{L}(C^{n_i}, C^{n_i})$, *detti blocchi di Jordan, sono del tipo:*

$$J_i = \left\| \begin{matrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & . & . & \\ & & & . & 1 \\ & & & & \lambda \end{matrix} \right\|$$ *con* λ *autovalore di* A *ed* $n_1+...+n_k=n$.

Ogni autovalore di molteplicità algebrica s *compare sulla diagonale di* J *esattamente* s *volte e può dar luogo ad un numero di blocchi di Jordan compreso tra* 1 *ed* s *. Precisamente, per ogni autovalore ci saranno tanti blocchi di Jordan quanto è la molteplicità geometrica di quell'autovalore.*

2-NORME DI VETTORI E DI MATRICI.

Indichiamo in generale con V uno spazio vettoriale .

Definizione: *Sullo spazio* V *si definisce* **norma** *ogni applicazione* $\| \ \|$ *a valori reali non negativi che gode delle proprietà:*

$\| x \| \geq 0$; $\quad \| x \| = 0 \quad \Leftrightarrow \quad x=0$ (elemento nullo di V)
$\| \alpha x \| = |\alpha| \, \| x \|$
$\| x+y \| \leq \| x \| + \| y \|$ per ogni x ed y di V, e per ogni scalare α.

In generale si possono definire più norme sullo stesso spazio vettoriale. Consideriamo dapprima lo spazio C^n e definiamo alcune possibili norme:

$\| x \|_1 = \sum |x_i|$

$\| x \|_2 = \sqrt{\sum |x_i|^2}$ è la norma Euclidea $\qquad \| x \|_2^2 = x^H x$

$\| x \|_p = \sqrt[p]{\sum |x_i|^p}$

$\| x \|_\infty = \max_i |x_i|$

$\| x \|_A = \| Ax \| \qquad\qquad |A| \neq 0$

$\| x \|_{\ell\ell} = \sqrt{x^H A x}$ con $A^H = A$, A def.pos. E' detta norma ellittica ed è ricavata dal prodotto scalare: $\langle x,y \rangle_A := x^H A y$

Si osservi che $\lim_{p\to\infty} \| x \|_p = \| x \|_\infty$ infatti $\sqrt[p]{\sum |x_i|^p} \geq \| x \|_\infty$ ed inoltre $\sqrt[p]{\sum |x_i|^p} \leq \sqrt[p]{n \max_i |x_i|^p} = \sqrt[p]{n} \, \| x \|_\infty \to \| x \|_\infty$ per $p \to \infty$.

D'ora in poi, nel considerare lo spazio vettoriale R^n , C^n, o in generale uno spazio V , lo penseremo dotato di una norma e lo indicheremo con $(V, \| \cdot \|)$. Ricordiamo che negli spazi vettoriali di dimensione finita, quali quelli che stiamo considerando, tutte le norme sono equivalenti, cioè per ogni coppia di norme $\|\ \|$, $\|\ \|^\wedge$ esistono due costanti m , M tali che

$$m\| x \| \leq \| x \|^\wedge \leq M\| x \| \qquad \forall\, x \in V.$$

In altre parole, se una successione converge in una norma, converge in ogni altra norma.

Un'altra osservazione che sarà utile nel seguito è che l'applicazione che ad un vettore x associa la sua norma $\| x \|$, è una applicazione continua di $(V,\| \,.\, \|)$ in $R^+ \cup \{0\}$.

Poichè ogni matrice A di dimensione $n{\times}m$ può essere vista come un vettore ad nm componenti, anche le matrici possono essere normate. Si avrà, per esempio:

$$\| A \|_{Euclidea} = \sqrt{\sum_{i\,j} |a_{i\,j}|^2}$$

$$\| A \|_{max} = \max_{i\,j} |a_{i\,j}|.$$

Un altro modo di definire delle norme per le matrici, è quello di normare l'operatore lineare che esse rappresentano. In questo caso la norma sarà indotta dalle norme assegnate agli spazi vettoriali tra i quali opera la matrice A.

Consideriamo, in generale, una matrice A , $m{\times}n$, rappresentativa di una applicazione lineare tra due spazi normati $(C^n,\| \,.\, \|^\wedge)$ e $(C^m,\| \,.\, \|^*)$. Se $n=m$ i due spazi sono da considerarsi comunque diversi inquanto dotati di norme diverse. L'applicazione di $\mathcal{L}(C^n, C^m)$ in R così definita:

$$\| A \| := \sup_{x \neq 0} \frac{\| Ax \|^*}{\| x \|^\wedge}. \qquad A \in \mathcal{L}(C^n, C^m)$$

gode delle seguenti proprietà (che dimostreremo tra poco):

1. $\| A \| \geq 0, \qquad \| A \| = 0 \Leftrightarrow A=0$ (matrice nulla)
2. $\|\alpha A \| = |\alpha|\, \| A \|$
3. $\| A+B \| \leq \| A \| + \| B \|$
4. $\| AB \| \leq \| A \|\ \| B \|$

e viene detta **norma indotta** dalle norme di vettore. In generale scriveremo:

$$\| A \| = \sup_{x\neq 0} \frac{\| Ax \|}{\| x \|}$$

tenendo però presente che $\| A \|$, $\| Ax \|$ e $\| x \|$ sono tre norme diverse, quella di $\mathcal{L}(C^n, C^m)$, di C^m e di C^n rispettivamente. Nel caso di applicazioni di uno spazio normato in sè, cioè di applicazioni tra due spazi vettoriali uguali e dotati della stessa norma $\| \ \|_i$, scriveremo

$$\| A \|_i = \sup_{x\neq 0} \frac{\| Ax \|_i}{\| x \|_i}$$ e la chiameremo **norma naturale .**

Osserviamo subito che per ogni norma naturale si ha $\| I \| =1$.

Prima di dimostrare che valgono le quattro proprietà delle norme, osserviamo che la norma indotta si può esprimere in altro modo:

$$\| A \| = \sup_{x\neq 0} \frac{\| Ax \|}{\| x \|} =\sup_{x\neq 0} \left\| \frac{Ax}{\| x \|} \right\| = \sup_{\| x \|=1} \| Ax \|$$

Poichè, come abbiamo già osservato, l'applicazione $x \to \| Ax \|$ è una applicazione continua , essa ammette massimo su ogni insieme chiuso e limitato. Poichè, infine, la sfera unitaria è un insieme chiuso e limitato, la nostra applicazione ammette massimo. Quindi:

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

In seguito sarà utile ricordare che esiste un x : $\| x \|=1$ e $\| Ax \|=\| A \|$.

Definizione: *Una norma di matrice di dice* **consistente** *o* **compatibile** *con una norma di vettore se* $\| Ax \|\leq\| A \| \| x \|$.

Ovviamente ogni norma naturale è compatibile con la norma di vettore dalla quale è dedotta, ma ci possono essere compatibilità anche per norme di matrici non naturali. Per esempio: la norma euclidea di matrice è compatibile con la norma 2 di vettore:

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

Verifichiamo ora che l'applicazione $A \to \| A \|$ verifica le 4 proprietà della norma precedentemente enunciate.

1. Parte della dimostrazione è ovvia; dimostriamo solo che $\| A \| = 0 \Rightarrow A=0$. Poichè $\| A \| =0$ allora $\| Ax \| = 0$ per ogni $\| x \| =1$. D'altra parte $\| Ax \| =0 \Rightarrow Ax=0$ e, per l'arbitrarietà di x nella sfera unitaria, si ha $A=0$.

2. Dimostrazione ovvia.

3. Sia x tale che $\| x \| =1$ e $\| (A+B)x \| =\| A+B \|$. Si ha $\| (A+B)x \| = \| Ax+Bx\| \leq \| Ax \| + \| Bx \| \leq \| A \| \| x \| + \| B \| \| x \| = \| A \| +\| B \|$.

4. Dimostrazione analoga.

La definizione di norma naturale non consente, in generale, di valutare facilmente la norma di una matrice in termini dei suoi coefficienti. Ciò risulta possibile in alcuni casi particolari.

TEOREMA 1.11. *Per ogni* $A \in \mathcal{L}(C^n, C^n)$, $\| A \|_\infty = \max_i \sum_j |a_{ij}|$.

Dim. Sia x tale che $\| x \|_\infty =1$ e $\| Ax \|_\infty = \| A \|_\infty$. Si ha:

$\| Ax \|_\infty = \max_i | \sum_j a_{ij} x_j | \leq \max_i \sum_j |a_{ij}||x_j| \leq \max_i \sum_j |a_{ij}| \| x \|_\infty = \max_i \sum_j |a_{ij}|$

Per dimostrare la disuguaglianza opposta, sia k l'indice per il quale:

$$\sum_j |a_{kj}| = \max_i \sum_j |a_{ij}| \text{ e sia } z = \left(\frac{a_{k1}}{|a_{k1}|}, \frac{a_{k2}}{|a_{k2}|}, \ldots, \frac{a_{kn}}{|a_{kn}|} \right)^T$$

Si ha ovviamente $\| z \|_\infty =1$ ed

$$Az = \left(\sum_j a_{1j} \frac{a_{kj}}{|a_{kj}|}, \sum_j a_{2j} \frac{a_{kj}}{|a_{kj}|}, \ldots, \sum_j a_{nj} \frac{a_{kj}}{|a_{kj}|} \right)^T$$

la cui norma è $\| Az \|_\infty = \max_i | \sum_j a_{ij} \frac{a_{kj}}{|a_{kj}|} |$. Per ogni indice i si ha:

$$| \sum_j a_{ij} \frac{a_{kj}}{|a_{kj}|} | \leq \sum_j |a_{ij}|.$$

mentre per l'indice i=k si ha

$$\left| \sum_j a_{kj} \frac{a_{kj}}{|a_{kj}|} \right| = \sum_j |a_{kj}|,$$

cosicchè si ottine:

$$\| Az \|_\infty = \max_i \sum_j |a_{ij}|$$

da cui $\| A \|_\infty \geq \max_i \sum_j |a_{ij}|$ e quindi la tesi.

Analogamente si dimostra il seguente teorema:

TEOREMA 1.12. *Per ogni* $A \in \mathcal{L}(C^n, C^n)$, $\| A \|_1 = \max_j \sum_i |a_{ij}|$

Dim.(analoga)

Dimostriamo ancora un teorema che fornisce il valore di ,, $\|A\|_2$ in funzione del raggio spettrale di A^HA. Cominciamo con l'osservare che, pur essendo in generale $A^HA \neq AA^H$, si ha $\rho(A^HA) = \rho(AA^H)$

TEOREMA 1.13. $\rho(A^HA) = \rho(AA^H)$.

Dim. Siccome A^HA è hermitiana semidefinita positiva i suoi autovalori sono reali e non negativi, quindi $\rho(A^HA)$ è un autovalore tale che $\rho(A^HA) \geq 0$. Supponiamo dapprima che sia $\rho(A^HA) > 0$ e sia x l'autovettore associato. Si ha:

$A^HAx = \rho(A^HA)x$ (da cui $Ax \neq 0$)

$AA^HAx = \rho(A^HA)Ax$.

Da ciò si deduce che $\rho(A^HA)$ è un autovalore, positivo, di AA^H e quindi:

$\rho(AA^H) \geq \rho(A^HA) > 0$.

Dall'ultima relazione si ricava $\rho(AA^H) > 0$ e, per quanto appena dimostrato,

$\rho(A^HA) \geq \rho(AA^H) > 0$

e quindi:

$$\rho(AA^H)=\rho(A^HA).$$

Sia ora $\rho(A^HA)=0$. Se fosse $\rho(AA^H)>0$, per le relazioni precedenti, sarebbe anche $\rho(A^HA)>0$. Il teorema è così dimostrato.

TEOREMA 1.14. $\| A \|_2 = \sqrt{\rho(A^HA)} = \sqrt{\rho(AA^H)}$

Dim. Per il corollario 1.6, gli autovalori di A^HA sono non negativi e per ogni $x\neq 0$ si ha $x^HA^HAx \leq \rho(A^HA)x^Hx = \rho(A^HA) \| x \|_2^2$ e quindi

$$\frac{\| Ax \|_2^2}{\| x \|_2^2} \leq \rho(A^HA),$$ dalla quale si deduce $\| A \|_2^2 \leq \rho(A^HA)$.

Per la disuguaglianza opposta, detto x l'autovettore di A^HA associato a $\rho(A^HA)$, si ha:

$$A^HAx = \rho(A^HA)x$$

$$x^HA^HAx = \rho(A^HA)x^Hx$$

$$\frac{\| Ax \|_2^2}{\| x \|_2^2} = \rho(A^HA),$$

dalla quale si deduce $\| A \|_2^2 \geq \rho(A^HA)$ e quindi la tesi.

E' immediato osservare che $\rho(A)$ è una minorazione per tutte le norme naturali di A. Infatti, detto λ l'autovalore di modulo massimo ed x il corrispondente autovettore normalizzato in una qualunque norma ($\| x \| =1$), si ha:

$$Ax = \lambda x \Rightarrow \| Ax \| = | \lambda | \, \| x \| = | \lambda | \Rightarrow | \lambda | \leq \| A \| \, \| x \| = \| A \|.$$

da cui si ricava $\rho(A)\leq \| A \|$. Per qualche matrice può accadere che il raggio spettrale coincida con qualche norma. Per esempio se A è hermitiana, dal teorema 1.14 segue che $\rho(A)= \|A\|_2$. Per una matrice arbitraria A, vale il seguente importante risultato

TEOREMA 1.15. *Per ogni* $A \in \mathcal{L}(C^n, C^n)$,

$$\rho(A) = \inf \|A\|,$$

dove l'estremo inferiore è inteso rispetto a tutte le norme naturali.

Dim. Dobbiamo dimostrare che, assegnato ad arbitrio un valore $\varepsilon>0$, esiste una norma tale che $\|A\| \leq \rho(A)+\varepsilon$.

Sia $D=\mathrm{diag}(1,\varepsilon,\varepsilon^2...,\varepsilon^{n-1})$ e sia $J=P^{-1}AP$ la forma canonica di Jordan della matrice A. Si vede, senza eccessive difficoltà, che la matrice $D^{-1}JD$ ha la stessa struttura a blocchi della matrice J, con la sola differenza che i blocchi sono del tipo:

$$J_i = \begin{Vmatrix} \lambda & \varepsilon & & & \\ & \lambda & \varepsilon & & \\ & & \cdot & \cdot & \\ & & & \cdot & \varepsilon \\ & & & & \lambda \end{Vmatrix}$$

Allora si ha, evidentemente $\|D^{-1}JD\|_\infty \leq \rho(A)+\varepsilon$.

Detto Q=PD, definiamo la seguente norma di vettore:

$$\|x\| := \|Q^{-1}x\|_\infty$$

e la corrispondente norma indotta di matrice:

$$\| A \| = \max_{\| x \|=1} \| Ax \| = \max_{\| Q^{-1}x \|_\infty=1} \| Q^{-1}Ax \|_\infty$$

Per tale norma si ha:

$$\| A \| = \max_{\| Q^{-1}x \|_\infty=1} \| Q^{-1}AQQ^{-1}x \|_\infty = \max_{\| y \|_\infty=1} \| Q^{-1}AQy \|_\infty$$

$$= \| Q^{-1}AQ \|_\infty = \| D^{-1}P^{-1}APD \|_\infty = \| D^{-1}JD \|_\infty \leq \rho(A)+\varepsilon.$$

Il teorema è così dimostrato.

<u>*Osservazione*</u>: Nella dimostrazione precedente, se i blocchi di Jordan relativi agli autovalori di modulo massimo di A hanno dimensione 1, allora le righe della matrice $D^{-1}JD$ in cui ci sono gli autovalori di modulo massimo, hanno coefficienti tutti nulli a parte l'autovalore stesso che sta sulla diagonale. Di conseguenza, per ε sufficientemente piccolo, si ha $\| D^{-1}JD \|_\infty = \rho(A)$ e quindi nella dimostrazione del teorema precedente si

sarebbe potuto concludere con $\| A \| = \rho(A)$. Da quanto detto si ricavano delle condizioni sufficienti affinchè, per una matrice A, esista una norma tale che $\| A \| = \rho(A)$. Precisamente, è sufficiente che A sia simile ad una matrice diagonale; oppure, più in generale, che gli autovalori di modulo massimo siano semplici; oppure, ancora più in generale, che gli autovalori di modulo massimo e multipli abbiano molteplicità geometrica uguale alla molteplicità algebrica.

Per lo studio dei metodi iterativi, di cui ci occuperemo nel prossimo capitolo, è importante il concetto di *matrice convergente* , cioè di matrice la cui successione di potenze A^m tende alla matrice nulla, per m che tende a infinito·
Il seguente teorema caratterizza le matrici convergenti.

TEOREMA 1.16. *Per ogni matrice* A *le seguenti tre affermazioni sono equivalenti* :

1. $\rho(A)<1$
2. $\| A^m \| \to 0$,
3. $(I - A)^{-1} = I + A + A^2 + ... + A^n + ...$ (*serie di Neumann*).

Dim. (1.$\Leftrightarrow$ 2.)
Se $\rho(A)<1$, esiste una norma tale che $\| A \|<1$. Quindi $\| A^m \| \leq \| A \|^m \to 0$. Viceversa, poiché $\rho^m(A) = \rho(A^m) \leq \| A^m \|$ ed inoltre $\| A^m \| \to 0$, allora $\rho^m(A) \to 0$ e quindi $\rho(A)<1$.
(2. $\Leftrightarrow$ 3.) Poichè la matrice è convergente, il raggio spettrale è minore di uno e quindi $\lambda=1$ non è autovalore di A. Di conseguenza $(I - A)$ è invertibile. Per ogni intero m si ha

$$(I - A)(I + A + A^2 + ... + A^m) = I - A^{m+1}$$

$$I + A + A^2 + ... + A^m = (I - A)^{-1} (I - A^{m+1}).$$

Poichè la matrice A è convergente, si può concludere, per la continuità della norma, che A^{m+1} tende alla matrice nulla e quindi il termine a destra dell'ultima uguaglianza tende a $(I - A)^{-1}$. Di conseguenza anche il termine a sinistra è convergente e converge allo stesso limite. Viceversa se la serie di Neumann converge allora il termine generale della serie deve essere infinitesimo.

Si osservi l'analogia tra la serie di Neumann e la serie geometrica $1+x+..+x^n+...$ che converge a $1/(1-x)$ per ogni x reale (o complesso) di modulo minore di uno. Ciò non è tanto sorprendente se si pensa che l'insieme delle matrici quadrate, con le usuali operazioni di somma e moltiplicazione, formano un anello con unità, cioè godono di quasi tutte le proprietà dei numeri reali o complessi. La sola proprietà mancante è la commutatività del prodotto, ma nella serie di Neumann i prodotti sono fatti solo sulla stessa matrice A.

Concludiamo questa parte introduttiva con una importante proprietà del raggio spettrale.

TEOREMA 1.17. *Per ogni matrice* A *e per ogni norma naturale si ha:*

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

Dim. Si ha $\rho^k(A) = \rho(A^k) \leq \| A^k \|$ e quindi, essendo $\rho(A) \geq 0$, si ha $\rho(A) \leq ||A^k||^{1/k}$. D'altra parte, per ogni $\varepsilon > 0$, la matrice $A_\varepsilon = \dfrac{A}{\rho(A)+\varepsilon}$ è tale che $\rho(A_\varepsilon) < 1$ e quindi, per il teorema precedente, $\lim_{k\to\infty} ||A_\varepsilon^k|| = 0$. Di conseguenza, esiste $n=n(\varepsilon)$ tale che, $\forall\, k>n$, $||A_\varepsilon^k|| <1$ cioè $\dfrac{||A^k||}{(\rho(A)+\varepsilon)^k} < 1$, e quindi $||A^k||^{1/k} < \rho(A)+\varepsilon$.