APPUNTI PER UN SECONDO CORSO DI ALGEBRA LINEARE

Prof. Stefano Capparelli

A mia madre

Ver. 19 marzo 07

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

Prefazione. Questi sono alcuni appunti delle lezioni da me tenute negli ultimi anni per il corso di Algebra II per la Laurea Specialistica in Telecomunicazioni, presso la Facoltà di Ingegneria dell'Università La Sapienza.

Il corso si pone come obbiettivi principali la riduzione a forma canonica di una matrice quadrata e il calcolo di funzioni di matrici; alcune proprietà e caratterizzazioni di vari tipi notevoli di matrici (ortogonali, hermitiane,...); varie fattorizzazioni di matrici.

Il corso ha come prerequisiti le conoscenze di base di Analisi Matematica e di Geometria che si ottengono di solito nei corsi del primo anno della Laurea Triennale di Ingegneria.

Questi appunti sono scritti volutamente in uno stile discorsivo e a volte decisamente informale. Per venire incontro in fretta ai bisogni e ai desideri degli studenti ho preferito questo tipo di esposizione piuttosto che una trattazione più completa e dotta, che comunque è possibile trovare altrove, ad esempio nei testi citati nella bibliografia. Sarei grato a chiunque volesse segnalare eventuali errori, imprecisioni od omissioni, o magari passi poco chiari che saranno inevitabilmente presenti in appunti di questo tipo. Ringrazio gli studenti che hanno "subito" le mie lezioni, e la collega Prof. Patrizia Macrì per spunti, suggerimenti e consigli utilissimi.

Introduzione. L'algebra lineare è senz'altro uno dei campi della matematica che trova maggiori applicazioni nelle scienze e nell'ingegneria. È allo stesso tempo un campo della matematica interessante di per sé. Per esempio, si potrebbe affermare che lo studio dell'algebra lineare costituisce un primo livello di introduzione ad uno dei campi più importanti della matematica: quello dedicato allo studio delle teorie di S. Lie. R. Howe infatti, nel suo interessante articolo [H], afferma che la decomposizione ai valori singolari corrisponde alla decomposizione di Cartan, che la decomposizione QR corrisponde alla decomposizione di Iwasawa, etc. Un altro merito dello studio dell'algebra lineare sta nel fatto di permettere agli studenti, un contatto accessibile, a volte il primo contatto, con vari aspetti della matematica: scoperta di risultati, dimostrazioni rigorose, applicazioni utili.

In questo corso supponiamo che lo studente abbia delle conoscenze elementari di algebra lineare, che di solito acquisisce nei corsi di primo anno, alcune delle quali sono qui richiamate brevemente all'inizio. L'enfasi di questo corso è sulle matrici, e, in maniera minore, sulle trasformazioni lineari. Storicamente lo studio delle matrici precede lo studio degli spazi vettoriali in astratto. Nei corsi di primo anno si vedono di solito alcuni aspetti generali degli spazi vettoriali e si stabilisce un linguaggio matriciale per alcune questioni. Si vedono alcune applicazioni a problemi geometrici (come ad esempio la classificazione delle coniche). Qui consideriamo le matrici come oggetto fondamentale dell'algebra lineare e non solo come strumento di calcolo.

Capitolo 0. Brevi richiami di algebra lineare.

In questo numero voglio solo indicare quegli argomenti di più immediata necessità che lo studente ha già acquisito nel corso di Geometria del primo anno.

Definizione di spazio vettoriale. Un insieme V si dice spazio vettoriale su un arbitrario campo $\mathbb{K}$ (ad esempio $\mathbb{K} = \mathbb{R}, \mathbb{C}, \mathbb{Q}$ ma non $\mathbb{N}, \mathbb{Z}$) se è definita una operazione binaria

$$V \times V \to V$$

di solito denotata con l'addizione

$$(v_1, v_2) \mapsto v_1 + v_2$$

ed una operazione "esterna"

$$\mathbb{K} \times V \to V$$

$$(a, v) \mapsto av$$

che soddisfano i seguenti assiomi

(1) $v_1 + v_2 = v_2 + v_1$
(2) Esiste $0 \in V$ tale che $0 + v = v \ \forall v \in V$
(3) Dato $v \in V$, esiste $v' \in V$ tale che $v + v' = 0$ (v' è di solito denotato $-v$)
(4) $(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3) \ \forall v_1, v_2, v_3 \in V$
(5) $a(bv) = (ab)v \ \forall a, b \in \mathbb{K}, v \in V$
(6) $1v = v, 1 \in \mathbb{K}, v \in V$
(7) $a(v_1 + v_2) = av_1 + av_2, \forall a \in \mathbb{K}, v_1, v_2 \in V$
(8) $(a_1 + a_2)v = a_1v + a_2v, \forall a_1, a_2 \in \mathbb{K}, v \in V$

Sottospazio. Un sottoinsieme $U \subset V$ si dice sottospazio di V se

(1) $v_1, v_2 \in U \implies v_1 + v_2 \in U$
(2) $a \in \mathbb{K}, v \in U \implies av \in U$

Base. Un insieme di vettori $\{e_1, \dots , e_n\}$ in uno spazio vettoriale V si dice base di V se ogni vettore $v \in V$ si scrive in maniera unica come combinazione lineare $v = \sum_{i=1}^n a_ie_i, a_i \in \mathbb{K}$. I coefficienti a_i si dicono coordinate di v rispetto alla base data.

Si dimostra che

Teorema 0.1. *ogni spazio vettoriale ammette una base e che due basi qualunque dello stesso spazio vettoriale sono equipotenti cioè è possibile stabilire una corrispondenza biunivoca tra esse.*

□

Dimensione finita. Se uno spazio V ha una base costituita da un numero finito di vettori si dice che è di dimensione finita. Il numero di vettori di una base si chiama **dimensione** di V ed è denotato da $dim\ V$. Se V non ha una base con un numero finito di elementi si dice che ha dimensione infinita.

Spazio generato. L'insieme di tutte le possibile combinazioni lineari di un insieme S di vettori di V si chiama lo spazio vettoriale generato da S e si denota con $\langle S\rangle$ oppure con $Span(S)$.

Linearmente indipendenti. Un insieme $\{e_i\}$ di vettori si dice insieme libero e i suoi vettori si dicono linearmente indipendenti se

$$\sum_{i=1}^{n} a_i e_i = 0 \implies a_i = 0 \quad \forall i$$

Teorema 0.2 (Estensione di una base). *Se E' è un insieme libero di V allora esiste una base di V che contiene E'.*

□

Applicazione Lineare. Se U e V sono due spazi vettoriali su $\mathbb{K}$, una applicazione

$$f : U \to V$$

si dice lineare se
(1) $f(av) = af(v)$
(2) $f(v_1 + v_2) = f(v_1) + f(v_2)$

Nucleo. L'insieme $ker\ f = \{u \in U | f(u) = 0\}$ si chiama nucleo di f.

È facile verificare che $ker\ f$ è un sottospazio di U.

Immagine. Si dice immagine di f l'insieme

$$Im\ f = \{v \in V | \exists u \in U, f(u) = v\}$$

Anche $Im\ f$ è un sottospazio di V.

Proposizione 0.3. *f è iniettiva $\iff$ $ker\ f = \{0\}$*

□

Teorema 0.4. *Se la dimensione di U è finita e $f : U \to V$ è lineare allora $ker\ f$ e $Im\ f$ hanno dimensione finita e*

$$dim\ ker\ f + dim\ Im\ f = dim\ U$$

□

Corollario 0.5. *Se U ha dimensione finita allora*

$$f \text{ è iniettiva} \iff dim\ U = dim\ Im\ f \iff f \text{ è suriettiva}$$

□

MATRICE DI UNA APPLICAZIONE LINEARE. Se U e V sono due spazi vettoriali su $\mathbb{K}$ e scegliamo una base $\{e_1, \ldots, e_n\}$ di U e una base $\{e'_1, \ldots, e'_m\}$ di V allora possiamo associare ad un'applicazione lineare $f : U \to V$ una matrice $m \times n$, diciamo A_f, come segue: Calcoliamo $f(e_k) = \sum_{i=1}^{n} a_{ik} e'_i$ e prendiamo

$$A_f = (a_{ik})$$

In altre parole, i coefficienti di queste combinazioni lineari sono le colonne di A_f.

OPERATORE. Un'applicazione lineare in cui $U = V$ si dice endomorfismo di V o operatore (lineare) su V.

PROPOSIZIONE 0.6. *Se $\{e_i\}$ è una base di V si può calcolare A_f, la relativa matrice di f. Analogamente per una base diversa $\{e'_i\}$ abbiamo un'altra matrice A'_f. Allora*

$$A'_f = B^{-1} A_f B$$

dove B è la matrice del cambiamento di base da $\{e_i\}$ a $\{e'_i\}$.

□

SIMILI. Due matrici A e A' per cui esiste una matrice invertibile B tale che

$$A' = B^{-1} A B$$

si dicono simili o coniugate.

PROPOSIZIONE 0.7. *Due matrici quadrate A e A' sono simili se e solo se esse rappresentano lo stesso operatore rispetto a due basi diverse.*

□

SOMMA DI SOTTOSPAZI. Se $U_1, \ldots, U_n$ sono sottospazi di V si definisce $U_1 + \cdots + U_n$ l'insieme

$$\{\sum_{i=1}^{n} u_i | u_i \in U_i\}$$

È facile verificare che $U_1 + \cdots + U_n$ è un sottospazio di V e che è il più piccolo sottospazio di V che contiene tutti gli U_i.

TEOREMA 0.8 (FORMULA DI GRASSMANN). *Se U_1 e U_2 sono due sottospazi di dimensione finita allora*

$$dim\ (U_1 \cap U_2) + \ dim\ (U_1 + U_2) = \ dim\ U_1 + \ dim\ U_2$$

□

SOMMA DIRETTA. Uno spazio V si dice somma diretta dei suoi sottospazi

$$U_1, \ldots, U_n$$

se ogni vettore $v \in V$ si scrive in maniera unica nella forma

$$v = u_1 + \ldots + u_n$$

$u_i \in U_i$. Si scrive allora

$$V = \bigoplus_{i=1}^{n} U_i$$

Teorema 0.9.
$$V = \bigoplus_{i=1}^{n} U_i$$
se e solo se $V = \sum_{i=1}^{n} U_i$ *e*
$$U_j \cap (V = \sum_{i \neq j} U_i)$$
per ogni $j = 1, \ldots, n.$

□

Sottospazio supplementare. Se V ha dimensione finita, allora per ogni sottospazio $U_1 \subset V$ esiste un altro sottospazio U_2 di V tale che $V = U_1 \oplus U_2$. (In generale il supplementare non è unico).

Operatore diagonalizzabile. Un operatore f su V si dice diagonalizzabile se esiste una base di V rispetto alla quale la matrice di f è diagonale.

Autovettore. Un vettore $v \neq 0$ si dice un autovettore per f se $f(v) = \lambda v$, $\lambda \in \mathbb{K}$. λ si dice autovalore di f.

Sottospazio invariante. Un sottospazio U di V si dice invariante (o stabile) per f se $f(U) \subset U$

Ad esempio, se v è un autovettore per f e $U = \langle v \rangle$ allora U è stabile per f.

Polinomio caratteristico. Se V ha dimensione finita, f un operatore su V e A una sua matrice in qualche base, si denota $p_T(t)$ il polinomio
$$p_T(t) = \det\,(A - tI)$$
dove I è la matrice identità e si chiama polinomio caratteristico di A.

Teorema 0.10. *Il polinomio caratteristico non dipende dalla scelta della matrice* A *che rappresenta* f. *Inoltre* p_T *è un polinomio di grado* $n = dim\ V$ *in cui il coefficiente direttore è* $(-1)^n$, *il termine noto è* $det\ A$ *e il coefficiente di* t^{n-1} *è* $(-1)^{n-1} Tr\ A$. *Infine* $\lambda_0 \in \mathbb{K}$ *è un autovalore di* T *se e solo se* $p_T(\lambda_0) = 0$. *Uno scalare* $\lambda \in \mathbb{K}$ *è un autovalore di* f *se e solo se* λ *è radice del polinomio caratteristico.*

□

L'insieme degli autovalori di un operatore T si dice spettro di T. L'insieme degli autovettori relativi ad uno stesso autovalore λ si dice autospazio e si indica con $V(\lambda)$ o V_λ.

Osservazione. Se $\lambda = 0$ allora $V_0 = ker\ T$ e in generale
$$V_\lambda = ker\ (T - \lambda I)$$
Dunque λ è un autovalore se e solo se $T - \lambda I$ è singolare.

Operatore Diagonalizzabile. *T è diagonalizzabile se e solo se esiste una base di V composta da autovettori di T.*

Teorema 0.11. *T è diagonalizzabile se e solo se*

$$V = V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_h}$$

ovvero se e solo se T ha tutti gli autovettori in $\mathbb{K}$ (cioè il polinomio caratteristico ha esattamente n radici in $\mathbb{K}$, contate con la relativa molteplicità, dove $n = dim\ V$) e per ciascun autovalore la molteplicità geometrica è uguale alla molteplicità algebrica.

□

Introduciamo ora un utile concetto che può non essere stato visto in un primo corso di geometria ma che risulta molto utile: il concetto di spazio quoziente.

Se U è un sottospazio di uno spazio vettoriale V e $v \in V$ un qualunque vettore, possiamo considerare l'insieme:

$$v + U = \{v + u | u \in U\}$$

Esso rappresenta "la traslazione" del sottospazio vettoriale U tramite il vettore v. Ad esempio, una qualunque retta del piano cartesiano è di questo tipo: se la retta passa per l'origine possiamo prendere $v = 0$ ed abbiamo un sottospazio vettoriale, altrimenti se la retta non passa per l'origine, ad esempio la retta di equazione $x + y = 1$ allora essa è del tipo $v + U$ dove U è il sottospazio di equazione $x + y = 0$ e v può essere preso come $v = (0, 1)$. Queste "parallele" chiaramente non sono sottospazi vettoriali (in quanto non sono chiusi rispetto alle operazioni), essi si dicono sottospazi affini.

Ci occorre un primo lemma

Lemma 0.12. *$v_1 + U_1 = v_2 + U_2$ se e solo se $U_1 = U_2$ e $v_1 - v_2 \in U_1 = U_2$.*

Dim. Se $U_1 = U_2$, chiamiamo U questo sottospazio ($U = U_1 = U_2$). Consideriamo $v_1 - v_2 = u_0 \in U$. Allora

$$v_1 + U = \{v_1 + u | u \in U\}$$

$$v_2 + U = \{v_1 + u - u_0 | u \in U\}$$

Al variare di u in U, $u - u_0$ varia tra tutti i vettori di U. Quindi $v_1 + U_1 = v_2 + U_2$.

Viceversa, sia $v_1 + U_1 = v_2 + U_2$. Sia $u_0 = v_1 - v_2$. È chiaro dalla definizione che $u_0 + U_1 = U_2$. Poiché $0 \in U_2$ deve essere $u_0 \in U_1$. Quindi $u_0 + U_1 = U_1$ secondo il ragionamento di prima, quindi $U_1 = U_2 = U$ come desiderato.

□

Definizione. Lo spazio quoziente V/U di uno spazio vettoriale V è l'insieme di tutti i sottospazi affini di V che sono traslazioni di U insieme alle seguenti operazioni:

(1) $(v_1 + U) + (v_2 + U) = (v_1 + v_2) + U$
(2) $a(v_1 + U) = av_1 + U$ per ogni $v_1, v_2 \in V$ e $a \in \mathbb{K}$.

Esercizio Queste definizioni sono ben poste e V/U con queste operazioni è uno spazio vettoriale.

Soluzione Per verificare che la definizione sia ben posta occorre verificare che la somma di due sottospazi affini, che è definita tramite l'uso dei due vettori v_1 e v_2 non dipende dalla scelta di questi due vettori (ché tale scelta non è unica). Sia allora $v_1 + U = v_1' + U$ e anche $v_2 + U = v_2' + U$. Allora il Lemma precedente implica che $v_1 - v_1' = u_1 \in U$ e $v_2 - v_2' = u_2 \in U$. Quindi, sempre a causa del Lemma precedente,

$$(v_1 + v_2) + U = (v_1' + v_2') + (u_1 + u_2) + U = (v_1' + v_2') + U$$

perché $u_1 + u_2 \in U$.

Inoltre, se $v_1 + U = v_1' + U$ allora posto $v_1 - v_1' = u \in U$, abbiamo $av_1 = av_1' = au \in U$ e il Lemma precedente ci dà la conclusione. Dunque le definizioni sono ben poste.

Occorre ora verificare che queste operazioni danno una struttura di spazio vettoriale. Verifichiamo la distributività.

$$\begin{aligned} a[(v_1 + U) + (v_2 + U)] &= a((v_1 + v_2) + U) = a(v_1 + v_2) + U = \\ &= av_1 + av_2 + U = (av_1 + U) + (av_2 + U) = a(v_1 + U) + a(v_2 + U). \end{aligned}$$

dove abbiamo usato in successione: la definizione di somma in V/U, la definizione di moltiplicazione per uno scalare in V/U, la distributività in V, di nuovo la definizione di somma e moltiplicazione per uno scalare in V/U.

In modo analogo si possono verificare le altre condizioni.

È molto importante osservare che esiste una applicazione naturale $\pi : V \to V/U$ che ad ogni vettore $v \in V$ associa il sottospazio affine $v + U$. È una applicazione suriettiva: la controimmagine del sottospazio affine $v + U$ sono tutti i vettori di V che sommati a U danno $v + U$; almeno v è dunque nella preimmagine. In effetti, ogni vettore di $v + U$ è nella preimmagine di $v + U$. Si può anche verificare che π è lineare e che il suo nucleo coincide con U.

Corollario 0.13. *Se V ha dimensione finita allora:*

$$dim\ V/U = dim\ V - dim\ U$$

Dim. Infatti dalla relazione $\dim Im\ f + \dim ker\ f = n$, data dal Teorema 4 precedente, applicata a questo caso abbiamo il corollario.

□

Esercizio.

Sia U il sottospazio di $\mathbb{R}^4$ generato da $u_1 = (1, 1, 0, 1)$ e $u_2 = (1, 0, 1, 0)$. Trovare due supplementari distinti di U.

Basta completare la base di U data da $\{u_1, u_2\}$ ad una base di $\mathbb{R}^4$ in due maniere distinte. Questo è equivalente a considerare una matrice 4×4 in cui le prime due righe sono costituite dalle componenti di $\{u_1, u_2\}$ e le ultime due righe sono scelte in modo che la matrice sia non singolare. Una maniera ad esempio è la seguente:

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

quindi $W_1 = \langle (0, 0, 1, 0), (0, 0, 0, 1) \rangle$ è un supplementare. Si faccia attenzione però. Scegliendo ad esempio la matrice

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

abbiamo si una matrice diversa dalla precedente ma lo spazio generato dalle ultime due righe coincide con W_1 e quindi questa non è una soluzione diversa dalla precedente. Se vogliamo effettivamente due supplementari distinti dobbiamo prendere opportunamente altri due vettori.

Ad esempio, la matrice

$$\begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 2 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

fornisce il sottospazio $W_2 = \langle (0,0,1,0),(2,1,1,2) \rangle$ che è effettivamente distinto da W_1.

Esercizio. Rappresentare l'endomorfismo

$$\frac{d}{dt} : \mathbb{R}_4[t] \to \mathbb{R}_4[t]$$

ove $\mathbb{R}_4[t]$ è lo spazio dei polinomi di grado minore di 4 con base

$$1, t, t^2, t^3$$

Calcoliamo le derivate dei polinomi della base:

$$\frac{d}{dt}1 = 0; \quad \frac{d}{dt}t = 1; \quad \frac{d}{dt}t^2 = 2t; \quad \frac{d}{dt}t^3 = 3t^2$$

la matrice allora diventa

$$\mathcal{D} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Osserviamo inoltre che $\frac{d}{dt}$ è un operatore con indice di nilpotenza 4 cioè

$$\mathcal{D}^4 = 0, \mathcal{D}^3 \neq 0$$

Esercizio. Fissata una base $\{e_i\}$ di uno spazio vettoriale V_n arbitrario ma di dimensione n si trovi la matrice di

$$A : V_n \to V_n$$

definita da

$$e_i \mapsto \begin{cases} e_{i-1} \text{ se } i-1 > 0 \\ 0 \text{ se } i-1 = 0 \end{cases}$$

Se A è una matrice $n \times n$ il suo polinomio caratteristico si può scrivere nel modo seguente

$$p_A(t) = c_0(A)t^n + c_1(A)t^{n-1} + c_2(A)t^{n-2} + \cdots + c_n(A)$$

Poiché matrici simili hanno lo stesso polinomio caratteristico dunque anche i coefficienti $c_i(A)$ sono invarianti per similitudine (cambiamento di base). Si verifica inoltre che

$$c_0(A) = (-1)^n,$$
$$c_1(A) = (-1)^{n-1}(a_{11} + a_{22} + \ldots + a_{nn}) = (-1)^{n-1} Tr\ A,$$
$$c_n(A) = det\ A$$

Esercizio

Calcolare esplicitamente il polinomio caratteristico di

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Osservazione. Non tutti gli operatori sono diagonalizzabili, ad esempio

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

non è diagonalizzabile. (Verificare)

Teorema 0.14. *Se U è un sottospazio di V invariante per un endomorfismo L e se indichiamo con L_U la restrizione di L ad U abbiamo che p_{L_U} è un fattore di p_L.*

Dim. Sia $u_1, \dots, u_r$ una base di U e la si completi ad una base di V aggiungendo dei vettori $v_{r+1}, \dots, v_n$. Allora

$$Lu_i = \sum_{j=1}^{r} a_{ji} u_j$$

mentre

$$Lv_i = \sum_{j=1}^{r} a_{ji} u_j + \sum_{j=r+1}^{n} a_{ji} v_j$$

e quindi la matice corrispondente ha la forma

$$A = \begin{pmatrix} A' & T \\ 0 & W \end{pmatrix}$$

(una matrice a blocchi, dove A' è la matrice di L_U). Sviluppando lungo le prime colonne si ha

$$det\ (A - \lambda I_n) = det\ (A' - \lambda I_r) det\ (W - \lambda I)$$

□

Nel caso in cui uno spazio V sia somma diretta di due suoi sottospazi

$$V = U \oplus W$$

ed entrambi i sottospazi sono stabili per L allora la matrice di L in una base adattata, cioè una base in cui i primi $h = dim\ U$ vettori sono in U e i restanti sono in W, ha la forma

$$A = \begin{pmatrix} A' & 0 \\ 0 & A'' \end{pmatrix}$$

In questo caso diremo che l'endomorfismo L è riducibile.

Capitolo I: Forma Canonica di Jordan.

In ciò che segue dimostreremo che per ogni matrice quadrata sui complessi esiste una forma canonica, essenzialmente unica, che, nel caso in cui la matrice sia diagonalizzabile, coincide con la forma diagonale. Si tratta quindi di una generalizzazione del concetto di matrice diagonalizzabile.

Abbiamo bisogno delle seguenti nozioni e definizioni.

Blocco di Jordan. Una matrice della forma

$$J_r(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & \lambda & \end{pmatrix}$$

si chiama un blocco di Jordan di ordine r con autovalore λ.

Ad esempio

$$J_3(5) = \begin{pmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{pmatrix}$$

$$J_5(3) = \begin{pmatrix} 3 & 1 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

Matrice di Jordan. Una matrice di Jordan è una matrice della forma

$$J = \begin{pmatrix} J_{r_1}(\lambda_1) & \dots & \dots & \dots \\ \dots & J_{r_2}(\lambda_2) & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

con zeri all'infuori dei blocchi di Jordan. Useremo anche la notazione

$$J = J_{r_1}(\lambda_1) \oplus J_{r_2}(\lambda_2) \oplus \cdots$$

Ad esempio

$$J = J_3(5) \oplus J_5(3) \oplus J_2(1)$$

è uguale a

$$\begin{pmatrix} 5 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Base di Jordan. Una base di Jordan per l'operatore L su V è una base di V rispetto alla quale la matrice di L è una matrice di Jordan ovvero, come si dice anche, la matrice ha la forma canonica di Jordan.

Esempio Sia $E_n(\lambda)$ lo spazio vettoriale di tutte le funzioni della forma $e^{\lambda x}f(x)$ dove $\lambda \in \mathbb{R}$ e $f(x)$ è un polinomio di grado minore o uguale a n. Ad esempio un tipico elemento di $E_3(\lambda)$ è

$$e^{\lambda x}(a_0 + a_1 x + a_2 x^2 + a_3 x^3)$$

Poiché

$$\frac{d}{dx}(e^{\lambda x}f(x)) = e^{\lambda x}(\lambda f(x) + f'(x))$$

abbiamo che l'operatore $\frac{d}{dx}$ è un operatore lineare in questo spazio. Ponendo

$$e_1 = x^0 e^{\lambda x}, \quad e_2 = x e^{\lambda x}, \quad e_3 = \frac{x^2}{2}e^{\lambda x}$$

abbiamo

$$\frac{d}{dx}(e_1) = \lambda e^{\lambda x}$$

$$\frac{d}{dx}(e_2) = e^{\lambda x} + \lambda x e^{\lambda x}$$

$$\frac{d}{dx}(e_3) = x e^{\lambda x} + \frac{x^2}{2}\lambda e^{\lambda x}$$

Dunque la matrice di questo operatore rispetto a questa base è $J_3(\lambda)$ e quindi

$$e_1 = x^0 e^{\lambda x}, \quad e_2 = x e^{\lambda x}, \quad e_3 = \frac{x^2}{2}e^{\lambda x}$$

è una base di Jordan di $\frac{d}{dx}$.

Se $p(t)$ è un polinomio a coefficienti in $\mathbb{K}$ allora $p(t) = \sum_{i=0}^{n} a_i t^i$ ed ha senso considerare

$$p(L) = \sum_{i=0}^{n} a_i L^i$$

dove per L^0 si intende l'operatore identità. Ad esempio se $p(t) = 3t^2 + 2t + 5$ e $L = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ allora

$$p(L) = 3\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^2 + 2\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} + 5\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 10 & 8 \\ 0 & 10 \end{pmatrix}$$

Nel caso in cui $p(L) = 0$ diremo che il polinomio $p(t)$ annulla l'operatore L.

Se V è uno spazio di dimensione finita allora per un qualunque operatore L esiste sempre un polinomio $p(t)$ che lo annulla. Infatti, se L è un operatore su uno spazio di dimensione n, è ben noto che lo spazio di tutte le matrici di ordine n ha dimensione n^2 (perché?), per cui se prendiamo le $n^2 + 1$ matrici

$$I, L, L^2, \ldots L^{n^2-2}, L^{n^2-1}, L^{n^2}$$

queste non possono essere tutte indipendenti (per questioni di dimensione) ed allora esistono dei coefficienti non tutti nulli tali che

$$a_0 I + a_1 L + \cdots + a_{n^2} L^{n^2} = 0$$

cioè il polinomio

$$a_0 + a_1 t + \cdots + a_{n^2} t^{n^2} = 0$$

annulla L. Questo ci dice che se L è un endomorfismo di uno spazio vettoriale V di dimensione n allora esiste un polinomio di grado minore o uguale a n^2 che annulla L. In effetti dimostreremo il Teorema di Cayley-Hamilton che ci garantisce l'esistenza di un polinomio anullatore di grado al più n (invece di n^2).

Polinomio Minimo. Per ora limitiamoci ad osservare che tra tutti i polinomi che annullano L ce ne saranno alcuni di grado minimo. Tra tutti i polinomi di grado minimo che annullano L ve n'è uno monico. Questo unico polinomio si dice polinomio minimo di L.

Proposizione 1.1. *Ogni polinomio che annulla L ha come fattore il polinomio minimo.*

Dim. Infatti sia $m(t)$ il polinomio minimo di L e sia $p(t)$ un qualunque polinomio che annulla L. Dividiamo $p(t)$ per $m(t)$. Abbiamo

$$p(t) = q(t)m(t) + r(t)$$

(divisione euclidea con il resto) in cui $r(t)$ ha grado strettamente minore di $m(t)$. Allora $r(t) = p(t) - q(t)m(t)$ per cui se calcoliamo $r(L)$ abbiamo

$$r(L) = p(L) - q(L)m(L) = 0$$

perché sia $p(t)$ che $m(t)$ annullano L per ipotesi. Allora $r(t)$ annulla L ma ha grado minore del polinomio minimo. Questo è possibile solo se $r(t)$ è il polinomio nullo. $\square$

Teorema di Cayley-Hamilton 1.2. *Il polinomio caratteristico $p_L(t)$ di un operatore L annulla L:*

$$p_L(L) = 0$$

Esempio $A = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}$ allora $p_A(t) = t^2 - 5$. Per calcolare $p_A(A)$ calcoliamo

$$A^2 = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}$$

e quindi

$$p_A(A) = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} - 5 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 0$$

Daremo due dimostrazioni di questo teorema.

Prima Dimostrazione. Facciamo qualche semplice osservazione preliminare. Una matrice quadrata $A(\lambda)$ i cui elementi sono polinomi in λ è un elemento di $M_n(\mathbb{C}[\lambda])$ ma può essere pensata anche come un polinomio a coefficienti matriciali cioè come elemento di $M_n(\mathbb{C})[\lambda]$ nel modo seguente

$$\begin{pmatrix} 1+\lambda & \lambda^2+\lambda \\ \lambda & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \lambda + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \lambda^2$$

In tal caso si dice che le due strutture sono naturalmente isomorfe: una matrice a coefficienti polinomiali è "la stessa cosa" di un polinomio a coefficienti matriciali.

Sia dato allora il polinomio a coefficienti matriciali

$$A(\lambda) = A_0\lambda^m + A_1\lambda^{m-1} + \ldots + A_m$$

Diremo che il polinomio ha grado m se $A_0 \neq 0$, che il suo ordine è n se le matrici coefficienti sono $n \times n$. Diremo che il polinomio è regolare se il determinante di A_0 è non nullo. Due polinomi dello stesso ordine possono essere sommati o moltiplicati con le solite regole, ricordando però che non essendo il prodotto tra matrici commutativo, non lo sarà neanche il prodotto tra questi polinomi. Converrà osservare anche che, contrariamente a quanto accade per i polinomi a coefficienti costanti, il grado del prodotto tra due polinomi può essere minore della somma dei gradi dei fattori. Il motivo sta nel fatto che il prodotto di due matrici non nulle può essere zero. Per quanto ci riguarda, ci preme sottolineare che è possibile effettuare una divisione tra questi polinomi matriciali esattamente come nel caso solito. Sussite cioè il seguente

Teorema 1.3. *Se $A(\lambda)$ e $B(\lambda)$ sono due polinomi matriciali e se $B(\lambda)$ è regolare allora esite un unico quoziente destro $Q(\lambda)$ ed un unico resto destro $R(\lambda)$ tali che*

$$A(\lambda) = Q(\lambda)B(\lambda) + R(\lambda)$$

Analogamente esiste un unico quoziente sinistro $\widehat{Q(\lambda)}$ ed un unico resto sinistro $\widehat{R(\lambda)}$ tali che

$$A(\lambda) = B(\lambda)\widehat{Q(\lambda)} + \widehat{R(\lambda)}$$

Esempio Calcolare la divisione destra e sinistra di

$$A(\lambda) = \begin{pmatrix} \lambda^3 + \lambda & 2\lambda^3 + \lambda^2 \\ \lambda^3 - 2\lambda^2 + 1 & 3\lambda^3 + \lambda \end{pmatrix}$$

per la matrice regolare

$$B(\lambda) = \begin{pmatrix} 2\lambda^2 + 3 & -\lambda^2 + 1 \\ \lambda^2 - 1 & \lambda^2 + 2 \end{pmatrix}$$

Procediamo in maniera analoga al solito algoritmo di divisione tra polinomi:

$$\begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}\lambda^3 + \begin{pmatrix} 0 & 1 \\ -2 & 0 \end{pmatrix}\lambda^2 + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\lambda + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

diviso

$$\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\lambda^2 + \begin{pmatrix} 3 & 1 \\ -1 & 2 \end{pmatrix}$$

Il primo passo della divisione richiede di "dividere" il primo coefficiente di A per il primo coefficiente di B, nel nostro esempio ciò corrisponde a calcolare

$$\begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 3 & 5 \\ 2 & 5 \end{pmatrix}$$

che sarà il coefficiente direttore del quoziente. Successivamente questo coefficiente (moltiplicato per λ) va moltiplicato per il divisore $B(\lambda)$ ottenendo

$$\begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix} \lambda^3 + \begin{pmatrix} 4 & 13 \\ 1 & 12 \end{pmatrix} \lambda$$

che viene sottratto da $A(\lambda)$. Si ottiene così il polinomio

$$\begin{pmatrix} 0 & 1 \\ -2 & 0 \end{pmatrix} \lambda^2 + \begin{pmatrix} -3 & -13 \\ -1 & -11 \end{pmatrix} \lambda + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

Ripetendo questo procedimento si ottiene il quoziente destro

$$Q(\lambda) = \begin{pmatrix} 3\lambda + 1 & 5\lambda + 2 \\ 2\lambda - 2 & 5\lambda - 2 \end{pmatrix}$$

e il resto destro

$$R(\lambda) = \begin{pmatrix} -3\lambda - 1 & -13\lambda - 5 \\ -\lambda + 5 & -11\lambda + 6 \end{pmatrix}$$

Il lettore è invitato a calcolare il quoziente sinistro $\widehat{Q(\lambda)}$ ed il resto sinistro $\widehat{R(\lambda)}$.

Quando calcoliamo un polinomio matriciale

$$F(\lambda) = F_0\lambda^m + F_1\lambda^{m-1} + \ldots + F_m$$

su una matrice A, abbiamo due possibili risultati: un valore destro $F(A) = F_0A^m + F_1A^{m-1} + \ldots + F_m$ e un valore sinistro $\widehat{F(A)} = A^mF_0 + A^{m-1}F_1 + \ldots + F_m$.

Come nel caso dei polinomi a coefficienti costanti, se dividiamo due polinomi, il grado del resto è minore del grado del divisore, per cui se dividiamo per un binomio di primo grado, cioè $\lambda I - A$ dove I è la matrice identità $n \times n$ e A una matrice costante, il resto sarà una costante. Questo ci permette di concludere che

Proposizione 1.4. *Un polinomio $F(\lambda)$ è divisibile per $\lambda I - A$ a destra (sinistra) se e solo se $F(A) = 0$ ($\widehat{F(A)} = 0$).*

Ricordiamo che se A è una matrice costante il suo polinomio caratteristico è definito da $\Delta(\lambda) = det(A - \lambda I)$. Ricordiamo che se $B(\lambda)$ è la matrice aggiunta di $(A - \lambda I)$ allora abbiamo le seguenti identità:

$$(A - \lambda I)B(\lambda) = \Delta(\lambda)I$$
$$B(\lambda)(A - \lambda I) = \Delta(\lambda)I$$

In altre parole $\Delta(\lambda)I$ è divisibile sia a sinistra che a destra per $A - \lambda I$. Per la Proposizione ciò implica che $\Delta(A)I = \Delta(A)$ è la matrice nulla.

$\square$

Seconda Dimostrazione. Questa seconda dimostrazione si basa su un calcolo esplicito.

Consideriamo la matrice $(A - tI)$ e la sua matrice aggiunta $Agg(A - tI)$. Allora sappiamo che

$$(A-tI)Agg(A-tI) = det\,(A-tI)I = (c_0(A)t^n + c_1(A)t^{n-1} + c_2(A)t^{n-2} + \ldots c_n(A))I$$

Osserviamo che $Agg(A - tI)$ è un polinomio di grado minore o uguale a $n - 1$ (ricordiamo infatti che nel calcolo della matrice aggiunta ad ogni passo si "cancella" una riga ed una colonna della matrice $A - tI$). Quindi esistono delle matrici B_i tali che

$$Agg(A - tI) = B_0 + B_1 t + \cdots + B_{n-1} t^{n-1}$$

e sostituendo

$$(A - tI)(B_0 + B_1 t + \cdots + b_{n-1} t^{n-1}) = (c_0(A)t^n + c_1(A)t^{n-1} + c_2(A)t^{n-2} + \ldots c_n(A))I$$

e sviluppando

$$\begin{aligned} AB_0 + AB_1 t + \cdots + AB_{n-1} t^{n-1} - B_0 t - B_1 t^2 - \cdots - B_{n-1} t^n \\ = (c_0(A)t^n + c_1(A)t^{n-1} + c_2(A)t^{n-2} + \ldots c_n(A))I \end{aligned}$$

Uguagliando i coefficienti abbiamo

$$AB_0 = c_n(A)I$$

$$AB_1 - B_0 = c_{n-1}(A)I$$

$$\vdots$$

$$AB_{n-1} - B_{n-2} = c_1(A)I$$

$$-B_{n-1} = c_0(A)I$$

Moltiplichiamo queste relazioni a sinistra per $I, A, A^2, \ldots, A^{n-1}, A^n$. Otteniamo

$$AB_0 = c_n(A)I$$

$$A(AB_1 - B_0) = c_{n-1}(A)A$$

$$A^2(AB_2 - B_1) = c_{n-2}(A)A^2$$

$$\vdots$$

$$A^{n-1}(AB_{n-1} - B_{n-2}) = c_1(A)A^{n-1}$$

$$-A^n B_{n-1} = c_0(A)A^n$$

Sommando i termini a primo membro abbiamo

$$AB_0 + A(AB_1 - B_0) + A^2(AB_2 - B_1) + \cdots + A^{n-1}(AB_{n-1} - B_{n-2}) + A^n B_{n-1} = 0$$

in quanto tutti i termini si cancellano a due a due. A secondo membro otteniamo

$$c_n(A)I + c_{n-1}(A)A + c_{n-2}(A)A^2 + \cdots + c_1(A)A^{n-1} + c_0(A)A^n$$

quindi $p_A(A) = 0$ $\square$

Corollario 1.5. *$A \in M_n(\mathbb{R})$ è una matrice nilpotente se e solo se*

$$p_A(t) = (-1)^n t^n$$

Dim. Se $p_A(t) = (-1)^n t^n$ allora il Teorema di Cayley-Hamilton ci dice che $p_A(A) = (-1)^n A^n = 0$ ossia $A^n = 0$ e quindi A è nilpotente per definizione.

Viceversa, se esiste un intero $k \geq 1$ tale che $A^k = 0$ allora per prima cosa 0 è un autovalore di A. Infatti, per il teorema di Binet

$$A^k = 0 \implies det\ (A^k) = 0 \implies (det\ A)^k = 0 \implies det\ A = 0$$

D'altra parte A non possiede autovalori diversi da 0. Infatti se $\lambda \in \mathbb{C}$ è un autovalore di A si ha $Ax = \lambda x$ per qualche $x \neq 0$ ma allora $A^k x = \lambda^k x$ e quindi $\lambda^k x = 0$. Essendo $x \neq 0$ segue che $\lambda^k = 0$ e quindi $\lambda = 0$. □

Il teorema ci fornisce anche una nuova maniera di calcolare l'inversa di una matrice.

Corollario 1.6. *Se $A \in M_n(\mathbb{R})$ è una matrice invertibile e*

$$p_A(t) = (-1)^n t^n + a_{n-1} t^{n-1} + \cdots a_1 t + a_0$$

è il suo polinomio caratteristico dove $a_0 \neq 0$ allora risulta

$$A^{-1} = -\frac{1}{a_0}((-1)^n A^{n-1} + a_{n-1} A^{n-2} + \cdots + a_2 A + a_1 I)$$

Dim.

$$p_A(A) = (-1)^n A^n + a_{n-1} A^{n-1} + \cdots a_1 A + a_0 I = 0$$

da cui

$$I = -\frac{1}{a_0}((-1)^n A^n + a_{n-1} A^{n-1} + \cdots + a_2 A^2 + a_1 A)$$

e moltiplicando ambo i membri per A^{-1} si ottiene il risultato. □

Calcolo del polinomio minimo in due semplici casi.

Se L è l'operatore identità su uno spazio di dimensione n allora il polinomio caratteristico di L è $(1-t)^n$ mentre il polinomio minimo è chiaramente $t-1$ (infatti $I - I = 0$, il polinomio $t-1$ è monico e nessun polinomio di grado inferiore annulla I).

Se L è un operatore rappresentato dal blocco di Jordan $J_r(\lambda)$ il suo polinomio caratteristico è $(t-\lambda)^r$. Sappiamo allora che il polinomio minimo, essendo un divisore di questo deve essere della forma $(t-\lambda)^k$ con k intero e $0 \leq k \leq r$. Osserviamo ora che, poiché

$$J_r(\lambda) = \lambda I_r + J_r(0)$$

nel calcolare il polinomio $(t-\lambda)^k$ su $J_r(\lambda)$ abbiamo

$$(J_r(\lambda) - \lambda I_r)^k = J_r(0)^k$$

ma poiché $J_r(0)^k \neq 0$ se $k < r$ ne segue che in questo caso il polinomio minimo coincide con il polinomio caratteristico.

Esempio

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

Calcoliamone l'inversa.

$$p_A(t) = -t^3 + 3t^2 - 3t + 2$$

da cui

$$A^{-1} = -\frac{1}{2}(-A^2 + 3A - 3I)$$

cioè

$$\frac{1}{2}\left[\begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 1 \end{pmatrix} - \begin{pmatrix} 3 & 3 & 0 \\ 0 & 3 & 3 \\ 3 & 0 & 3 \end{pmatrix} + \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}\right] =$$

$$= \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Esempio
Consideriamo

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

Calcoliamo

$$A^{2005}$$

Poiché $p_A(t) = t^4 - 1$ abbiamo $A^4 = I$ da cui $A^{2005} = A^{4 \cdot 501 + 1} = A$.

Vettore radicale o autovettore generalizzato. Un vettore $v \in V$ si dice vettore radicale corrispondente a λ per l'operatore L se esiste un intero r tale che

$$(L - \lambda I)^r v = 0$$

Chiaramente quando $r = 1$ abbiamo autovettori, in altre parole gli autovettori sono vettori radicali. Per questo motivo i vettori radicali sono anche detti autovettori generalizzati (di rango r).

Proposizione 1.10. *Denotiamo $V(\lambda)$ l'insieme degli autovettori generalizzati rispetto a λ. Allora $V(\lambda)$ è un sottospazio vettoriale di V e $V(\lambda) \neq 0$ se e solo se λ è un autovalore di L.*

Dimostrazione. Supponiamo che $v_1, v_2 \in V(\lambda)$, diciamo

$$(L - \lambda I)^{r_1} v_1 = 0$$

$$(L - \lambda I)^{r_2} v_2 = 0$$

allora se $r = max(r_1, r_2)$ abbiamo che

$$(L - \lambda I)^r (v_1 + v_2) = (L - \lambda I)^r v_1 + (L - \lambda I)^r v_2 = 0$$

e quindi $v_1 + v_2 \in V(\lambda)$. Inoltre se $k \in \mathbb{K}$

$$(L - \lambda I)^r (kv) = k(L - \lambda I)^r v = 0$$

quindi $V(\lambda)$ è un sottospazio.

Se λ è un autovalore di L allora esiste un autovettore v relativo a λ e quindi $V(\lambda) \neq 0$. Viceversa, sia $v \in V(\lambda)$, $v \neq 0$ e prendiamo il minimo r per cui

$$(L - \lambda I)^r v = 0$$

Ovviamente $r \geq 1$. Se $r = 1$ allora λ è un autovalore. Se $r \geq 2$ consideriamo

$$v' = (L - \lambda I)^{r-1} v \neq 0$$

e prendiamo

$$(L - \lambda I)v' = (L - \lambda I)^r v = 0$$

allora v' è un autovettore per λ e dunque λ è un autovalore. $\square$

Esercizio: Dimostrare che $V(\lambda)$ è invariante per L.

Ci proponiamo di dimostrare il seguente importante Teorema di decomposizione:

Teorema di Decomposizione 1.11.

$$V = \bigoplus V(\lambda_i)$$

dove λ_i *varia tra tutti gli autovalori di* L

Corollario 1.12. *Se lo spettro di un operatore* L *è semplice allora* L *è diagonalizzabile.*

Dim. Uno spettro si dice semplice se è costituito da autovalori distinti. In tal caso il numero di autovalori distinti di L è proprio $n = dim\ V$. Dunque nella decomposizione

$$V = \bigoplus V(\lambda_i)$$

tutti gli spazi $V(\lambda_i)$ hanno dimensione 1 e poiché ciascuno di essi contiene un autovettore, l'operatore L possiede una base di autovettori ed è quindi diagonalizzabile. $\square$

Diamo ora la dimostrazione che si basa sulla decomposizione di Fitting.

Decomposizione di Fitting.

Se V è uno spazio di dimensione finita, come supporremo d'ora in poi salvo avviso del contrario, sappiamo che per ogni endomorfismo f si ha $dim\ ker\ f + dim\ Im\ f = dimV$. Non è però vero, in generale, che $ker\ f \oplus Im\ f = V$. Tuttavia dimostreremo che se prendiamo una potenza opportuna di f, diciamo f^q, allora si avrà $V = ker\ f^q \oplus Im\ f^q$. Questa si chiama decomposizione di Fitting di V. Il nostro scopo attuale è dimostrare che questa decomposizione esiste.

Cominciamo con l'osservare i seguenti fatti.

Proposizione 1.13. *Se* $i \in \mathbb{N}$ *è un intero maggiore o uguale a 1 allora*

$$ker\ f^i \subset ker\ f^{i+1}$$

$$Im\ f^i \supset Im\ f^{i+1}$$

Dimostrazione. Esercizio. $\square$

Dunque calcolando le potenze di un endomorfismo si ottengono due successioni di sottospazi: una crescente, l'altra decrescente. È inoltre anche un esercizio verificare che questi sottospazi sono invarianti per f. Essendo la dimensione dello spazio V finita queste successioni devono ad un certo punto stabilizzarsi, cioè esisterà una potenza f^h per cui $ker\ f^h = ker\ f^{h+1}$ e una potenza f^k per cui $Im\ f^k = Im\ f^{k+1}$. Dimostreremo che in effetti $h = k$, cioè quando si stabilizza una successione si stabilizza necessariamente anche l'altra. Non solo: ma per quell'esponente particolare che indicheremo con q, avremo la decomposizione di Fitting desiderata.

Il fatto che le due successioni si stabilizzano simultaneamente, cioè $h = k = q$, segue dal Teorema 0.4. È inoltre facile verificare che se $ker\ f^q = ker\ f^{q+1}$ allora $ker\ f^i = ker\ f^{i+1}$ per ogni $i \geq q$. Dimostriamo ora il seguente

Teorema 1.14: Decomposizione di Fitting. *Lo spazio vettoriale V si decompone nella somma diretta*

$$V = ker\ f^q \oplus Im\ f^q$$

se e solo se $ker\ f^q = ker\ f^{q+1}$. Queste condizioni sono anche equivalenti al fatto che la restrizione di f a $Im\ f^q$ è un isomorfismo.

Dimostrazione. Supponiamo che $ker\ f^q = ker\ f^{q+1}$, vogliamo dimostrare che $V = ker\ f^q \oplus Im\ f^q$. Per il Teorema 0.4, basta verificare che l'intersezione $ker\ f^q \cap Im\ f^q = \{0\}$. Sia dunque $x \in ker\ f^q \cap Im\ f^q = \{0\}$. Esiste allora $y \in V$ tale che $f^q(y) = x$ essendo $x \in Im\ f^q$. Allora $0 = f^q(x) = f^{2q}(y)$, essendo anche $x \in ker\ f^q$. Dunque $y \in ker\ f^{2q}$ Ma poiché la successione dei nuclei e delle immagini è stazionaria dopo il passo q, $ker\ f^{2q} = ker\ f^q$. Quindi $y \in ker\ f^{2q} = ker\ f^q$, cioè $x = f^q(y) = 0$, come desiderato.

Viceversa, supponiamo che $V = ker\ f^q \oplus Im\ f^q$. Vogliamo dimostrare che allora le successioni sono stazionarie al passo q. Poiché $ker\ f^q \subset ker\ f^{q+1}$ in generale, per dimostrare l'uguaglianza dobbiamo solo dimostrare l'inclusione inversa. Sia quindi $x \in ker f^{q+1}$. Ciò comporta che $0 = f^{q+1}(x) = f(f^q(x))$ ossia $f^q(x) \in ker\ f \subset ker\ f^q$. D'altra parte è ovvio dalla definizione di immagine che $f^q(x) \in Im\ f^q$ quindi $f^q(x) \in ker\ f^q \cap Im\ f^q = \{0\}$ quindi necessariamente $f^q(x) = 0$ cioè $x \in ker\ f^q$, e l'inclusione inversa è dimostrata.

Rimane ora da dimostrare che queste condizioni sono equivalenti al fatto che la restrizione di f a $Im\ f^q$ è un isomorfismo. Chiamiamo $\bar{f}$ questa restrizione. Supponiamo dunque che

$$V = ker\ f^q \oplus Im\ f^q$$

ovvero, equivalentemente, che $ker\ f^q = ker\ f^{q+1}$.

Essendo $\bar{f}$ un endomorfismo è sufficiente verificare che $\bar{f}$ è iniettivo. Se $x \in ker\ \bar{f}$ allora $x \in ker\ f$ e quindi anche in $ker\ f^q$. Ma allora $x \in ker\ f^q \cap Im\ f^q = \{0\}$, cioè $x = 0$ quindi $\bar{f}$ è iniettivo e quindi è un isomorfismo.

Viceversa, se $\bar{f}$ è un isomorfismo, $Im\ f^q \simeq Im\ f^{q+1}$ e poichè $Im\ f^q \supset Im\ f^{q+1}$ essi devono coincidere. Quindi la successione delle immagini è stazionaria e quindi abbiamo la conclusione per la prima parte della dimostrazione. □

Definizione Chiameremo *indice* di un endomorfismo f (o di una matrice quadrata A) il minimo intero non negativo q per cui vale la decomposizione di Fitting. Se una matrice è non singolare il suo indice è zero. Se una matrice è nilpotente il suo indice coincide con il suo indice di nilpotenza. Se una matrice ha autovalore λ diremo che λ ha indice k se k è l'indice della matrice $A - \lambda I$.

Dimostrazione del Teorema di decomposizione. Sia allora f un endomorfismo e siano $\lambda_1, \dots \lambda_k$ gli autovalori di f. Sia $V = V(\lambda_1) \oplus V'(\lambda_1)$ la decomposizione di Fitting di V relativa a $f - \lambda_1 I$. Si verifica che $V(\lambda_1)$ coincide con l'autospazio generalizzato relativo a λ_1. Rispetto ad una base adattata la matrice di f diventa una matrice a blocchi:

$$\begin{pmatrix} A_1 & 0 \\ 0 & B \end{pmatrix}$$

dove A_1 è la matrice della restrizione di f a $V(\lambda_1)$ mentre B è la matrice di f ristretta a $V'(\lambda_1)$. B ha autovalori $\lambda_2, \dots, \lambda_k$, infatti gli autovalori di B sono anche autovalori di A e λ_1 non può essere autovalore di B. Dopo un numero finito di iterazioni questo procedimento porta a decomporre V nella somma diretta dei suoi autospazi generalizzati. □

Forma canonica di Jordan: Caso Nilpotente.

Vogliamo studiare gli operatori lineari di uno spazio vettoriale V. La classe di operatori più semplice è quella degli operatori diagonalizzabili. Ci sono varie ragioni che impediscono ad un operatore di essere diagonalizzabile. Un motivo può essere che il polinomio caratteristico non ha abbastanza radici in $\mathbb{K}$. Ad esempio la matrice

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

ha polinomio caratteristico t^2+1 e dunque non ha autovalori reali. L'operatore non è quindi diagonalizzabile come operatore reale. Geometricamente questo è ovvio in quanto la matrice in questione rappresenta la rotazione del piano xy di 90 gradi in senso antiorario e quindi nessun vettore rimane parallelo a se stesso (che non sia nullo). Per evitare questo tipo di problemi supponiamo che il nostro campo sia algebricamente chiuso, cioè supponiamo che ogni polinomio a coefficienti in $\mathbb{K}$ abbia una radice in $\mathbb{K}$. Possiamo pensare che $\mathbb{K} = \mathbb{C}$. In questo caso, qualunque L possiede sicuramente un autovalore, tuttavia potrebbe non essere ancora diagonalizzabile. Ad esempio, la matrice vista in precedenza:

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

Per questo operatore abbiamo $p_A(t) = (1-t)^2$ quindi l'unico autovalore di A è $t = 1$ con molteplicità algebrica 2. Per calcolare la sua molteplicità geometrica dobbiamo calcolare l'autospazio relativo a questo autovalore. Denotiamolo V_1. Dobbiamo risolvere il sistema

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$$

ovvero $(A - \lambda I)v = 0$ cioè

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

quindi $y = 0$. Si tratta di una retta in $\mathbb{R}^2$ e quindi ha dimensione 1. Ciò significa che la molteplicità geometrica è 1. Poiché la molteplicità geometrica è diversa dalla molteplicità algebrica questo operatore non potrà essere diagonalizzabile in quanto non sarà possibile trovare una base di autovettori.

In un certo senso, l'ostruzione essenziale alla diagonalizzazione delle matrici risiede nell'esistenza di matrici *nilpotenti*, cioè, per definizione, matrici una cui potenza è nulla. Tali matrici sono, per così dire, agli "antipodi" delle matrici diagonalizzabili.

Mostreremo che ogni endomorfismo nilpotente è rappresentabile tramite una matrice di Jordan in una base opportuna. In secondo luogo vedremo che il teorema è vero anche per gli endomorfismi che possiedono un unico autovalore di molteplicità uguale alla dimensione dello spazio, riducendoci al caso nilpotente. Infine, dimostreremo il teorema nel caso generale, riducendoci al caso di endomorfismi con un solo autovalore.

Nel caso in cui un operatore abbia il solo autovalore nullo, osserviamo che una base di Jordan è composta da un insieme di vettori

$$\mathcal{B} = \{v_1, \dots, v_{h_1}, v_{h_1+1}, \dots, v_{h_1+h_2}, v_{h_1+h_2+1}, \dots, v_n\}$$

con la proprietà che

$$\begin{array}{ccc} v_1 \mapsto 0 & v_{h_1+1} \mapsto 0 & v_{h_1+h_2+1} \mapsto 0 \\ v_2 \mapsto v_1 & v_{h_1+2} \mapsto v_{h_1+1} & \vdots \\ \cdots & \cdots & \cdots \\ v_{h_1} \mapsto v_{h_1-1} & v_{h_1+h_2} \mapsto v_{h_1+h_2-1} & v_n \mapsto v_{n-1} \end{array}$$

Definiamo allora un diagramma associato a questa base prendendo tanti punti, indicati nel testo graficamente con asterischi, quanti sono i vettori della base e collegati tramite delle frecce verticali che indicano la trasformazione che li lega: Ad esempio, il diagramma che segue illustra il caso di un endomorfismo con il solo autovalore nullo in uno spazio di dimensione 38.

$$\begin{array}{ccccccccccccc}
* & * & * & * & * & * & & & & & & & \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & & & & & & & \\
* & * & * & * & * & * & * & * & * & & & & \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & & & & \\
* & * & * & * & * & * & * & * & * & * & & & \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & & & \\
* & * & * & * & * & * & * & * & * & * & * & * & *
\end{array}$$

Ogni punto del diagramma rappresenta un vettore di una base di Jordan che viene trasformato, tramite l'operatore nel vettore corrispondente al punto più in basso come indicato dalla freccia. L'ultima riga, quella "terminale", più in basso, rappresenta una base del nucleo dell'operatore.

Consideriamo ad esempio un caso molto speciale. Immaginiamo di avere un operatore T il quale su una qualche base $\{v_1, \dots, v_n\}$ di uno spazio vettoriale V agisce nella maniera seguente:

$$Tv_n = v_{n-1}, Tv_{n-1} = v_{n-2}, \dots, Tv_2 = v_1, Tv_1 = 0$$

(Si pensi, ad esempio, ad un operatore di derivazione $\frac{d}{dx}$ che agisce sui polinomi $\{x^n, nx^{n-1}, n(n-1)x^{n-2}, \dots\}$). L'azione di questo operatore, sicuramente nilpotente, può essere riassunta con un diagramma:

$$\begin{array}{c}
* \\ \downarrow \\ * \\ \downarrow \\ * \\ \downarrow \\ \vdots \\ * \\ \downarrow \\ *
\end{array}$$

Supponiamo per il momento che L possieda un unico autovalore e, per di più, supponiamo che questo autovalore sia nullo. In altre parole, supponiamo che l'operatore L sia nilpotente con polinomio caratteristico $P(t) = t^n$. Abbiamo il

Teorema 1.15. *Un operatore nilpotente L su uno spazio vettoriale V di dimensione finita n possiede una base di Jordan e quindi la sua matrice in questa base è una matrice di Jordan.*

Per dare una dimostrazione più facilmente comprensibile ad una prima lettura ci permetteremo alcune libertà di linguaggio.

Ricordiamoci della Proposizione 1.13 e che quindi calcolando le potenze di un endomorfismo si ottengono due successioni di sottospazi: una crescente, l'altra decrescente. Supponiamo quindi, per concretezza, di avere un endomorfismo nilpotente su uno spazio di dimensione 38 corrispondente al diagramma visto prima che qui ripetiamo:

$$\begin{array}{ccccccccccccc}
* & * & * & * & * & * & & & & & & & \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & & & & & & & \\
* & * & * & * & * & * & * & * & * & & & & \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & & & & \\
* & * & * & * & * & * & * & * & * & * & & & \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & & & \\
* & * & * & * & * & * & * & * & * & * & * & * & *
\end{array}$$

Consideriamo la successione di sottospazi come segue: $K_0 = Ker\ T, K_1 = Ker\ T \cap Im\ T$ e in generale $K_i = Ker\ T \cap Im\ T^i$. Questo sono sottospazi del nucleo di T e sono uno dentro l'altro. Nel nostro diagramma abbiamo che K_0 è lo spazio generato dai vettori della riga più in basso del diagramma. Nella fattispecie ha dimensione 13. K_1 è generato dai primi 10 vettori a sinistra di quei 13 vettori. K_2 è generato dai primi 9, K_3 generato dai primi 6, $K_4 = \{0\}$, infatti l'indice di nilpotenza di T è 4.

Vogliamo ora dimostrare che ogni operatore nilpotente possiede un diagramma di questo tipo. Inoltre, tale diagramma corrisponde ad una base di Jordan. Di conseguenza avremo dimostrato il teorema.

Faremo prima il ragionamento in ipotesi restrittive per semplificare le notazioni. Cominciamo prendendo il più piccolo sottospazio non nullo della successione K_i definita sopra. Chiamiamolo K_{i_0}. Supponiamo che abbia dimensione 2. Troviamone una base, diciamo $\{u_1^1, u_2^1\}$. Passiamo poi a considerare il precedente sottospazio della successione, quello strettamente più grande. Chiamiamolo K_{i_1}. Estendiamo la base trovata prima ad una base di tutto K_{i_1}. Sia essa, per semplicità, $\{u_1^1, u_2^1, u_3^1, u_4^1, u_5^1\}$. Dopo un numero finito di passi arriveremo all'intero nucleo K_0. Estendiamo ancora la base, sia essa $\{u_1^1, u_2^1, u_3^1, u_4^1, u_5^1, u_6^1, u_7^1\}$. (Nel diagramma di sopra abbiamo K_3 di dimensione 6, che è il più piccolo, poi K_2 di dimensione 9, quindi K_1 di dimensione 10, infine K_0 di dimensione 13.) In corrispondenza di ciascuno dei vettori trovati, che costituiscono una base di K_0, determiniamo ora una stringa di vettori che "giacciono sopra" essi. Consideriamo, per iniziare, u_1^1. Essendo u_1^1 un vettore di K_{i_0} e dunque di $Im\ T^{i_0}$, esisterà un vettore, chiamiamolo $u_1^{i_0+1}$ tale che $T^{i_0}(u_1^{i_0+1}) = u_1^1$. Supponiamo per esempio che $i_0 = 3$ allora $T^3(u_1^4) = u_1^1$. Questo è il vettore più in alto della stringa sopra u_1^1. Gli altri vettori, "ai piani intermedi", sono ottenuti applicando T varie volte. Ripetendo questo ragionamento per ciascuno dei vettori della base del nucleo si trova un insieme di vettori che sono rappresentati da un diagramma a punti come sopra.

Ci rimane da dimostrare ora che tale insieme di vettori costituisce una base dello spazio V.

Essi sono indipendenti. Infatti immaginiamo di avere i nostri vettori schierati come segue

$$\begin{array}{ccccc}
u_1^4 & u_2^4 & u_3^4 & & \\
u_1^3 & u_2^3 & u_3^3 & u_4^3 & \\
u_1^2 & u_2^2 & u_3^2 & u_4^2 & \\
u_1^1 & u_2^1 & u_3^1 & u_4^1 & u_5^1
\end{array}$$

e consideriamo una combinazione lineare di tutti questi vettori con opportuni coefficienti $a_i^j \in \mathbb{C}$,

$$\sum_{i,j} a_i^j u_i^j = 0 \tag{1}$$

Applichiamo a questa combinazione lineare una opportuna potenza di T. Nel nostro esempio basta prendere T^3. Allora

$$T^3(\sum_{i,j} a_i^j u_i^j) = 0 \tag{2}$$

Quando si applica T^3 ai vettori di questa combinazione lineare quasi tutti vengono annullati tranne quelli di "livello" più alto che vengono invece trasformati in vettori di primo livello. La (2) diventa allora

$$a_1^4 u_1^1 + a_2^4 u_2^1 + a_3^4 u_3^1 = 0$$

Poiché i vettori di livello 1 sono stati scelti come facenti parte di una base di K_i, essi sono linearmente indipendenti. Quindi i coefficienti $a_1^4 = a_2^4 = a_3^4 = 0$. Ciò significa allora che nella combinazione lineare (1) i vettori di livello 4 non appaiono. Applichiamo allora a (1) l'operatore T^2. Con analogo ragionamento si dimostra che allora anche i coefficienti dei vettori di livello 3 sono nulli. Dopo un numero finito di passi troviamo che nella (1) solo i coefficienti dei vettori di livello 1 potrebbero essere non nulli. Ma in tal caso la (1) sarebbe una combinazione lineare tra i vettori di livello 1, ma questi costituiscono una base. Quindi anche questi coefficienti sono nulli. In definitiva abbiamo dimostrato che la (1) è possibile solo con coefficienti tutti nulli, cioè i vettori sono linearmente indipendenti.

Dimostriamo ora che essi sono dei generatori.

Prendiamo un vettore qualunque $v \in V$. Essendo T nilpotente sarà certamente possibile trovare una opportuna potenza di T che annulla v. Potrebbe essere $T(v) = 0$ se v è nel nucleo di T, oppure qualche T^h con $h > 1$. Supponiamo ad esempio che $T^3(v) = 0$, mentre $T^2(v) \neq 0$. Allora $T^2(v) \in Ker\ T$ Ma allora $T^2(v) = \sum_i a_i^1 u_i^1$. Consideriamo allora $T^2(v - \sum_i a_i^1 u_i^3)$. Avremo

$$T^2(v - \sum_i a_i^1 u_i^3) = T^2(v) - T^2(\sum_i a_i^1 u_i^3) =$$
$$T^2(v) - \sum_i a_i^1 T^2(u_i^3) = T^2(v) - \sum_i a_i^1 u_i^1 = 0$$

In altre parole, $T^2(v - \sum_i a_i^1 u_i^3) = 0$. Posto $w = v - \sum_i a_i^1 u_i^3$ abbiamo allora un vettore per cui $T^2(w) = 0$ mentre $T(w) \neq 0$. Ma allora, ripetiamo il ragionamento: $T(w) \in Ker\,T$. $T(w) = \sum_i b_i^1 u_i^1$ e quindi $T(w - \sum_i b_i^1 u_i^2) = 0$. Quindi $w - \sum_i b_i^1 u_i^2 \in Ker\ T$. In definitiva,

$$w - \sum_i b_i^1 u_i^2 = \sum_i c_i^1 u_i^1$$

Quindi

$$w = \sum_i b_i^1 u_i^2 + \sum_i c_i^1 u_i^1$$

e

$$v - \sum_i a_i^1 u_i^3 = \sum_i b_i^1 u_i^2 + \sum_i c_i^1 u_i^1$$

cioè

$$v = \sum_i a_i^1 u_i^3 + \sum_i b_i^1 u_i^2 + \sum_i c_i^1 u_i^1$$

ossia v è combinazione lineare dei vettori dati. Questo conclude la dimostrazione.

Forma canonica di Jordan: caso generale.
Per il caso generale resta poco da fare. Procederemo in questo modo:

(1) Se l'endomorfismo f ha un unico autovalore λ di molteplicità algebrica pari alla dimensione n dello spazio, il suo polinomio caratteristico è allora $p_f(t) = (t-\lambda)^n$. Per il Teorema di Cayley-Hamilton allora $(f-\lambda I)^n = 0$, ossia $f-\lambda I$ è un endomorfismo nilpotente. Esiste quindi una base di Jordan per essa. È facile vedere allora che questa stessa base è una base di Jordan anche per A.

(2) Nel caso generale, cioè quello di un operatore avente vari autovalori con varie molteplicità algebriche, a causa del Teorema di Decomposizione, lo spazio vettoriale V si decompone nella somma diretta di sottospazi stabili, rispetto ai quali la restrizione dell'operatore possiede un unico autovalore come nel punto precedente. Ciscuna delle componenti possiede una base di Jordan e quindi l'intero spazio ha una base di Jordan

Teorema 1.16. *Un operatore lineare L su uno spazio vettoriale complesso V di dimensione finita n possiede una base di Jordan e quindi la sua matrice in questa base è una matrice di Jordan.*

□

Osservazione Data una matrice A si può dimostrare che l'indice dell'autovalore λ coincide con l'esponente del fattore $(t-\lambda)$ del polinomio minimo di A.

Alcuni esempi.
Esempio 1.

$$A = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}$$

Il suo polinomio caratteristico è t^2 dunque si tratta di una matrice nilpotente. Il rango di A è chiaramente 1 e quindi la dimensione del nucleo è $2 - 1 = 1$. Ciò significa che c'è un unico blocco di Jordan che deve quindi essere necessariamente

$$J_2(0) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

Per determinare esplicitamente una base di Jordan:
1. Scegliamo un supplementare di $ker\ L$. $ker\ L$ è il sottospazio delle soluzioni del sistema lineare omogeneo (SLO)

$$\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

e dunque l'equazione cartesiana del nucleo è $x - y = 0$. Un suo supplementare è, per esempio, l'asse delle y con base $e_2 = (0, 1)$.
2. A partire da tale vettore, applichiamo L per ottenere un secondo vettore $v_1 = L(e_2)$, cioè

$$\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix} = -e_1 - e_2$$

e poniamo $v_2 = e_2$;
3. I vettori cosí ottenuti, $\{v_1, v_2\}$, sono una base di $\mathbb{R}^2$ ed in questa base abbiamo la matrice:

$$J_2(0) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

poiché $L(v_1) = 0$ e $L(v_2) = v_1$.

Si può riassumere la costruzione con un diagramma composto da due punti (asterischi) ciascuno dei quali rappresenta un vettore della base di Jordan. I punti più in basso rappresentano i vettori della base che sono nel nucleo dell'operatore. Sopra ci sono quelli del nucleo del quadrato dell'operatore (ma che non sono nel nucleo dell'operatore).

$$\begin{array}{c} * \\ \downarrow \\ * \end{array}$$

dove $v_2 \in ker\ L^2 - ker\ L, \quad v_1 \in ker\ L$.
Esempio In $\mathbb{R}^5$

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

Ha polinomio caratteristico $-t^5$ ed è dunque nilpotente. Calcoliamo l'indice di nilpotenza, che ci fornisce la dimensione massima dei blocchi di Jordan, come segue.

$$e_1 \mapsto -e_2 + 3e_3 + 2e_5 \mapsto 3(2e_2) + 2e_2 = 8e_2 \mapsto 0$$

$$e_2 \mapsto 0$$

$$e_3 \mapsto 2e_2 \mapsto 0$$

$$e_4 \mapsto e_2 + e_3 \mapsto 2e_2 \mapsto 0$$

$$e_5 \mapsto e_2 \mapsto 0$$

dunque $A^3 = 0$ mentre $A^2 \neq 0$, l'indice di nilpotenza è 3 e questo è dunque il massimo ordine dei blocchi di Jordan. Andiamo a studiare la successione di sottospazi del nucleo:

$$K_0 = ker\ L$$

$$K_1 = ker\ L \cap Im\ L$$

$$K_2 = ker\ L \cap Im\ L^2$$

e così via, sottospazi sempre più piccoli.

Il nucleo $ker\ L$ è l'insieme delle soluzioni del SLO $AX = 0$ e poiché $rg(A) = 3$ il teorema di Rouché-Capelli ci dice che la molteplicità geometrica dell'autovalore è $dim\ ker\ L = 2$. Precisamente le equazioni di $ker\ L$ sono

$$\begin{cases} x_1 = 0 \\ x_4 = 0 \\ 2x_3 + x_5 = 0 \end{cases}$$

ed una base è data da

$$\{(0, 1, 0, 0, 0), (0, 0, 1, 0, -2)\}$$

Questi due vettori rappresentano i due punti più in basso del diagramma a punti, rappresentano anche il numero dei blocchi di Jordan. Essendo l'indice di nilpotenza 3 possiamo senz'altro scrivere il diagramma

$$\begin{array}{cc} * & \\ \downarrow & \\ * & * \\ \downarrow & \downarrow \\ * & * \end{array}$$

e quindi la relativa forma di Jordan è

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = J_3(0) \oplus J_2(0)$$

Per calcolare esplicitamente l'intera base di Jordan occorre continuare il calcolo. Posto

$$v_1 = (0, 1, 0, 0, 0), v_4 = (0, 0, 1, 0, -2)\}$$

(il motivo della numerazione sarà chiaro fra poco) Cerchiamo ora dei vettori v_2 tale che $Av_2 = v_1$ e v_5 tale che $Av_5 = v_4$. Per far ciò occorre risolvere i seguenti sistemi

$$AX = v_1$$

ossia
$$\begin{cases} -x_1 + 2x_3 + x_4 + x_5 = 1 \\ 3x_1 + x_4 = 0 \\ 2x_1 = 0 \end{cases}$$
una soluzione è ad esempio
$$v_2 = (0, 0, 1, 0, -1)$$
e
$$AX = v_4$$
ossia
$$\begin{cases} -x_1 + 2x_3 + x_4 + x_5 = 0 \\ 3x_1 + x_4 = 1 \\ 2x_1 = -2 \end{cases}$$
una soluzione è ad esempio
$$v_3 = (-1, 0, 1, 4, -7)$$
Infine
$$AX = v_2$$
ossia
$$\begin{cases} -x_1 + 2x_3 + x_4 + x_5 = 0 \\ 3x_1 + x_4 = 1 \\ 2x_1 = -1 \end{cases}$$
una soluzione è ad esempio
$$v_3 = (-\frac{1}{2}, 0, 0, \frac{5}{2}, -3)$$
Rispetto alla base ordinata
$$\{v_1, v_2, v_3, v_4, v_5\}$$
la matrice è
$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = J_3(0) \oplus J_2(0)$$
abbiamo cioè trovato una base di Jordan (si ha infatti $L(v_1) = 0$, $L(v_2) = v_1$, $L(v_3) = v_2$,$L(v_4) = 0$, $L(v_5) = v_4$).

Osservazione. L'altezza delle colonne del diagramma a punti ci dà l'ordine dei blocchi di Jordan. La lunghezza delle righe del diagramma fornice la dimensione dei supplementari di $ker\ L^k$ in $ker\ L^{k+1}$. Infine la riga più in basso ci dà la dimensione del nucleo di L e corrisponde al numero dei blocchi di Jordan.

Esempio In $\mathbb{R}^6$
$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 2 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Ha polinomio caratteristico t^6 e quindi è nilpotente. Calcoliamone l'indice di nilpotenza.

$$e_1 \mapsto e_2 + 2e_4 + e_6 \mapsto 0$$
$$e_2 \mapsto 0$$
$$e_3 \mapsto e_2 \mapsto 0$$
$$e_4 \mapsto 0$$
$$e_5 \mapsto 2e_2 + e_3 + e_4 \mapsto e_2 \mapsto 0$$
$$e_6 \mapsto 0$$

L'indice di nilpotenza è 3. Studiamo $ker\ L$: $rg(A) = 3$ quindi $dim\ ker\ L = 6-3 = 3$ con equazioni

$$\begin{cases} x_1 = 0 \\ x_3 = 0 \\ x_5 = 0 \end{cases}$$

come si può vedere, ad esempio, riducendo a gradini la matrice A, ed una base del nucleo è

$$\{(0,1,0,0,0,0),(0,0,0,1,0,0),(0,0,0,0,0,1)\}$$

Consideriamo poi A^2:

$$A^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Ha rango 1 e dunque $dim\ ker\ L^2 = 5$ con equazione $x_5 = 0$. Una base di $ker\ L^2$ è

$$\{(1,0,0,0,0,0),(0,1,0,0,0,0),(0,0,1,0,0,0),(0,0,0,0,1,0),(0,0,0,0,0,1)\}$$

Allora

$$ker\ L \subset ker\ L^2 \subset ker\ L^3 = \mathbb{R}^6$$

Per trovare una base di Jordan: sia $v_3 = (0,0,0,0,1,0) \in ker\ L^3 - ker\ L^2$ e $v_2 = L(v_3) = (0,2,1,1,0,0)$, $v_1 = L(v_2) = (0,1,0,0,0,0)$. Abbiamo il diagramma

$$\begin{array}{c} * \\ \downarrow \\ * \\ \downarrow \\ * \end{array}$$

Prendiamo poi il supplementare di $ker\ L$ in $ker\ L^2$. Questo supplementare ha dimensione 2. Sia v_5 un vettore che insieme a v_2 genera un supplementare di $ker\ L$ in $ker\ L^2$ ad esempio $v_5 = (1,0,0,0,0,0)$ e quindi $v_4 = L(v_5) = (0,1,0,2,0,1)$. Dunque riassumendo abbiamo il diagramma

$$\begin{array}{cc} * & \\ \downarrow & \\ * & * \\ \downarrow & \downarrow \\ * & * \end{array}$$

Un ultimo vettore lo troviamo in $ker\ L$ perché $ker\ L$ ha dimensione 3 e quindi possiamo trovare $v_6 = (0,0,0,1,0,0)$ in modo che v_1, v_4, v_6 siano una base di $ker\ L$. Allora rispetto alla base formata da quasti 6 vettori la matrice è

$$J_3(0) \oplus J_2(0) \oplus J_1(0) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

In generale la situazione è illustrata dalla seguente figura, ove $q = 4$ è l'indice di nilpotenza

$$\begin{array}{ccccccccccccc} * & * & * & * & * & * & & & & & & & \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & & & & & & & \\ * & * & * & * & * & * & * & * & * & & & & \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & & & & \\ * & * & * & * & * & * & * & * & * & * & & & \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & & & \\ * & * & * & * & * & * & * & * & * & * & * & * & * \end{array}$$

Ogni punto del diagramma rappresenta un vettore di una base di Jordan di $ker\ L^q = \mathbb{R}^n$. Se tralasciamo la riga alla sommità del diagramma, i vettori rimanenti danno una base di $ker\ L^{q-1}$. La seconda riga è costituita dai vettori trasformati della prima riga più eventuali altri vettori di un opportuno supplementare. Continuando in questa maniera si ottiene una base di Jordan.

Vediamo ora un esempio di un endomorfismo non nilpotente. Sia

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

il suo polinomio caratteristico è $t^3(1-t)^2$. Ha un autovalore 0 con molteplicità algebrica 3 e un autovalore 1 con molteplicità algebrica 2. Chiaramente non è nilpotente. Sappiamo allora che $\mathbb{R}^5$ si spezza nella somma diretta degli autospazi generalizzati

$$\mathbb{R}^5 = V_0 \oplus V_1$$

dove V_0 ha dimensione 3 e V_1 ha dimensione 2. Se prendiamo una base di $\mathbb{R}^5$ composta da una base di V_0 e da una di V_1 la matrice di L diventa

$$\begin{pmatrix} A' & \\ & A'' \end{pmatrix}$$

dove A' è la matrice della restrizione a V_0 e A'' è la matrice della restrizione a V_1. Studiamo la restrizione L_{V_0} che è nilpotente. Consideriamo la successione

$$ker\ L \subset ker\ L^2 \subset ker\ L^3 \subset \dots$$

finché non si stabilizza. Abbiamo che $ker\ L$ è l'insieme delle soluzioni di un sistema lineare omogeneo che ha A come matrice e quindi, poiché il rango di A è 4, tale nucleo ha dimensione 1. Abbiamo poi:

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

che ha rango 3 e quindi *dim ker* $L^2 = 2$. Poi

$$A^3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 5 & 0 \\ 1 & 0 & 0 & 2 & 0 \\ 3 & 0 & 0 & 1 & 0 \\ 6 & 0 & 0 & 3 & 0 \end{pmatrix}$$

che ha rango 2 e quindi *dim ker* $L^3 = 3$. Infine

$$A^4 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 5 & 0 \\ 3 & 0 & 0 & 2 & 0 \\ 4 & 0 & 0 & 1 & 0 \\ 9 & 0 & 0 & 3 & 0 \end{pmatrix}$$

ha ancora rango 2 e qundi la successione si è stabilizzata. Abbiamo dunque la catena

$$ker\ L \subset ker\ L^2 \subset ker\ L^3$$

Dunque esiste un'unica catena di Jordan corrispondente al blocco $J_3(0)$

$$\begin{array}{c} * \\ \downarrow \\ * \\ \downarrow \\ * \end{array}$$

Studiamo ora la restrizione di L a V_1. Per questo consideriamo $(L - I)_{V_1}$, la restrizione a V_1 della differenza tra L e l'identità. Per costruzione questo endomorfismo è nilpotente e si ha

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

Questa ha chiaramente rango 4 (ci sono 4 righe indipendenti) e quindi la dimensione del suo nucleo è 1. Quindi anche qui avremo una sola catena di autovettori generalizzati associata all'autovalore 1 che sarà necessariamente di lunghezza 2 essendo la dimensione di V_1 uguale a 2. La matrice di $(L - I)_{V_1}$ è dunque $J_2(0)$ e quindi quella di L_{V_1} è $J_2(1)$. In definitiva la matrice di Jordan in questa base è

$$J_3(0) \oplus J_2(1) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Esempio

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

ha come autovalore -1 con molteplicità algebrica 4 e 2 con molteplicità algebrica 2. La matrice $A+I$ ha rango 4 e quindi la dimensione del suo nucleo è 2. $(A+I)^2$ ha rango 3 ed il suo nucleo ha dimensione 3. $(A+I)^3$ ha rango 2 e quindi il suo nucleo ha dimensione 4. Infine $(A+I)^4$ ha ancora rango 2 quindi la successione si è stabilizzata. Abbiamo il diagramma

$$\begin{array}{cc} * & \\ \downarrow & \\ * & \\ \downarrow & \\ * & * \end{array}$$

Una base di $Ker\ (A+I)^3$ è

$$\{(1,0,0,0,0,0),(0,1,0,0,0,0),(0,0,1,-3,0,0),(0,0,1,0,3,-9)\}$$

Dobbiamo trovare una combinazione lineare di questi vettori che non appartenga a $ker\ (A+I)^2$. Si può prendere $v_3 = (0,0,1,0,3,-9)$. Calcoliamo quindi $v_2 = (A+I)v_3 = (5,0,3,-9,0,0)$ e $v_1 = (A+I)v_2 = (6,0,0,0,0,0) \in ker\ (A+I)$. Un supplementare di $\langle(6,0,0,0,0,0)\rangle$ in $ker\ (A+I)$ è $\langle(0,1,0,0,0,0)\rangle$.

Infine dobbiamo studiare $A-2I$ che ha rango 4 e dunque nucleo di dimensione 2. $(A-2I)^2$ ha rango ancora 4 e quindi la successione si stabilizza. Possiamo prendere $v_5 = (2,0,3,0,0,0)$ e $v_6 = (1,0,0,0,3,0)$. Otteniamo in totale una base di Jordan e la matrice di Jordan è allora

$$J_3(-1) \oplus J_1(-1) \oplus J_1(2) \oplus J_1(2)$$

Esercizio

Trovare la forma canonica di

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

Troveremo il diagramma

$$\begin{array}{ccc} + & * & \\ \downarrow & \downarrow & \\ + & * & * \end{array}$$

($+$ indica autovettori per 1 mentre $*$ indica autovettori per 0).

Osservazione. Gli autospazi generalizzati di un operatore L sono invarianti. Infatti se $v \in ker\ (A-\lambda I)^r$ considero Av e voglio verificare che Av è ancora un autovettore generalizzato. Calcolo $(A-\lambda I)^r Av$. Per questo, osserviamo che si ha

$$A(A-\lambda I) = A^2 - \lambda A = (A-\lambda I)A$$

ossia A e $A-\lambda I$ commutano. Quindi

$$A(A-\lambda I)^2 = (A-\lambda I)^2 A$$

e

$$A(A-\lambda I)^r = (A-\lambda I)^r A$$

Allora nel nostro calcolo abbiamo

$$(A-\lambda I)^r Av = A(A-\lambda I)^r v = A0 = 0$$

Osservazione. Sappiamo che la molteplicità algebrica di un autovalore λ è sempre maggiore o uguale alla molteplicità geometrica (= dimensione dell'autospazio). Essa è comunque uguale alla dimensione dell'autospazio generalizzato relativo a λ.

Ad esempio, se

$$V = V(\lambda_1) \oplus V(\lambda_2) \oplus V(\lambda_3)$$

è la decomposizione di V nella somma diretta degli autospazi generalizzati (relativi ad autovalori distinti) avremo chiaramente, se $d_i = dim\ V(\lambda_i)$,

$$d_1 + d_2 + d_3 = n = dim\ V$$

Sia

$$p_A(t) = (t-\lambda_1)^{r_1}(t-\lambda_2)^{r_2}(t-\lambda_3)^{r_3}$$

avremo anche $r_1+r_2+r_3 = n$. Sappiamo anche che $r_i \le d_i$ (infatti $p_A = p_{A_1}p_{A_2}p_{A_3}$ dove A_i è la matrice della restrizione dell'operatore su $V(\lambda_i)$, e λ_1 non è radice di p_{A_2} né di p_{A_3} perché la somma è diretta. Dunque la molteplicità algebrica di λ_1 è minore o uguale alla dimensione dell'autospazio generalizzato.) Quindi

$$d_1 + d_2 + d_3 - r_1 - r_2 - r_3 = 0$$

$$(d_1 - r_1) + (d_2 - r_2) + (d_2 - r_2) = 0$$

Quindi $d_i - r_i = 0\ i = 1, 2, 3$.

Osservazione. Se due matrici sono simili esse hanno lo stesso polinomio caratteristico. Il viceversa non è però vero. Ad esempio, se

$$p_A(t) = (t-\lambda_1)^2(t-\lambda_2)^3$$

con $\lambda_1 \ne \lambda_2$ le seguenti matrici, tra loro non simili, hanno tutte questo polinomio caratteristico:

$$J_2(\lambda_1) \oplus J_3(\lambda_2)$$

$$J_2(\lambda_1) \oplus J_2(\lambda_2) \oplus J_1(\lambda_2)$$

$$J_2(\lambda_1) \oplus J_1(\lambda_2) \oplus J_1(\lambda_2) \oplus J_1(\lambda_2)$$

$$J_1(\lambda_1) \oplus J_1(\lambda_1) \oplus J_3(\lambda_2)$$

$$J_1(\lambda_1) \oplus J_1(\lambda_1) \oplus J_2(\lambda_2) \oplus J_1(\lambda_2)$$

$$J_1(\lambda_1) \oplus J_1(\lambda_1) \oplus J_1(\lambda_1) \oplus J_1(\lambda_2) \oplus J_1(\lambda_2)$$

Queste matrici corrispondono rispettvamente ai seguenti diagrammi:

$$\begin{array}{cc} & * \\ & \downarrow \\ + & * \\ \downarrow & \downarrow \\ + & * \end{array}$$

$$\begin{array}{ccc} + & * & \\ \downarrow & \downarrow & \\ + & * & * \end{array}$$

$$\begin{array}{cccc} + & & & \\ \downarrow & & & \\ + & * & * & * \end{array}$$

$$\begin{array}{ccc} & & * \\ & & \downarrow \\ & & * \\ & & \downarrow \\ + & + & * \end{array}$$

$$\begin{array}{cccc} & & * & \\ & & \downarrow & \\ + & + & * & * \end{array}$$

$$\begin{array}{ccccc} + & + & * & * & * \end{array}$$

L'ultimo caso, quello in cui tutti gli autovettori generalizzati hanno rango 1 cioè sono autovettori nel senso vecchio del termine, abbiamo il caso di un operatore diagonalizzabile. In corrispondenza di ciascuno di questi casi abbiamo diversi polinomi minimi, rispettivamente

$$(t-\lambda_1)^2(t-\lambda_2)^3$$

$$(t-\lambda_1)^2(t-\lambda_2)^2$$

$$(t-\lambda_1)^2(t-\lambda_2)$$

$$(t-\lambda_1)(t-\lambda_2)^3$$

$$(t-\lambda_1)(t-\lambda_2)^2$$

$$(t-\lambda_1)(t-\lambda_2)$$

Il caso diagonalizzabile corrisponde al caso in cui il polinomio minimo ha radici semplici.

Osservazione. Occorre osservare che è anche possibile che due matrici abbiano lo stesso polinomio minimo senza essere simili. Ad esempio,

$$J_3(\lambda) \oplus J_2(\lambda)$$

$$J_3(\lambda) \oplus J_1(\lambda) \oplus J_1(\lambda)$$

sono due matrici non simili ma hanno lo stesso polinomio minimo $(t-\lambda)^3$.

Proposiszione 1.17. *Se L è un operatore riducibile su V e la matrice di L in una base adattata è del tipo*

$$A = A_1 \oplus A_2 \oplus \cdots \oplus A_k$$

allora il polinomio minimo di A è il minimo comune multiplo dei polinomi minimi di ciascuna A_i.

Dimostrazione. Cominciamo osservando che L induce, per restrizione, un endomorfismo L_i su ciascuna componente V_i con matrice A_i. Osserviamo che se $q(t)$ è un polinomio e consideriamo l'operatore $q(L)$, allora gli spazi V_i sono invarianti anche per $q(L)$ e la restrizione di $q(L)$ a V_i è $q(L_i)$. In particolare se $q(L) = 0$ anche $q(L_i) = 0$. Questo accade nel caso del polinomio minimo $m_A(t)$, infatti $m_A(L) = 0$ e dunque $m_A(L_i) = 0$. Ciò significa che $m_A(t)$ è un multiplo di ciascun $m_{A_i}(t)$ e quindi è un multiplo del m.c.m dei polinomi $m_{A_i}(t)$. Chiamiamo $q(t)$ questo m.c.m. Sia $v_i \in V_i$ e calcoliamo $q(L)v_i = q(L_i)v_i = 0$ perché $q(t) = a(t)m_{A_i}(t)$ per qualche $a(t)$. Poiché un v qualsiasi si scrive come $v = \sum v_i$ ne segue che $q(L)v = \sum q(L)v_i = 0$ cioè $q(L) = 0$ Ma allora $q(t)$ è un multiplo di $m_A(t)$. In definitiva dunque $m_A(t) = q(t)$ $\square$

Esempio
Nel caso in cui si abbia una matrice di Jordan

$$J_3(\lambda_1) \oplus J_2(\lambda_1) \oplus J_4(\lambda_2) \oplus J_1(\lambda_2)$$

essa corrisponde ad una decomposizione in sottospazi invarianti:

$$V_3(\lambda_1) \oplus V_2(\lambda_1) \oplus V_4(\lambda_2) \oplus V_1(\lambda_2)$$

Il polinomio caratteristico di ciascun blocco è rispettivamente

$$(t - \lambda_1)^3, (t - \lambda_1)^2, (t - \lambda_2)^4, (t - \lambda_2)$$

e il m.c.m. è

$$(t - \lambda_1)^3 (t - \lambda_2)^4$$

Ancora sul calcolo del polinomio minimo. Abbiamo visto come calcolare il polinomio minimo di una matrice di Jordan. Poichè matrici simili hanno lo stesso polinomio minimo questo ci permette di calcolare il polinomio minimo di qualunque matrice possegga una forma di Jordan. Infatti, se due matrici sono simili allora

$$B = C^{-1}AC$$

e quindi $B^k = C^{-1}A^kC$ e dunque anche per un qualunque polinomio $q(t)$, che è combinazione lineare di potenze dell'indeterminata, abbiamo $q(C^{-1}AC) = C^{-1}q(A)C$. Questo significa in particolare che un polinomio si annulla su A se e solo se si annulla su B. Essendo dunque uguali gli insiemi dei polinomi annullatori saranno uguali anche i polinomi minimi.

Ora se A è una matrice a coefficienti in $\mathbb{C}$ essa ammette una forma di Jordan e quindi il suo polinomio minimo si può calcolare come sopra. Lo stesso vale per una matrice reale che abbia tutti gli autovalori reali. In tali casi A è simile a

$$A' = J_{r_1}(\lambda_1) \oplus \ldots \oplus J_{r_s}(\lambda_s)$$

I polinomi minimi dei vari addendi sono

$$(t - \lambda_1)^{r_1}, (t - \lambda_2)^{r_2}, \ldots, (t - \lambda_s)^{r_s}$$

dove $\lambda_1, \lambda_2, \ldots, \lambda_s$ non sono tutti distinti. Per avere $m_A(t)$ occorrerà prendere il massimo esponente che corrisponde al massimo blocco con quell'autovalore. Qusto esponente si chiama indice dell'autovalore λ.

Se la matrice reale non ha tutti gli autovalori nel campo reale, la pensiamo come matrice complessa, ne calcoliamo il polinomio minimo complesso. Vogliamo dimostrare che in effetti questo polinomio ha coeffficienti reali e quindi esso è il polinomio minimo desiderato.

Ad esempio, se una matrice A ha polinomio caratteristico

$$p_A(t) = (t-i)^3(t+i)^3(t-2)^2$$

il polinomio minimo sarà del tipo

$$m_A(t) = (t-i)^{k_1}(t+i)^{k_2}(t-2)^{k_3}$$

dove $0 \le k_1 \le 3$, $0 \le k_2 \le 3$, $0 \le k_3 \le 2$. A priori, non è detto che $k_1 = k_2$. Osserviamo però che la matrice $A - \lambda I$ è coniugata alla matrice $A - \overline{\lambda} I$ e quindi tali sono anche tutte le loro potenze. Ne segue che, poiché matrici coniugate hanno lo stesso rango, che per ogni k

$$\dim\, Ker\ (A - \lambda I)^k = \dim\, Ker\ (A - \overline{\lambda} I)^k$$

Poiché la successione dei nuclei $Ker\ (A-\lambda I)^k$ diventa stazionaria quando la dimensione di $Ker\ (A-\lambda I)^k$ è uguale alla molteplicità algebrica di λ allora le due successioni relative a λ e a $\overline{\lambda}$ diventano stazionarie allo stesso passo. Infine, poiché l'esponente k per cui si ha

$$Ker\ (A-\lambda I)^{k-1} \subset Ker\ (A-\lambda I)^k = Ker\ (A-\lambda I)^{k+1}$$

è la lunghezza massima delle catene di autovettori generalizzati associate a λ, gli autovalori λ e $\overline{\lambda}$ hanno indici uguali. In definitiva, nel polinomio minimo complesso ogni fattore complesso $(t-\lambda)^r$ appare insieme al suo coniugato $(t-\overline{\lambda})^r$ e quindi $m_A(t)$ è a coefficienti reali.

Esempio Calcolare il polinomio minimo di

$$A = \begin{pmatrix} 1 & 0 & 0 & 3 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 1 & 5 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix}$$

$p_A(t) = (1-t)^2(5-t)^3$ quindi tutti gli autovalori sono in $\mathbb{R}$: 1 con molteplicità algebrica 2 e 5 con molteplicità algebrica 3.

$A - I$ ha rango 4 e dunque il suo nucleo ha dimensione 1. Quindi c'è un solo blocco di Jordan con autovalore 1, e poiché la molteplicità algebrica è 2, 2 deve anche essere la lunghezza della catena di vettori:

$$\begin{matrix} * \\ * \end{matrix}$$

La matrice $A - 5I$ ha rango 3 e dunque il suo nucleo ha dimensione 2. Quindi A' contiene due blocchi associati all'autovalore 5. Poiché la molteplicità algebrica è 3 i due blocchi devono essere di ordine 2 e 1. In fin dei conti,

$$m_A(t) = (t-1)^2(t-5)^2$$

Esercizio

Determinare i polinomi di grado minore o uguale a 5 e di grado minore o uguale a 3 che si annullano sulla matrice A appena studiata.

Sappiamo che i polinomi annullatori di A sono multipli di $m_A(t)$ e quindi sono del tipo

$$q(t)(t-1)^2(t-5)^2$$

se lo vogliamo di grado minore o uguale a 5 occorre che $q(t)$ sia un polinomio lineare. Si capisce anche che non possiamo avere polinomi di grado minore o uguale a 3 che si annullano su A.

Esercizi

Determinare tutti i polinomi che su A hanno valore $2A^3 + 7A^2 - 9I$.

Ovviamente $2t^3 + 7t^2 - 9$ è un tale polinomio. Osserviamo ora che due polinomi $f(t), g(t)$ sono tali che $f(A) = g(A)$ se e solo se essi danno lo stesso resto quando vengono divisi per $m_A(t)$. Infatti se $f(t) = q_1(t)m_A(t)+r_1(t)$ e $g(t) = q_2(t)m_A(t)+r_2(t)$ con $r_2(t) = r_1(t)$ allora

$$f(A) = r_1(A) = r_2(A) = g(A)$$

Viceversa, se $f(A) = g(A)$ allora

$$f(t) - g(t) = (q_1(t) - q_2(t))m_A(t) + r_1(t) - r_2(t)$$

che calcolato in A ci dà

$$0 = (q_1(A) - q_2(A))m_A(A) + r_1(A) - r_2(A)$$

allora $r_1(A)-r_2(A) = 0$ ossia $r_1(t)-r_2(t)$ annulla A. Essendo il grado di $r_1(t)-r_2(t)$ minore del grado di $m_A(t)$ ne segue che $r_1(t) = r_2(t)$. Per concludere l'esercizio, tutti i polinomi desiderati hanno la forma

$$q(t)(t-1)^2(t-5)^2 + 2t^3 + 7t^2 - 9$$

con $q(t)$ un qualunque polinomio.

Esempio

Calcolare il polinomio minimo della matrice reale

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

Poiché si tratta di una matrice a blocchi, il suo polinomio minimo è ottenuto dai polinomi minimi dei singoli blocchi. Ora

$$A_1 = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix}$$

ha polinomio caratteristico t^2+1 quindi possiede due autovalori complessi coniugati: $i, -i$, ciascuno con molteplicità algebrica 1 e quindi necessariamente con indice 1. Il suo polinomio minimo è $t^2 + 1$. Il secondo blocco A_2 ha polinomio caratteristico $(1-t)(t-1)(t+1)$ dunque ha autovalore 1 con molteplicità algebrica 2 e -1 con molteplicità algebrica 1. Chiaramente -1 ha indice 1. Per vedere l'indice di 1, studiamo $A_2 - I$ che ha rango 1. Dunque il suo nucleo ha dimensione 2. Ci sono quindi due blocchi necessariamente di ordine 1 e l'indice è quindi 1. Il polinomio minimo è quindi $(t-1)(t+1)$ la matrice $A_2 - I$ è quindi diagonalizzabile. Infine,

$$m_A(t) = (t^2+1)(t^2-1).$$

Esercizio

Verificare che $A^{97} = A^{93}$.

Due polinomi t^{97} e t^{93} hanno lo stesso valore su A se e solo se danno lo stesso resto quando li dividiamo per $m_A(t)$, ovvero se la loro differenza è un multiplo di

$$m_A(t) = (t^2+1)(t^2-1)$$

Infatti:

$$t^{97} - t^{93} = t^{93}(t^4-1) = t^{93}(t^2-1)(t^2+1).$$

Capitolo II.

Funzioni di matrici. Il problema che vogliamo affrontare è quello di estendere una funzione $f(t)$ di variabile reale o complessa e a valori reali o complessi, ad una funzione chiamata ancora f su matrici quadrate $A \in M_n(K)$, dove K è il campo reale o complesso.

Nel caso in cui la funzione $f(t)$ è una funzione polinomiale, questa estensione è immediata e l'abbiamo usata già molte volte. Ma vogliamo, se possibile, poter calcolare $f(A)$, A una matrice, anche quando f è una funzione non polinomiale, come ad esempio *log* t, e^t, *sen* t, etc.

Supponiamo di voler affrontare il problema nella maniera più ingenua e naturale possibile, il che è certamente la strada da seguire ad un primo approccio, salvo soffermarsi, e magari tornare indietro qualora si dovessero incontrare delle difficoltà. Potremmo pensare di definire, ad esempio, l'esponenziale di una matrice tramite una serie di potenze. Dopotutto, per la funzione esponenziale abbiamo lo sviluppo in serie

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Se pensiamo in maniera ingenua a questa serie come ad una sorta di "polinomio infinito" possiamo sostituire alla variabile scalare x una "variabile matriciale" A e prendere questa come definizione di e^A. Vedremo che ciò è effettivamente possibile. Tuttavia, questo modo di procedere presta il fianco a numerose, e giustificate, critiche! Per cominciare, una serie di potenze richiede la giustificazione di una convergenza di una serie, o equivalentemente, di una successione. Cosa significa la convergenza di una successione di matrici? Qual è la metrica o la topologia sullo spazio delle matrici? Inoltre, come calcolare questi limiti? Supponiamo di spazzare queste obiezioni, per il momento, sotto il tappeto, e tiriamo avanti per un po' in questa maniera. Facciamo, una prima ipotesi semplificativa, e supponiamo che la nostra matrice A di cui vogliamo calcolare l'esponenziale sia diagonalizzabile. Anzi, supponiamo direttamente, per prima cosa che la matrice A sia diagonale. Se A è quadrata, diagonale, supponiamo

$$A = \begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & 0 & \dots & \alpha_n \end{pmatrix}$$

poichè

$$A^k = \begin{pmatrix} \alpha_1^k & 0 & \dots & 0 \\ 0 & 0 & \dots & \alpha_n^k \end{pmatrix}$$

possiamo pensare di definire

$$e^A = \begin{pmatrix} e^{\alpha_1} & 0 & \dots & 0 \\ 0 & 0 & \dots & e^{\alpha_n} \end{pmatrix}$$

senza far ricorso a serie di potenze e quindi evitando questioni di convergenza. Ora se A non è diagonale ma è diagonalizzabile, allora $A = CDC^{-1}$ per qualche matrice invertibile C e possiamo definire

$$e^A = Ce^DC^{-1}$$

Questo ci suggerisce l'idea di definire, qualunque sia la funzione f, non solamente per la funzione esponenziale,

$$f(D) = \begin{pmatrix} f(\alpha_1) & 0 & \dots & 0 \\ 0 & 0 & \dots & f(\alpha_n) \end{pmatrix}$$

e
$$f(A) = Cf(D)C^{-1}$$
Questo evita la considerazione di problemi di convergenza tuttavia presenta altri problemi: e se la matrice non è diagonalizzabile? Forse possiamo affrontare il problema tramite la sua forma di Jordan. Un altro tipo di problema è dato dal fatto che, come sappiamo, la matrice C non è unica e quindi, a priori chi ci garantisce che la $f(A)$ sia ben definita, cioè non dipenda da C? Questo modo di affrontare il problema ci ha fatto capire dove sono i possibili problemi e nel prosieguo cercheremo di affrontarli e risolverli.

Nel caso di una matrice A qualunque sui complessi, sia J la sua forma di Jordan. Abbiamo $A = CJC^{-1}$ e per completare la nostra definizione basterà dare la definizione di $f(J)$, ed essendo J somma diretta di blocchi di Jordan, basterà definire $f(J_k(\lambda))$. Per dare questa definizione, prendiamo ispirazione da una eventuale serie di Taylor della funzione f centrata nel punto λ:

$$f(t) = f(\lambda) + f'(\lambda)(t-\lambda) + \frac{f''(\lambda)}{2!}(t-\lambda)^2 + \frac{f'''(\lambda)}{3!}(t-\lambda)^3 + \cdots$$

con un certo raggio di convergenza. Una eventuale $f(J_k(\lambda)$ dovrà allora soddisfare

$$\begin{aligned} f(J_k(\lambda)) &= f(\lambda)I + f'(\lambda)(J_k(\lambda)-\lambda I) + \frac{f''(\lambda)}{2!}(J_k(\lambda)-\lambda I)^2 \\ &+ \frac{f'''(\lambda)}{3!}(J_k(\lambda)-\lambda I)^3 + \cdots \end{aligned}$$

ci accorgiamo allora con piacere che questa serie non è infinita! Infatti, essendo la matrice $J_k(\lambda) - \lambda I$ una matrice nilpotente di ordine k, infatti

$$J_k(\lambda) - \lambda I = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & 1 & \ddots \\ 0 & 0 & \ddots & 1 \\ 0 & 0 & & 0 \end{pmatrix}$$

la serie è in effetti una somma finita fino al termine $k-1$-esimo. Quindi questa definizione di $f(J_k(\lambda))$ ha senso non appena esistano le derivate $f^{(i)}(\lambda)$. Rimane aperta la questione dell'unicità, e si potrebbe proseguire su questa strada. Ma la discussione precedente è stata data al solo scopo di fornire una qualche motivazione alla definizione che ora daremo.

Supponiamo che A sia una matrice con polinomio minimo del tipo

$$m_A(t) = (t-\lambda_1)^{n_1}(t-\lambda_2)^{n_2}\cdots(t-\lambda_p)^{n_p}$$

dove gli autovalori $\lambda_i \in K$, e $\lambda_i \neq \lambda_j$. Una funzione $f : K \to K$ si dice definita sullo spettro di una matrice A se sono definiti gli scalari

$$\frac{d^k f}{dt^k}|_{t=\lambda_i}$$

per ogni autovalore λ_i e $0 \leq k \leq n_i - 1$.

L'idea è quella di ricondursi al caso dei polinomi. Cominciamo osservando che il valore di due distinti polinomi su una matrice A è univocamente determinato dai valori assunti sullo spettro di A. Ossia:

Proposizione 2.1. *Due polinomi $f(t)$ e $g(t)$ assumono gli stessi valori sullo spettro di una matrice A se e solo se $f(A) = g(A)$.*

Dim. Supponiamo che $f(A) = g(A)$. Allora $f(t) - g(t)$ è un multiplo del polinomio minimo $m_A(t)$ e quindi $f(t)$ e $g(t)$ danno lo stesso resto nella divisione per $m_A(t)$:

$$f(t) = q_1(t)m_A(t) + r_1(t) \tag{1}$$

$$g(t) = q_2(t)m_A(t) + r_1(t) \tag{2}$$

Osserviamo ora che se un polinomio ha una radice α con molteplicità m allora α è radice anche di tutte le sue derivate fino alla $(m-1)$-esima. Infatti se $p(t) = (t-\alpha)^m q(t)$ allora $p'(t) = m(t-\alpha)^{m-1}q(t) + (t-\alpha)^m q'(t)$ e quindi α è radice anche di $p'(t)$ etc. Calcolando allora le derivate di $f(t)$ e di $g(t)$ nelle espressioni (1) e (2) e quindi il valore di queste derivate sugli autovalori di A abbiamo che la parte $q_1(t)m_A(t)$ e la parte $q_2(t)m_A(t)$ si annulla. Quindi queste derivate coincidono con le derivate di $r_1(t)$ calcolate sugli autovalori. Dunque sono anche uguali tra loro.

Viceversa, se tutte le derivate sono uguali tra loro, in ogni caso abbiamo

$$f(t) = q_1(t)m_A(t) + r_1(t) \tag{3}$$

$$g(t) = q_2(t)m_A(t) + r_2(t) \tag{4}$$

con r_1 e r_2 a priori distinti e di grado minore del grado di $m_A(t)$. Nel calcolo delle derivate abbiamo comunque che

$$f^{(k)}(\lambda_k) = r_1^{(k)}(\lambda_k)$$

$$g^{(k)}(\lambda_k) = r_2^{(k)}(\lambda_k)$$

e dunque i polinomi $r_1(t)$ e $r_2(t)$ coincidono insieme alle loro derivate su ogni radice di $m_A(t)$. Poiché il grado di r_1 e r_2 è minore del grado di $m_A(t)$ allora $r_1(t) = r_2(t)$. Questa ultima conclusione si basa sul principio che un polinomio non nullo di grado n su $\mathbb{C}$ non può avere più di n radici.

Se $f(t)$ è una qualunque funzione di t si dice che f è definita sullo spettro di A se sono definiti i valori

$$f(\lambda_k), f'(\lambda_k) \dots, f^{n_k-1}(\lambda_k)$$

per ogni radice λ_k di molteplicità n_k di $m_A(t)$.

Se questi valori sono definiti porremo allora per definizione $f(A)$ uguale alla matrice $p(A)$ dove $p(t)$ è un qualunque polinomio che assume sullo spettro di A gli stessi valori di f. L'insieme dei polinomi siffatti, detti polinomi interpolatori, è non vuoto. Il polinomio di grado inferiore al grado di m_A si chiama polinomio interpolatore di Lagrange-Sylvester. Vedremo in qualche esempio come determinare il polinomio interpolatore di Lagrange-Sylvester. Si tratta di trovare un polinomio $r(t)$ tale che $f(\lambda_k) = r(\lambda_k), f'(\lambda_k) = r'(\lambda_k) \dots, f^{n_k-1}(\lambda_k) = r^{n_k-1}(\lambda_k)$

Esempio Sia

$$J_5(0) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Questa ha polinomio minimo t^5. I valori di $f(t)$ sullo spettro sono

$$f(0), f'(0), f''(0), f'''(0), f^{(iv)}(0)$$

Cerchiamo un polinomio di grado minore di 5 che abbia le prime 5 derivate (dalla 0-esima alla quarta) uguali a questi valori di f (che per il momento non abbiamo specificato).

$$r(0) = f(0); \quad r'(0) = f'(0); \quad r''(0) = f''(0);$$
$$r^{(3)}(0) = f^{(3)}(0); \quad r^{(4)}(0) = f^{(4)}(0)$$

Dato

$$r(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4$$
$$r(0) = a_0 = f(0); \quad r'(0) = a_1 = f'(0); \quad r''(0) = 2a_2 = f''(0);$$
$$r^{(3)}(0) = 6a_3 = f^{(3)}(0); \quad r^{(4)}(0) = 24a_4 = f^{(4)}(0)$$

abbiamo

$$r(t) = f(0) + f'(0)t + \frac{f^{(2)}(0)}{2}t^2 + \frac{f^{(3)}(0)}{3!}t^3 + \frac{f^{(4)}(0)}{4!}t^4$$

cioè il polinomio di Taylor. Dunque

$$f(J_5) = r(J_5) = \begin{pmatrix} f(0) & f'(0) & \frac{f^{(2)}(0)}{2} & \frac{f^{(3)}(0)}{3!} & \frac{f^{(4)}(0)}{4!} \\ 0 & f(0) & f'(0) & \frac{f^{(2)}(0)}{2} & \frac{f^{(3)}(0)}{3!} \\ 0 & 0 & f(0) & f'(0) & \frac{f^{(2)}(0)}{2} \\ 0 & 0 & 0 & f(0) & f'(0) \\ 0 & 0 & 0 & 0 & f(0) \end{pmatrix}$$

Esempio Sia

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

Calcolare la funzione $\sqrt{x^2 + 3x}$ su A. A ha autovalori $0, -3, 2, 1$ (ciascuno di molteplicità 1). I valori della funzione sullo spettro sono $f(0) = 0, f(-3) = 0, f(2) = \sqrt{10}, f(1) = 2$. Dunque cerchiamo un polinomio di terzo grado che soddisfi le condizioni

$$p(0) = a_0 = 0, p(-3) = -27a_3 + 9a_2 - 3a_1 + a_0 = 0,$$
$$p(2) = 8a_3 + 4a_2 + 2a_1 + a_0 = \sqrt{10}, p(1) = a_0 + a_1 + a_2 + a_3 = 2$$

Questo è un sistema di equazioni lineari la cui soluzione è

$$(a_0, a_1, a_2, a_3) = (0, \frac{30 - 3\sqrt{10}}{10}, \frac{-5 + 2\sqrt{10}}{10}, \frac{-5 + \sqrt{10}}{10})$$

Infine essendo

$$A^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -2 & 9 & 0 & -2 \\ 0 & 0 & 4 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

e

$$A^3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 7 & -27 & 0 & 7 \\ 0 & 0 & 8 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

abbiamo $f(A) = p(A) = a_3A^3 + a_2A^2 + a_1A =$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \sqrt{10} & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix}$$

Osservazione. Il polinomio

$$\frac{2}{-4}t(t+3)(t-2) + \frac{\sqrt{10}}{10}t(t+3)(t-1)$$
$$\frac{0}{-60}t(t-1)(t-2) + \frac{\sqrt{0}}{6}(t+3)(t-1)(t-2)$$

è un polinomio di grado ≤ 3 che soddisfa le stesse condizioni. In generale, se il polinomio minimo ha radici distinte allora il polinomio di Lagrange-Sylvester è

$$\frac{f(\alpha_1)}{\phi_1(\alpha_1)}\phi_1(t) + \frac{f(\alpha_2)}{\phi_2(\alpha_2)}\phi_2(t)$$
$$+ \cdots + \frac{f(\alpha_h)}{\phi_h(\alpha_h)}\phi_h(t)$$

ove $\phi_i(t) = \frac{m_A(t)}{(t-\alpha_i)}$.

Esercizio

Calcolare, se possibile, la funzione e^t sulla matrice $J = J_4(5)$.

Poiché e^t è una funzione C^∞ su tutto l'asse reale, la funzione e^A è definita per qualunque matrice. Nel caso di un blocco di Jordan di ordine p sappiamo che il polinomio interpolatore è il polinomio di Taylor di ordine di ordine $p-1$, dunque, nel nostro caso

$$e^J = e^5I + e^5(J-5I) + \frac{1}{2}e^5(J-5I)^2 + \frac{1}{3!}e^5(J-5I)^3$$

cioè

$$\begin{pmatrix} e^5 & e^5 & \frac{e^5}{2} & \frac{e^5}{6} \\ 0 & e^5 & e^5 & \frac{e^5}{2} \\ 0 & 0 & e^5 & e^5 \\ 0 & 0 & 0 & e^5 \end{pmatrix}$$

Esercizio

Calcolare se possibile la funzione $log\ x$ sulla matrice $J = J_3(1)$. Lo spettro è $\{1\}$. Le derivate sono: $log\ x, \frac{1}{x}, -\frac{1}{x^2}$ che calcolate sullo spettro danno: 0,1,-1. Abbiamo quindi

$$f(J) = log\ J = 0I + (J-I) + \frac{1}{2}(-1)(j-I)^2 = \begin{pmatrix} 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Esercizio
$f(x) = \frac{1}{x}$, $J = J_4(2)$ Si trova

$$f(A) = \begin{pmatrix} \frac{1}{2} & -\frac{1}{4} & \frac{1}{8} & -\frac{1}{16} \\ 0 & \frac{1}{2} & -\frac{1}{4} & \frac{1}{8} \\ 0 & 0 & \frac{1}{2} & -\frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

Si può osservare che $f(x)x = 1$ e quindi $f(A)a = I$ cioè se A è non singolare allora

$$(\frac{1}{x})_{|A} = A^{-1}$$

Esprimiamo ora in maniera più formale alcune delle idee descritte all'inizio di questo capitolo.

Se f è una funzione analitica della variabile complessa z e sia

$$f(z) = \sum_{i=0}^{\infty} a_i (z - z_0)^i$$

la sua rappresentazione in serie di potenze nell'intorno dell'origine con raggio di convergenza ρ. Se $X = (x_{ij})$ è una matrice quadrata si può considerare l'espressione

$$\sum_{i=0}^{\infty} a_i (X - z_0 I)^i$$

Ha senso una somma infinita di matrici? In effetti, si può dare una nozione di convergenza nell'insieme delle matrici e verificare se e quando una serie di matrici converge. Si dimostra:

Teorema 2.2. *Se tutti gli autovalori di X hanno valore assoluto minore del raggio di convergenza ρ della serie di potenze, allora la serie di matrici converge.*

□

Osservazione. Si verifica anche che il valore della funzione $f(X)$ così definito, coincide col valore definito nell'altra maniera. In particolare, si può definire

$$e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots$$

per ogni A, e nel caso in cui gli autovalori di A siano in valore assoluto minori di 1, si può definire anche

$$log\ (I + A) = A - \frac{A^2}{2} + \frac{A^3}{3} - \cdots$$

Proposizione 2.3. *Se A e B sono due matrici che commutano, ossia $AB = BA$, allora*

$$e^A e^B = e^{A+B}$$

Dim.

$$e^A e^B = (\sum_{i=0}^{\infty} \frac{A^i}{i!})(\sum_{j=0}^{\infty} \frac{B^j}{j!})$$

$$I + (A + B) + AB + \frac{A^2}{2} + \frac{B^2}{2} + \cdots$$

L'espressione

$$AB + \frac{A^2}{2} + \frac{B^2}{2}$$

coincide con $(A + B)^2$ se e solo se $AB = BA$ etc. □

Proposizione 2.4. *Se A è una matrice di costanti allora*

$$\frac{d}{dt}e^{tA} = Ae^{tA}$$

Dim. Infatti derivando termine a termine (la serie converge assolutamente e uniformemente) si ha la conclusione. □

Allora possiamo calcolare $e^{tJ_r(\lambda)}$ scrivendo

$$A = \lambda I + (J_r(\lambda) - \lambda I) = \lambda I + J_r(0)$$

e poiché λI commuta con ogni matrice possiamo scrivere

$$e^{tA} = e^{t\lambda I}e^{tJ_r(0)}$$

Ora

$$e^{tJ_r(0)}$$

è una serie finita. $e^{t\lambda I} = e^{t\lambda}I$ In definitiva

$$e^{tA} = \begin{pmatrix} e^{t\lambda} & te^{t\lambda} & \cdots & \frac{t^{r-1}}{(r-1)!}e^{t\lambda} \\ & e^{t\lambda} & & \\ & & \ddots & \end{pmatrix}$$

Diamo due proprietà delle funzioni di matrici.

Siano A e B matrici simili e T la matrice tale che $B = T^{-1}AT$ allora anche $f(B) = T^{-1}f(A)T$. Abbiamo già visto che questo è vero per polinomi. Ma siccome le due matrici simili hanno lo stesso spettro e quindi un polinomio interpolatore va bene per calcolare f sia su A che su B e quindi

$$f(B) = r(B) = T^{-1}r(A)T = T^{-1}f(A)T$$

Se A è una matrice diagonale a blocchi $A = A_1 \oplus \cdots \oplus A_k$ allora $f(A) = f(A_1) \oplus \cdots \oplus f(A_k)$ Infatti, f è definita sullo spettro di A se e solo se f è definita sullo spettro di ciascun A_i e un polinomio interpolatore di f su A è anche un polinomio interpolatore di f per A_i, diciamo $r(t)$. Poiché per un polinomio la conclusione è vera questo risulta vero anche per f.

Con queste due osservazioni possiamo calcolare il valore matriciale di una funzione nel modo seguente.

Se A è una matrice e sia J la sua forma di Jordan, diciamo

$$J = J_1 \oplus J_2 \oplus \cdots \oplus J_h$$

con $A = M^{-1}JM$ allora

$$f(J) = f(J_1) \oplus f(J_2) \oplus \cdots \oplus f(J_h)$$

e

$$f(A) = M^{-1}f(J)M$$

Esercizio

Calcolare $log(x+2)$ su

$$\begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 1 \end{pmatrix}$$

Si ottiene

$$\begin{pmatrix} log2 & \frac{1}{2} & -5log\ 2 - \frac{3}{2} + 5log\ 3 \\ 0 & log\ 2 & -3log\ 2 + 3log\ 3 \\ 0 & 0 & log\ 3 \end{pmatrix}$$

La maniera di dare un senso alla nozione di convergenza di serie di matrici è quella di introdurre una nozione di metrica sullo spazio delle matrici.

Spazio metrico. Un insieme non vuoto X su cui è definita una funzione $d : X \times X \to \mathbb{R}$ tale che

(1) $d(x, y) \geq 0$ e $d(x, y) = 0 \iff x = y$
(2) $d(x, y) = d(y, x)$
(3) $d(x, z) \leq d(x, y) + d(x, y)$

si dice spazio metrico.

Convergenza. Se $\{x_n\}$ è una successione di elementi di uno spazio metrico X e $x_0 \in X$ si dice che la successione converge ad x_0 se dato comunque $\epsilon > 0$ esiste n_ϵ tale che per $n > n_\epsilon$ si abbia $d(x_n, x_0) < \epsilon$.

Uno stesso spazio X può essere dotato di diverse funzioni metriche. Una maniera di definire una metrica è attraverso la definizione di norma, cioè una applicazione su uno spazio vettoriale V, denotata $\|v\|$ per $v \in V$ tale che

(1) $\|v\| \geq 0$ e $\|v\| = 0 \iff v = 0$
(2) $\|v\| + \|w\| \leq \|v + w\|$
(3) $\|av\| = |a|\|v\|$ per ogni scalare a.

Uno spazio vettoriale con una norma si dice spazio normato. Nello spazio vettoriale delle matrici $M_{m,n}(\mathbb{R})$ si possono definire le seguenti norme

(1) $\|A\|_1 = \sum i, j|a_{ij}|$ (norma L^1)
(2) $\|A\|_2 = (\sum i, j|a_{ij}|^2)^{\frac{1}{2}}$ (norma L^2)
(3) $\|A\|_\infty = max\ \{|a_{ij}|\}$ (norma L^∞)

La seconda delle norme elencate qui sopra, ad esempio, non è altro che la "solita" norma euclidea applicata alla matrice A pensata come elemento di uno spazio vettoriale $\mathbb{R}^{nm}$. In questo contesto si chiama *norma di Frobenius*. È utile osservare che questa norma si può esprimere anche nel modo seguente:

$$\|A\|_2{}^2 = Tr(A^T A)$$

Si può verificare, per esercizio, che questa è effettivamente una norma, cioè verifica le proprietà date sopra per un qualunque spazio vettoriale. Nel caso delle matrici richiediamo anche che la norma soddisfi la proprietà

$$\|AB\| \leq \|A\|\|B\|$$

ogni volta che è possibile moltiplicare due matrici. Si può verificare che la norma di Frobenius soddisfa anche questa proprietà.

Se V è uno spazio normato diremo che una serie converge se la successione delle somme parziali converge. La nozione di convergenza dipende dalla metrica usata, tuttavia in uno spazio di dimensione finita una successione converge in una metrica se solo se converge nell'altra metrica.

2.2 Un'applicazione dell'esponenziale di una matrice alla soluzione di sistemi di equazioni differenziali. Consideriamo un sistema di equazioni differenziali omogenee, lineari, a coefficienti costanti

$$\begin{cases} x_1' = a_{11}x_1(t) + \cdots + a_{1n}x_n(t) \\ x_2' = a_{21}x_1(t) + \cdots + a_{2n}x_n(t) \\ \dots \\ x_n' = a_{n1}x_1(t) + \cdots + a_{nn}x_n(t) \end{cases}$$

dove $x_1, \dots, x_n$ sono funzioni incognite e i coefficienti a_{ij} sono costanti. Possiamo riscrivere il sistema come

$$\mathbf{x}' = A\mathbf{x}$$

dove $\mathbf{x}$ è un vettore di funzioni. Vogliamo trovare soluzioni di questi sistemi. Ricordiamoci i casi già noti dal corso di Analisi del primo anno. Se $n = 1$ il sistema si riduce all'equazione
$$y' = ay$$
con $a \in \mathbb{R}$ che ha per soluzione la funzione esponenziale
$$y(t) = Ce^{at}$$
con C costante dipendente dalle condizioni iniziali. Nel corso di analisi si sono anche studiate le equazioni lineari del secondo ordine a coefficienti costanti:
$$x'' + px' + qx = 0 \tag{3}$$
Questa equazione è equivalente al sistema
$$\begin{cases} y_1' = y_2 \\ y_2' = -qy_1 - py_2 \end{cases} \tag{4}$$
Infatti se $x(t)$ soddisfa (3) e prendiamo $y_1 = x$, e $y_2 = x'$ allora la prima delle (4) è soddisfatta per costruzione, mentre la seconda è soddisfatta in quanto
$$y_2' = x''(t) = -px'(t) - qx(t) = -py_2 - qy_1$$
Viceversa, se y_1 e y_2 sono due soluzioni di (4) allora posto $x(t) = y_1(t)$ avremo $x'(t) = y_1'(t) = y_2(t)$ e $y_2' = -qy_1 - py_2$ cioè $x''(t) = -qx(t) - px'(t)$ cioè una soluzione di (3).

È facile verificare che
$$\mathbf{x}(t) = \begin{pmatrix} x(t) \\ x'(t) \end{pmatrix} = e^{\lambda t} \begin{pmatrix} 1 \\ \lambda \end{pmatrix} \tag{5}$$
è soluzione di (4).(Questa soluzione si può pensare geometricamente come una curva del piano data in forma parametrica).

Data l'analogia con la semplice equazione (2) sembra ragionevole andare a cercare soluzioni del sistema
$$\mathbf{y}' = A\mathbf{y} \tag{6}$$
tra le funzioni esponenziali. Per analogia con (5) potremmo prendere una soluzione del tipo
$$\mathbf{y}'(t) = e^{\lambda t}\mathbf{v} \tag{7}$$
dove λ e $\mathbf{v}$ sono, rispettivamente, un vettore e uno scalare da determinarsi. Sostituiamo (7) in (6). Il primo membro dà $\mathbf{y}'(t) = \lambda e^{\lambda t}\mathbf{v}$ mentre il secondo membro è
$$A\mathbf{y} = A(e^{\lambda t}\mathbf{v}) = e^{\lambda t}(A(\mathbf{v}))$$
Perchè questa sia una soluzione occorre che
$$\lambda e^{\lambda t}\mathbf{v} = e^{\lambda t}A\mathbf{v}$$
e poiché $e^{\lambda t} \neq 0$, questa funzione $y(t)$ è una soluzione se e solo se $A\mathbf{v} = \lambda\mathbf{v}$, cioè λ è un autovalore e $\mathbf{v}$ è un autovettore di A.

EsempioTrovare le soluzioni del sistema

$$\mathbf{y}' = A\mathbf{y}$$

con

$$A = \begin{pmatrix} -4 & 6 \\ -3 & 5 \end{pmatrix}$$

Troviamo gli autovalori di A:

$$|A - \lambda I| = (-4 - \lambda)(5 - \lambda) + 18 = (\lambda - 2)(\lambda + 1)$$

Quando $\lambda = 2$

$$A - 2I = \begin{pmatrix} -6 & 6 \\ -3 & 3 \end{pmatrix}$$

quindi l'autospazio ha equazione $-x + y = 0$ e $(1, 1)$ è un autovettore. Quando $\lambda = -1$ l'autospazio ha invece equazione $-x + 2y = 0$ e $(2, 1)$ è un autovettore. In corrispondenza di questi due autovettori abbiamo due soluzioni

$$\mathbf{y}_1(t) = e^{2t} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} e^{2t} \\ e^{2t} \end{pmatrix}$$

$$\mathbf{y}_2(t) = e^{-t} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2e^{-t} \\ e^{-t} \end{pmatrix}$$

In generale, in un sistema di n equazioni in n incognite cerchiamo n soluzioni linearmente indipendenti. Per ogni autovalore λ di A ed un relativo auutovettore $\mathbf{v}$, abbiamo una soluzione. Se queste sono n soluzioni lineramente indipendenti abbiamo concluso. Ma non sempre è così, cioè, come ben sappiamo, non sempre è possibile trovare una base di autovettori. Il problema dunque è quello di vedere come trovare "abbastanza" soluzioni nel caso di una matrice non diagonalizzabile, ed anche quando la matrice è diagonalizzabile solo sui complessi, dobbiamo vedere come trovare delle soluzioni mediante funzioni reali del nostro sistema differenziale. Vediamo da vicino alcuni esempi di dimensione 2. Sia dunque

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

Al variare di t la soluzione $t \mapsto \mathbf{y}(t)$ descrive una curva parametrica nel piano $\mathbb{R}^2$. Vogliamo cercare come sopra una soluzione esponenziale. L'equazione caratteristica di A è

$$t^2 - Tt + D = 0$$

dove $T = a_{11} + a_{22}$ è la traccia di A e D è il determinante di A. Allora gli autovalori sono dati da

$$\lambda = \frac{T \pm \sqrt{T^2 - 4D}}{2}$$

Distinguiamo dunque 3 casi:

(1) $T^2 - 4D > 0$, in cui abbiamo due radici distinte;
(2) $T^2 - 4D < 0$ in cui abbiamo due radici complesse coniugate;
(3) $T^2 - 4D = 0$ in cui abbiamo una radice doppia.

Il caso (1) è quello visto nell'esempio. La soluzione generale del sistema è

$$\mathbf{y}(t) = C_1\mathbf{y}_1(t) + C_2\mathbf{y}_2(t)$$

Se prendiamo invece la matrice

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

allora il polinomio caratteristico è t^2-2t+2 e quindi ha radici $\lambda = 1+i$ e $\overline{\lambda} = 1-i$. Si trova allora che $\mathbf{w} = \begin{pmatrix} 1 \\ 1+i \end{pmatrix}$ è un autovettore per $\lambda = 1+i$ mentre $\overline{\mathbf{w}} = \begin{pmatrix} 1 \\ 1-i \end{pmatrix}$ lo è per $\lambda = 1-i$. Allora possiamo prendere le soluzioni complesse

$$\mathbf{z}(t) = e^{\lambda t}\mathbf{w} = e^{(1+i)t}\begin{pmatrix} 1 \\ 1+i \end{pmatrix}$$

$$\overline{\mathbf{z}}(t) = e^{(1-i)t}\begin{pmatrix} 1 \\ 1-i \end{pmatrix}$$

Se ora prendiamo la parte reale e la parte immaginaria di $\mathbf{z}(t) = \mathbf{x}(t) + i\mathbf{y}(t)$ allora $\mathbf{x}(t)$ e $\mathbf{y}(t)$ sono soluzioni reali del sistema. Nel nostro esempio abbiamo

$$\mathbf{z}(t) = e^t(cos\ t + isen\ t)\left[\begin{pmatrix} 1 \\ 1 \end{pmatrix} + i\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right] =$$
$$e^t\left\{\left[cos\ t\begin{pmatrix} 1 \\ 1 \end{pmatrix} - sen\ t\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right] + i\left[cos\ t\begin{pmatrix} 0 \\ 1 \end{pmatrix} + sen\ t\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right]\right\}$$

e dunque

$$\mathbf{x}(t) = e^t\begin{pmatrix} cos\ t \\ cos\ t - sen\ t \end{pmatrix}$$

$$\mathbf{y}(t) = e^t\begin{pmatrix} sen\ t \\ cos\ t + sen\ t \end{pmatrix}$$

Esempio
Come altro esempio prendiamo

$$A = \begin{pmatrix} 1 & 4 \\ -1 & -3 \end{pmatrix}$$

e supponiamo di volere la soluzione che ha per condizione iniziale $\mathbf{x}(0) = \mathbf{v}$.

Abbiamo il polinomio caratteristico $p_A(t) = (t+1)^2$ e dunque -1 è un autovalore doppio. In questo caso l'autospazio è il nucleo di

$$A + I = \begin{pmatrix} 2 & 4 \\ -1 & -2 \end{pmatrix}$$

Poiché questa matrice $A + I$ ha rango 1 il suo nucleo ha dimensione $2 - 1 = 1$ che è anche la molteplicità dell'autovalore. Quindi non troviamo due autovettori indipendenti, e quindi dove trovare la seconda soluzione dell'equazione differenziale? In precedenza siamo stati guidati dall'analogia con il caso unidimensionale a cercare una soluzione esponenziale del tipo $\mathbf{x}(t) = e^{\lambda t}\mathbf{v}$. Il problema sembra essere che, nell'usare soltanto gli autovalori di A, perdiamo delle informazioni su A (perlomeno qiuando la matrice A non è diagonalizzabile). Forse, allora l'analogo preciso di $x(t) = x_0 e^{at}$ del caso unidimensionale è

$$\mathbf{x}(t) = e^{tA}\mathbf{v}$$

usando l'esponenziale della matrice che abbiamo recentemente definito. Osserviamo che nel caso in cui $\mathbf{v}$ è un autovettore di A per l'autovalore λ abbiamo in effetti che

$$e^{tA}\mathbf{v} = e^{t\lambda}\mathbf{v}$$

e quindi l'esponenziale della matrice nei casi visti sopra non è effettivamente necessario.

Immaginiamo dunque di scrivere $A = \lambda I + (A - \lambda I)$ e osserviamo che

$$e^{tA} = e^{t\lambda I}e^{t(A-\lambda I)}$$

(questo richiede una giustificazione ma la daremo in seguito ed è dovuta al fatto che la matrice $t\lambda I$ commuta con $t(A - \lambda I)$. Ora osserviamo che

$$e^{t\lambda I} = \sum_{n=0}^{\infty} \frac{(t\lambda I)^n}{n!} = I \sum_{n=0}^{\infty} \frac{t^n \lambda^n}{n!} = e^{t\lambda} I$$

mentre

$$e^{t(A-\lambda I)} = \sum_{n=0}^{\infty} \frac{t^n (A_\lambda I)^n}{n!}$$

Nel nostro caso l'autovalore è $\lambda = -1$ e quindi $A + I = \begin{pmatrix} 2 & 4 \\ -1 & -2 \end{pmatrix}$ e $(A + I)^2 = 0$ dunque la serie esponenziale, a priori infinita, è infatti molto breve:

$$e^{t(A-\lambda I)} = I + t(A_\lambda I) = \begin{pmatrix} 1 + 2t & 4t \\ -t & 1 - 2t \end{pmatrix}$$

In definitiva, la soluzione desiderata è quindi

$$\mathbf{x}(t) = e^{-t} \begin{pmatrix} 1 + 2t & 4t \\ -t & 1 - 2t \end{pmatrix} \mathbf{v}$$

dove $\mathbf{v}$ sono le condizioni iniziali.

Vediamo ancora qualche esempio in dimensione 3.

Esercizio

Risolvere il sistema di equazioni differenziali in cui

$$A = \begin{pmatrix} 0 & -1 & 0 \\ 4 & 4 & 0 \\ -1 & -1 & 2 \end{pmatrix}$$

Il polinomio caratteristico è $(t-2)^3$. Si verifica che $(A-2I)^3 = 0$ mentre $(A-2I)^2 \neq 0$ e quindi

$$\begin{aligned} e^{tA} &= e^{2t}(I + t(A - 2I) + \frac{t^2}{2}(A - 2I)^2) \\ &= e^{2t} \begin{pmatrix} 1 - 2t & -t & 0 \\ 4t & 1 + 2t & 0 \\ -t - t^2 & -t - \frac{t^2}{2} & 1 \end{pmatrix} \end{aligned}$$

Per trovare tre soluzioni indipendenti calcoliamo

$$e^{tA}\mathbf{v}$$

ove v varia in una base di $\mathbb{R}^3$ per esempio

$$\mathbf{x}_1(t) = e^{tA}\mathbf{e}_1 = e^{2t}\begin{pmatrix} 1-2t \\ 4t \\ -t-t^2 \end{pmatrix}$$

$$\mathbf{x}_2(t) = e^{tA}\mathbf{e}_2 = e^{2t}\begin{pmatrix} -t \\ 1+2t \\ -t-\frac{t^2}{2} \end{pmatrix}$$

$$\mathbf{x}_3(t) = e^{tA}\mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ e^{2t} \end{pmatrix}$$

Esercizio
Risolvere il sistema di equazioni differenziali in cui

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

Gli autovalori sono $\lambda_1 = -1$, con molteplicità algebrica 1 e molteplicità geometrica 1 con autovettore $v_1 = (1, 1, -2)$, e $\lambda_2 = -2$, con molteplicità algebrica 2 e molteplicità geometrica 1. Un suo autovettore è $v_2 = (1, 0, -1)$. Così possiamo trovare due soluzioni

$$x_1(t) = e^{-t}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

$$x_2(t) = e^{-2t}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

Per cercare una terza soluzione calcoliamo

$$e^{tA}\mathbf{v} = e^{t\lambda_2 I + t(A-\lambda_2 I)}\mathbf{v}$$

su un vettore $\mathbf{v}$ da determinarsi

$$e^{\lambda_2}t(\mathbf{v} + t(A-\lambda_2 I)\mathbf{v} + \frac{t^2}{2!}(A-\lambda_2)^2 + \cdots)$$

occorre e basta determinare un vettore $\mathbf{v}$ per cui la serie tra parentesi sia tronca cioè un vettore per cui $(A-\lambda_2 I)^k\mathbf{v} = 0$ per qualche k, cioè ci occorre un autovettore generalizzato! Possiamo verificare che

$$(A-\lambda_2 I)^2 = (A-\lambda_2 I) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 0 \end{pmatrix}$$

ha nucleo e dimensione 2. Oltre a $v_2 = (1, 0, -1)$ contiene ad esempio $v_3 = (1, 0, 0)$ e quindi $(A + 2I)^2 v = 0$ per cui la serie termina. Quindi

$$\mathbf{x}_3(t) = e^{tA}\mathbf{v}_3 = e^{t\lambda_2}(\mathbf{v}_3 + t(A + 2I)v_3) = e^{-2t}\begin{pmatrix} 1+t \\ 0 \\ -t \end{pmatrix}$$

Possiamo avere un'idea del comportamento delle soluzioni guardando ai grafici nel piano delle fasi. Ci sono vari casi con comportamenti diversi. Ne esamineremo un paio a titolo di esempio senza esaurire tutte le possibilità.

Consideriamo dunque il sistema

$$\mathbf{y}' = A\mathbf{y}$$

dove

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

e

$$\mathbf{y}(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}$$

L'equazione caratteristica è

$$t^2 - Tt + D = 0$$

dove $T = a_{11} + a_{22}$ è la traccia di A e D è il determinante di A.

I punti di equilibrio del sistema sono i punti $v \in \mathbb{R}^2$ per cui $Av = 0$. In altre parole, i punti di equilibrio corrispondono al nucleo di A.

Supponiamo ad esempio che la matrice abbia due autovalori reali e distinti e che, per di più, i due autovalori, λ_1 e λ_2 siano discordi. Per esempio, sia $\lambda_2 > 0$. La soluzione generale del sistema è

$$\mathbf{y}(t) = C_1 e^{\lambda_1 t}\mathbf{v}_1 + C_2 e^{\lambda_2 t}\mathbf{v}_2$$

dove C_1, C_2 sono opportune costanti che dipendono dalle condizioni iniziali, e $\mathbf{v}_1, \mathbf{v}_2$ sono due autovettori indipendenti. La particolare soluzione che si ottiene ponendo $C_2 = 0$ è una soluzione in cui ogni valore della funzione soluzione $\mathbf{y}(t)$ è un multiplo positivo di $C_1\mathbf{v}_1$, cioè giace sulla semiretta individuata da $\mathbf{v}_1$ se $C_1 > 0$ o in quella opposta se $C_1 < 0$. Inoltre la soluzione si avvicina asintoticamente a zero. Analogamente se $C_1 = 0$ abbiamo altre due possibili soluzioni che si svolgono sulla retta generata dall'altro autovettore, ma questa volta il moto parte dall'origine e si allontana per $t \leftrightarrow \infty$. In generale, una soluzione è data dalla composizione di questi moti. Le traiettorie sono simili a delle iperboli. iIn questa situazione si dice che l'origine è un punto di sella ed è un punto di equilibrio instabile. Ad esempio, questo è il caso con

$$A = \begin{pmatrix} 1 & 4 \\ 2 & -1 \end{pmatrix}$$

Nel caso in cui i due autovalori siano reali e distinti ma entrambi negativi, supponiamo $\lambda_1 < \lambda_2 < 0$, entrambe le soluzioni particolari convergono a zero lungo le corrispondenti semirette. Per capire meglio cosa succede nel caso di una soluzione generica scriviamola nella forma

$$\mathbf{y}(t) = e^{\lambda_2 t}(C_1 e^{(\lambda_1 - \lambda_2)t}\mathbf{v}_1 + C_2\mathbf{v}_2)$$

Il primo fattore tende a zero perché $\lambda_2 < 0$. Il secondo fattore, essendo $\lambda_1 - \lambda_2 < 0$ tende a $C - 2\mathbf{v}_2$. Di conseguenza il prodotto, che è la nostra soluzione, converge a zero ma lo fa diventando tangente alla semiretta generata da $C_2\mathbf{v}_2$. In tale caso l'origine si dice nodo asintoticamente stabile. Questo è il caso, ad esempio, di

$$A = \begin{pmatrix} -3 & -1 \\ -1 & -3 \end{pmatrix}$$

2.3 Alcuni Esercizi.
Esercizio 1 Si considerino i sottospazi di $\mathbb{R}^4$

$$S_1 = \{(a, 2a - c - d, c, d), a, c, d \in \mathbb{R}\}$$

e

$$S_2 = \{(\alpha, \alpha - \gamma, \gamma, 2\alpha + \gamma), \alpha, \gamma \in \mathbb{R}\}$$

(1) Determinare il generico vettore ed una base di $S_1 \cap S_2$.
(2) Determinare il generico vettore ed una base di $S_1 + S_2$.

Soluzione(1) È facile verificare che i vettori $(1, 2, 0, 0), (0, -1, 1, 0), (0, -1, 0, 1)$ costituiscono una base di S_1 e che quindi la dimensione di S_1 è 3. Cerchiamo i vettori di S_2 che risultano combinazione lineare dei vettori della base di S_1. Tali vettori devono verificare

$$rg \begin{pmatrix} 1 & 0 & 0 & \alpha \\ 2 & -1 & -1 & \alpha - \gamma \\ 0 & 1 & 0 & \gamma \\ 0 & 0 & 1 & 2\alpha + \gamma \end{pmatrix} = 3$$

dovrà allora essere $det = \gamma + \alpha = 0$. Ne segue che $S_1 \cap S_2 = \{(\alpha, 2\alpha, -\alpha, \alpha), \alpha \in \mathbb{R}\}$

(2) Una base di S_2 è data da $(1, 1, 0, 2), (0, -1, 1, 1)$ e $dim\ S_2 = 2$ Quindi per la formula di Grassmann

$$dim\ (S_1 + S_2) = dim\ S_1 + dim\ S_2 - dim\ S_1 \cap S_2 = 3 + 2 - 1 = 4$$

e quindi $S_1 + S_2 = \mathbb{R}^4$ una sua base è ad esempio quella canonica.

Esercizio
Dimostrare che la condizione necessaria e sufficiente affinché $V = ker\ L^r \oplus Im\ L^r$ è che la successione dei nuclei sia stabile al passo r cioè $ker\ L^r = ker\ L^{r+1}$.

Infatti, se $ker\ L^r = ker\ L^{r+1}$ andiamo a considerare un vettore $v \in ker\ L^r \cap Im\ L^r$. Da un lato, poiché $v \in Im\ L^r$ deve esistere un vettore $w \in V$ tale che $L^r(w) = v$. Ma essendo $v \in ker\ L^r$ avremo anche

$$0 = L^r(v) = L^r(L^r(w))$$

Questo significa che $w \in ker\ L^{2r}$ ma se la successione dei nuclei si è stabilizzata allora $ker\ L^{2r} = ker\ L^r$. Dunque $w \in ker\ L^r$ ma allora $L^r(w) = 0$ ossia $v = 0$. Abbiamo dunque dimostrato che se un vettore appartiene all'intersezione esso è necessariamente 0. Dunque la somma è diretta.

Viceversa, supponiamo che la somma sia diretta. Vogliamo dimostrare che allora

$$ker\ L^{r+1} = ker\ L^r$$

Sia dunque $v \in ker\ L^{r+1}$. Allora $L^{r+1}(v) = 0$, in altre parole $L^r(v) \in ker\ L \subset ker\ L^r$. Poiché chiaramente $L^r(v) \in Im\ L^r$ abbiamo trovato un vettore che è contenuto sia in $ker\ L^r$ che in $Im\ L^r$. Ma questa intersezione è nulla perché per ipotesi la somma è diretta. Quindi necessariamente $L^r(v) = 0$. Ossia $v \in ker\ L^r$. Abbiamo quindi visto che se un vettore v appartiene a $ker\ L^{r+1}$ esso appartiene anche a $ker\ L^r$. Poiché l'inclusione inversa è nota, abbiamo che $ker\ L^r = ker\ L^{r+1}$.

Esercizio
Sia L l'operatore di $\mathbb{R}^4$ rappresentato, rispetto alla base canonica, dalla matrice

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

(1) Si determinino tutti i sottospazi S di $\mathbb{R}^4$ tali che $L_{|S} = 0$;
(2) Si determinino tutti gli $r \in \mathbb{N}$ tali che $\mathbb{R}^4 = ker\ L^r \oplus Im\ L^r$.

Soluzione

(1) I sottospazi richiesti sono necessariamente sottospazi di $ker\ L$. Si può verificare che il rango di A è 3 e dunque $dim\ ker\ L = 1$. Più precisamente $ker\ L$ è l'insieme delle soluzioni del sistema lineare omogeneo (SLO) con matrice A, o, equivalentemente, con una sua forma ridotta a gradini e quindi abbiamo il seguente sistema

$$\begin{cases} x_1 + \frac{1}{2}x_3 = 0 \\ x_2 + x_3 = 0 \\ x_4 = 0 \end{cases}$$

ossia $ker\ L$ è lo spazio generato da $(1, 2, -2, 0)$. Dunque gli unici sottospazi che soddisfano la richiesta sono lo spazio $\{0\}$ e tutto il nucleo.

(2) Nell'esercizio precedente abbiamo visto che la condizione necessaria e sufficiente affinché $\mathbb{R}^4 = ker\ L^r \oplus Im\ L^r$ è che $ker\ L^r = ker\ L^{r+1}$.

Andiamo allora a studiare la successione dei nuclei. Abbiamo osservato nella prima parte dell'esercizio che il rango di A è 3 e la dimensione del nucleo è 1. Poiché si verifica anche che il rango di A^2 è 3 allora la successione dei nuclei si è già stabilizzata:

$$ker\ L = ker\ L^2$$

Allora $\mathbb{R}^4 = ker\ L^r \oplus Im\ L^r$ per qualunque $r \geq 1$.

Esercizio 3

Si considerino i due sottospazi di $\mathbb{R}^4$, S_1 di equazioni cartesiane $x_1 - x_3 = 0, x_2 = 0$ e S_2 di equazioni $x_1 + x_2 = 0, x_4 = 0$.

(1) Si dimostri che $S_1 + S_2 = \mathbb{R}^4$;

(2) dato l'endomorfismo $L_h : \mathbb{R}^4 \to \mathbb{R}^4$ rappresentato nella base canonica da

$$\begin{pmatrix} 5-2h & 3-2h & h-3 & -1 \\ -3+h & -1+h & 2 & 0 \\ 1-h & 1-h & h & 2h-3 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

con $h \in \mathbb{R}$, determinare h in modo che S_1 e S_2 siano invarianti per L_h. Per tale valore di h determinare, se esiste, una forma di Jordan di L_h.

Soluzione

(1) Una base di S_1 è data da $v_1 = (0, 0, 0, 1), v_2 = (1, 0, 1, 0)$ e una base di S_2 è $v_3 = (1, -1, 0, 0), v_4 = (0, 0, 1, 0)$. Poiché il rango di

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

è 4 la somma $S_1 + S_2$ è diretta e coincide con $\mathbb{R}^4$.

(2) Si ha $L_h(v_1) = (-1, 0, 2h-3, 2)$, $L_h(v_2) = (2-h, -1+h, 1, 0)$, $L_h(v_3) = (2, -2, 0, 0)$, $L_h(v_4) = (h-3, 2, h, 0)$. Affinché S_1 sia invariante: $L_h(v_1) \in S_1$. Per questo occorre e basta che $2h - 3 = -1$ e quindi $h = 1$. Ora verifichiamo che per $h = 1$ si ha

$$L_1(v_1) = 2v_1 - v_2, \quad L_1(v_2) = v_2,$$

$$L_1(v_3) = 2v_3, \quad L_1(v_4) = -2v_3 + v_4$$

e quindi S_1 e S_2 sono invarianti se e solo se $h = 1$. Nella base $\{v_1, v_2, v_3, v_4\}$ la matrice di L ha la forma a blocchi

$$\begin{pmatrix} 2 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

La forma di Jordan di questa matrice si può calcolare lavorando sui due blocchi separatamente e si ottiene

$$\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Esercizio
Sia data la matrice

$$A = \begin{pmatrix} \alpha & 1 & 0 & 2 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & \alpha & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

con $\alpha \in R$.

(1) Per quali valori di α A è diagonalizzabile?
(2) Determinare, se esiste, una forma di Jordan reale per $\alpha = 1$.

Soluzione
(1) Calcoliamo $p_A(t) = (t-\alpha)^2(t-1)^3$. Supponiamo $\alpha \neq 1$. Allora lo spettro di A è composto da 2 autovalori distinti: α con molteplicità algebrica 2, e 1 con molteplicità algebrica 3. Affiché A sia diagonalizzabile dovrà essere $rg(A - \alpha I) = 3$ e $rg(A - I) = 2$. Si verifica ora che la matrice $A - I$ ha rango 2 qualunque sia α. Inoltre $A - \alpha I$ ha due colonne nulle quindi basta guardare il rango di

$$\begin{pmatrix} 1 & 2 & -1 \\ 1-\alpha & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 1-\alpha & 0 \\ 0 & 0 & 1-\alpha \end{pmatrix}$$

Se $\alpha \neq 1$ il rango è certamente 3. In questo caso, quindi, A è diagonalizzabile. Se $\alpha = 1$ il polinomio caratteristico è $p_A(t) = (t-1)^5$ e $A - I$ ha rango 2 dunque il nucleo $ker\ (A - I)$ ha dimensione 3 e A non è diagonalizzabile. Poiché però A ammette l'autovalore 1 con molteplicità algebrica 5, e quindi ammette tutti autovalori reali, esiste una forma di Jordan reale. Per determinarla: abbiamo già osservato che $dim\ ker\ (A - I) = 3$ e quindi ci saranno 3 blocchi di Jordan. Si verifica che A ha indice di nilpotenza 2 e quindi l'indice dell'autovalore è 2. La forma di Jordan, a meno di permutazioni dei blocchi, deve quindi essere la seguente: $J_2(1) \oplus J_2(1) \oplus J_1(1)$.

Capitolo III: Spazi euclidei e spazi unitari.

Definizioni e concetti fondamentali.

Negli spazi vettoriali finora considerati non abbiamo introdotto nozioni di distanza, perpendicolarità, angoli etc. Queste nozioni infatti non sono conseguenza degli otto assiomi che definiscono gli spazi vettoriali. Per poter trattare queste questioni occorre introdurre un **prodotto interno o scalare**.

Cominciamo esaminando il caso di uno spazio vettoriale V definito su $\mathbb{R}$. Diremo che V è uno spazio vettoriale euclideo se è definita una applicazione

$$V \times V \to \mathbb{R}$$

che ad ogni coppia di vettori (u, v) di V associa uno scalare indicato con (u, v), oppure $u \cdot v$ oppure $(u|v)$ oppure $\langle u, v\rangle$, e tale che essa soddsfi le seguenti proprietà:

(1) $(\alpha u, v) = (u, \alpha v) = \alpha(u, v)$ per ogni $u, v \in V$, $\alpha \in \mathbb{R}$;
(2) $(w, u + v) = (w, u) + (w, v)$ per ogni $u, v, w \in V$;
(3) $(u, v) = (v, u)$ per ogni $u, v \in V$;
(4) $(u, v) = 0$ per ogni $v \in V$ implica $u = 0$.

Una applicazione che soddisfa solo le proprietà (1) e (2) si dice bilineare. Se soddisfa anche la (3) si dice bilineare simmetrica, se soddisfa anche la (4) si dice bilineare simmetrica non degenere. Se invece della (4) si richiede la proprietà

(4')
$$(u, u) \geq 0 \forall u \in V \text{ and } (u, u) = 0 \implies u = 0$$

allora si dice che la forma bilineare simmetrica è definita positiva.

Esercizio. Verificare che (4') implica (4) ma non viceversa.

Una forma bilineare simmetrica definita positiva si dice **prodotto scalare** e lo spazio euclideo si dice allora **proprio**.

Esempi

(1) $\mathbb{R}^n$ con il prodotto scalare canonico è uno spazio euclideo.
(2) Lo spazio vettoriale $C^0([a, b])$ delle funzioni continue sull'intervallo reale $[a, b]$ è uno spazio euclideo se definiamo $(f, g) = \int_a^b f(t)g(t)\, dt$.

Se V ha dimensione finita n e se $\{v_1, \dots , v_n\}$ è una base di V allora siano $u = \sum_i a_i v_i$ e $v = \sum_j b_j v_j$ le rispettive espressioni. Per le proprietà date sopra avremo

$$(u, v) = \sum_{i,j} a_i b_j (v_i, v_j)$$

che mostra come il prodotto scalare di due vettori sia determinato una volta che siano dati i prodotti scalari dei vettori di una base. Se allora poniamo $g_{ij} = (v_i, v_j)$ potremo scrivere

$$(u, v) = \sum_{i,j} a_i b_j g_{ij}$$

La quarta proprietà nella definizione di prodotto scalare si traduce allora come segue:

$$\sum_{i,j} a_i b_j g_{ij} = 0$$

per ogni scelta delle a_i, $i = 1, \dots , n$ implica che

$$b_j = 0, j = 1, \dots , n$$

In particolare, cioò significa che per ogni $i = 1, \dots , n$

$$(u, v_i) = \sum_{i,j} b_j g_{ij} = 0$$

Questo equivale a dire che il sistema lineare omogeneo

$$\sum_{i,j} b_j g_{ij} = 0$$

ammette solo la soluzione banale. È ben noto che condizione necessaria e sufficiente affinché questo accada è che il determinante della matrice (g_{ij}), che si dice matrice della forma, sia diverso da zero, condizione che quindi traduce analiticamente la quarta proprietà.

Diremo che due vettori sono ortogonali se il loro prodotto scalare è zero.

Il prodotto scalare usuale su $\mathbb{R}$ è definito positivo. Nel caso di uno spazio euclideo proprio, diremo norma del vettore v lo scalare $\|v\| = \sqrt{(v,v)}$. Questa norma induce su $\mathbb{R}^n$ una metrica. Chiameremo **unitario** oppure **normale** oppure **versore** un vettore di norma 1.

Possiamo ottenere un gran numero di esempi di forme bilineari prendendo $V = \mathbb{R}^n$ e considerando una qualunque matrice A $n \times n$ a coefficienti reali con la definizione

$$\langle x, y \rangle = x^t A y$$

Ad esempio, la matrice

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

definisce la forma bilineare su $\mathbb{R}^2$ data da

$$(x_1, x_2) \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$= x_1y_1 + 2x_1y_2 + 3x_2y_1 + 4x_2y_2$$

È facile anche vedere che la forma è simmetrica (o antisimmetrica) se e solo se la matrice è simmetrica (antisimmetrica). Se abbiamo una forma su uno spazio vettoriale V e fissiamo una base $\mathcal{B}$ di V allora possiamo considerare la matrice della forma rispetto alla base scelta. Ovviamente in tal caso, la matrice varia al variare della base scelta. Tuttavia, come nel caso degli operatori lineari, si può calcolare la relazione che intercorre tra due matrici diverse di una stessa forma bilineare. Precisamente abbiamo:

Proposizione 3.1. *Se A e A' sono due matrici che rappresentano una stessa forma bilineare rispetto a due basi diverse allora*

$$A = QA'Q^t$$

dove Q è una qualunque matrice $n \times n$ invertibile.

Dim.. Infatti se x, y sono le coordinate di due vettori u, v rispetto ad una base $\mathcal{B}$ mentre $x' = Qx$ e $y' = Qy$ sono le coordinate di u, v rispetto a $\mathcal{B}'$ allora

$$\langle u, v \rangle = x'^T A' y' = x^t Q^T A' Q y = x^T A y$$

deve quindi essere

$$Q^T A' Q = A$$

e abbiamo la conclusione. $\square$

Osserviamo che il prodotto scalare ”solito”, detto canonico o standard, su $\mathbb{R}^2$ (o su $\mathbb{R}^n$, $n \geq 2$) cioè

$$\langle x, y \rangle = x_1y_1 + x_2y_2$$

corrisponde alla matrice identità. Dalla proposizione precedente segue allora che le matrici che rappresentano una forma bilineare equivalente al prodotto scalare canonico sono quelle matrici A per cui

$$A = Q^T Q$$

dove Q è una matrice invertibile.

In generale, affinché una matrice rappresenti un prodotto scalare deve essere simmetrica e tale che

$$x^t A x > 0$$

per ogni $x \neq 0$. Una matrice siffatta si dice **definita positiva**.

TEOREMA 3.2. *Le seguenti proprietà di una matrice $n \times n$ reale A sono equivalenti:*

(1) *A rappresenta il prodotto scalare canonico di $\mathbb{R}^n$ rispetto ad una opportuna base di $\mathbb{R}^n$.*
(2) *Esiste una matrice invertibile P tale che $A = Q^T Q$.*
(3) *A è simmetrica e definita positiva.*

DIM.. Abbiamo visto sopra che (1) e (2) sono equivalenti e che (1) implica (3). Dobbiamo solo dimostrare che (3) implica (1). Questo segue dal procedimento di ortonormalizzazione di Gram-Schmidt che richiamiamo più avanti. Infatti se A è definita positiva, allora possiamo applicare il procedimento di Gram-Schmidt allo spazio euclideo relativo alla forma data da A e determinare quindi una base $\mathcal{B}'$ ortonormale. Rispetto a $\mathcal{B}'$ la nuova matrice A' della forma sarà tale che $Q^T A' Q = A$. Ma essendo $\mathcal{B}'$ ortonormale la matrice A' sarà uguale all'identità. Quindi avremo $Q^T I Q = A$ ossia la (2) che sappiamo equivalente a (1). □

Ma come riconoscere se una matrice simmetrica è definita positiva? A tale scopo dimostriamo delle altre condizioni equivalenti.

TEOREMA. *Le seguenti proprietà di una matrice $n \times n$ reale A sono equivalenti:*

(1) *$x^T A x > 0$ per ogni vettore non nullo x.*
(2) *Tutti gli autovalori di A sono positivi.*
(3) *Tutti i minori principali del tipo*

$$A_1 = (a_{11}) \quad A_2 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

sono positivi

PROPOSIZIONE 3.3 (DISUGUAGLIANZA DI SCHWARTZ). *Sussiste la seguente importante disuguaglianza:*

$$|(u, v)| \leq \|u\| \, \|v\|$$

ove $u, v \in V$, V è uno spazio euclideo, e il segno di uguaglianza vale se e solo se i due vettori sono proporzionali.

DIM. Infatti: se uno dei due vettori è nullo, non c'è niente da dimostrare. Altrimenti, sia t un numero reale arbitrario e consideriamo il vettore $tu + v$. Per la definizione di prodotto scalare definito positivo abbiamo

$$(tu + v, tu + v) \geq 0$$

per ogni valore di t. Calcolando abbiamo:

$$t^2(u, u) + 2t(u, v) + (v, v) \geq 0$$

ossia

$$\|u\|^2t^2 + 2t(u, v) + \|v\|^2 \geq 0$$

Ora, un trinomio di secondo grado è sempre nonnegativo se e solo se il discriminante è negativo o nullo, ossia

$$(u, v)^2 - \|u\|^2\|v\|^2 \leq 0$$

quindi

$$|(u, v)| \leq \|u\|\|v\|$$

Inoltre se vale l'ugualglianza significa che il discriminante è nullo e quindi esiste un valore t_0 che è soluzione dell'equazione di secondo grado cioè

$$\|u\|^2t_0^2 + 2t_0(u, v) + \|v\|^2 = 0$$

ossia

$$(t_0u + v, t_0u + v) = 0$$

ma questo implica che $t_0u + v = 0$ ossia u e v sono linearmente dipendenti. □

Esercizio
Dato il prodotto scalare canonico su $\mathbb{R}^2$. Calcolarne la matrice che lo rappresenta rispetto alla base canonica di $\mathbb{R}^2$, e rispetto alla base $(1, 1)^T, (1, 2)^T$.

Nello spazio vettoriale n-dimensionale $\mathbb{R}^n$ con prodotto scalare canonico definiamo componente ortogonale di u secondo v il vettore

$$\frac{(u, v)}{(v, v)}v$$

Osserviamo che questa definizione coincide, nel caso del solito prodotto scalare tra vettori euclidei del piano, con il prendere il modulo di u moltiplicato per il coseno dell'angolo tra u e v e infine moltiplicare questo scalare per il versore della direzione v. Il numero

$$c = \frac{(u, v)}{(v, v)}$$

si chiama coefficiente di Fourier di u rispetto a v. È facile verificare che $u - cv$ è ortogonale a v:

$$(u - cv, v) = (u, v) - c(v, v) = (u, v) - \frac{(u, v)}{(v, v)}(v, v) = 0$$

Dunque possiamo chiamare

$$\frac{(u, v)}{(v, v)}v$$

la proiezione ortogonale di u su v.

Come corollario della disuguaglianza di Schwarz abbiamo la seguente disuguaglianza triangolare:

$$\|u + v\| \leq \|u\| + \|v\|.$$

La disuguaglianza di Schwartz si può riscrivere

$$\frac{|(u, v)|}{\|u\| \ \|v\|} \leq 1$$

e dunque esiste uno ed un solo angolo θ compreso tra 0 e π tale che questo rapporto sia uguale a $cos\ \theta$. Questo angolo, per definizione, si chiama angolo tra i vettori u e v. Come già detto, il prodotto scalare permette di definire una metrica sullo spazio vettoriale euclideo ponendo

$$d(u, v) = \|u - v\|.$$

Questo permette di definire su V il concetto di limite e quindi di trasferire su tale spazio molti dei concetti e degli strumenti dell'analisi matematica.

Sistemi di vettori ortonormali. Un insieme di vettori di V di norma 1 e a due a due ortogonali si chiama *sistema di vettori ortonormali.* Un tale sistema $\{v_1, \dots, v_n\}$ è caratterizzato dalle condizioni

$$(v_i, v_j) = \begin{cases} 0 \text{ se } i \neq j \\ 1 \text{ se } i = j \end{cases}$$

TEOREMA 3.4. *I vettori di un sitema ortonormale sono linearmente indipendenti.*

□

Se dunque la dimensione di V è n ed il sistema ortonormale ha n vettori parleremo di base ortonormale.

TEOREMA 3.5. PROCEDIMENTO DI GRAM-SCHMIDT. *Dato un qualunque insieme di r vettori linearmente indipendenti $\{v_1, \dots, v_r\}$ è possibile costruire un sistema ortonormale di r vettori $\{e_1, \dots, e_r\}$ che generano lo stesso sottospazio U generato da $\{v_1, \dots, v_r\}$.*

Precisamente:

$$u_1 = v_1$$

$$u_2 = v_2 - \frac{(v_2, u_1)}{(u_1, u_1)} u_1$$

$$u_3 = v_3 - \frac{(v_3, u_2)}{(u_2, u_2)} u_2 - \frac{(v_3, u_1)}{(u_1, u_1)} u_1$$

In generale,

$$u_r = v_r - \frac{(v_r, u_{r-1}2)}{(u_{r-1}, u_{r-1})} u_{r-1} - \dots - \frac{(v_r, u_1)}{(u_1, u_1)} u_1$$

Dividendo ciascuno dei vettori u_i per il suo modulo otteniamo un sistema ortonormale di vettori. Se $r = n$ abbiamo una base ortonormale.

Esempio

Sia $\mathbb{R}_3[t]$ l'insieme di grado minore o uguale a 2 con prodotto scalare

$$(p(t), q(t)) = \int_{-1}^{1} p(t)q(t)\ dt$$

Determinare una base ortonormale a partire da $\{1, t, t^2\}$.

$$u_1 = 1, u_2 = t - \frac{t \cdot 1}{1 \cdot 1} 1 = t - \frac{\int_{-1}^{1} t\ dt}{\int_{-1}^{1} dt} = t$$

$$u_3 = t^2 - \frac{\int_{-1}^{1} t^3\ dt}{\int_{-1}^{1} t^2\ dt} - \frac{\int_{-1}^{1} t^2\ dt}{\int_{-1}^{1} dt} = t^2 - \frac{1}{3}$$

Questi, a meno di un fattore moltiplicativo costante, sono i cosidetti polinomi di Legendre.

Esercizio

Consideriamo l'insieme T di tutti i polinomi del tipo

$$\frac{a_0}{2} + a_1 cos\ t + b_1 sen\ t + a_2 cos\ 2t + b_2 sen\ 2t + \dots + a_n cos\ nt + b_n sen\ nt$$

con n intero fissato, $a_i, b_i \in \mathbb{R}$ e $0 \leq t \leq 2\pi$. Questo T è un sottopsazio vettoriale di dimensione $2n+1$ dello spazio vettoriale di tutte le funzioni continue definite nell'intervallo $0 \leq t \leq 2\pi$. Su T si può definire un prodotto interno

$$a_1(t)a_2(t) = \int_{-\pi}^{\pi} a_1(t)a_2(t)\ dt$$

Verificare che l'insieme

$$\{1, cos\ t, sen\ t, cos\ 2t, sen\ 2t, \dots, cos\ nt, sen\ nt\}$$

è una base ortogonale di T. (Suggerimento: occorre ricordarsi delle formule del tipo $\int sen\ kt\ sen\ ht\ dt = \frac{sen\ (k-h)t}{2(h-k)} - \frac{sen\ (k+h)t}{2(h+k)} + c$. Trovare una base ortonormale.

Data una funzione $f \in C(-\pi, \pi)$ con il prodotto scalare appena descritto, supponiamo che $\{\phi_1, \dots, \phi_{2n+1}\}$ sia la base ortonormale trovata sopra. Se $f \notin T$ possiamo comunque cercare di determinare l'elemento $g \in T$ che abbia distanza minima da f. Consideriamo allora un generico elemento di T:

$$g = \sum_{h=1}^{2n+1} a_h \phi_h$$

allora dobbiamo minimizzare

$$\|f - g\|^2$$

Ossia l'espressione

$$\|f - g\|^2 = (f, f) - 2 \sum_{h=1}^{2n+1} a_h (f, \phi_h) + \sum_{h=1}^{2n+1} a_h^2$$

Osserviamo che i coefficienti $c_h = (f, \phi_h)$ sono i coefficienti di Fourier di f rispetto alla base $\{\phi_h\}$, ossia sono le componenti ortogonali di f nella direzione ϕ_h. Allora

$$\|f\|^2 - 2 \sum_{h=1}^{2n+1} a_h c_h + \sum_{h=1}^{2n+1} a_h^2$$
$$\|f\|^2 + \sum_{h=1}^{2n+1} (a_h - c_h)^2 - \sum_{h=1}^{2n+1} c_h^2$$

Da questa espressione è evidente che la scelta dei coefficienti a_h che minimizza questa espressione è quella per cui il termine

$$\sum_{h=1}^{2n+1} (a_h - c_h)^2$$

si annulla cioè quando $a_h = c_h$. In altre parole la soluzione del nostro problema è

$$g = \sum_{h=1}^{2n+1} (f, \phi_h) \phi_h$$

Questa soluzione è l'approssimazione di una funzione continua mediante polinomi trigonometrici ortonormali. Un'altra maniera di pensare a questa soluzione è come la proiezione ortogonale di f sul sottospazio T.

Se W è un sottospazio vettoriale V sappiamo che ci sono infiniti supplementari. Se V è uno spazio euclideo allora esiste un supplementare univocamente determinato, il supplementare ortogonale

$$W^{\perp} = \{v \in V | (v, w) = 0, \forall w \in W\}$$

Si può definire una nozione di norma ed angoli anche in uno spazio vettoriale sui complessi, con qualche variazione:

(1) $(u, v) = \overline{(v, u)}$
(2) $(\lambda u + \mu v, w) = \lambda(u, w) + \mu(v, w)$
(3) $(u, u) \geq 0$ e $(u, u) = 0 \iff u = 0$

Uno spazio vettoriale su $\mathbb{C}$ con un tale prodotto interno si dice uno spazio vettoriale unitario e il prodotto interno si dice hermitiano.

Osservazione. La richiesta (1) nella definizione di prodotto hermitiano è necessaria perché se definissimo il prodotto interno con la richiesta che il prodotto sia simmetrico avremmo un contraddizione con la richiesta (3). Infatti vorremmo che $(u, u) \geq 0$ per qualunque u. In particolare anche per iu. Ma allora avremmo $(iu, iu) = -1(u, u)$ che non può essere ancora positivo.

Se L è un operatore su uno spazio unitario V si definisce l'operatore aggiunto L^* mediante la relazione

$$(L(x), y) = (x, L^*(y))$$

Un tale operatore se esiste è unico: Se ci fossero due operatori B e C con questa proprietà avremmo

$$(L(x), y) = (x, B(y)) = (x, C(y))$$

per ogni $x, y \in V$ allora

$$(x, B(x)) - (x, C(y)) = 0$$

ossia

$$(x, (B - C)(y)) = 0$$

ma allora il vettore $(B - C)(y)$ sarebbe ortogonale a tutti i vettori di V, il che implica che $(B - C)(y) = 0$ qualunque sia y, in altre parole $B(y) = C(y)$ per ogni $y \in V$ ossia $B = C$.

Vogliamo ora dimostrare l'esistenza di un tale operatore. Fissiamo una base ortonormale di V: $\mathcal{B} = \{u_1, \dots u_n\}$ e sia A la matrice di L rispetto a questa base. Prendiamo ora l'operatore T rappresentato in questa base dalla matrice $\overline{A}^T$ (coniugata e trasposta di A). Mostreremo che questo operatore soddisfa le condizioni di aggiunzione ossia

$$(L(u_i), u_j) = (u_i, T(u_j))$$

Calcoliamo $L(u_j)$, le cui componenti sono date dalla j-esima colonna della matrice A:

$$L(u_j) = a_{1j}u_1 + \cdots + a_{ij}u_i + \cdots + a_{nj}u_n$$

ma in una base ortonormale queste componenti sono le "proiezioni ortogonali sugli assi" ossia

$$= (L(u_j), u_1)u_1 + \cdots + (L(u_j), u_i)u_i + \cdots + (L(u_j), u_n)u_n$$

cioè $a_{ij} = (L(u_j), u_i)$. Un calcolo analogo per l'operatore T ci dice che, se indichiamo con b_{ij} i coefficienti della sua matrice, avremo

$$b_{ij} = (T(u_j), u_i) = \overline{a}_{ji}$$

Ma allora

$$(L(u_i), u_j) = a_{ji} = \overline{b}_{ij}$$

(per definizione di T)

$$= \overline{(T(u_j), u_i)} = (u_i, T(u_j))$$

(per la (1) del prodotto hermitiano) e dunque $T = L^*$.

Proposizione 3.6. *Se U è un sottospazio invariante per L allora $U^\perp$ è invariante per L^*.*

Dim. Sia $v \in U^\perp$ e calcoliamo $L^*(v)$. Per verificare che $L^*(v) \in U^\perp$ occorre verificare che $(L^*(v), u) = 0$ per ogni $u \in U$. Ora $(L(u), v) = 0$ perché $v \in U^\perp$ e $L(u) \in U$. Ma allora $(u, L^*(v)) = 0$ ossia $L^*(v) \in U^\perp$ e quindi $U^\perp$ è stabile per L^*. □

Può darsi il caso speciale e importante di un operatore che soddisfi $L = L^*$ in tal caso diremo che l'operatore è autoaggiunto o hermitiano cioè tale che

$$(L(u), v) = (u, L(v))$$

per ogni $u, v \in V$. Nel caso reale parliamo anche di operatore simmetrico. Per esempio, la matrice

$$\begin{pmatrix} 2 & i \\ -i & 1 \end{pmatrix}$$

è una matrice hermitiana. Ovviamente una matrice reale è hermitiana se e solo se è simmetrica.

Teorema 3.7. *Gli autovalori di un operatore hermitiano H sono reali*

Dim. Se v è un autovettore di H con autovalore λ allora $v \neq 0$ e

$$(Hv, v) = (v, Hv) \iff (\lambda v, v) = (v, \lambda v) \iff \lambda(v, v) = \overline{\lambda}(v, v)$$

e poiché $(v, v) \neq 0$ allora $\lambda = \overline{\lambda}$. □

Teorema 3.8. *La forma canonica di Jordan di un operatore hermitiano è diagonale.*

Dim. Basta dimostrare che ogni autovalore generalizzato ha rango 1. Sia v un autovettore generalizzato di rango $k \geq 2$ cioè $(H-\lambda I)^k v = 0$ mentre $(H-\lambda I)^{k-1} v \neq 0$. Allora

$$\big((H - \lambda I)^k v, (H - \lambda I)^{k-2} v\big) = 0$$

Poiché $H - \lambda I$ è autoaggiunto (infatti H è autoaggiunto e λ è reale) allora questo prodotto è uguale a

$$\big((H - \lambda I)^{k-1} v, (H - \lambda I)^{k-1} v\big) = 0$$

Ne segue che $(H - \lambda I)^{k-1} v = 0$ contro l'ipotesi. □

Teorema 3.9. *Gli autovettori di una matrice hermitiana corrispondenti ad autovalori distinti sono tra loro ortogonali*

Dim. Siano
$$Hv_i = \lambda_i v_i, \quad Hv_j = \lambda_j v_j$$
con $\lambda_i \neq \lambda_j$. Allora
$$(v_j, Hv_i) = (v_j, \lambda_i v_i) = \lambda_i (v_j, v_i)$$
(ricordiamo che $\lambda_i \in \mathbb{R}$). D'altra parte abbiamo anche
$$(v_j, Hv_i) = (Hv_j, v_i) = \lambda_j (v_j, v_i)$$
Dunque
$$0 = (\lambda_i - \lambda_j)(v_j, v_i)$$
ed essendo $(\lambda_i - \lambda_j) \neq 0$ ne segue che necessariamente $(v_j, v_i) = 0$. □

Esempio Sia V uno spazio unitario e sia W un suo sottospazio. Si consideri l'applicazione $P : V \to V$, proiezione ortogonale su W, allora P è un operatore hermitiano.

Infatti, sia $\{w_1, \dots, w_r\}$ una base ortonormale di W. Allora la proiezione ortogonale di un generico $v \in V$ su W è
$$w = \sum_{i=0}^{r} (v, w_i) w_i$$
(è anche la "migliore approssimazione" di v su W). Dunque l'operatore P è quello che associa ad ogni v la sua proiezione ortogonale su W.

Ci sarà utile osservare anche che anche $I - P$ è una proiezione ortogonale, ma sul sottospazio $W^\perp$. Infatti, $V = W \oplus W^\perp$ e se $v = w + w'$ allora $w = P(v)$ e quindi $w' = P'(v) = (I - P)(v)$. Dunque se $v_1, v_2 \in V$ abbiamo
$$(Pv_1, v_2) = (Pv_1, Pv_2 + v_2 - Pv_2) = (Pv_1, Pv_2) + (Pv_1, v_2 - Pv_2) = (Pv_1, Pv_2)$$
perché $(Pv_1, v_2 - Pv_2) = 0$ essendo questi due vettori ortogonali tra loro. Analogamente, osserviamo che $((I - P)v_1, Pv_2) = 0$ perché sono su spazi tra loro ortogonali, quindi posso aggiungere questa quantità:
$$\begin{aligned}(Pv_1, v_2) &= (Pv_1, Pv_2) = (Pv_1, Pv_2) + ((I - P)v_1, Pv_2) \\ &= (Pv_1 + v_1 - Pv_1, Pv_2) = (v_1, Pv_2)\end{aligned}$$
Dunque $P = P^*$ cioè P è hermitiano.

Esempio

$\mathbb{R}^3$ con il prodotto scalare canonico. Calcolare la proiezione ortogonale di $v = (-10, 2, 8)$ sul sottospazio $W = Span(3, 12, -1)$.
$$P_W(v) = \frac{(v, w)}{(w, w)} w = \frac{-30 + 24 - 8}{9 + 144 + 1}(3, 12, -1) = \frac{-14}{154}(2, 12, -1)$$
In generale l'espressione ortogonale $P : \mathbb{R}^3 \to W$ è data da
$$(x_1, x_2, x_3) \mapsto \frac{3x_1 + 12x_2 - x_3}{154}(3, 12, -1)$$
Ovviamente il rango di questa applicazione (che coincide con la dimensione della sua immagine) è 1 e dunque il suo nucleo ha dimensione 2. Precisamente *ker P*

ha equazione $3x_1 + 12x_2 - x_3 = 0$. Questa è l'equazione cartesiana del piano $W^\perp$. Con questa formula possiamo calcolare la matrice di P nella base canonica di $\mathbb{R}^3$.

$$P(1,0,0) = \frac{1}{154}3(3,12,-1)$$

$$P(0,1,0) = \frac{12}{154}(3,12,-1)$$

$$P(0,0,1) = \frac{-1}{154}(3,12,-1)$$

e la matrice è

$$A = \frac{1}{154}\begin{pmatrix} 9 & 36 & -3 \\ 36 & 144 & -12 \\ -3 & -12 & 1 \end{pmatrix}$$

che è una matrice simmetrica come ci aspettiamo.

Supponiamo di ripetere l'esercizio calcolando la matrice di P nella base data dai vettori $v_1 = (154,0,0), v_2 = (145,-36,3), v_3 = (-36,10,12)$ allora $Pv_1 = (9,36,-3) = (154,0,0) - (145,-36,3)$, $Pv_2 = (0,0,0)$, $Pv_3 = (0,0,0)$. Allora la matrice di P in questa base è

$$B = \begin{pmatrix} -1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

che non è né simmetrica né hermitiana. Attenzione dunque, la matrice di un operatore hermitiano non è sempre hermitiana, ma solo se è la matrice rispetto ad una base ortonormale.

3.2 Operatori Unitari. Tra tutti gli operatori possibili su spazi vettoriali ci sono certamente gli isomorfismi cioè applicazioni lineari biunivoche. Nel caso di spazi unitari richiederemo anche che un isomorfismo T abbia la proprietà che

$$(Tu, Tv) = (u, v)$$

cioè T sia una isometria ("movimento rigido"). Un isomorfismo di uno spazio unitario V in sé si dirà un operatore unitario.

Proposizione 3.10. *T è unitario se e solo se T trasforma basi ortonormali in basi ortonormali.*

Dim. Se T è unitario e $\{u_1, \dots, u_n\}$ è una base ortonormale di V allora $(u_i, u_j) = \delta_{ij}$. Ma

$$(Tu_i, Tu_j) = (u_i, u_j) = \delta_{ij}$$

e dunque $\{Tu_1, \dots, Tu_n\}$ è una base ortonormale.

Viceversa, se $\{u_1, \dots, u_n\}$ e $\{Tu_1, \dots, Tu_n\}$ sono basi ortonormali allora

$$(Tu_i, Tu_j) = (u_i, u_j) = \delta_{ij}$$

ma allora se v_1, v_2 sono due qualunque vettori di V abbiamo

$$(v_1, v_2) = (\sum a_i u_i, \sum b_j u_i) = \sum a_i \bar{b}_j$$

mentre

$$(Tv_1, Tv_2) = (\sum a_i Tu_i, \sum b_j Tu_i) = \sum a_i \bar{b}_j$$

essendo i risultati uguali concludiamo che T è un'isometria.

$\square$

Teorema 3.11. *U è unitario se e solo se $UU^* = U^*U = I$*

Dim. Se U è unitario allora U è invertibile e

$$(Uv, w) = (Uv, UU^{-1}w) = (v, U^{-1}w)$$

quindi $U^{-1} = U^*$

Viceversa, supponiamo che $UU^* = U^*U = I$, dobbiamo mostrare che U conserva il prodotto interno. Infatti,

$$(Uv, Uw) = (v, U^*Uw) = (v, Iw) = (v, w)$$

□

Matrice unitaria. Una matrice complessa si dice unitaria se $A^*A = I$ dove $A^* = \overline{A}^T$. (nel caso reale si dice che una matrice è ortogonale se $AA^T = I$.

Come sono fatte le matrici che rappresentano un operatore unitario? Una risposta semplice la si trova se si prendono le matrici che rappresntano un operatore unitario rispetto ad una base ortonormale. Infatti in tal caso le matrici sono esse stesse unitarie. Questo vuol dire in particolare che

$$\sum a_{ik}\overline{a}_{ik} = 1$$

mentre

$$\sum a_{ik}\overline{a}_{rk} = 0$$

se $i \neq r$.

Proposizione 3.12. *Se λ è un autovalore di un operatore unitario U allora $|\lambda| = 1$*

Dim. Infatti se x è un autovettore per λ allora

$$(x, x) = (Ux, Ux) = (\lambda x, \lambda x) = \lambda\overline{\lambda}(x, x)$$

quindi $\lambda\overline{\lambda} = 1$ ossia $|\lambda| = 1$. □

Raccogliamo ora in una proposizione alcune proprietà delle matrici unitarie, alcune delle quali ci sono già note. Ricordiamo la definizione: una matrice U si dice unitaria se $UU^* = I$.

Proposizione 3.13.

(1) *Una matrice U è unitaria se e solo se le sue colonne e le sue righe formano insiemi di vettori ortonormali.*
(2) *L'insieme delle matrici unitarie è chiuso rispetto ai prodotti ed alle inverse, cioè il prodotto di due matrici unitarie è ancora una matrice unitaria e l'inversa di una matrice unitaria è una matrice unitaria.*
(3) *Se λ è un autovalore di una matrice unitaria U allora $|\lambda| = 1$.*
(4) *Se U è unitaria allora $|det\ U| = 1$.*
(5) *U è unitaria se e solo se $(Ux, Uy) = (x, y)$ per ogni $x, y \in C^n$.*
(6) *U è unitaria se e solo se $\|Ux\| = \|x\|$ per ogni $x \in \mathbb{C}^n$.*

Dim. La prima affermazione è ovvia se ripensiamo al fatto che la verifica che una base è ortonormale significa verificare che il prodotto scalare fra due vettori della base è 0 se i vettori sono diversi ed è 1 se i vettori sono uguali. Ma calcolare tale prodotto scalare è esattamente la stessa cosa che moltiplicare le righe di U per le colonne di U^*. Dire che il risultato di tale prodotto è la matrice identità significa dire esattamente che il prodotto righe per colonne dà 0 quando gli indici di riga e colonna sono diversi mentre è 1 quando hanno lo stesso indice di riga e colonna.

(2) Siano U e V due matrici unitarie allora

$$(UV)^* = V^*U^* = V^{-1}U^{-1} = (UV)^{-1}$$

e

$$(U^{-1})^* = (U^*)^{-1} = (U^{-1})^{-1} = U$$

(3) Se $Uv = \lambda v$, $v \neq 0$, allora $\|Uv\| = \|\lambda v\|$. Ma allora calcolando il primo membro si ha:

$$\|Uv\| = (Uv)^T\overline{(Uv)} = (Uv)^*Uv = v^*U^*Uv = v^*v = \|v\|$$

Il secondo membro dà

$$\|\lambda v\| = (\lambda v)^*(\lambda v) = \overline{\lambda}\lambda v^*v = |\lambda|^2 v^*v = |\lambda|^2\|v\|$$

Uguagliando ne ricaviamo che $|\lambda|^2 = 1$ e quindi $|\lambda| = 1$.

(4) Ricordiamo che il determinante di una qualunque matrice quadrata è uguale al prodotto degli autovalori della matrice (si pensi ad esempio alla sua forma di Jordan che è una forma triangolare) allora per il punto precedente è ovvio che $|det\ U| = 1$.

(5) Se U è unitaria abbiamo

$$(Ux, Uy) = x^*U^*Uy = x^*y = (x, y).$$

Viceversa, se $(Ux, Uy) = (x, y)$ per ogni $x, y \in \mathbb{C}^n$ allora se u_i è la i-esima colonna di U dovrà essere $u_i = Ue_i$ (dove e_i è l'i-esimo vettore della base canonica e abbiamo applicato la definizione di matrice associata ad un operatore in una base data) quindi

$$(u_i, u_j) = (Ue_i, Ue_j) = (e_i, e_j) = \delta_{ij}$$

dunque le colonne sono ortonormali e U è unitaria per il punto (1).

(6) Se U è unitaria allora per (5) abbiamo $(Ux, Ux) = (x, x)$ e quindi $\|Ux\| = \|x\|$. Il viceversa, non è a priori così ovvio, anzi è abbastanza sorprendente. Esso sembra dire che se un operatore conserva le lunghezze allora esso conserva anche gli angoli. In effetti, è proprio così. Supponiamo che, per ogni $x \in \mathbb{C}^n$ si abbia $\|Ux\| = \|x\|$. Allora

$$(U(x-y), U(x-y)) = (Ux, Ux) - (Ux, Uy) - (Uy, Ux) + (Uy, Uy)$$

Per ipotesi questo è uguale a

$$(x-y, x-y) = (x, x) - (x, y) - (y, x) + (y, y)$$

Uguagliando e semplificando abbiamo

$$(Ux, Uy) + (Uy, Ux) = (x, y) + (y, x)$$

il che implica, essendo $(Uy, Ux) = \overline{(Ux, Uy)}$

$$(Ux, Uy) + \overline{(Ux, Uy)} = (x, y) + \overline{(x, y)}$$

Ossia che le parti reali di questi due numeri complessi coincidono. Possiamo poi ripetere questo calcolo con

$$(U(x+iy),U(x+iy)) = (Ux,Ux) - i(Ux,Uy) + i(Uy,Ux) + (Uy,Uy)$$

Per ipotesi questo è uguale a

$$(x+iy,x+iy) = (x,x) - i(x,y) + i(y,x) + (y,y)$$

Uguagliando e semplificando abbiamo

$$-i(Ux,Uy) + i\overline{(Ux,Uy)} = -i(x,y) + i\overline{(x,y)}$$

ossia

$$(Ux,Uy) - \overline{(Ux,Uy)} = (x,y) - \overline{(x,y)}$$

e quindi anche le parti immaginarie coincidono. Dunque i due numeri sono uguali

$$(Ux,Uy) = (x,y)$$

□

Esercizio
Verificare che le matrici reali ortogonali di ordine 2 sono del tipo

$$\begin{pmatrix} cos\ \theta & -sen\theta \\ sen\ \theta & cos\theta \end{pmatrix}$$

oppure

$$\begin{pmatrix} cos\ \theta & sen\theta \\ sen\ \theta & -cos\theta \end{pmatrix}$$

Il primo tipo ha determinante uguale a 1 e rappresenta una rotazione di un angolo θ intorno all'origine in senso antiorario. Il secondo tipo ha invece determinante -1 e rappresenta una riflessione intorno ad una retta per l'origine insieme ad una rotazione.

Vogliamo studiare il problema di stabilire sotto quali condizioni due matrici siano unitariamente equivalenti ossia quando rapresentino uno stesso operatore rispetto a due diverse basi ortonormali.

Il risultato di tipo generale in questo contesto è dato dal seguente

Teorema di Schur. *Una qualunque matrice $n \times n$ A a coefficienti complessi è unitariamente equivalente ad una matrice triangolare superiore.*

Dim. Siano $\lambda_1, \dots, \lambda_n$ gli autovalori di A. Sia λ_1 un autovalore di A e sia v_1 un relativo autovettore di norma 1. Consideriamo la base $\{v_1, v_2, \dots, v_n\}$ di $\mathbb{C}^n$ ottenuta estendendo $\{v_1\}$. Pur di applicare il procedimento di Gram-Schmidt possiamo supporre che questa sia una base ortonormale. Formiamo una matrice $n \times n$ che ha questi vettori come colonne: in tal modo otteniamo una matrice unitaria U_1. Operando il cambiamento di base relativo abbiamo

$$U_1^* A U_1 = \begin{pmatrix} \lambda_1 & * \\ 0 & A_1 \end{pmatrix}$$

una matrice a blocchi in cui A_1 è $(n-1) \times (n-1)$. La matrice A_1 ha autovalori $\lambda_2, \dots, \lambda_n$. Sia $v_2' \in \mathbb{C}^{n-1}$ un autovettore relativo a λ_2 e ripetiamo lo stesso ragionamento fatto sopra per trovare una matrice unitaria U_2 tale che

$$U_2^* A U_2 = \begin{pmatrix} \lambda_1 & * \\ 0 & A_2 \end{pmatrix}$$

una matrice a blocchi in cui A_2 è $(n-2)\times(n-2)$. Formiamo la matrice $n \times n$ unitaria

$$V_2 = \begin{pmatrix} 1 & 0 \\ 0 & U_2 \end{pmatrix}$$

La matrice U_1V_2 è ancora unitaria in quanto prodotto di matrici unitarie e, scritta come matrici a blocchi, $V_2^*U_1^*AU_1V_2$ ha la forma

$$\begin{pmatrix} \lambda_1 & * & \\ 0 & \lambda_2 & * \\ & 0 & A_2 \end{pmatrix}$$

In un numero finito di passi trovaimo dunque una matrice unitaria

$$U = U_1V_2V_3\cdots V_{n-1}$$

che trasforma A in una matrice triangolare superiore.

Esercizio Verificare che le matrici

$$T_1 = \begin{pmatrix} 1 & 1 & 4 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}$$

e

$$T_2 = \begin{pmatrix} 2 & -1 & 3\sqrt{2} \\ 0 & 1 & \sqrt{2} \\ 0 & 0 & 3 \end{pmatrix}$$

sono unitariamente equivalenti mediante

$$U = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix}$$

Questo esercizio ci mostra come due matrici triangolari superiori diverse possano essere equivalenti, e come non sia facile stabilire facilmente se due matrici sono unitariamente equivalenti o meno. Un risultato che ci aiuta a questo scopo è il seguente:

TEOREMA. *Se due matrici A e B sono unitariamente equivalenti allora*

$$\sum_{i,j=1}^{n} |b_{ij}|^2 = \sum_{i,j=1}^{n} |a_{ij}|^2$$

DIM. Basta osservare che $\sum_{i,j=1}^{n} |a_{ij}|^2 = TrA^*A$ e che quindi basta verificare che

$$TrA^*A = TrB^*B$$

Se $B = U^*AU$ allora $TrB^*B = TrU^*A^*UU^*AU = TrU^*A^*AU = TrA^*A$ dove l'ultima uguaglianza è dovuta al fatto che se due matrici sono simili allora hanno l stessa traccia.

Il viceversa non è vero. Per trovare delle condizioni sufficienti occorre considerare la traccia di tutti i possibili prodotti tra A e A^*. Più precisamente, occorre introdurre il concetto di monomio non commutativo nelle due variabili x e y: diremo che

$$W(x,y) = x^{m_1}y^{n_1}x^{m_2}y^{n_2}\cdots x^{m_k}y^{n_k}$$

è un monomio o una parola nelle variabili non commutative x e y di grado $m_1 + n_1 + m_2 + n_2 + \ldots m_k + n_k$. Abbiamo allora il seguente

Teorema (Specht)[S]. *Due matrici $A, B \in M_n(\mathbb{C})$ sono unitariamente equivalenti se e solo se*

$$TrW(A, A^*) = TrW(B, B^*)$$

per ogni parola $W(x, y)$ in due variabili non commutative.

Questo Teorema è stato successivamente migliorato da C. Pearcy [Pe]:

Teorema. *Due matrici $A, B \in M_n(\mathbb{C})$ sono unitariamente equivalenti se e solo se*

$$TrW(A, A^*) = TrW(B, B^*)$$

per ogni parola $W(x, y)$ in due variabili non commutative di grado al più $2n^2$.

Basta quindi verificare un numero finito di condizioni anche se questo numero è ancora molto elevato. In casi particolari è possibile ridurlo sensibilmente. Ad esempio per matrici 2×2 si dimostra che

Teorema. *Due matrici $A, B \in M_2(\mathbb{C})$ sono unitariamente equivalenti se e solo se*

$$\begin{aligned} &TrA = TrB; \\ &TrA^2 = TrB^2; \\ &TrA^*A = TrB^*B \end{aligned}$$

In generale, non ci si può aspettare in questo caso, una risposta semplice analoga al caso della forma canonica di Jordan. Ad esempio si può verificare che per $n > 2$, tutte le matrici $n \times n$ della forma

$$\begin{pmatrix} 1 & 1 & a_{13} & a_{14} & a_{15} & \dots & a_{1n} \\ 0 & 2 & 1 & a_{24} & a_{25} & \dots & a_{2n} \\ 0 & 0 & 3 & 1 & a_{35} & \dots & a_{3n} \\ 0 & 0 & 0 & 4 & 1 & \dots & a_{4n} \\ & & & \vdots & & & \\ 0 & 0 & 0 & 0 & 0 & \dots & n \end{pmatrix}$$

dove $a_{ij} \in \mathbb{C}$ per ogni coppia di indici, sono tutte simili tra loro e hanno forma di Jordan diagonale uguale alla matrice diagonale $diag(1, 2, 3, 4, \dots, n)$. Tuttavia si può verificare che queste matrici sono tutte a due a due unitariamente non equivalenti. [R].

Ci poniamo dunque una domanda un po' più abbordabile: sotto quali condizioni un operatore A su uno spazio unitario è unitariamente diagonalizzabile cioè ammette una base ortonormale di autovettori.

Osserviamo che se A è un operatore che ammette una base ortonormale di autovettori allora

$$A = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}$$

e quindi

$$A^* = \begin{pmatrix} \overline{\lambda}_1 & 0 & \dots & 0 \\ 0 & \overline{\lambda}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \overline{\lambda}_n \end{pmatrix}$$

è chiaro che $AA^* = A^*A$. Cioè se A ammette una base ortonormale di autovettori allora A e A^* commutano. La cosa sorprendente è che è vero anche il viceversa, cioè se $AA^* = A^*A$ allora A ammette una base ortonormale di autovettori. Chiamiamo normale un operatore che commuta con il suo aggiunto. Per esempio, se U è unitario allora il suo aggiunto è il suo inverso e quindi U certamente commuta con il suo aggiunto. Anche se A è autoaggiunto ovviamente allora A sarà normale.

Lemma 3.14. *Due operatori che commutano, in particolare A e A^* nel caso di un operatore normale, hanno un autovettore in comune.*

Dim. Sia $V(\lambda)$ un autospazio per A. Vediamo che $V(\lambda)$ è invariante per A^* cioè se $v \in V(\lambda)$ allora $A^*v \in V(\lambda)$. Infatti

$$A(A^*v) = A^*(Av)$$

perché A è normale, e

$$= A^*(\lambda v) = \lambda A^*v$$

cioè $A^*v \in V(\lambda)$. Ma allora la restrizione di A^* su $V(\lambda)$ possiede certamente un autovettore v_1 che quindi è un autovettore in comune per i due operatori. □

Osserviamo ora che per questo autovettore in comune si avrà $Av_1 = \lambda v_1$ e $A^*v_1 = \mu v_1$ ma essendo

$$(Av_1, v_1) = (v_1, A^*v_1)$$

$$(\lambda v_1, v_1) = (v_1, \mu v_1)$$

$$\lambda(v_1, v_1) = \overline{\mu}(v_1, v_1)$$

Dunque $\lambda = \overline{\mu}$

Sia ora S_1 lo spazio generato da v_1 e $T_1 = S_1^\perp$. Poiché S_1 è invariante sia per A che per A^* (vedi una proposizione precedente) anche T_1 lo è. Dunque possiamo ripetere il ragionamento visto prima e trovare un vettore $v_2 \in T_1$ che è un autovettore in comune per A e per A^*. Ovviamente $v_2 \perp v_1$. Anche qui si avrà

$$Av_2 = \lambda_2 v_2$$

$$A^*v_2 = \overline{\lambda}_2 v_2$$

Consideriamo quindi S_2 lo spazio generato da v_1 e v_2, e $T_2 = S_2^\perp$ e cosí via. Otterremo in tal modo una base ortogonale di autovettori che potrà facilmente essere normalizzata. Ciò dimostra che un operatore normale ammette una base ortonormale di autovettori.

Abbiamo anche dimostrato

Proposizione 3.15. *Se A è normale allora A e A^* hanno gli stessi autovettori ma autovalori tra loro coniugati.*

Possiamo allora enunciare il

Teorema spettrale 3.16. *Se A è normale in uno spazio di dimensione finita e se $\lambda_1, \dots, \lambda_r$ sono gli autovalori distinti allora*

1. *$V(\lambda_i) \perp V(\lambda_j)$ se $i \neq j$;*
2. *$V = V(\lambda_1) \oplus \cdots \oplus V(\lambda_r)$;*
3. *$A = \lambda_1 P_1 + \cdots + \lambda_r P_r$ dove P_i è l'operatore di proiezione ortogonale su $V(\lambda_i)$. ("risoluzione spettrale di A")*

Dim. La proprietà (1) è stata dimostrata per operatori autoaggiunti, nel Teorema 3.9. È un semplice esercizio modificare quella dimostrazione per ottenere lo stesso risultato per un operatore normale: con notazioni naturali analoghe a quelle del Teorema 3.9, partiamo da

$$\langle Av_i, v_j\rangle = \langle v_i, A^* v_j\rangle$$

quindi

$$\langle \lambda_i v_i, v_j\rangle = \langle v_i, \bar{\lambda}_j v_j\rangle$$

e

$$\lambda_i\langle v_i, v_j\rangle = \lambda_j\langle v_i, v_j\rangle$$

$$(\lambda_i - \lambda_j)\langle v_i, v_j\rangle = 0$$

dunque

$$\langle v_i, v_j\rangle = 0$$

La proprietà (2) è equivalente a dire che l'operatore è diagonalizzabile. Questo è stato dimostrato nella pagina precedente.

Infine la proprietà (3) è una semplice conseguenza delle prime due, in quanto le prime due ci garantiscono l'esistenza di una base ortonormale di autovettori, quindi per ogni $v \in V$ abbiamo una decomposizione $v = v_1 + \ldots + v_k$ e dunque $A(v) = \lambda_1 v_1 + \ldots + \lambda_k v_k$. Ciascuna delle componenti v_i coincide con $P_i(v)$ ove P_i è l'operatore proiezione ortogonale sull'autospazio $V(\lambda_i)$

□

Esercizio Trovare, se possibile, una base ortonormale di autovettori per la matrice

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & 2 & 1 \end{pmatrix}$$

Soluzione A è una matrice simmetrica (essendo reale è anche hermitiana). Sappiamo dunque che esiste una base ortonormale di autovettori di $\mathbb{R}^3$. Il polinomio caratteristico è

$$p_A(t) = (1-t)(t^2 - 2t - 4)$$

Ci sono dunque tre autovalori distinti: $t = 1, 1+\sqrt{5}, 1-\sqrt{5}$ Per l'autospazio $V(1)$ studiamo la matrice

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix}$$

che è equivalente per righe a

$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Dunque le equazioni dell'autospazio sono:

$$\begin{cases} x + 2y = 0 \\ z = 0 \end{cases}$$

Una base di questa retta è $\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$

Per $V(1+\sqrt{5})$:

$$\begin{pmatrix} -\sqrt{5} & 0 & 1 \\ 0 & -\sqrt{5} & 2 \\ 1 & 2 & -\sqrt{5} \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & -\sqrt{5} \\ -\sqrt{5} & 0 & 1 \\ 0 & -\sqrt{5} & 2 \end{pmatrix}$$

$$\sim \begin{pmatrix} 1 & 2 & -\sqrt{5} \\ 0 & 2\sqrt{5} & -4 \\ 0 & -\sqrt{5} & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & -\sqrt{5} \\ 0 & 2\sqrt{5} & -4 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\sim \begin{pmatrix} 1 & 2 & -\sqrt{5} \\ 0 & 1 & -\frac{2}{\sqrt{5}} \\ 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -\frac{1}{\sqrt{5}} \\ 0 & 1 & -\frac{2}{\sqrt{5}} \\ 0 & 0 & 0 \end{pmatrix}$$

Le equazioni sono dunque

$$\begin{cases} x = \frac{1}{\sqrt{5}} z \\ y = \frac{2}{\sqrt{5}} z \end{cases}$$

con base $\begin{pmatrix} 1 \\ 2 \\ \sqrt{5} \end{pmatrix}$

Infine: $V(1-\sqrt{5})$:

$$\begin{pmatrix} \sqrt{5} & 0 & 1 \\ 0 & \sqrt{5} & 2 \\ 1 & 2 & \sqrt{5} \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & \sqrt{5} \\ \sqrt{5} & 0 & 1 \\ 0 & \sqrt{5} & 2 \end{pmatrix}$$

$$\sim \begin{pmatrix} 1 & 2 & \sqrt{5} \\ 0 & -2\sqrt{5} & -4 \\ 0 & \sqrt{5} & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & \sqrt{5} \\ 0 & -2\sqrt{5} & -4 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\sim \begin{pmatrix} 1 & 2 & \sqrt{5} \\ 0 & 1 & \frac{2}{\sqrt{5}} \\ 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & \frac{1}{\sqrt{5}} \\ 0 & 1 & \frac{2}{\sqrt{5}} \\ 0 & 0 & 0 \end{pmatrix}$$

Le equazioni sono dunque

$$\begin{cases} x = -\frac{1}{\sqrt{5}} z \\ y = -\frac{2}{\sqrt{5}} z \end{cases}$$

con base $\begin{pmatrix} 1 \\ 2 \\ -\sqrt{5} \end{pmatrix}$

I tre autovettori trovati sono tra di loro ortogonali, dunque normalizzando troviamo una base ortonormale:

$$\mathcal{B} = \left\{ \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{10}} \\ \frac{2}{\sqrt{10}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{10}} \\ \frac{2}{\sqrt{10}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \right\}$$

Osserviamo che se indichiamo con P_1, P_2,P_3, le proiezioni ortogonali sugli autospazi abbiamo

$$A = P_1 + (1+\sqrt{5})P_2 + (1-\sqrt{5})P_3$$

la risoluzione spettrale di A.

A proposito di questo esercizio, osserviamo che esistono operatori diagonalizzabili che non sono simmetrici. In tal caso però la base di autovettori non sarà ortonormale. Ad esempio la matrice

$$A = \begin{pmatrix} 2 & 1 \\ 0 & 4 \end{pmatrix}$$

è chiaramente diagonalizzabile in quanto ha due autovalori distinti (poiché la matrice è triangolare i suoi autovalori sono gli elementi sulla diagonale principale). Eppure A non è simmetrica. Le due direzioni degli autovettori sono $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ che però non sono ortogonali.

Un altro esempio interessante è dato da

$$A = \begin{pmatrix} i & 2 \\ 2 & -i \end{pmatrix}$$

che è una matrice simmetrica ma non è ortogonalmente diagonalizzabile. Il punto, chiaramente, è che questa matrice non è hermitiana e non è neanche normale, (verificare). Si può comunque vedere che è diagonalizzabile ma gli autospazi non sono tra loro ortogonali.

Infine la matrice

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

non è hermitiana ma è normale e quindi è unitariamente diagonalizzabile. Verificare e determinare una decomposizione spettrale di questo operatore.

3.3 Il procedimento di Gram-Schmidt e la fattorizzazione QR.

Una matrice ortogonale è caratterizzata dall'avere delle colonne che costituiscono una base ortonormale di $\mathbb{R}^n$. Se A è una matrice $n \times n$ reale con colonne linearmente indipendenti (quindi non singolare) allora l'applicazione del procedimento di Gram-Schmidt alle colonne di A fornisce una fattorizzazione di A nel prodotto di una matrice Q con colonne ortonormali ed una matrice triangolare superiore R. Questa è la cosiddetta fattorizzazione QR. Per capire in che modo si ottiene questa fattorizzazione, siano $a_1, \dots, a_n$ le n colonne indipendenti di A e siano $q_1, \dots, q_n$ i vettori ortonormali ottenuti applicando Gram-Schmidt. Sappiamo che ad ogni passo $i = 1, \dots, n$

$$W_i = span(a_1, \dots, a_i) = span(q_1, \dots, q_i)$$

Ci sono quindi degli scalari $r_{1i}, r_{2i}, \dots, r_{ii}$ tali che

$$a_i = r_{1i}q_1 + r_{2i}q_2 + \cdots + r_{ii}q_i$$

per $i = 1, \dots, n$. In altre parole,

$$\begin{aligned}
a_1 &= r_{11}q_1 \\
a_2 &= r_{12}q_1 + r_{22}q_2 \\
&\vdots \\
a_n &= r_{1n}q_1 + r_{2n}q_2 + \cdots + r_{nn}q_n
\end{aligned}$$

ovvero in notazione matriciale:

$$A = [a_1 a_2 \dots a_n] = [q_1, q_2, \dots, q_n] \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ 0 & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & r_{nn} \end{pmatrix} = QR$$

Chiaramente la matrice Q ha colonne ortonormali ed è quindi una matrice ortogonale. Osserviamo inoltre che i coefficienti sulla diagonale principale di R sono tutti non nulli. Infatti se fosse $r_{ii} = 0$ allora a_i sarebbe una combinazione lineare di $q_1, \dots q_{i-1}$ e quindi si trova in W_{i-1}. Ma allora a_i sarebbe combinazione lineare di $a_1, \dots, a_{i-1}$ il che è impossibile perché i vettori $a_1, \dots, a_i$ sono linearmente indipendenti. Essendo R triangolare superiore ne segue che R è invertibile. Abbiamo quindi dimostrato il seguente teorema:

Teorema 3.17. *Se A è una matrice $m \times n$ con colonne linearmente indipendenti allora A si può fattorizzare come*

$$A = QR$$

dove Q è una matrice $m \times n$ con colonne ortonormali e R è una matrice invertibile triangolare superiore.

Osserviamo che si può sempre assumere che i coefficienti r_{ii} siano positivi. Infatti, se fosse $r_{ii} < 0$ possiamo sostituire q_i con $-q_i$ e r_{ii} con $-r_{ii}$.

È necessario che le colonne di A siano linearmente indipendenti. Infatti, se $A = QR$, essendo R invertibile abbiamo $Q = AR^{-1}$. In questa situazione si può dimostrare, per esercizio, che il rango di A è uguale al rango di Q. Ma il rango di Q è n perché le colonne di Q sono ortonormali e quindi linearmente indipendenti, e quindi anche il rango di A è n e quindi le sue colonne sono linearmente indipendenti.

Esempio. Trovare la fattorizzazione QR di

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

Applicando il procedimento di Gram-Schmidt alle colonne otteniamo

$$q_1 = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \quad q_2 = \begin{pmatrix} \frac{3\sqrt{5}}{10} \\ \frac{3\sqrt{5}}{10} \\ \frac{\sqrt{5}}{10} \\ \frac{\sqrt{5}}{10} \end{pmatrix}, \quad q_3 = \begin{pmatrix} -\frac{\sqrt{6}}{6} \\ 0 \\ \frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{3} \end{pmatrix}$$

Queste sono quindi le colonne di Q. Per trovare R sfruttiamo il fatto che $Q^TQ = I$, essendo Q ortogonale, e quindi

$$Q^TA = Q^TQR = IR = R$$

Quindi

$$R = Q^TA = \begin{pmatrix} 2 & 1 & \frac{1}{2} \\ 0 & \sqrt{5} & \frac{3\sqrt{5}}{2} \\ 0 & 0 & \frac{\sqrt{6}}{2} \end{pmatrix}$$

Quando la matrice A non ha colonne linearmente indipendenti, non possiamo più applicare il procedimento di Gram-Schmdit. Possiamo comunque ottenere una fattorizzazione QR di tipo più generale utilizzzando un metodo che ricorda il metodo di eliminazione di Gauss in cui una matrice è ridotta ad una forma triangolare (la forma gradini) mediante una successione di operazioni elementari.

Le matrici che corrispondono alle operazioni elementari dell'algoritmo di Gauss sono le matrici di Householder che introdurremo nella prossima sezione.

3.4 Matrici di Householder.

Sia $w \in \mathbb{R}^n$ un vettore di norma euclidea 1, ad esempio $w = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}$, e si consideri la matrice

$$H_w = I_3 - 2ww^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - 2 \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}})$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - 2 \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}$$

Questa si dice la matrice di Householder relativa al vettore w. Geometricamente essa rappresenta la riflessione rispetto all'iperpiano ortogonale a w. Nel nostro esempio, sia π il piano ortogonale a w, cioè $\pi = \{v | v^T w = 0\}$ che ha equazione cartesiana $x + z = 0$. Infatti $H_w(w) = w - 2ww^T w = w - 2w = -w$ (osserviamo che $w^T w = \|w\|^2 = 1$) quindi w viene trasformato nel suo opposto, mentre se $v \in \pi$ allora $H_w(v) = v - 2ww^T v = v$ (in quanto $w^T v = 0$), cioè v è fissato da H_w. È chiaro allora che H_w è un operatore ortogonale.

Altri esempi. Se e_1 è il primo vettore della base canonica di $\mathbb{R}^3$ allora $H_{e_1} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ inoltre $H_{e_2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ e $H_{e_3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$

È facile anche capire e verificare che se v e w sono proporzionali allora $H_v = H_w$.

Essendo H_w una riflessione è anche chiaro che $H_w^2 = I$. Ciò significa dunque che $H_w = H_w^{-1} = H_w^T$ (perché H_w è ortogonale) dunque H_w è anche simmetrica.

Proposizione 3.18. *Se x e y sono due vettori di $\mathbb{R}^n$ di uguale norma, allora, posto $w = \frac{x-y}{\|x-y\|}$ si ha $H_w(x) = y$*

Dim.

$$H_w(x) = x - 2ww^T x = x - 2\frac{x-y}{\|x-y\|}\frac{(x-y)^T}{\|x-y\|}x$$

$$= x - \frac{2}{(x-y)^T(x-y)}[(x-y)^T x](x-y)$$

dove abbiamo usato il fatto che $(x-y)^T(x-y)$ è la norma al quadrato mentre $[(x-y)^T x]$ è uno scalare. Questa espressione risulta essere uguale a $x - \frac{2}{2}(x-y) = y$ una volta che abbiamo verificato che

$$\frac{(x-y)^T x}{(x-y)^T(x-y)} = \frac{1}{2}$$

Infatti

$$\frac{(x-y)^T x}{(x-y)^T(x-y)} = \frac{x^T x - y^T x}{x^T x - y^T x - x^T y + y^T y} = \frac{x^T x - y^T x}{x^T x - 2x^T y + y^T y}$$

(essendo il prodotto scalare simmetrico). Essendo inoltre $x^T x = y^T y$, perché i vettori hanno per ipotesi la stessa norma, si ha allora

$$\frac{x^T x - y^T x}{2x^T x - 2x^T y} = \frac{1}{2}$$

□

Osservazione. Osserviamo che essendo H_w ortogonale, nel caso in cui x e y siano di norma differente non può esistere alcun H_w tale che $H_w(x) = y$.

Corollario 3.19. *Dato $x \in \mathbb{R}^n$ esiste sicuramente $w \in \mathbb{R}^n$ tale che $H_w(x) = \|x\|e_1$.*

Dim. La proposizione garantisce l'esistenza tranne nel caso in cui $x = \|x\|e_1$ ma in questo caso si può prendere H_w con w un qualunque vettore ortogonale a x. □

Le matrici di Householder si possono usare per ottenere la cosiddetta decomposizione QR di una matrice A. Ad esempio, se Q è quadrata, si può determinare una matrice ortogonale Q ed una matrice triangolare superiore R con elementi sulla diagonale principale non negativi tali che

$$A = QR$$

Ad esempio, se abbiamo da risolvere il sistema lineare $AX = B$ possiamo sostituire $QRX = B$. Se ora moltiplichiamo ambo i membri per $Q^{-1} = Q^T$ otteniamo

$$RX = Q^T B$$

e poiché RX è un sistema triangolare esso è facilmente risolubile per sostituzione.

Esercizio Sia

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

Si determini una decomposizione QR di A, con Q un prodotto di matrici di Householder ed R triangolare superiore con elementi diagonali non negativi.

Soluzione Si prenda $A_0 = A$. Procederemo per passi costruendo ad ogni passo delle matrici A_i ottenute modificando la matrice A. Per calcolare la matrice $A_1 = Q_0 A_0$ occorre determinare Q_0. Poiché la prima colonna di A è $\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}$ cioè un multiplo positivo di e_1 allora prendiamo Q_0 una matrice di Householder relativa ad un qualunque vettore ortogonale a e_1 che quindi lascia fissa la prima colonna ("la prima colonna è già a posto")(La matrice identità non è una matrice di Householder). Dunque $A_1 = A_0$. Fissiamo ora l'attenzione sulla sottomatrice $C = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$. Per calcolare $A_2 = Q_1 A_1$, occorre determinare Q_1. Fissiamo l'attenzione sulla prima colonna di C che è $c_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, poiché non è un multiplo positivo di $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ "dobbiamo aggiustarla". Si prenda allora

$$Q_1 = \begin{pmatrix} 1 & 0_{1\times 2} \\ 0_{2\times 1} & H_w \end{pmatrix}$$

dove H_w è la matrice di Householder relativa al vettore w così determinato: poiché $\|c_1\| = \sqrt{2}$ allora

$$w = \begin{pmatrix} -1 \\ 1 \end{pmatrix} - \sqrt{2}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1-\sqrt{2} \\ 1 \end{pmatrix}$$

Allora troviamo

$$H_w = \begin{pmatrix} \frac{-1-\sqrt{2}}{2+\sqrt{2}} & \frac{1+\sqrt{2}}{2+\sqrt{2}} \\ \frac{1+\sqrt{2}}{2+\sqrt{2}} & \frac{1+\sqrt{2}}{2+\sqrt{2}} \end{pmatrix} = \begin{pmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

Abbiamo infine

$$A_2 = Q_1 A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 2 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & \sqrt{2} & -\sqrt{2} \\ 0 & 0 & 0 \end{pmatrix}$$

Dunque

$$A_2 = Q_1 A_1 = \begin{pmatrix} 2 & 1 & 0 \\ 0 & \sqrt{2} & -\sqrt{2} \\ 0 & 0 & 0 \end{pmatrix}$$

è triangolare con elementi non negativi sulla diagonale principale.

Riassumendo: $A = A_0 = A_1, A_2 = Q_1 A_1 = Q_1 A$. Osserviamo che Q_1 è simmetrica e ortogonale per cui $Q_1 Q_1 = I$. Quindi

$$A = Q_1 A_2$$

è la decomposizione QR desiderata.

Osserviamo che non è necessario che la matrice A sia quadrata. In generale, se A è una matrice reale $m \times n$ con $m \geq n$ allora esiste una successione di al più n matrici di Householder di ordine m, $H_1, H_2, \dots, H_r$, tali che

$$H_r \cdots H_2 H_1 A = \begin{pmatrix} R \\ 0 \end{pmatrix}$$

con R triangolare superiore con elementi diagonali positivi o nulli.

Esercizio Sia $W = Span\left\{ v_1 = \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 5 \\ -1 \end{pmatrix} \right\}$ un sottospazio di $\mathbb{R}^3$ dotato di prodotto scalare canonico, determinare una base ortogonale di W mediante il procedimento di Gram-Schmidt.

Soluzione Otteniamo una base ortogonale scegliendo $u_1 = v_1$, $u_2 = v_2 - \frac{(v_2,u_1)}{(u_1,u_1)} u_1$ ossia

$$u_2 = \begin{pmatrix} 0 \\ 5 \\ -1 \end{pmatrix} - \frac{5+2}{9} \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} -\frac{14}{9} \\ \frac{38}{9} \\ \frac{5}{9} \end{pmatrix}$$

Normalizzando otteniamo una base ortonormale

$$e_1 = \frac{1}{3} u_1 = \begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \\ -\frac{2}{3} \end{pmatrix}, \frac{2}{\sqrt{185}} \begin{pmatrix} -\frac{14}{9} \\ \frac{38}{9} \\ \frac{5}{9} \end{pmatrix}$$

Osserviamo dall'esercizio precedente che è possibile scrivere i vettori v_1, v_2 in funzione dei vettori e_1, e_2 semplicemente "ricavandoli" dalle espressioni scritte sopra

$$v_1 = 3e_1, v_2 = \frac{\sqrt{185}}{2} e_2 + \frac{7}{3} e_1$$

Si può verificare allora che

$$\begin{pmatrix} 2 & 0 \\ 1 & 5 \\ -2 & -1 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & -\frac{14}{3\sqrt{185}} \\ \frac{1}{3} & \frac{38}{3\sqrt{185}} \\ -\frac{2}{3} & \frac{5}{3\sqrt{185}} \end{pmatrix} \begin{pmatrix} 3 & \frac{7}{3} \\ 0 & \frac{\sqrt{185}}{3} \end{pmatrix}$$

È possibile cioè ottenere una fattorizzazione QR di A tramite una matrice Q le cui colonne sono ortonormali e una matrice R che è triangolare superiore e con coefficienti non negativi sulla diagonale principale. Questa è un tipo di fattorizzazione QR. Come detto in precedenza una tale fattorizzazione si può ottenere anche tramite le matrici di Householder. Ad esempio, partendo dalla matrice

$$A = \begin{pmatrix} 2 & 0 \\ 1 & 5 \\ -2 & -1 \end{pmatrix}$$

fissiamo l'attenzione sulla prima colonna che vogliamo trasformare in una colonna del tipo $\begin{pmatrix} * \\ 0 \\ 0 \end{pmatrix}$ dove $*$ è uno scalare non negativo. Per fare questo tramite una matrice di Householder occorre che questo "nuovo" vettore abbia la stessa norma di quello "vecchio". Vogliamo trasformare

$$\begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix} \rightsquigarrow 3 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

basterà prendere $w = \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix} - 3 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ -2 \end{pmatrix}$. Quindi

$$H_w = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \frac{2}{6} \begin{pmatrix} 1 & -1 & 2 \\ -1 & 1 & -2 \\ 2 & -2 & 4 \end{pmatrix} = \begin{pmatrix} \frac{4}{6} & \frac{2}{6} & -\frac{4}{6} \\ \frac{2}{6} & \frac{4}{6} & \frac{4}{6} \\ -\frac{4}{6} & \frac{4}{6} & -\frac{2}{6} \end{pmatrix}$$

Chiamiamo questa matrice Q_1 e moltiplichiamo per A. Otteniamo

$$A_1 = Q_1 A = \begin{pmatrix} \frac{4}{6} & \frac{2}{6} & -\frac{4}{6} \\ \frac{2}{6} & \frac{4}{6} & \frac{4}{6} \\ -\frac{4}{6} & \frac{4}{6} & -\frac{2}{6} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 1 & 5 \\ -2 & -1 \end{pmatrix} = \begin{pmatrix} 3 & \frac{7}{3} \\ 0 & \frac{8}{3} \\ 0 & \frac{11}{3} \end{pmatrix}$$

Ora fissiamo l'attenzione sulla colonna $\begin{pmatrix} \frac{8}{3} \\ \frac{11}{3} \end{pmatrix}$ (che è solo una "parte" della seconda colonna) che vogliamo trasformare in

$$\begin{pmatrix} \frac{8}{3} \\ \frac{11}{3} \end{pmatrix} \rightsquigarrow \begin{pmatrix} \frac{\sqrt{185}}{3} \\ 0 \end{pmatrix}$$

Prendiamo $w = \begin{pmatrix} \frac{8-\sqrt{185}}{3} \\ \frac{11}{3} \end{pmatrix}$ Allora

$$H_w = \begin{pmatrix} 1 - \frac{2(8-\sqrt{185})^2}{9w^T w} & -\frac{22(8-\sqrt{185})}{9w^T w} \\ -\frac{22(8-\sqrt{185})}{9w^T w} & 1 - \frac{242}{9w^T w} \end{pmatrix}$$

Prendendo allora $Q_2 = 1 \oplus H_w$ abbiamo

$$Q_2 \begin{pmatrix} 3 & \frac{7}{3} \\ 0 & \frac{8}{3} \\ 0 & \frac{11}{3} \end{pmatrix} = \begin{pmatrix} 3 & \frac{7}{3} \\ 0 & \frac{\sqrt{185}}{3} \\ 0 & 0 \end{pmatrix}$$

Per verificare questo calcolo:

$$\frac{8}{3}(1 - \frac{2(8-\sqrt{185})^2}{9w^T w}) - \frac{11}{3}\frac{22(8-\sqrt{185})}{9w^T w} = \\ \frac{370\sqrt{185} - 2960}{1110 - 48\sqrt{185}}$$

che razionalizzando dà $\frac{\sqrt{185}}{3}$. Inoltre

$$-\frac{8}{3}\frac{22(8-\sqrt{185})}{9w^T w} + \frac{11}{3}(1 - \frac{242}{9w^T w}) = 0$$

Dunque

$$Q_2 Q_1 A = \begin{pmatrix} 3 & \frac{7}{3} \\ 0 & \frac{\sqrt{185}}{3} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} R \\ 0 \end{pmatrix}$$

e quindi

$$A = Q_1^T Q_2^T \begin{pmatrix} R \\ 0 \end{pmatrix}$$

è una decomposizione QR.

Diamo ora una interessante applicazione delle matrici di Householder che dimostra come esse siano i "mattoni" costituenti delle matrici ortogonali, ovvero, detto in maniera geometrica, la proposizione seguente mostra come ogni "movimento rigido" di uno spazio euclideo sia composto da riflessioni elementari attorno ad un iperpiano.

Proposizione 3.20. *Ogni matrice reale ortogonale Q di ordine n è il prodotto di al più n matrici di Householder.*

Dim. Applichiamo la fattorizzazione QR alla matrice $A = Q$ Allora otteniamo

$$Q = H_1 H_2 \cdots H_r R$$

dove $r \leq n$, R è triangolare superiore con elementi diagonali non negativi e H_i sono matrici di Householder. Ma allora

$$R = H_r \cdots H_2 H_1 Q$$

è una matrice ortogonale in quanto è prodotto di matrici ortogonali.

Poiché R è triangolare superiore anche la sua inversa, R^{-1}, è triangolare superiore. Ma essendo R ortogonale in quanto prodotto di matrici ortogonali allora $R^{-1} = R^T$. Ma R^T essendo la trasposta di una matrice triangolare superiore, deve essere triangolare inferiore. Dunque R^T deve essere contemporaneamente triangolare superiore e triangolare inferiore, dunque è diagoanle. Ma allora $RR^T = R^2 = I$ implica che

$$\begin{pmatrix} a_1^2 & & & \\ & a_2^2 & & \\ & & \ddots & \\ & & & a_n^2 \end{pmatrix} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}$$

e poiché $a_i \geq 0$ ne segue che $a_i = 1$ e quindi $R = I$. Cioè

$$Q = H_1 H_2 \cdots H_r$$

□

Altro esempio. Vogliamo costruire una base ortonormale di $\mathbb{R}^3$ con prodotto scalare canonico a partire dai vettori

$$v_1 = \begin{pmatrix} 3 \\ 0 \\ 4 \end{pmatrix}, v_2 = \begin{pmatrix} -1 \\ 0 \\ 7 \end{pmatrix}, v_3 = \begin{pmatrix} 2 \\ 9 \\ 11 \end{pmatrix}$$

Abbiamo allora $u_1 = \begin{pmatrix} 3 \\ 0 \\ 4 \end{pmatrix}$, $u_2 = v_2 - \frac{(v_2,u_1)}{(u_1,u_1)} u_1 = \begin{pmatrix} -4 \\ 0 \\ 3 \end{pmatrix}$ infine $u_3 = \begin{pmatrix} 0 \\ 9 \\ 0 \end{pmatrix}$

Normalizzando abbiamo $e_1 = \frac{1}{5} \begin{pmatrix} 3 \\ 0 \\ 4 \end{pmatrix}$, $e_2 = \frac{1}{5} \begin{pmatrix} -4 \\ 0 \\ 3 \end{pmatrix}$, $e_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$.

Consideriamo ora

$$A = \begin{pmatrix} 3 & -1 & 2 \\ 0 & 0 & 9 \\ 4 & 7 & 11 \end{pmatrix} \quad Q = \begin{pmatrix} \frac{3}{5} & -\frac{4}{5} & 0 \\ 0 & 0 & 1 \\ \frac{4}{5} & \frac{3}{5} & 0 \end{pmatrix}$$

(Q è ortogonale). Esprimendo e_i in funzione di v_i abbiamo

$$\begin{cases} v_1 = 5e_1 \\ v_2 = 5e_1 + 5e_2 \\ v_3 = 10e_1 + 5e_2 + 9e_3 \end{cases}$$

che ci dà la matrice $R = \begin{pmatrix} 5 & 5 & 10 \\ 0 & 5 & 5 \\ 0 & 0 & 9 \end{pmatrix}$ che è triangolare superiore con elementi diagonali non negativi. Allora $A = QR$:

$$\begin{pmatrix} 3 & -1 & 2 \\ 0 & 0 & 9 \\ 4 & 7 & 11 \end{pmatrix} \begin{pmatrix} \frac{3}{5} & -\frac{4}{5} & 0 \\ 0 & 0 & 1 \\ \frac{4}{5} & \frac{3}{5} & 0 \end{pmatrix} = \begin{pmatrix} 5 & 5 & 10 \\ 0 & 5 & 5 \\ 0 & 0 & 9 \end{pmatrix}$$

Esercizio Consideriamo i sottospazi V_1 e V_2 di $\mathbb{R}^4$ di equazioni, rispettivamente,

$$V_1 : \begin{cases} x_1 - x_2 = 0 \\ x_1 + x_3 = 0 \\ x_2 + x_4 = 0 \end{cases} \quad V_2 : \begin{cases} x_1 + x_4 = 0 \\ x_1 + x_2 + x_3 + x_4 = 0 \\ 2x_1 - x_2 + x_3 = 0 \end{cases}$$

Calcolare $V_1 + V_2$ e verificare se $V_1 + V_2$ è stabile rispetto all'operatore L definito da

$$\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} 0 \\ 0 \\ 0 \\ \sqrt{2} \end{pmatrix} \quad \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} \sqrt{2} \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$$

Verificare infine se L è ortogonale.

Soluzione Si può verificare che V_1 è una retta generata dal vettore $\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$ e che V_2 coincide con V_1 (si può vedere ad esempio che i sistemi sono equivalenti). Dunque $V_1 + V_2 = V_1$. Possiamo allora calcolare

$$L \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} = L \left[\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \right] = \begin{pmatrix} -\sqrt{2} \\ 0 \\ 0 \\ \sqrt{2} \end{pmatrix} \notin V_1$$

quindi V_1 non è stabile. Per vedere se L è ortogonale, calcoliamo la matrice di L rispetto ad una base ortogonale, per esempio quella canonica. $L \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}$,

$L \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}$, $L \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$, $L \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$

Allora la matrice desiderata è

$$\begin{pmatrix} 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \end{pmatrix}$$

la quale è costituita da colonne ortonormali, dunque è ortonormale. Quindi L è ortogonale.

Esercizio Si consideri in $\mathbb{C}^3$ l'operatore rappresentato rispetto alla base canonica da

$$C = \begin{pmatrix} 0 & \frac{1-i}{\sqrt{2}} & 0 \\ \frac{i}{\sqrt{2}} & 0 & \frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \end{pmatrix}$$

ed il sottospazio $S = \langle \begin{pmatrix} i \\ i \\ 1 \end{pmatrix} \rangle$. Determinare $S^\perp, CS, CS^\perp$. Per definizione

$$S^\perp = \{ \begin{pmatrix} a \\ b \\ c \end{pmatrix} \in \mathbb{C}^3 : (a, b, c) \begin{pmatrix} -i \\ -i \\ 1 \end{pmatrix} = 0 \} = \{ -ia - ib + c = 0 \}$$

quindi $S^\perp$ è lo spazio generato da $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -i \\ 0 \\ 1 \end{pmatrix}$. Allora

$$CS = \langle \begin{pmatrix} 0 & \frac{1-i}{\sqrt{2}} & 0 \\ \frac{i}{\sqrt{2}} & 0 & \frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} i \\ i \\ 1 \end{pmatrix} \rangle = \langle \begin{pmatrix} i+1 \\ i-1 \\ -1-i \end{pmatrix} \rangle$$

e

$$CS^{\perp} = \langle \begin{pmatrix} 0 & \frac{1-i}{\sqrt{2}} & 0 \\ \frac{i}{\sqrt{2}} & 0 & \frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} -1 & -i \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \rangle = \langle \begin{pmatrix} 1-i \\ -i \\ -i \end{pmatrix}, \begin{pmatrix} 0 \\ i+1 \\ -1-i \end{pmatrix} \rangle$$

Osserviamo anche che essendo C una matrice unitaria perché $CC^* = I$ si ha che allora C conserva il prodotto hermitiano e quindi essendo $S \perp S^{\perp}$ anche $CS \perp CS^{\perp}$.

Esercizio Data la matrice

$$A = \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{pmatrix}$$

stabilire se esiste una matrice unitaria U tale che $U^{-1}AU$sia diagonale e in caso affermativo determinarla.

Soluzione Basta verificare che la matrice sia normale ossia che commuta con A^*.

$$A^* = \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{pmatrix} = A^T$$

perché A è reale. Si verfica che $AA^* = I$ cioè A è ortogonale quindi normale. Dunque U esiste, Calcoliamo una base ortonormale di autovettori.

$$det \begin{pmatrix} \frac{\sqrt{2}}{2} - \lambda & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1-\lambda & 0 \\ -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} - \lambda \end{pmatrix}$$

$$= (\frac{\sqrt{2}}{2} - \lambda)^2(1-\lambda) - (-\frac{\sqrt{2}}{2})(\frac{\sqrt{2}}{2})(1-\lambda) = (1-\lambda)(\lambda^2 - \sqrt{2}\lambda + 1)$$

Quindi $\lambda = \frac{\sqrt{2}}{2}(1 \pm i)$. I tre autospazi sono a 2 a 2 ortogonali. Gli autovettori sono $V(1)$, generato da $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $V(\frac{\sqrt{2}}{2}(1+i))$, generato da $\begin{pmatrix} 1 \\ 0 \\ i \end{pmatrix}$ e $V(\frac{\sqrt{2}}{2}(1-i))$, generato da $\begin{pmatrix} 1 \\ 0 \\ -i \end{pmatrix}$. Una base ortonormale di $\mathbb{C}^3$ è $u_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $u_2 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ 0 \\ i\frac{\sqrt{2}}{2} \end{pmatrix}$ e $u_3 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ 0 \\ -i\frac{\sqrt{2}}{2} \end{pmatrix}$. Posto

$$U = \begin{pmatrix} 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 1 & 0 & 0 \\ 0 & i\frac{\sqrt{2}}{2} & -i\frac{\sqrt{2}}{2} \end{pmatrix}$$

si ha

$$U^{-1}AU = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{2}}{2}(1+i) & 0 \\ 0 & 0 & \frac{\sqrt{2}}{2}(1-i) \end{pmatrix}$$

Esercizio Dimostrare che la matrice di rotazione $R_{\frac{\pi}{2}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ di $\frac{\pi}{2}$ antioraria, è il prodotto di due riflessioni di Householder.

Soluzione Procediamo come nella fattorizzazione QR. $\begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ dunque prendiamo la differenza $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Allora

$$I_2 - 2\frac{vv^T}{v^Tv} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - 2\frac{\begin{pmatrix} -1 \\ 1 \end{pmatrix}(-1,1)}{(-1,1)\begin{pmatrix} -1 \\ 1 \end{pmatrix}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{2}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

che rappresenta la riflessione rispetto alla retta $x = y$. Prendiamo poi la riflessione rispetto all'asse y, $u = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, e $H_u = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ otteniamo

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = H_v = H_u$$

Osserviamo che $H_uH_v = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ è la rotazione di $\frac{\pi}{2}$ in senso orario.

Esercizio Data la matrice reale $B = \begin{pmatrix} 1 & -1 & 0 \\ c & 0 & -1 \\ 0 & d & 1 \end{pmatrix}$, $c, d \in \mathbb{R}$ determinare i valori di c, d in modo che esista in $\mathbb{R}^3$ una base ortonormale di autovettori di B. Per tali valori di c, d decomporre $\mathbb{R}^3$

$$\mathbb{R}^3 = S_1 \oplus S_2 \oplus S_3$$

dove i sottospazi S_i sono invarianti e $S_i \perp S_j, i \neq j$.

Soluzione Una base ortonormale esiste se e solo se B è simmetrica, dunque $c = -1, d = -1$ Per tali valori abbiamo $B = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 0 & -1 \\ 0 & -1 & 1 \end{pmatrix}$ e $|B - \lambda I| = (1-\lambda)[(1-\lambda)(-\lambda)-1]+[-(1-\lambda)] = (1-\lambda)(\lambda^2-\lambda-2) = 0$ dunque $\lambda = 2$ oppure $\lambda = -1$. Gli autovalori sono $1, 2, -1$.

$V(1)$: la matrice è $\begin{pmatrix} 0 & -1 & 0 \\ -1 & -1 & -1 \\ 0 & -1 & 0 \end{pmatrix}$ e le equazioni sono

$$\begin{cases} x + z = 0 \\ y = 0 \end{cases}$$

Una base è costituita da $\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$

$V(-1)$: la matrice è $\begin{pmatrix} 2 & -1 & 0 \\ -1 & 1 & -1 \\ 0 & -1 & 2 \end{pmatrix} \sim \begin{pmatrix} -1 & 1 & -1 \\ 0 & 1 & -2 \\ 0 & -1 & 2 \end{pmatrix}$ e le equazioni sono

$$\begin{cases} x - y + z = 0 \\ y - 2z = 0 \end{cases}$$

Una base è costituita da $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$

$V(2)$: la matrice è $\begin{pmatrix} -1 & -1 & 0 \\ -1 & -2 & -1 \\ 0 & -1 & -1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 0 \\ 0 & -1 & -1 \\ 0 & -1 & -1 \end{pmatrix}$ e le equazioni sono

$$\begin{cases} x + y = 0 \\ y + z = 0 \end{cases}$$

Una base è costituita da $\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$

Questi tre sottospazi danno la decomposizione desiderata.

Esercizio Data la matrice $A = \begin{pmatrix} 0 & h & 2 \\ 1 & k & 0 \\ 2 & 2k & 0 \end{pmatrix}$, $h, k \in \mathbb{R}$ determinare i valori di h, k in modo che esista una matrice ortogonale reale $M \in O_n(\mathbb{R})$ (l'insieme di tutte le matrici ortogonali reali di ordine n) tale che $M^{-1}AM$ sia diagonale. Determinare tale M per questi valori di h e k.

Soluzione Poiché A è reale, una tale matrice ortogonale esiste se e solo se A è simmetrica quindi $h = 1, k = 0$. Abbiamo dunque $A = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix}$ il cui polinomio caratteristico è $\lambda(5 - \lambda^2)$. Gli autovalori sono dunque $0, \sqrt{5}, -\sqrt{5}$. Gli autospazi sono:

$$V(0) = \langle \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix} \rangle$$

$$V(\sqrt{5}) = \langle \begin{pmatrix} 1 \\ \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \rangle$$

$$V(-\sqrt{5}) = \langle \begin{pmatrix} 1 \\ -\frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{pmatrix} \rangle$$

Per trovare la matrice M occorre ricordarsi che M è la matrice del cambiamento di base dalla base canonica alla base ortonormale di autovettori e quindi

$$M = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{10}} & -\frac{2}{\sqrt{10}} \end{pmatrix}$$

Si può verificare che

$$M^{-1}AM = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sqrt{5} & 0 \\ 0 & 0 & -\sqrt{5} \end{pmatrix}$$

Esercizio Data la matrice hermitiana

$$A = \begin{pmatrix} 1 & i & 0 & 0 \\ -i & 1 & 0 & 0 \\ 0 & 0 & 2 & 1+i \\ 0 & 0 & 1-i & 1 \end{pmatrix}$$

determinare una matrice unitaria U ed una matrice diagonale D tale che $A = U^{-1}DU$.

Soluzione Una tale matrice esiste perché A è hermitiana e dunque normale ($A = A^*$ e quindi A commuta con A^*.) Per calcolare gli autovalori:

$$\begin{vmatrix} 1-\lambda & i & & \\ -i & 1-\lambda & & \\ & & 2-\lambda & 1+i \\ & & 1-i & 1-\lambda \end{vmatrix}$$

$$= \begin{vmatrix} 1-\lambda & i \\ -i & 1-\lambda \end{vmatrix} \cdot \begin{vmatrix} 2-\lambda & 1+i \\ 1-i & 1-\lambda \end{vmatrix} =$$

$$= [(1-\lambda)^2 - 1][(2-\lambda)(1-\lambda) - 2] = (\lambda^2 - 2\lambda)(\lambda^2 - 3\lambda) = \lambda^2(\lambda-2)(\lambda-3)$$

Gli autovalori sono dunque: 0 con molteplicità algebrica 2, 2 con molteplicità algebrica 1, 3 con molteplicità algebrica 1. Per calcolare gli autospazi: $V(0)$, la matrice è

$$\begin{pmatrix} 1 & i & 0 & 0 \\ -i & 1 & 0 & 0 \\ 0 & 0 & 2 & 1+i \\ 0 & 0 & 1-i & 1 \end{pmatrix}$$

la quale, tramite le operazioni elementari sulle righe: $iR_1 + R_2 \mapsto R_2, \frac{1}{2}R_3 \mapsto R_3, (i-1)R_3 \mapsto R_4$ diventa

$$\begin{pmatrix} 1 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \frac{1+i}{2} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Le equazioni dell'autospazio sono dunque $x_1 + ix_2 = 0, x_3 + (\frac{1+i}{2})x_4 = 0$, una base è dunque

$$\begin{pmatrix} -i \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1-i \\ 2 \end{pmatrix}$$

È una base ortogonale di $V(0)$. Normalizziamo:

$$(-i, 1, 0, 0) \begin{pmatrix} i \\ 1 \\ 0 \\ 0 \end{pmatrix} = 2$$

e

$$(0,0,-1-i,2)\begin{pmatrix} 0 \\ 0 \\ -1+i \\ 2 \end{pmatrix} = 6$$

Una base ortonormale di $V(0)$ è

$$\begin{pmatrix} \frac{-i}{2} \\ \frac{1}{2} \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ \frac{-1-i}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}$$

Analogamente per $V(2)$:

$$\begin{pmatrix} -1 & i & & \\ -i & -1 & & \\ & & 0 & 1+i \\ & & 1-i & -1 \end{pmatrix}$$

Le equazioni: $-x_1 + ix_2 = 0$ con base $\begin{pmatrix} i \\ 1 \\ 0 \\ 0 \end{pmatrix}$ di norma $\sqrt{2}$.

Infine per $V(3)$:

$$\begin{pmatrix} -2 & i & & \\ -i & -2 & & \\ & & -1 & 1+i \\ & & 1-i & -2 \end{pmatrix}$$

Le equazioni: $-x_3 + (1+i)x_4 = 0$ con base $\begin{pmatrix} 0 \\ 0 \\ 1+i \\ 1 \end{pmatrix}$ di norma $\sqrt{3}$. La matrice unitaria desisderata è

$$U = \begin{pmatrix} -\frac{i}{\sqrt{2}} & 0\frac{i}{\sqrt{2}} & 0 & \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{(-1-i)}{\sqrt{6}} & 0 & \frac{(1+i)}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{pmatrix}$$

e

$$D = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & 2 & \\ & & & 3 \end{pmatrix}$$

Esercizio Si consideri, al variare di $h, k \in \mathbb{C}$ la matrice

$$A_{h,k} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ h-i & 0 & 0 & -ik \\ h-i & 0 & 1 & 0 \\ 0 & ik & 0 & 0 \end{pmatrix}$$

(1) Determinare per quali valori di h e k in $\mathbb{C}$ esiste una matrice invertibile M tale che $M^{-1}AM$ sia diagonale.
(2) Per $k \neq 0$, stabilire per quali valori di $h \in \mathbb{C}$ e $k \in \mathbb{C} - \{0\}$ esiste una matrice unitaria U tale che $U^{-1}AU$ sia diagonale.

Soluzione Studiamo il polinomio caratteristico

$$p_A(t) = \begin{vmatrix} 1-t & 0 & 0 & 0 \\ h-i & -t & 0-ik & \\ h-i & 0 & 1-t & 0 \\ 0 & ik & -t & \end{vmatrix} = (1-t)[t^2(1-t)+(ik)^2(1-t)] = (1-t)^2(t^2-k^2).$$

Gli autovalori sono dunque "1 con molteplicità algebrica 2 e k con molteplicità algebrica 1". In effetti, la frase tra virgolette non è corretta finché non precisiamo il valore di k. Essa risulta corretta purché $k \neq 0, \pm 1$. Infatti, se $k = \pm 1$ allora 1 è un autovalore con molteplicità algebrica 3 mentre -1 ha molteplicità algebrica 1. Se $k = 0$ allora 1 ha molteplicità algebrica 2 e 0 ha molteplicità algebrica 2. Studiamo questi casi separatamente.

Sia $k = 1$ allora

$$A_{h,1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ h-i & 0 & 0 & -i \\ h-i & 0 & 1 & 0 \\ 0 & i & 0 & 0 \end{pmatrix}$$

Occorre studiare il rango di $A_{h,1} - I$. Affinché la matrice sia diagonalizzabile occorre che il rango di questa matrice sia 1, cosicché l'autospazio abbia dimensione 3. Osservando le colonne di

$$A_{h,1} - I = \begin{pmatrix} 0 & 0 & 0 & 0 \\ h-i & -1 & 0 & -i \\ h-i & 0 & 0 & 0 \\ 0 & i & 0 & -1 \end{pmatrix}$$

vediamo che la seconda e la quarta sono proporzionali. Se $h-i \neq 0$ la prima colonna è indipendente dalla seconda. Occorre dunque che $h = i$. In tal caso il rango è 1. Abbiamo allora l'equazione $ix_2 - x_4 = 0$ ed una base è data da

$$\begin{pmatrix} 0 \\ 1 \\ 1 \\ i \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ i \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ i \end{pmatrix}$$

Se $k = -1$ abbiamo una situazione del tutto analoga e la matrice è diagonalizzabile se e solo se $h = i$.

Se $k = 0$ il rango di

$$A_{h,0} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ h-i & 0 & 0 & 0 \\ h-i & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

è 2 qualunque sia h e il rango di

$$A_{h,0} - I = \begin{pmatrix} 0 & 0 & 0 & 0 \\ h-i & -1 & 0 & 0 \\ h-i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

è 2 se e solo se $h = i$. In tal caso la matrice è diagonalizzabile.

Supponiamo infine che $k \neq \pm 1, 0$. Studiamo

$$A_{h,k} - I = \begin{pmatrix} 0 & 0 & 0 & 0 \\ h-i & -1 & 0 & -ik \\ h-i & 0 & 0 & 0 \\ 0 & ik & 0 & -1 \end{pmatrix}$$

Questa ha rango 2 se e solo se $h = i$. In conclusione se $h = i$ la matrice è diagonalizzabile qualunque sia k.

Se vogliamo una base ortonormale di autovettori: intanto $h = i$ per quanto precede. Dunque

$$A_{i,k} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & ik \\ 0 & 0 & 1 & 0 \\ 0 & -ik & 0 & 0 \end{pmatrix}$$

Dobbiamo imporre che $A_{i,k}$ sia normale, cioè

$$A_{i,k} A^*_{i,k} = A^*_{i,k} A_{i,k}$$

Poiché

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & ik \\ 0 & 0 & 1 & 0 \\ 0 & -ik & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \overline{-ik} \\ 0 & 0 & 1 & 0 \\ 0 & \overline{ik} & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & |ik|^2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & |ik|^2 \end{pmatrix}$$

la matrice è normale qualunque sia k.

3.5 Metodo dei Minimi Quadrati. Mostriamo come usare alcune tecniche apprese nelle pagine precedenti per risolvere un comune problema di approssimazione che è simile ad un problema già risolto in precedenza: quello della approssimazione di una funzione continua tramite polinomi trigonometrici, che in apparenza sembra invece molto diverso.

Supponiamo di avere dei dati sperimentali e di voler predire l'andamento futuro dell'esperimento, ci occorrerà avere una funzione che approssimi questa evoluzione. Il metodo dei minimi quadrati che ci apprestiamo ad illustrare, serve a questo scopo ed è dovuto a Gauss che lo utilizzò per calcolare l'orbita dell'asteroide Cerere scoperto nel 1801.

Come accennato sopra, ci ricordiamo che, in un esercizio precedentemente svolto a lezione, abbiamo visto che se W è un sottospazio di uno spazio euclideo V e $v \in V$ allora la migliore approssimazione di v in W è il vettore $\bar{v} \in W$ tale che

$$\|v - \bar{v}\| < \|v - w\|$$

per ogni vettore $w \in W$ differente da $\bar{v}$. Inoltre sappiamo che $\bar{v} = P_W(v)$ dove con $P_W(v)$ abbiamo indicato la proiezione ortogonale di v su W.

Supponiamo di avere tre punti $(1, 2), (2, 2), (3, 4)$ ottenuti per esempio da un ipotetico esperimento. Supponiamo di aver motivo di credere che i valori di x e y siano legati da una relazione lineare, cioè ci aspettiamo una relazione del tipo $y = a + bx$ con a e b da determinarsi. Avremo allora

$$2 = a + b \cdot 1$$
$$2 = a + b \cdot 2$$
$$4 = a + b \cdot 3$$

ossia

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix}$$

che "purtroppo" non ha soluzione. Cerchiamo allora una retta che sia la più vicina possibile alla soluzione ideale. È chiaro che qualunque retta prendiamo non potrà passare per i tre punti dati. Definiamo allora gli errori $\epsilon_1, \epsilon_2, \epsilon_3$ in corrispondenza dei tre punti, dove $\epsilon_1 = 2-(a+b)$, $\epsilon_2 = 2-(a+2b)$, $\epsilon_3 = 4-(a+3b)$, e consideriamo il vettore errore

$$\mathbf{e} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix}$$

Vogliamo minimizzare $\|\mathbf{e}\| = \sqrt{\epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2}$

Esercizio Quale tra le seguenti tre rette fornisce l'errore più piccolo per i punti dati? $y = 1 + x$, $y = -2 + 2x$, $y + \frac{2}{3}$.

In generale, supponiamo di avere n punti $(x_1, y_1), \dots, (x_n, y_n)$ ed una retta $y = a + bx$. Il vettore errore è

$$\mathbf{e} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

dove $\epsilon_i = y_i - (a + bx_i)$. La retta $y = a + bx$ che minimizza $\|\mathbf{e}\| = \epsilon_1^2 + \ldots + \epsilon_n^2$ si dice retta ai minimi quadrati.

Ci si propone di risolvere il sistema

$$\begin{aligned} a + bx_1 &= y_1 \\ a + bx_2 &= y_2 \\ &\vdots \\ a + bx_n &= y_n \end{aligned}$$

ovvero

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

cioè $Ax = b$. Il vettore errore è $b - Ax = \mathbf{e}$ ed è questo che vogliamo minimizzare.

Def. Chiamiamo una soluzione ai minimi quadrati del sistema $Ax = b$ il vettore $\bar{x} \in \mathbb{R}^n$ tale che

$$\|b - A\bar{x}\| \leq \|b - Ax\|$$

per ogni $x \in \mathbb{R}^n$.

Al variare di x in $\mathbb{R}^n$, Ax varia nell'immagine di A, o, detto in altra maniera, Ax varia nello spazio generato dalle colonne di A. Se indichiamo con W lo spazio generato dalle colonne di A, stiamo cercando la migliore approssimazione a b all'interno di W. Per quanto visto prima la soluzione desiderata è

$$A\bar{x} = P_W(b) \tag{1}$$

Quindi sembra che per risolvere il problema posto dobbiamo prima calcolare la proiezione ortogonale $P_W(b)$ e poi risolvere il sistema (1).

C'è però un metodo migliore. Osserviamo che

$$b - A\bar{x} = b - P_W(b) = perp_W(b)$$

cioè, la differenza $b - A\bar{x}$ è ortogonale a W, ossia ortogonale alle colonne di A. In definitiva, quindi, se a_i è la i-esima colonna di A abbiamo:

$$a_i^T(b - A\bar{x}) = 0$$

per ogni i. Equivalentemente:

$$A^T(b - A\bar{x}) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

ovvero

$$A^T b - A^T A\bar{x} = 0$$

ossia

$$A^T A\bar{x} = A^T b$$

Questo nuovo sistema di equazioni si dice sistema di equazioni normali per $\bar{x}$.

Avremo bisogno di un

Lemma 3.21. *Sia A una matrice $m \times n$. Il rango di A è uguale al rango della matrice $A^T A$.*

Dim. Osserviamo che $A^T A$ è una matrice quadrata $n \times n$. Il teorema della dimensione, ovvero, equivalentemente, il Teorema di Rouché-Capelli, ci dice che $n = rg(A) + dim\ KerA$, ma anche $n = rg(A^T A) + dim\ KerA^T A$. I ranghi delle due matrici saranno senz'altro uguali se dimostriamo che i nuclei sono uguali.

Per far ciò, supponiamo che $x \in KerA$. Allora $Ax = 0$ e quindi $A^T Ax = A^T 0 = 0$ cioè $x \in KerA^T A$. Viceversa: se $A^T Ax = 0$ allora $x^T A^T Ax = 0$ ma questo è il prodotto scalare di Ax per se stesso e quindi è zero se e solo se $Ax = 0$. $\square$

Corollario 3.22. *La matrice $A^T A$ è invertibile se e solo se le colonne di A sono linearmente indipendenti.*

Dim. Le colonne di A sono linearmente indipendenti vuol dire che il rango di A è n. Per il lemma, questo significa che anche $A^T A$ ha rango n, cioè invertibile. $\square$

Abbiamo quindi il seguente

Teorema 3.23.

a) *$\bar{x}$ è una soluzione ai minimi quadrati di $Ax = b$ se e solo se $\bar{x}$ è una soluzione del sistema di equazioni normali*

$$A^T A\bar{x} = A^T b$$

b) *Se A ha colonne linearmente indipendenti la soluzione ai minimi quadrati è unica ed è data da*

$$\bar{x} = (A^T A)^{-1} A^T b$$

Avendo osservato in precedenza che $A\bar{x} = P_W(b)$ e che $\bar{x} = (A^T A)^{-1} A^T b$, nel caso in cui le n colonne di A sono linearmente indipendenti, avremo allora che

$$A(A^T A)^{-1} A^T b = P_W(b)$$

In altre parole la matrice

$$A(A^T A)^{-1} A^T$$

è la matrice della trasformazione lineare

$$P_W : \mathbb{R}^m \to \mathbb{R}^m$$

che proietta ortogonalmente $\mathbb{R}^m$ sul sottospazio W delle colonne di A.

Esempio. Calcolare la matrice della proiezione ortogonale di $\mathbb{R}^3$ sul sottospazio W costituito dal piano di equazione cartesiana $x - y + 2z = 0$.

Soluzione. Troviamo una base di W:

$$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$$

e consideriamo la matrice rettangolare che ha questi vettori per colonne:

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Allora

$$A^T A = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

Quindi

$$(A^T A)^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}$$

e infine

$$A(A^T A)^{-1} A^T = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3}3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{5}{6} & \frac{1}{6} & -\frac{1}{3} \\ \frac{1}{6} & \frac{5}{6} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

è la matrice desiderata.

Def. La matrice $A^+ = (A^T A)^{-1} A^T$ si dice la pseudoinversa di A.

Possiamo allora dire che la soluzione ai minimi quadrati è semplicemente

$$\bar{x} = A^+ b$$

e che la matrice della proiezione ortogonale è AA^+, questo evidenzia il parallelismo di A^+ con la matrice A^{-1} in quanto permette di risolvere un sistema di equazioni lineari ("regola di Cramer") e il prodotto di A per la sua pseudoinversa è "quasi" l'identità, pensando ad una proiezione ortogonale come una identità "parziale".

Nell'esercizio che segue cerchiamo una parabola che meglio approssima quattro punti dati. Il problema, come vedremo, è del tutto analogo al precedente.

Siano dati $P_1 = (-2, 4), P_2 = (-1, 1), P_3 = (1, 1), P_4 = (2, -6)$. Cerchiamo una curva di equazione

$$y = a_0 + a_1 x + a_2 x^2$$

tale che

$$\begin{cases} a_0 - 2a_1 + 4a_2 = 4 \\ a_0 - a_1 + a_2 = 1 \\ a_0 + a_1 + a_2 = 1 \\ a_0 + 2a_1 + 4a_2 = -6 \end{cases}$$

Il sistema ha matrice

$$\begin{pmatrix} 1 & -2 & 4 & 4 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & -6 \end{pmatrix} \sim \begin{pmatrix} 1 & -2 & 4 & 4 \\ 0 & 1 & -3 & -3 \\ 0 & 3 & -3 & -3 \\ 0 & 4 & 0 & -10 \end{pmatrix}$$

$$\sim \begin{pmatrix} 1 & -2 & 4 & 4 \\ 0 & 1 & -3 & -3 \\ 0 & 0 & 6 & 6 \\ 0 & 0 & 12 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & -2 & 4 & 4 \\ 0 & 1 & -3 & -3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -10 \end{pmatrix}$$

Il che dimostra che il sistema è incompatibile.

Consideriamo allora

$$\overline{X} = (A^T A)^{-1} A^T B$$

con

$$A = \begin{pmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix}$$

$$A^T A = \begin{pmatrix} 2 & 0 & 5 \\ 0 & 5 & 0 \\ 5 & 0 & 17 \end{pmatrix}$$

$$(A^T A)^{-1} = \frac{1}{90} \begin{pmatrix} 85 & 0 & -25 \\ 0 & 9 & 0 \\ -25 & 0 & 10 \end{pmatrix}$$

quindi

$$\overline{X} = \frac{1}{90} \begin{pmatrix} 85 & 0 & -25 \\ 0 & 9 & 0 \\ -25 & 0 & 10 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ -2 * -1 & 1 & 2 & \\ 4 & 1 & 1 & 4 \end{pmatrix} \begin{pmatrix} 4 \\ 1 \\ 1 \\ -6 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 5 \\ -6 \\ -2 \end{pmatrix}$$

dunque $a_0 = \frac{5}{3}$ $a_1 = -2$, $a_2 = -\frac{2}{3}$ e la parabola

$$y^2 = \frac{5}{3} - 2x - \frac{2}{3} x^2$$

è la richiesta parabola ai minimi quadrati.

3.6 Decomposizione ai valori singolari. Sappiamo che se A è una matrice simmetrica allora A si fattorizza in $A = PDP^T$ dove P è una matrice ortogonale, e D è diagonale. Se A non è simmetrica, il meglio che possiamo fare, se A è diagonalizzabile, è fattorizzare $A = PDP^{-1}$ dove P non è necessariamente ortogonale. Risulta allora alquanto sorprendente apprendere che una matrice A qualunque, in generale rettangolare, è fattorizzabile nella forma $A = PDQ^T$ dove P e Q sono matrici ortogonali e D è diagonale.

Se pensiamo alle matrici come trasformazioni lineari, allora le matrici quadrate corrispondono ad endomorfismi e la fattorizzazione indicata sopra corrisponde, come sappiamo, a determinare una base ortonormale del dominio rispetto alla quale l'endomorfismo si rappresenta come una somma di dilatazioni $Au_i = \lambda u_i$ (o compressioni) in alcune direzioni (quelle degli autovettori). Nel caso di una matrice rettangolare, abbiamo una trasformazione tra due spazi di dimensioni, in generale, diverse. Quindi occorre trovare non una, bensì due basi ortonormali, una nel dominio e una nel codominio, se possibile, in modo che la trasformazione sia ancora una dilatazione in varie direzioni, cioè $Av_i = \lambda u_i$, dove vogliamo che non solo l'insieme $\{v_i\}$ sia ortonormale ma anche $\{u_i\}$.

Sia dunque A una matrice $m \times n$ allora $A^T A$ è una matrice $n \times n$ simmetrica e quindi può essere diagonalizzata ortogonalmente. Osserviamo che gli autovalori di $A^T A$ non solo sono tutti reali ma anche non negativi. Infatti, se v è un autovettore per $A^T A$ di norma 1, con autovalore λ, cioè $A^T Av = \lambda v$,

$$0 \leq \|Av\|^2 = (Av)^T Av = v^T A^T Av = v^T \lambda v = \lambda(v^T v) = \lambda\|v\|^2 = \lambda$$

Def. Se A è una matrice $m \times n$ definiamo **valori singolari** di A le radici quadrate degli autovalori di $A^T A$.

Esempio Trovare i valori singolari di

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

Calcoliamo

$$A^T A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

che ha autovalori $3, 1$:

$$\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = (2-\lambda)^2 - 1 = 4 - 4\lambda + \lambda^2 - 1 = \lambda^2 - 4\lambda + 3 = 0$$

I valori singolari sono quindi, ordinati in maniera decrescente:

$$\sigma_1 = \sqrt{3}, \sigma_2 = 1$$

Osserviamo anche che se $\{v_1, \ldots, v_n\}$ è una base ortonormale di autovettori di $A^T A$ allora, per ogni i,

$$\|Av_i\|^2 = (Av_i)^T Av_i = v_i^T A^T Av_i = v_i^T \lambda_i v_i = \lambda_i(v_i^T v_i) = \lambda_i\|v_i\|^2 = \lambda_i$$

e quindi $\sigma_i = \|Av_i\|$ è la lunghezza di Av_i.

Teorema 3.24 SVD. *Ogni matrice A è il prodotto*

$$A = P\Sigma Q^T$$

Dove P e Q sono matrici ortogonali (in generale di dimensioni diverse) e Σ è una matrice pseudodiagonale delle stesse dimensioni di A.

Dim. Supponiamo infatti che A sia una matrice $m \times n$ e siano $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ i valori singolari non nulli di A e $\sigma_{r+1} = \sigma_{r+2} = \ldots = \sigma_n = 0$ i restanti valori singolari. Sia $\{v_1, \ldots, v_n\}$ una base ortonormale di $\mathbb{R}^n$ formata da autovettori di $A^T A$ allora consideriamo la matrice ortogonale Q

$$Q = (v_1, \ldots v_n)$$

che ha i vettori v_i per colonne: essa è una matrice ortogonale $n \times n$. Osserviamo ora che $\{Av_1, \ldots, Av_n\}$ è un sistema ortogonale di vettori di $\mathbb{R}^m$. Infatti, se $i \neq j$

$$(Av_i)^T Av_j = v_i^T A^T Av_j = v_i^T \lambda_j v_j = \lambda_j (v_i^T v_j) = 0$$

Ricordando allora che $\sigma_i = \|Av_i\|$ e che i primi r sono non nulli, possiamo normalizzare $\{Av_1, \ldots, Av_r\}$ ottenendo $\{u_1, \ldots, u_r\}$ dove

$$u_i = \frac{1}{\sigma_i} Av_i$$

Allora $\{u_1, \ldots, u_r\}$ è un sistema ortonormale di $\mathbb{R}^m$, ma se $r < m$ non sarà una base di $\mathbb{R}^m$. Possiamo tuttavia completare $\{u_1, \ldots, u_r\}$ fino ad ottenere una base ortonormale $\{u_1, \ldots, u_m\}$ di $\mathbb{R}^m$. Sia dunque

$$P = (u_1, \ldots, u_m)$$

la matrice che ha per colonne i vettori u_i: anch'essa una matrice ortogonale $m \times m$. Vogliamo ora dimostrare che $A = P\Sigma Q^T$ dove Σ è una matrice delle stesse dimensioni di A e che è *pseudodiagonale* cioè non è necessariamente quadrata ma i coefficienti di posto (i, j) per $i \neq j$ sono nulli. Per far ciò, mostreremo che $AQ = P\Sigma$.

$$\begin{aligned} AQ &= A(v_1, \ldots v_n) = (Av_1, \ldots Av_n) = (Av_1, \ldots Av_r, 0 \ldots 0) \\ &= (\sigma_1 u_1, \ldots \sigma_r u_r, 0 \ldots 0) = (u_1, \ldots u_r, u_{r+1} \ldots u_m) \begin{pmatrix} \sigma_1 & \ldots & 0 \\ & \ddots & \\ 0 & \ldots & 0 \end{pmatrix} = P\Sigma \end{aligned}$$

□

Questa è la cosiddetta fattorizzazione ai valori singolari di A. Le colonne di P si dicono vettori singolari a sinistra di A e le colonne di Q si dicono vettori singolari a destra di A.

Esempio Trovare la decomposizione ai valori singolari di

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

Soluzione

$$A^T A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{vmatrix} 1-\lambda & 1 & 0 \\ 1 & 1-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{vmatrix} = (1-\lambda)\begin{vmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix}$$
$$= (1-\lambda[(1-\lambda)^2-1] = (1-\lambda)(\lambda-2)\lambda$$

Dunque gli autovalori sono $\lambda_1 = 0, \lambda_2 = 1, \lambda_3 = 2$ con autovettori rispettivamente:

$$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$$

Questi tre vettori sono automaticamente ortogonali. Normalizzando abbiamo

$$\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$$

i valori singolari di A sono, ordinati in modo decrescente,

$$\sigma_1 = \sqrt{2}, \sigma_2 = 1, \sigma_3 = 0$$

Quindi

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Per trovare P:

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2} \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

quindi

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

Infine la decomposizione ai valori singolari è

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{pmatrix} = P\Sigma Q^T$$

Osservazione. La decomposizione può essere riscritta in maniera differente come segue

$$A = P\Sigma Q^T = (u_1 \dots u_r u_{r+1} \dots u_m) \begin{pmatrix} \sigma_1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} v_1^T \\ \vdots \\ v_r^T \\ v_{r+1}^T \\ \vdots \\ v_n^T \end{pmatrix}$$

$$= (u_1 \dots u_r) \begin{pmatrix} \sigma_1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & \sigma_r \end{pmatrix} \begin{pmatrix} v_1^T \\ \vdots \\ v_r^T \end{pmatrix} = (\sigma_1 u_1 \dots \sigma_r u_r) \begin{pmatrix} v_1^T \\ \vdots \\ v_r^T \end{pmatrix}$$

$$= \sigma_1 u_1 v_1^T + \dots + \sigma_r u_r v_r^T$$

che possiamo pensare come una generalizzazione del terzo punto del teorema spettrale.

La decomposizione ai valori singolari, spesso abbreviata con SVD ("singular value decomposition"), è importante sia dal punto di vista teorico che pratico.

Esercizio. Calcolare la decomposizione ai valori singolari di

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

Soluzione.

$$A^T A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

che ha polinomio caratteristico $\lambda^2 - \lambda$ e quindi autovalori $\lambda = 2$ e $\lambda = 0$. I valori singolari sono dunque $\sigma_1 = \sqrt{2}$ e $\sigma_2 = 0$. A ha rango 1. Gli autospazi di $A^T A$ hanno equazioni $x - y = 0$ e $x + y = 0$ (due rette perpendicolari). La matrice Q è quindi

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

Per calcolare P:

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2} \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

estendiamo $\{u_1\}$ ad una base ortonormale di $\mathbb{R}^2$ aggiungendo $u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ e quindi

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

ed infine

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

Abbiamo visto che nel caso in cui $A^T A$ è invertibile allora la matrice $A^+ = (A^T A)^{-1} A^T$ è una pseudo inversa di A la quale permette di risolvere il problema ai minimi quadrati. Nel caso in cui la matrice $A^T A$ non è invertibile la soluzione ai minimi quadrati non è più unica, tuttavia, tra le infinite soluzioni, ne esiste una di lunghezza minima. Si dimostra che tale soluzione si trova attraverso la generalizzazione del concetto di matrice pseudo inversa. Come segue.

Def. Se A è una qualunque matrice e $A = P\Sigma Q^T$ una sua SVD definiamo la sua pseudo inversa di Moore e Penrose la matrice

$$A^+ = Q\Sigma^+ P^T$$

dove

$$\Sigma^+ = \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

essendo

$$\Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$$

Ad esempio con la matrice

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

si ha

$$\begin{aligned} A^+ &= \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

Infine abbiamo

Teorema. *Il sistema $Ax = b$ possiede un'unica soluzione $\bar{x}$ ai minimi quadrati di lunghezza minima data da*

$$\bar{x} = A^+ b$$

dove A^+ è la matrice pseudoinversa.

Esercizio. Supponiamo che sia A una matrice 3×2 a colonne ortonormali, diciamo $\mathbf{a}_1$ e $\mathbf{a}_2$. Calcolare la sua SVD e A^+.

Soluzione. Osserviamo che $A^T A = I_2$ la matrice identità 2×2, quindi i valori singolari sono $\sigma_1 = \sigma_2 = 1$ e dunque

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

la matrice Q le cui colonne sono una base ortonormale di autovettori di $A^T A$ coincide con I_2, mentre per calcolare P dobbiamo prendere come colonne i vettori

$$A \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

che non è altro che la prima colonna di A, $\mathbf{a}_1$, e

$$A \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

che è la seconda colonna di A, $\mathbf{a}_2$. Dobbiamo infine estendere questi due vettori ad una base ortonormale di $\mathbb{R}^3$, chiamiamo $\mathbf{a}_3$ questo terzo vettore. Abbiamo allora che $P = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$. La SVD cercata è allora

$$(\mathbf{a}_1, \mathbf{a}_2) = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3) \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Ora possiamo quindi calcolare A^+:

$$A^+ = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)^T = A^T$$

3.7 Esercizi.

1. Dare un esempio di matrice a coefficienti complessi 3×3 non diagonale ma diagonalizzabile.
2. Dare un esempio di matrice complessa 3×3 non diagonalizzabile.
3. Enunciare il Teorema di Cayley-Hamilton, illustrarlo con un esempio e dare un rapido abbozzo della sua dimostrazione.
4. Dare la definizione di matrice nilpotente e spiegare come essa può essere caratterizzata come conseguenza del Teorema di Cayley-Hamilton. Illustrare con un esempio.
5. Scrivere ogni possibile forma di Jordan di un operatore nilpotente su uno spazio vettoriale di dimensione 5. Quante ce ne sono? Quali sono?
6. Trovare la forma canonica di

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

7. Sia $\mathcal{F}$ l'insieme delle successioni di tipo Fibonacci. Dimostrare che $\mathcal{F}$ è uno spazio vettoriale di dimensione 2. Dimostrare che esiste una base di $\mathcal{F}$ costituita da successioni geometriche. Dedurne la formula di Binet:

$$f_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n$$

8. Illustrare la questione dell'estensione di variabile reale o complessa ad una funzione di variabile matriciale. Enunciare il problema ed illustrare la soluzione spiegando i risultati (lemmi, proposizioni) principali e fornendo qualche esempio significativo. Calcolare $sen\pi A$ dove

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

9. Dare la definizione di prodotto scalare e/o hermitiano e illustrare qualche risultato notevole con opportuni esempi.
10. Data la matrice

$$A = \begin{pmatrix} 3 & -\frac{i}{\sqrt{2}} & -\frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{5}{2} & \frac{1}{2} \\ \frac{i}{\sqrt{2}} & \frac{1}{2} & \frac{5}{2} \end{pmatrix}$$

Verificare che A è hermitiana. Determinare una matrice unitaria U che la diagonalizzi. Scrivere A come combinazione lineare di matrici di proiezione ortogonale.

11. Sia data la matrice

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

Trovare due soluzioni reali indipendenti del sistema di equazioni differenziali $\mathbf{y}' = A\mathbf{y}$.

12. Determinare la trasformazione di Householder di $\mathbb{R}^3$ che trasforma $\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$ in $\begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix}$

13. Quali delle seguenti matrici è unitaria? Trovare l'inversa della matrice unitaria.
$$A = \begin{pmatrix} \frac{i}{\sqrt{2}} & -\frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{i}{\sqrt{2}} \end{pmatrix}$$
$$B = \begin{pmatrix} 1+i & 1+i \\ 1-i & -1+i \end{pmatrix}$$
14. Siano date le matrici
$$A = \begin{pmatrix} 2 & 8 & 2 \\ 1 & 7 & -1 \\ -2 & -2 & 1 \end{pmatrix}$$
e
$$Q = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{pmatrix}$$
e supponiamo di aver verificato che le colonne di Q sono ottenute applicando il procedimento di Gram-Schmidt alle colonne di A. Determinare una matrice triangolare superiore R tale che $A = QR$
15. Calcolare il polinomio minimo di
$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$
16. Determinare i valori singolari della matrice $\begin{pmatrix} 3 \\ 4 \end{pmatrix}$
17. Trovare la forma canonica J di
$$A = \begin{pmatrix} 5 & -2 & -6 & 1 & -13 \\ \frac{9}{2} & -1 & -\frac{9}{2} & 0 & -\frac{27}{2} \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & -2 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$
e calcolare $cos\ \pi J$

Bibliografia

[A] Abate, M., *Algebra Lineare* McGraw-Hill 2000

[Ab] Abeasis, S., *Algebra Lineare e Geometria.* Zanichelli 1990

[Ar] Artin, M., *Algebra* Bollati-Boringhieri 1997.

[B] Beardon A.F., *Algebra and Geometry* Cambridge University Press 2005.

[C] Cattaneo Gasparini, I. *Complementi di Algebra e Geometria* Veschi ed.

[G] Gantmacher, F.R., *The Theory of Matrices* Chelsea Publishing Company, 1959.

[H] R. Howe, *Very Basic Lie Theory* The American Mathematical Monthly, Vol. 90, No. 9, (Nov. 1983), 600-623.

[KM] Kostrikin, A.I., Manin, Yu. I. *Linear Algebra and Geometry* Gordon and Breach Science Publishers

[M1] Maroscia, P., *Geometria e Algebra Lineare* Zanichelli 2002.

[M2] Maroscia, P., *Problemi di Geometria* Zanichelli 1999.

[P] Poole, D. *Linear Algebra, a Modern Introduction* Thomson, 2006.

[S] Strang, G., *Introduction to Linear Algebra,*Wellesley-Cambridge Press Third Edition 2003.

[Pe] Pearcy, C. "A complete Set of Unitary Invariants for Operators Generating Finite W^*-Algebras of Type I", *Pacific J. Math.* 12 (1962), 1405-1416.

[Pe1] Pearcy, C. "A complete Set of Unitary Invariants for 3×3 Complex Matrices", *Trans. Amer. Math. Soc.* 104 (1962), 425-429.

[R] Radjavi, H. "On unitary Equivalence of Arbitrary Matrices", *Trans. Amer. Math. Soc.* 104, (1962), 363-373.

[S] Specht,W. "Zur Theorie der Matrizen II," *Jahresbericht der Deutschen Mathematiker Vereinung* 50 (1940), 19-23.