Geometria 2
La forma canonica di Jordan

Corso di Laurea in Matematica

Alberto Albano

30 novembre 2017

In tutto quello che segue, se non altrimenti indicato, gli spazi vettoriali considerati sono di dimensione finita. I campi degli scalari saranno sempre il campo reale $\mathbb{R}$ o il campo complesso $\mathbb{C}$ (solo nella discussione del teorema di Cayley-Hamilton il campo sarà completamente arbitrario). Useremo la notazione K per indicare il campo degli scalari quando i risultati sono validi (con la stessa dimostrazione) sia nel caso reale che nel caso complesso.

Indice

1 Diagonalizzazione simultanea

1.1 Autovalori e autovettori

Ricordiamo alcune definizioni e alcuni importanti teoremi sugli autovalori e autovettori di una applicazione lineare (o di una matrice quadrata). Per maggiori dettagli e le dimostrazioni si può guardare un qualunque testo di Geometria I.

Sia V uno spazio vettoriale di dimensione finita sul campo K e $f : V \to V$ un *endomorfismo*, cioè un'applicazione lineare da V in sé.

Definizione 1.1. Uno scalare $\lambda \in K$ è un *autovalore* di f se esiste un vettore non nullo $v \in V$ tale che $f(v) = \lambda v$. Il vettore v viene detto *autovettore* di f relativo all'autovalore λ.

Definizione 1.2. Sia $\lambda \in K$ un autovalore di f.

$$V_\lambda = \{v \in V \mid f(v) = \lambda v\}$$

viene detto *autospazio* di f relativo all'autovalore λ.

Si dimostra facilmente che un autospazio è un sottospazio vettoriale, e poiché λ è un autovalore, V_λ non è mai il sottospazio nullo.

Per una matrice si danno le stesse definizioni, considerando la matrice come la matrice associata ad un'applicazione lineare espressa usando la stessa base in partenza e in arrivo.

λ è un autovalore di f se e solo se l'endomorfismo $f - \lambda \cdot \mathrm{id}_V$ ha nucleo non nullo, dato da V_λ. Dunque se A è la matrice che esprime f (in una qualche base), allora λ è un autovalore di f se e solo se $\det(A - \lambda I) = 0$. Il polinomio

$$p_A(t) = \det(A - tI)$$

viene detto *polinomio caratteristico* della matrice A. Il polinomio caratteristico in effetti dipende solo da f. Infatti due matrici A e B sono le matrici dello stesso endomorfismo f in due basi diverse se e solo se esiste una matrice invertibile P (la matrice di passaggio che esprime il cambiamento di base) tale che $B = PAP^{-1}$. In tal caso le matrici A e B sono dette *simili*. Si ha allora

$$\begin{aligned} p_B(t) &= \det(B - tI) = \det(PAP^{-1} - P(tI)P^{-1}) = \det(P \cdot (A - tI) \cdot P^{-1}) \\ &= \det(P)\det(A - tI)\det(P^{-1}) = \det(A - tI) = p_A(t) \end{aligned}$$

Concludiamo perciò che gli autovalori di f sono le radici del polinomio caratteristico e dunque dipendono dal campo K. Per esempio, la matrice

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

ha polinomio caratteristico $p_A(t) = t^2 + 1$ e quindi non ci sono autovalori reali, mentre ci sono due autovalori complessi, i e $-i$.

Osserviamo anche che poiché il polinomio caratteristico ha un numero finito di radici, ci sono solo un numero finito di autovalori, cosa non immediatamente ovvia dalla definizione.

Definizione 1.3. Un endomorfismo $f : V \to V$ si dice *diagonalizzabile* (o *semisemplice*) se esiste una base di V formata da autovettori di f.

È chiaro che la matrice di un endomorfismo in una base formata da autovettori è diagonale e, viceversa, se la matrice dell'endomorfismo in una base è diagonale, allora la base è formata da autovettori. Gli autovalori si trovano sulla diagonale.

Diciamo perciò che una matrice A è *diagonalizzabile* se è simile a una matrice diagonale, e cioè se esiste una matrice P tale che PAP^{-1} è una matrice diagonale.

Teorema 1.4. *Autovettori relativi ad autovalori diversi sono linearmente indipendenti.*

Questo teorema implica nuovamente che il numero di autovalori è finito.

Teorema 1.5. *La somma di autospazi è sempre diretta.*

Teorema 1.6. *Siano $\lambda_1, \dots, \lambda_k$ gli autovalori di f e siano $V_1, \dots, V_k$ i relativi autospazi. Allora f è diagonalizzabile se e solo se*

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$$

cioè se V è la somma diretta dei suoi autospazi.

1.2 Diagonalizzazione simultanea

Se f e g sono due endomorfismi diagonalizzabili, esiste una base in cui f è diagonale e un'altra base in cui g è diagonale, ma in generale non c'è una base in cui entrambi sono rappresentati da matrici diagonali. Per esempio, le matrici

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

hanno come basi di autovettori rispettivamente $\{(1,0),(1,1)\}$ e $\{(1,0),(0,1)\}$ e non è possibile trovare altre basi se non prendendo multipli non nulli dei vettori indicati. Non si può quindi trovare una base in cui entrambe le matrici diventano diagonali.

Si può però dimostrare che se A e B commutano (e sono diagonalizzabili) allora esiste una base comune di autovettori. Cominciamo con un

Lemma 1.7. *Siano A e B matrici tali che $AB = BA$ e sia $W \subseteq V$ un autospazio di B. Allora A lascia W invariato, cioè $AW \subseteq W$.*

Dimostrazione. Sia $w \in W$. Dobbiamo dimostrare che $Aw \in W$. Poiché W è un autospazio di B (di autovalore λ), basta dimostrare che Aw è un autovettore di B di autovalore λ. Si ha:

$$B(Aw) = A(Bw) = A(\lambda w) = \lambda Aw$$

e quindi la tesi. □

Dimostriamo ora il

Teorema 1.8. *Siano A e B matrici diagonalizzabili tali che $AB = BA$. Allora esiste una base comune di autovettori. Equivalentemente, esiste una matrice invertibile P tale che PAP^{-1} e PBP^{-1} siano entrambe diagonali.*

Dimostrazione. Sia $\{v_1, \ldots, v_n\}$ una base di autovettori di A. Poiché anche B è diagonalizzabile possiamo scrivere

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$$

dove i W_i sono gli autospazi di B. Da questa decomposizione possiamo scrivere

$$v_1 = w_{11} + w_{12} + \cdots + w_{1k}$$

in modo unico, dove $w_{1i} \in W_i$. Applicando A ad entrambi i membri dell'uguaglianza e ricordando che $Av_1 = \lambda_1 v_1$ si ottiene:

$$Aw_{11} + Aw_{12} + \cdots + Aw_{1k} = \lambda_1 w_{11} + \lambda_1 w_{12} + \cdots + \lambda_1 w_{1k}$$

e cioè

$$(Aw_{11} - \lambda_1 w_{11}) + (Aw_{12} - \lambda_1 w_{12}) + \cdots + (Aw_{1k} - \lambda_1 w_{1k}) = 0$$

Per il lemma precedente, $Aw_{1j} \in W_j$ e quindi anche $Aw_{1j} - \lambda_1 w_{1j} \in W_j$. Ma poiché la scrittura è unica e la somma è il vettore nullo dobbiamo avere:

$$Aw_{11} = \lambda_1 w_{11}, \quad Aw_{12} = \lambda_1 w_{12}, \quad \ldots, \quad Aw_{1k} = \lambda_1 w_{1k}$$

e quindi i vettori w_{1j} *non nulli* sono autovettori di A. Osserviamo che almeno uno fra questi è non nullo, in quanto la loro somma dà il vettore non nullo v_1.

Al vettore v_1 associamo allora i vettori non nulli che compaiono fra $w_{11}, \ldots, w_{1k}$. Procedendo allo stesso modo con v_2, …, v_n, otteniamo un insieme di vettori $\{w_{ij}\}$ tali che:

1. sono autovettori comuni di A e B per costruzione, e
2. generano lo spazio V perché tramite loro è possibile scrivere i vettori v_1, …, v_n, che formano una base di V.

Per ottenere la tesi basta allora estrarre da questi vettori una base. □

Osserviamo che vale anche il viceversa, e cioè

Proposizione 1.9. *Supponiamo che A e B abbiano una base comune di autovettori. Allora $AB = BA$*

Dimostrazione. Sia P la matrice del cambiamento di base. Allora $D_1 = PAP^{-1}$ e $D_2 = PBP^{-1}$ sono entrambe diagonali. Calcolando

$$\begin{aligned} AB &= (P^{-1}D_1P)(P^{-1}D_2P) = P^{-1}D_1D_2P \\ &= P^{-1}D_2D_1P = (P^{-1}D_2P)(P^{-1}D_1P) \\ &= BA \end{aligned}$$

perché le matrici D_1 e D_2, essendo diagonali, commutano. □

1.3 Esempi

Esempio 1.10. Si considerino le matrici

$$A = \begin{pmatrix} 3 & 1 & 1 \\ 0 & 4 & 0 \\ 1 & -1 & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} 3 & 1 & 1 \\ 3 & 1 & 3 \\ -2 & 2 & 0 \end{pmatrix}$$

È immediato verificare che $AB = BA$. Calcolando, si ha che gli autovalori di A sono

$$\lambda_1 = 2, \qquad \lambda_2 = 4, \qquad \lambda_3 = 4$$

(A ha un autovalore di molteplicità 2) mentre quelli di B sono

$$\mu_1 = -2, \qquad \mu_2 = 2, \qquad \mu_3 = 4$$

Dunque B è diagonalizzabile e poiché i suoi autospazi hanno dimensione 1 c'è una sola base possibile di autovettori (a meno di multipli scalari). Dunque l'unica base possibile di autovettori comuni è una base di autovettori di B.

Questo però non basta per concludere che A e B si diagonalizzano simultaneamente perché A potrebbe non essere diagonalizzabile. Per verificare quest'ultimo fatto osserviamo che il rango di

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

è 1 e quindi l'autospazio di autovalore 4 ha dimensione 2. Allora A è diagonalizzabile in quanto ha due autospazi di dimensione 1 e 2 e quindi A e B hanno una base comune di autovettori, che è l'unica base (a meno di multipli) di B.

Esempio 1.11. Si considerino le matrici

$$A = \begin{pmatrix} 3 & 1 & 1 \\ 0 & 4 & 0 \\ 1 & -1 & 3 \end{pmatrix}, \qquad C = \begin{pmatrix} 4 & 2 & -2 \\ -1 & 7 & -1 \\ -1 & 1 & 5 \end{pmatrix}$$

È immediato verificare che $AC = CA$. Calcolando, si ha che gli autovalori di A sono (A è la stessa matrice dell'esempio precedente)

$$\lambda_1 = 2, \qquad \lambda_2 = 4, \qquad \lambda_3 = 4$$

mentre quelli di C sono

$$\mu_1 = 4, \qquad \mu_2 = 6, \qquad \mu_3 = 6$$

e sia A che C hanno autovalori multipli. Poiché

$$C - 6I = \begin{pmatrix} -2 & 2 & -2 \\ -1 & 1 & -1 \\ -1 & 1 & -1 \end{pmatrix}$$

ha rango 1 anche C è diagonalizzabile e dunque A e C hanno una base comune di autovettori. Seguiamo il procedimento usato nella dimostrazione del Teorema 1.8 per trovare una base di autovettori comuni. Una base di autovettori di A è

$$v_1 = (1, 0, 1) \qquad v_2 = (1, 1, 0) \qquad v_3 = (-1, 0, 1)$$

dove v_1 e v_2 sono relativi all'autovalore 4 e v_3 ha autovalore 2. Gli autospazi di C sono

$$W_1 = \{(2z, z, z) \mid z \in K\}, \text{ di dimensione } 1$$

e

$$W_2 = \{(x, y, z) \mid x - y + z = 0\}, \text{ di dimensione } 2$$

Decomponendo i vettori della base di A secondo gli autospazi di C si ha

$$\begin{aligned} v_1 &= (1,0,1) = (2,1,1) - (1,1,0) = w_{11} + w_{12} \\ v_2 &= (1,1,0) = (0,0,0) + (1,1,0) = w_{21} + w_{22} \\ v_3 &= (-1,0,1) = (0,0,0) + (-1,0,1) = w_{31} + w_{32} \end{aligned}$$

e otteniamo quindi una base formata dagli autovettori comuni

$$w_{11} = (2,1,1), \quad w_{22} = (1,1,0), \quad w_{32} = (-1,0,1)$$

1.4 Esercizi

1. Si considerino le matrici

$$A = \begin{pmatrix} 2 & 1 & -1 & 1 \\ 0 & 3 & -1 & 1 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

   Dimostrare che A e B sono simultaneamente diagonalizzabili (attenzione: non basta dimostrare che $AB = BA$) e trovare una base comune di autovettori.

2. Consideriamo la situazione: A e B due matrici che commutano e A diagonalizzabile. È vero che anche B è diagonalizzabile? La risposta è no, in generale. Basta per esempio considerare $A = I$, la matrice unità e B non diagonalizzabile. Quindi potremmo considerare: A e B commutano, A diagonalizzabile, ma $A \neq I$. Anche qui c'è una risposta semplice: $A = \lambda I$ è una matrice scalare (multiplo dell'identità) e B non diagonalizzabile. Osserviamo anche che essere un multiplo dell'identità è una condizione indipendente dalla base usata (cambiando base, la matrice resta sempre multiplo dell'identità). Formuliamo quindi la domanda:

   Siano A e B due matrici che commutano, A diagonalizzabile e A non un multiplo dell'identità. Sotto queste condizioni, è vero che anche B è diagonalizzabile?

2 Il teorema di Cayley-Hamilton

2.1 Funzioni di matrici.

Sia $f(t)$ una funzione e A una matrice quadrata. Ci chiediamo quando ha senso considerare l'espressione $f(A)$, cioè valutare la funzione con un argomento matriciale. Non sempre questo è possibile, almeno non elementarmente. Per esempio, se $f(t) = \sqrt{t}$, non è chiaro cosa voglia dire $\sqrt{A}$, la radice quadrata di

una matrice. Anche se intendiamo una matrice B il cui quadrato sia A, occorre fare attenzione: per esempio, se $A = I_2$, la matrice identità di ordine 2, ci sono (almeno) quattro matrici il cui quadrato è A:

$$B_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad B_2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \qquad B_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad B_4 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$$

e non solo due come ci aspetteremmo.

Esercizio 2.1. Trovare infinite matrici 2×2 a elementi reali il cui quadrato è la matrice identità I_2. Siete in grado di trovare *tutte* le matrici 2×2 il cui quadrato è l'identità? La risposta cambia se usiamo matrici complesse oppure matrici razionali?

In questa parte studieremo il caso in cui $f(t)$ è un *polinomio*. Vedremo in seguito il caso della funzione esponenziale, che sarà di importanza fondamentale nella teoria dei sistemi di equazioni differenziali lineari.

Sia A una matrice quadrata a elementi in K e sia $f(t) \in K[t]$ un polinomio. Se

$$f(t) = b_n t^n + b_{n-1} t^{n-1} + \cdots + b_1 t + b_0$$

poniamo per definizione

$$f(A) = b_n A^n + b_{n-1} A^{n-1} + \cdots + b_1 A + b_0 I$$

dove ogni addendo è il prodotto di uno scalare per una matrice e la somma è la somma di matrici. Dunque $f(A)$ è una matrice quadrata dello stesso ordine di A.

È immediato dalla definizione che se $f(t)$, $g(t)$ sono polinomi e $f + g$ e fg sono la somma e il prodotto come polinomi, allora

$$(f + g)(A) = f(A) + g(A), \qquad (fg)(A) = f(A) \cdot g(A)$$

La prima uguaglianza è ovvia, la seconda è meno ovvia ed è vera perché tutte le matrici coinvolte sono potenze di A, e queste commutano fra loro.

Se A è una matrice quadrata, consideriamo l'insieme di polinomi che si annullano su A e cioè

$$I_A = \{f(t) \in K[t] \mid f(A) = 0\}$$

Per prima cosa osserviamo che I_A contiene dei polinomi non nulli: considerando le potenze della matrice A:

$$I, \quad A, \quad A^2, \quad \ldots, \quad A^m, \quad \ldots$$

osserviamo che queste matrici non possono essere tutte linearmente indipendenti nello spazio vettoriale $M(n \times n, K)$ delle matrici a coefficienti in K, che ha dimensione finita pari a n^2. Dunque c'è sicuramente una relazione lineare non nulla:

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

e il polinomio $f(t) = a_0 + a_1 t + \cdots + a_{n^2} t^{n^2} \in I_A$ non è il polinomio nullo.

2.2 Il polinomio minimo di una matrice

Studiamo adesso la struttura di I_A come sottoinsieme dell'anello $K[t]$. Si vede subito che I_A è un *sottogruppo* rispetto alla somma ed è anche un *ideale*, e cioè è un sottogruppo e in più vale la proprietà: se $f(t) \in I_A$ allora $f(t) \cdot g(t) \in I_A$ per ogni polinomio $g(t)$.

Ricordiamo il ben noto fatto che ogni ideale di $K[t]$ è principale:

Proposizione 2.2. *Sia I un ideale nell'anello dei polinomi $K[t]$. Allora esiste un polinomio $p(t)$ tale che ogni polinomio appartenente a I è un multiplo di $p(t)$. Se prendiamo $p(t)$ monico (cioè con coefficiente direttore 1), allora $p(t)$ è unico.*

Dimostrazione. Sia $p(t)$ un polinomio di grado minimo fra i polinomi di I. Se $f(t) \in I$ è un polinomio qualunque, si può fare la divisione fra polinomi, ottenendo

$$f(t) = q(t) \cdot p(t) + r(t)$$

dove il resto $r(t)$ ha grado strettamente minore del grado di $p(t)$. Poiché I è un ideale, $q(t) \cdot p(t) \in I$ e allora

$$r(t) = f(t) - q(t) \cdot p(t) \in I$$

e poiché il suo grado è minore del grado minimo, deve essere $r(t) = 0$. Dunque $f(t)$ è multiplo di $p(t)$.

Se adesso $p_1(t) = t^n + a_{n-1}t^{n-1} + \dots$ e $p_2(t) = t^n + b_{n-1}t^{n-1} + \dots$ sono due polinomi monici di grado minimo, allora $p_1 - p_2$ appartiene ancora ad I e ha grado inferiore al minimo, dunque $p_1 - p_2 = 0$ e cioè $p_1 = p_2$. □

Definizione 2.3. Sia A una matrice quadrata e sia I_A l'ideale dei polinomi che si annullano in A. L'unico polinomio monico di grado minimo di I_A si dice *polinomio minimo di A*.

Non è chiaro quale sia il grado del polinomio minimo. L'esempio alla fine del paragrafo predecente mostra che il polinomio minimo ha grado minore o uguale a n^2. In realtà il grado è minore o uguale a n, come si ottiene immediatamente dal teorema di Cayley-Hamilton.

2.3 Il teorema di Cayley-Hamilton.

Sia A una matrice quadrata, e sia $c_A(t) = \det(tI - A)$ il suo polinomio caratteristico. Il teorema di Cayley-Hamilton afferma che:

Teorema 2.4 (Cayley-Hamilton)**.** *$c_A(t)$ appartiene a I_A, cioè $c_A(A) = 0$.*

Poiché $\deg c_A(t) = n$ e il polinomio minimo ha grado minimo fra i polinomi in I_A dal teorema si ottiene in particolare che il grado del polinomio minimo è minore o uguale a n.

Vi sono molte dimostrazioni di questo teorema, basate su varie proprietà degli spazi vettoriali e delle matrici. Quella che vedremo si basa sulle proprietà dei determinanti, in particolare le regole di Laplace sullo sviluppo di un determinante. In particolare questo teorema vale per matrici a coefficienti in un campo K qualunque (non solo per $K = \mathbb{R}$ o $K = \mathbb{C}$).

Sia M una matrice quadrata. Poniamo

$$M_{ij} = (-1)^{i+j} \cdot \det(\text{matrice ottenuta cancellando la riga } i \text{ e la colonna } j)$$

M_{ij} viene detto il *complemento algebrico* dell'elemento m_{ij} e la matrice il cui elemento di posto (i, j) è M_{ji} (attenzione: notare lo scambio di indici) è detta l'*aggiunta classica* di M e si indica con $\text{adj}(M)$. Le regole di Laplace sullo sviluppo dei determinanti danno allora la formula:

$$\text{adj}(M) \cdot M = M \cdot \text{adj}(M) = \det(M)I$$

Questa non è nient'altro che la formula della matrice inversa, scritta però senza dividere per il determinante, che potrebbe essere nullo.

Vediamo ora la dimostrazione del teorema di Cayley-Hamilton.

Dimostrazione. Poniamo $M = tI - A$ e scriviamo la formula precedente:

$$(tI - A) \cdot \text{adj}(tI - A) = c_A(t) \cdot I$$

Poniamo

$$c_A(t) = t^n + b_{n-1}t^{n-1} + \cdots + b_1 t + b_0$$

La matrice $\text{adj}(tI-A)$ ha per elementi polinomi in t di grado al massimo $n-1$, perché i suoi elementi sono determinanti di sottomatrici di $tI-A$ di ordine $n-1$. Raccogliendo i coefficienti, possiamo scrivere

$$\text{adj}(tI - A) = A_{n-1}t^{n-1} + A_{n-2}t^{n-2} + \cdots + A_1 t + A_0$$

dove le A_i sono opportune matrici. Eseguendo la moltiplicazione e uguagliando i coefficienti nella formula dell'aggiunta si ottengono le seguenti relazioni:

$$\begin{aligned}
A_{n-1} &= I \\
-A \cdot A_{n-1} + A_{n-2} &= b_{n-1}I \\
-A \cdot A_{n-2} + A_{n-3} &= b_{n-2}I \\
&\cdots \\
-A \cdot A_{n-j} + A_{n-j-1} &= b_{n-j}I \\
&\cdots \\
-A \cdot A_1 + A_0 &= b_1 I \\
-A \cdot A_0 &= b_0 I
\end{aligned}$$

Sostituendo la prima relazione nella seconda si ottiene

$$A_{n-2} = A + b_{n-1}I$$

sostutuendo questa nella terza si ottiene

$$A_{n-3} = A^2 + b_{n-1}A + b_{n-2}I$$

e continuando così a sostituire si ottiene (al penultimo passo)

$$A_0 = A^{n-1} + b_{n-1}A^{n-2} + \cdots + b_1 I$$

e finalmente all'ultimo

$$A^n + b_{n-1}A^{n-1} + \ldots b_1 A + b_0 I = 0$$

che è la tesi. □

2.4 Polinomio minimo e polinomio caratteristico.

Dal teorema di Cayley-Hamilton si ottiene che il polinomio caratteristico di una matrice A è un multiplo del polinomio minimo. Se denotiamo con $m_A(t)$ il polinomio minimo e con $c_A(t)$ il polinomio caratteristico, possiamo scrivere

$$c_A(t) = m_A(t) \cdot q(t)$$

Dunque le radici del polinomio minimo, essendo anche radici del polinomio caratteristico, sono autovalori di A. Poiché il polinomio minimo è individuato da A e non da altre condizioni, sarebbe strano che alcuni autovalori fossero radici del polinomio minimo e altri no.

In effetti vale il

Teorema 2.5. *Sia $m_A(t)$ il polinomio minimo di A. Allora $m_A(\alpha) = 0$ se e solo se α è un autovalore di A.*

Dimostrazione. La discussione che precede l'enunciato dimostra che le radici di $m_A(t)$ sono autovalori. Viceversa, sia α un autovalore di A e sia $v \neq 0$ un autovettore di autovalore α. Allora

$$Av = \alpha v, \quad A^2v = \alpha^2 v, \quad \ldots, \quad A^k v = \alpha^k v, \quad \ldots$$

e per un polinomio $p(t)$ si ha $p(A)v = p(\alpha)v$. In particolare $m_A(A)v = m_A(\alpha)v$, ma $m_A(A) = 0$ e quindi $m_A(\alpha)v = 0$ Poiché $v \neq 0$ deve essere $m_A(\alpha) = 0$. □

Abbiamo quindi che $m_A(t)$ si scompone negli stessi fattori lineari (su $\mathbb{C}$) di $c_A(t)$, con esponenti minori o uguali a quelli presenti nel polinomio caratteristico.

A questo punto vogliamo vedere alcuni esempi per cominciare a capire la differenza fra polinomio minimo e polinomio caratteristico.

Esempio 2.6. Sia

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

Allora $c_A(t) = (t-3)^2$. Il polinomio minimo è un sottomultiplo e perciò ci sono due possibilità: $m_A(t) = t - 3$ oppure $m_A(t) = (t-3)^2$. Poiché

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

il polinomio minimo è $m_A(t) = (t-3)^2$ (verificare che in effetti $(A-3I)^2 = 0$).

Osserviamo che in questo caso A non è diagonalizzabile e $m_A(t) = c_A(t)$.

Esempio 2.7. Sia

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

Allora $c_B(t) = (t-2)(t-3)$. Il polinomio minimo è un sottomultiplo e perciò ci sono tre possibilità: $m_B(t) = t-2$, $m_B(t) = t-3$ oppure $m_B(t) = (t-2)(t-3)$. Poiché

$$B - 2I = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \neq 0 \qquad B - 3I = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \neq 0$$

il polinomio minimo è $m_B(t) = (t-2)(t-3)$ (verificare che in effetti $(B - 2I)(B - 3I) = 0$).

Osserviamo che in questo caso B è diagonalizzabile (ha autovalori distinti) e $m_B(t) = c_B(t)$.

Esempio 2.8. Sia

$$C = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$$

Allora $c_C(t) = (t-3)^2$. Il polinomio minimo è un sottomultiplo e perciò ci sono due possibilità: $m_C(t) = t - 3$ oppure $m_C(t) = (t-3)^2$. Poiché

$$C - 3I = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0$$

il polinomio minimo è $m_C(t) = (t-3)$.

Osserviamo che in questo caso C è diagonalizzabile e $m_C(t) \neq c_C(t)$.

Esempio 2.9. Sia

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

Allora $c_D(t) = (t-2)^2(t-3)^2$. Il polinomio minimo è un sottomultiplo e ci sono varie possibilità. Analizzando le possibilità, anche alla luce degli esempi precedenti, è facile concludere che $m_D(t) = (t-2)(t-3)^2$.

Osserviamo che in questo caso D non è diagonalizzabile e $m_D(t) \neq c_D(t)$.

Concludiamo che non c'è relazione fra la diagonalizzabilità di una matrice e l'uguaglianza fra polinomio minimo e polinomio caratteristico. Osserviamo però che per le matrici diagonalizzabili (almeno negli esempi precedenti) i fattori del polinomio minimo sono tutti a primo grado, mentre per le matrici non diagonalizzabili è presente almeno un fattore con esponente maggiore o uguale a 2. Questo fatto è vero in generale e può essere dimostrato come corollario immediato delle proprietà della decomposizione di Jordan che studieremo nei prossimi paragrafi. Vi è però una dimostrazione diretta, che non usa tutta la teoria necessaria per la decomposizione di Jordan.

2.5 Polinomio minimo e diagonalizzabilità

L'enunciato preciso del teorema di cui abbiamo parlato alla fine del paragrafo precedente è:

Teorema 2.10. *Una matrice è diagonalizzabile se e solo se il suo polinomio minimo ha tutte radici di molteplicità* 1.

Per la dimostrazione utilizziamo il

Lemma 2.11. *Siano $f : U \to V$ e $g : V \to W$ due applicazioni lineari. Allora*

$$\dim \ker(g \circ f) \leq \dim \ker f + \dim \ker g$$

Dimostrazione. Osserviamo che $\ker(g \circ f) = T = f^{-1}(\ker g)$, il sottospazio di U controimmagine del sottospazio $\ker g$ di V. Consideriamo la funzione lineare

$$h : T \to \ker g$$

data dalla restrizione di f a T. Si ha $\ker h \subseteq \ker f$. Infatti $h(v) = 0$ vuol dire $v \in T$ e $f(v) = 0$, e quindi $v \in \ker f$. Dunque:

$$\dim \ker(g \circ f) = \dim T = \dim \ker h + \dim \operatorname{Im} h \leq \dim \ker f + \dim \ker g$$

che è la tesi. □

Una semplice induzione dimostra che, più in generale,

$$\dim \ker(f_1 \circ f_2 \circ \cdots \circ f_k) \leq \dim \ker f_1 + \dim \ker f_2 + \cdots + \dim \ker f_k$$

Dimostrazione del teorema 2.10. Se una matrice è diagonalizzabile, è chiaro che il polinomio minimo ha tutte la radici di molteplicità 1.

Viceversa, sia $m_A(t) = (t - \lambda_1) \cdot (t - \lambda_2) \cdots (t - \lambda_k)$, dove $\lambda_1, \ldots, \lambda_k$ sono gli autovalori *distinti* di A. Il fatto che $m_A(A) = 0$ significa che la composizione delle applicazioni lineari

$$(A - \lambda_1 I) \circ (A - \lambda_2 I) \circ \cdots \circ (A - \lambda_k I)$$

è l'applicazione lineare nulla. Usando il lemma si ha

$$\begin{aligned} \dim V &= \dim \ker (A - \lambda_1 I) \circ (A - \lambda_2 I) \circ \cdots \circ (A - \lambda_k I) \\ &\leq \dim \ker (A - \lambda_1 I) + \dim \ker (A - \lambda_2 I) + \cdots + \dim \ker (A - \lambda_k I) \\ &= \dim \ker (A - \lambda_1 I) \oplus \ker (A - \lambda_2 I) \oplus \cdots \oplus \ker (A - \lambda_k I) \end{aligned}$$

perché i nuclei scritti sono gli autospazi di A e sappiamo che la somma di autospazi è diretta. Allora la dimensione di V è minore o uguale alla dimensione della somma degli autospazi e poiché è anche maggiore o uguale (gli autospazi stanno dentro V), deve essere uguale. Allora

$$V = \ker (A - \lambda_1 I) \oplus \ker (A - \lambda_2 I) \oplus \cdots \oplus \ker (A - \lambda_k I)$$

e cioè V è la somma diretta degli autospazi di A. Ma questo vuol dire A diagonalizzabile. □

2.6 Esercizi.

1. Per ognuna delle matrici seguenti calcolare il polinomio caratteristico, il polinomio minimo e dire se è diagonalizzabile oppure no:

$$A = \begin{pmatrix} 3 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \qquad B = \begin{pmatrix} 3 & 7 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 5 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

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

2. Per le seguenti coppie di polinomi, scrivere una matrice che li ha come polinomio caratteristico e minimo, rispettivamente:

$$\begin{aligned}
&c_A(t) = (t-2)^3(t-3), && m_A(t) = (t-2)(t-3)\\
&c_B(t) = (t-2)^2(t-3)(t-4), && m_B(t) = (t-2)^2(t-3)(t-4)\\
&c_C(t) = (t-2)^3(t-3)^2, && m_C(t) = (t-2)^2(t-3)\\
&c_D(t) = (t-2)(t-3)(t-4), && m_D(t) = (t-2)(t-3)\\
&c_E(t) = (t-2)^3(t-3)^3, && m_E(t) = (t-2)^2(t-3)^3
\end{aligned}$$

Quali matrici sono diagonalizzabili? Quali sono le dimensioni degli autospazi?

3 La forma canonica di Jordan

3.1 La forma canonica di Jordan.

Sia A una matrice quadrata, che può essere pensata come la matrice di un endomorfismo $f : V \to V$. Non è sempre possibile trovare una base di autovettori di f, cioè non tutte le matrici sono diagonalizzabili. È allora importante trovare delle basi che rendano la matrice "più semplice possibile", in qualche senso.

Supponiamo che esistano due sottospazi complementari $W, T \subseteq V$ invarianti per A, e cioè $AW \subseteq W$ e $AT \subseteq T$ e tali che $V = W \oplus T$. Allora si può trovare una base di V prendendo una base di W e una base di T. In questa base la matrice A diventa "a blocchi", del tipo

$$A = \begin{bmatrix} A_W & 0 \\ 0 & A_T \end{bmatrix}$$

dove A_W e A_T sono le matrici delle restrizioni $A|_W : W \to W$ e $A|_T : T \to T$. Per esempio, un autovettore di f genera un sottospazio invariante di dimensione 1, e il blocco corrispondente è uno scalare (l'autovalore corrispondente) sulla diagonale.

Quando esistono più sottospazi invarianti la cui somma diretta dà lo spazio, allora la matrice si decompone in tanti blocchi lungo la diagonale quanti sono i sottospazi invarianti. Per esempio, se la matrice è diagonalizzabile, prendendo una base di autovettori si ha che ogni autovettore genera un sottospazio invariante di dimensione 1 e la somma di questi sottospazi è diretta in quanto generati dai vettori di una base. In questo caso tutti i blocchi hanno dimensione 1 e cioè la matrice è in forma diagonale.

L'esempio più semplice di matrice non diagonalizzabile è del tipo

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

In questo caso c'è un solo autovalore $\lambda = 1$ di molteplicità 2, e un solo autovettore $v_1 = (1, 0)$. Il sottospazio invariante generato da v_1 è l'asse x ed è un semplice esercizio verificare che non ci sono altri sottospazi invarianti non banali, cioè diversi dal sottospazio nullo e da V. Dunque non è possibile scrivere V, che

in questo caso ha dimensione 2, come somma di sottospazi invarianti e quindi la matrice A non può essere ulteriormente semplificata.

Un altro esempio simile è dato dalla matrice

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

Anche in questo caso c'è un solo autovettore, $v_1 = (1, 0, 0)$, che genera un sottospazio invariante W di dimensione 1 e cerchiamo un sottospazio invariante complementare T per decomporre lo spazio V. Sia $ax + by + cz = 0$ l'equazione del piano T. La condizione $W \cap T = \{0\}$ implica $a \neq 0$ e quindi possiamo supporre $a = 1$. T invariante significa che $v \in T \implies Av \in T$. Sia dunque $v = (x, y, z) \in T$, cioè $x + by + cz = 0$. $Av = (x + y, y + z, z)$ e quindi la condizione $Av \in T$ diventa $(x + y) + b(y + z) + cz = 0$ che, usando la condizione $v \in T$, si semplifica in $y + bz = 0$. È allora immediato trovare un vettore $v \in T$ tale che $Av \notin T$, per esempio $v = (-b, 1, 0)$. Dunque T non è invariante e perciò di nuovo la matrice A non può essere scritta in blocchi più piccoli.

La matrice A è un *blocco di Jordan*, e cioè è triangolare superiore, con tutti gli elementi sulla diagonale uguali (un autovalore ripetuto) e sopra la diagonale ha degli 1. Altri esempi di blocchi di Jordan sono

$$B = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad D = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

di dimensioni rispettivamente 3, 3, e 4. La matrice

$$E = \left[\begin{array}{cc|cc} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ \hline 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{array}\right]$$

non è un blocco di Jordan di dimensione 4, ma piuttosto è formata da due blocchi di dimensione 2.

Poniamo la seguente:

Definizione 3.1. Un *blocco di Jordan* di autovalore a e dimensione k è una matrice quadrata di ordine k della forma

$$J_k(a) = \begin{bmatrix} a & 1 & 0 & \dots & 0 \\ 0 & a & 1 & \dots & 0 \\ & \dots & & \dots & \\ 0 & 0 & \dots & a & 1 \\ 0 & 0 & \dots & 0 & a \end{bmatrix}$$

dove a è uno scalare (reale o complesso) e k è un intero positivo.

Definizione 3.2. Una matrice (quadrata) A è *in forma di Jordan* se ha blocchi di Jordan lungo la diagonale e altrove è nulla.

Per esempio, la matrice

$$A = \left[\begin{array}{ccc|cc|ccc|c} 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{array}\right]$$

è in forma di Jordan, con blocchi $J_3(2)$, $J_2(3)$, $J_3(3)$ e $J_1(4)$.

Il risultato a cui vogliamo arrivare è che ogni matrice complessa è simile a una matrice in forma di Jordan. Inoltre i blocchi sono unici, nel senso che due matrici in forma di Jordan sono simili fra loro se e solo se hanno gli stessi blocchi, al più in ordine diverso. Enunceremo con precisione i teoremi nel seguito.

Analizziamo ora in dettaglio la base che dà un blocco di Jordan

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & \dots & & \dots & \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} = J_k(0)$$

di autovalore 0. Se chiamiamo $\{e_1, e_2, \dots, e_k\}$ la base dello spazio V notiamo che e_1 è l'unico autovettore di A. Inoltre si ha

$$A(e_k) = e_{k-1}, \quad A(e_{k-1}) = e_{k-2}, \quad \dots, \quad A(e_2) = e_1$$

e cioè cominciando da e_k si ottengono tutti i vettori della base applicando ripetutamente A. Questa osservazione sarà la base del procedimento che seguiremo per trovare la forma di Jordan di una matrice.

Fissiamo uno spazio vettoriale V. Nel seguito scriveremo A per indicare sia un endomorfismo di V sia la matrice associata quadrata all'endomorfismo in una base fissata. Con questa convenzione $A(v)$ può significare sia l'immagine del vettore v mediante l'endomorfismo A sia il vettore colonna ottenuto moltiplicando la matrice A con il vettore colonna delle coordinate di v nella base fissata. Questo uso non causerà nessuna confusione.

L'ultima osservazione è che nella forma di Jordan di una matrice compaiono gli autovalori. È quindi essenziale che tutti gli autovalori esistano e per questo dimostreremo il teorema seguente nel caso complesso. Discuteremo in seguito la situazione nel caso reale.

3.2 Il teorema di esistenza e unicità

Teorema 3.3 (**Forma canonica di Jordan**). *Sia V uno spazio vettoriale complesso e sia $A : V \to V$ un endomorfismo. Allora esiste una base di V in cui la matrice di A è in forma di Jordan.*

La dimostrazione è per induzione sulla dimensione di V. Se $\dim V = 1$ non c'è niente da dimostrare.

Sia quindi $\dim V = n > 1$, sia λ un autovalore di A e sia $B = A - \lambda I$. Se troviamo una base in cui B è in forma di Jordan, $A = B + \lambda I$ è in forma di Jordan nella stessa base.

Poniamo $M_i = \operatorname{Im} B^i$, e in particolare $M_0 = V$. Poiché $B^{i+1}v = B^i(Bv) \in M_i$ abbiamo che $M_{i+1} \subseteq M_i$. Allora la sequenza di sottospazi

$$V = M_0 \supseteq M_1 \supseteq M_2 \supseteq \cdots \supseteq M_p \supseteq \dots$$

stabilizza. Infatti, se due sottospazi successivi sono diversi la dimensione diminuisce strettamente e la dimensione di V è finita. Dunque esiste un indice p tale che

$$M_{p-1} \supsetneq M_p = M_{p+1} = M_{p+2} = \dots$$

cioè da M_p in poi tutti i sottospazi sono uguali e p è il minimo indice per cui questo accade.

Poiché $B(M_i) = M_{i+1} \subseteq M_i$, allora M_i è un sottospazio invariante per B e possiamo considerare la restrizione $B|_{M_i} : M_i \to M_i$. In particolare si ha che $\ker B|_{M_i} = M_i \cap \ker B$. Poiché $B(M_p) = M_{p+1} = M_p$, la restrizione $B|_{M_p}$ è suriettiva e quindi anche iniettiva, e cioè:

$$M_p \cap \ker B = \{0\}$$

e più in generale, $B^i(M_p) = M_p$ per tutti gli $i \geq 1$, e cioè

$$\forall\, i \geq 1 \qquad M_p \cap \ker B^i = \{0\} \tag{1}$$

Per prima cosa proviamo che

Lemma 3.4. *Si ha la decomposizione in somma diretta*

$$V = \ker B^p \oplus \operatorname{Im} B^p$$

Dimostrazione. Ricordiamo che $M_p = \operatorname{Im} B^p$. Dalla (1) abbiamo che la somma di $\operatorname{Im} B^p$ e $\ker B^p$ è diretta e la somma delle dimensioni dà $\dim V$ (sono le dimensioni di nucleo e immagine della stessa applicazione lineare) e quindi si ha la tesi. □

Abbiamo già osservato che $\operatorname{Im} B^p$ è un sottospazio invariante per B, e allo stesso modo anche $\ker B^p$ lo è. Dunque la matrice di B può essere scritta nella forma

$$B = \begin{bmatrix} B_K & 0 \\ 0 & B_I \end{bmatrix}$$

dove B_K è la matrice di B ristretta a $\ker B^p$ e B_I la matrice di B ristretta a $\operatorname{Im} B^p$. Poiché $\ker B^p$ contiene $\ker B$, che è non triviale in quanto contiene almeno un autovettore di A di autovalore λ, si ha $\dim \operatorname{Im} B^p < \dim V = n$ e quindi, per ipotesi induttiva, è possibile trovare una base di $\operatorname{Im} B^p$ in cui B_I sia in forma di Jordan. Per concludere la dimostrazione basta quindi trovare una base di $\ker B^p$ rispetto a cui B_K sia in forma di Jordan.

Poniamo adesso

$$S_i = M_{i-1} \cap \ker B$$

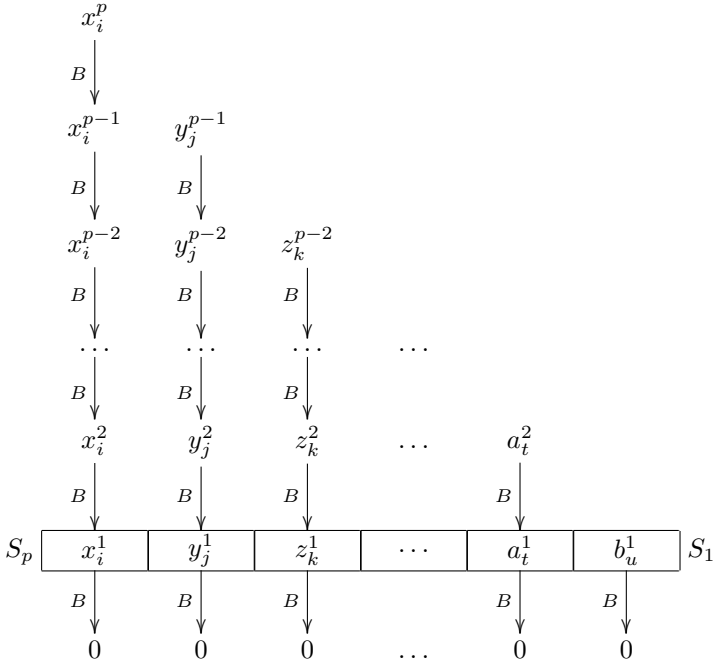
e osserviamo che $S_{p+1} = \{0\}$. Inoltre $S_1 = \ker B$ e abbiamo come prima una sequenza discendente di sottospazi

$$\ker B = S_1 \supseteq S_2 \supseteq \cdots \supseteq S_p$$

Consideriamo il sottospazio S_p e sia $\{x_1^1, x_2^1, \ldots, x_r^1\}$ una base di S_p. Poiché $S_p = \operatorname{Im} B^{p-1} \cap \ker B$, esistono dei vettori x_i^p tali che

$$B^{p-1}(x_i^p) = x_i^1, \qquad i = 1, 2, \ldots, r$$

Disegnamo un diagramma che sarà utile nel capire la struttura della costruzione:



C'è una colonna come la prima per ogni vettore della base di S_p. I vettori sulla colonna si ottengono ognuno applicando B al vettore che sta subito sopra nella colonna, cioè $x_i^l = B(x_i^{l+1})$. L'ultima applicazione di B dà 0 perché i vettori x_i^1 appartengono a S_p, che è contenuto in $\ker B$ per definizione.

Adesso prendiamo dei vettori $\{y_1^1, y_2^1, \ldots, y_s^1\}$ in modo che i vettori $\{x_i^1\}$ e $\{y_j^1\}$ formino una base di S_{p-1}. Come prima, poiché $S_{p-1} = \operatorname{Im} B^{p-2} \cap \ker B$, esistono dei vettori y_j^{p-1} tali che

$$B^{p-2}(y_j^{p-1}) = y_j^1, \qquad j = 1, 2, \ldots, s$$

Vi sono quindi $\dim S_p = r$ colonne con vettori x_i, $\dim S_{p-1} - \dim S_p = s$ colonne con vettori y_j, e continuiamo così scegliendo vettori z_k^1 per ottenere una base di S_{p-2}, …, fino a prendere vettori b_u^1 in modo che tutti i vettori scelti presi insieme diano una base di $S_1 = \ker B$.

Osserviamo che è possibile che due sottospazi consecutivi siano uguali. In tal caso non c'è bisogno di aggiungere vettori alla base, e si prosegue considerando il sottospazio successivo.

I vettori scritti nella riga più in basso (non nulla) formano una base di S_1 per costruzione, ma tutti gli altri vettori non stanno in S_1. Osserviamo anche che poiché i vettori si ottengono tutti a partire da quelli "più in alto" applicando ripetutamente B, e i vettori "più in basso" formano una base e sono quindi non nulli, tutti i vettori scritti sono non nulli.

Adesso contiamo quanti vettori abbiamo determinato: sulla riga più in basso c'è una base di S_1, quindi ce ne sono $\dim S_1$. Sulla penultima riga ce ne sono tanti quanti i vettori di una base di S_2: infatti i vettori x_i^2, y_j^2, …, a_t^2 sono tanti quanti i vettori x_i^1, y_j^1, …, a_t^1 che formano una base di S_2, e quindi sulla penultima riga ci sono $\dim S_2$ vettori. Continuando così e sommando tutte le righe si ottiene che nel diagramma ci sono

$$N = \sum_{i=1}^{p} \dim S_i$$

vettori.

Calcoliamo ora $\dim S_i$. Abbiamo bisogno di un lemma:

Lemma 3.5. *Siano $f : U \to V$ e $g : V \to W$ due applicazioni lineari. Allora*

$$\dim(\operatorname{Im} f \cap \ker g) = \dim \operatorname{Im} f - \dim \operatorname{Im}(g \circ f) = \dim \ker(g \circ f) - \dim \ker f$$

Dimostrazione. Poniamo $h : \operatorname{Im} f \to W$ la restrizione di g all'immagine di f. Il teorema su nucleo e immagine dà:

$$\dim \ker h = \dim \operatorname{Im} f - \dim \operatorname{Im} h$$

e poiché $\ker h = \operatorname{Im} f \cap \ker g$ e $\operatorname{Im} h = \operatorname{Im}(g \circ f)$ si ha la prima uguaglianza.

Poiché $\dim \operatorname{Im} f = \dim U - \dim \ker f$ e $\dim \operatorname{Im}(g \circ f) = \dim U - \dim \ker(g \circ f)$, sostituendo si ha la seconda uguaglianza. □

Applichiamo adesso il lemma alla situazione: $f = B^{i-1}$, $g = B$ e otteniamo

$$\dim S_i = \dim(\operatorname{Im} B^{i-1} \cap \ker B) = \dim \ker B^i - \dim \ker B^{i-1}$$

Sommando su i otteniamo finalmente:

$$\begin{aligned} N = \sum_{i=1}^{p} \dim S_i &= \sum_{i=1}^{p} \dim \ker B^i - \dim \ker B^{i-1} \\ &= \dim \ker B^p - \dim \ker B^0 \\ &= \dim \ker B^p \end{aligned}$$

Osserviamo anche che tutti i vettori nel diagramma precedente appartengono a $\ker B^p$ (applicando p volte B si arriva alla riga tutta nulla).

Dunque abbiamo costruito un insieme di vettori $\{x_i^\alpha, y_j^\beta, \dots\}$ che appartengono tutti al sottospazio $\ker B^p$ e sono tanti quanti la dimensione del sottospazio. Se questi vettori sono linearmente indipendenti, allora formano una base e in questa base la matrice di $B_K = B|_{\ker B^p}$ è in forma di Jordan, come richiesto.

Supponiamo allora di avere una relazione di dipendenza lineare della forma

$$\sum \alpha_i x_i^p + \sum \beta_i x_i^{p-1} + \cdots + \sum \gamma_j y_j^{p-1} + \cdots + \sum \delta_u b_u^1 = 0$$

Applicando B^{p-1} tutti i vettori con indice minore o uguale a $p-1$ si annullano, mentre i vettori di indice p diventano i corrispondenti vettori di indice 1. Si ottiene quindi

$$\sum \alpha_i x_i^1 = 0$$

e poiché i vettori $\{x_i^1\}$ sono una base di S_p, i coefficienti α_i sono tutti nulli.

Applicando adesso B^{p-2}, i vettori di indice minore o uguale a $p-2$ si annullano, e i vettori di indice $p-1$ diventano i corrispondenti vettori di indice 1. Si ottiene quindi:

$$\sum \beta_i x_i^1 + \sum \gamma_j y_j^1 = 0$$

e poiché i vettori $\{x_i^1, y_j^1\}$ sono una base di S_{p-1}, i coefficienti β_i e γ_j sono tutti nulli.

Applicando successivamente B^{p-3}, ..., B^2 si ottiene infine che tutti i coefficienti sono nulli e quindi i vettori sono linearmente indipendenti. Questo conclude la dimostrazione.

La forma canonica di Jordan che abbiamo ottenuto è sostanzialmente unica. Più precisamente:

Teorema 3.6. *La forma di Jordan di una matrice A è unicamente determinata a meno dell'ordine dei blocchi di Jordan.*

Dimostrazione. Sulla diagonale della forma di Jordan compaiono gli autovalori di A, tante volte quanto è la loro molteplicità come radici del polinomio caratteristico, e questo dipende solo da A.

Per un dato autovalore λ, i blocchi di Jordan corrispondenti si ottengono considerando la matrice $B = A - \lambda I$. Il numero di blocchi di dimensione $k \times k$ è pari a $\dim S_k - \dim S_{k+1}$ che dipende solo dalla matrice B (e quindi solo da A) e non dal procedimento. Dunque la forma di Jordan è unica. □

Una conseguenza immediata è:

Corollario 3.7. *Due matrici in forma di Jordan sono simili se e solo se hanno gli stessi blocchi (a meno dell'ordine).*

3.3 Esercizi

1. Trovare la forma di Jordan delle matrici:
$$A = \begin{bmatrix} 0 & -2 & -2 & -2 \\ -3 & 1 & 1 & -3 \\ 1 & -1 & -1 & 1 \\ 2 & 2 & 2 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

2. Dimostrare che la matrice:
$$A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$
è *idempotente* (cioè $A^2 = A$) e determinare la sua forma di Jordan. (La forma di Jordan di una matrice idempotente è ancora idempotente? Se sì, come può un blocco di Jordan essere idempotente?)

3. Sia A una matrice 5×5 con polinomio caratteristico $p_A(t) = (t-\alpha)^5$. Se il rango della matrice $A - \alpha I$ è 2, quali sono le possibili forme di Jordan di A?

4. Sia A una matrice con polinomio caratteristico $p_A(t) = (t-2)^2(t-5)^3$ e tale che l'autospazio di autovalore 2 ha dimensione 1 e l'autospazio di autovalore 5 ha dimensione 2. Qual è la forma di Jordan di A?

3.4 Funzione esponenziale sui numeri complessi.

Una delle applicazioni più importanti della forma di Jordan riguarda il calcolo dell'esponenziale di una matrice. Per cominciare, ricordiamo la definizione della funzione esponenziale per un argomento complesso.

Per ogni numero complesso $z \in \mathbb{C}$ poniamo

$$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \tag{2}$$

Osserviamo che per i valori di z reali questa non è altro che l'usuale funzione esponenziale. La serie dei valori assoluti è evidentemente convergente (converge a $e^{|z|}$) e quindi la serie converge assolutamente. Inoltre converge uniformemente su ogni sottoinsieme limitato: infatti se $S \subset \mathbb{C}$ è un sottoinsieme limitato, allora esiste un numero reale positivo a tale che $|z| < a$ per ogni $z \in S$. Allora i termini della (2) sono maggiorati in valore assoluto dai termini di una serie numerica convergente e quindi la convergenza è uniforme.

Questo significa che su ogni sottoinsieme limitato di $\mathbb{C}$ la somma della serie rappresenta una funzione continua. Poiché possiamo coprire $\mathbb{C}$ con una successione crescente di insiemi limitati, per esempio i cerchi D_n di centro l'origine e raggio $n \in \mathbb{N}$, la somma rappresenta una funzione continua su tutto il piano complesso.

La convergenza assoluta implica che possiamo riordinare i termini della serie e ottenere ancora una serie convergente allo stesso valore. Si ha:

$$\left(\sum_{k=0}^{\infty} \frac{z^k}{k!}\right) \cdot \left(\sum_{m=0}^{\infty} \frac{w^m}{m!}\right) = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} z^k w^{n-k} = \sum_{n=0}^{\infty} \frac{(z+w)^n}{n!} \tag{3}$$

e dunque per ogni z, $w \in \mathbb{C}$ vale l'importante formula

$$\exp(z)\exp(w) = \exp(z+w) \tag{4}$$

Osserviamo che la seconda uguaglianza è vera per la formula del binomio di Newton, che è valida in quanto il prodotto è commutativo: $zw = wz$. Poiché $\exp(1) = e$, scriveremo di solito e^z piuttosto che $\exp(z)$.

Il seguente teorema riassume le principali proprietà della funzione esponenziale complessa.

Teorema 3.8.

1. *per ogni numero complesso z si ha $e^z \neq 0$;*
2. exp *coincide con la sua derivata:* $\exp'(z) = \exp(z)$*;*
3. *se $t \in \mathbb{R}$, allora $e^{it} = \cos(t) + i\sin(t)$;*

Dimostrazione. Per la (4) si ha $e^z \cdot e^{-z} = e^0 = 1$ e questo implica la prima affermazione.

Calcolando la derivata come limite del rapporto incrementale si ha:

$$\exp'(z) = \lim_{h\to 0} \frac{\exp(z+h) - \exp(z)}{h} = \exp(z) \lim_{h\to 0} \frac{\exp(h) - 1}{h} = \exp(z)$$

dove la prima uguaglianza è la definizione di derivata, la seconda segue dalla (4) e la terza dalla (2).

Ricordiamo che $i^2 = -1$, $i^3 = -i$ e $i^4 = 1$. Sostituendo it nella serie che definisce l'esponenziale e raccogliendo i termini reali e quelli immaginari si ha:

$$e^{it} = \sum_{n=0}^{\infty}(-1)^n \frac{t^{2n}}{(2n)!} + i\sum_{n=0}^{\infty}(-1)^n \frac{t^{2n+1}}{(2n+1)!}$$

e riconosciamo immediatamente che la prima serie è quella del coseno e la seconda quella del seno. □

La formula

$$e^{it} = \cos(t) + i\sin(t)$$

viene di solito detta *identità di Eulero*. Osserviamo anche che per ogni $t \in \mathbb{R}$

$$|e^{it}| = \cos^2 t + \sin^2 t = 1$$

Se scriviamo un numero complesso $z = x + iy$ con le sue parti reale e immaginaria, il teorema precedente dà:

$$e^z = e^{x+iy} = e^x(\cos y + i\sin y)$$

e quindi otteniamo la forma polare del numero complesso e^z. In particolare il modulo di e^z è e^x, mentre l'argomento è y.

Ponendo $t = \pi$ nella formula di Eulero si ottiene la famosa identità

$$e^{i\pi} + 1 = 0$$

fra i cinque numeri più importanti in Matematica.

3.5 Esponenziale di una matrice.

Abbiamo parlato nella Lezione 2 di polinomi applicati a matrici. Vogliamo ora definire l'esponenziale di una matrice. In analogia con la definizione per numeri reali e per numeri complessi poniamo:

Definizione 3.9. Sia A una matrice quadrata (reale o complessa). L'esponenziale di A è la matrice

$$e^A = \sum_{k=0}^{\infty}\frac{A^k}{k!} \tag{5}$$

La serie scritta è una serie di matrici, e converge se e solo se convergono tutte le serie che danno gli elementi della matrice somma. Per dimostrare la convergenza, poniamo:

$$\|A\|_\infty = \max|a_{ij}|$$

Vale le seguente diseguaglianza:

Lemma 3.10. *Per A, B matrici quadrate di ordine n si ha:*

$$\|A \cdot B\|_\infty \leq n \cdot \|A\|_\infty \cdot \|B\|_\infty$$

Dimostrazione. Sia $AB = (c_{ij})$. Per ogni elemento c_{ij} si ha

$$|c_{ij}| = \left| \sum_{k=0}^{n} a_{ik} b_{kj} \right| \leq \sum_{k=0}^{n} |a_{ik}||b_{kj}| \leq n \max |a_{ik}| \max |b_{kj}| \leq n \cdot \|A\|_\infty \cdot \|B\|_\infty$$

e dunque

$$\|A \cdot B\|_\infty = \max |c_{ij}| \leq n \cdot \|A\|_\infty \cdot \|B\|_\infty$$

□

Con una semplice induzione si ottiene allora:

$$\|A^k\|_\infty \leq n^{k-1} \|A\|_\infty^k$$

e quindi tutte le serie che compaiono negli elementi della definizione di e^A sono maggiorate dalla serie

$$\sum_{k=0}^{\infty} \frac{n^{k-1} \|A\|_\infty^k}{k!} = \frac{1}{n} \sum_{k=0}^{\infty} \frac{(n\|A\|_\infty)^k}{k!}$$

che converge a $\frac{1}{n} e^{n\|A\|_\infty}$. Dunque tutte le serie convergono e e^A è una matrice ben definita.

Osserviamo che, in generale,

$$e^{A+B} \neq e^A \cdot e^B$$

poiché se $AB \neq BA$ non si può usare la formula del binomio di Newton per esprimere $(A + B)^n$ in termini di monomi del tipo $A^k B^{n-k}$. Si ha però

$$AB = BA \implies e^{A+B} = e^A \cdot e^B$$

con la stessa dimostrazione che abbiamo usato per l'esponenziale complesso (formula (4)).

Vi sono varie formule che esprimono la relazione fra e^{A+B} e $e^A \cdot e^B$. Una è la *formula di Lie* (*Lie product formula* nei testi in inglese)

$$e^{A+B} = \lim_{m \to +\infty} \left(e^{A/m} \cdot e^{B/m} \right)^m$$

Una formula più generale è la *formula di Baker-Campbell-Hausdorff* (usualmente detta *BCH formula* nei testi in inglese), che spiega il ruolo giocato dalla non commutatività del prodotto. BCH esprime $e^A \cdot e^B$ mediante l'esponenziale di una serie che comincia con $A + B$ e prosegue con termini che coinvolgono solo i commutatori $[A, B] = AB - BA$. Ponendo

$$e^A \cdot e^B = e^Z$$

i primi termini della serie che esprime Z sono

$$Z = A + B + \frac{1}{2}[A, B] + \frac{1}{12}\Big([A, [A, B]] + [B, [B, A]]\Big) - \frac{1}{24}[B, A, [A, B]]] + \dots$$

Per alcune informazioni ulteriori su questi argomenti si può vedere come al solito Wikipedia o, meglio, leggere un libro sui gruppi di Lie.

3.6 Calcolo di e^A mediante la forma di Jordan.

Se A e B sono due matrici simili esiste una matrice invertibile P tale che $A = P^{-1}BP$. Allora $A^k = P^{-1}B^kP$ e quindi dalla definizione di e^A come serie si ha che

$$e^A = P^{-1} \cdot e^B \cdot P$$

e in particolare $\det(e^A) = \det(e^B)$.

Il calcolo di e^A può essere dunque ricondotto al calcolo dell'esponenziale di una matrice in forma di Jordan. Poiché una matrice in forma di Jordan è una matrice a blocchi, tutte le sue potenze avranno ancora la stessa struttura a blocchi (stesso numero di blocchi e stesse dimensioni) e quindi basta calcolare l'esponenziale di un blocco di Jordan.

Sia allora $J = \lambda I + N$ un blocco di Jordan di dimensione p. La matrice N è tutta nulla tranne la sovradiagonale, dove ci sono degli 1. N è nilpotente ed è immediato dimostrare (esercizio) che $N^m = 0$ per $m \geq p$. Le matrici λI e N commutano e perciò possiamo scrivere

$$e^J = e^{\lambda I + N} = e^{\lambda I} \cdot e^N$$

Il primo termine è facile: l'esponenziale di una matrice diagonale è ancora diagonale e i suoi termini sulla diagonale sono gli esponenziali dei termini corrispondenti nella matrice di partenza. Dunque

$$e^{\lambda I} = e^\lambda I$$

Il secondo termine è ancora più facile: poiché N è nilpotente, la serie esponenziale è un polinomio e si ha:

$$e^N = I + N + \frac{1}{2!}N^2 + \cdots + \frac{1}{(p-1)!}N^{p-1} = \sum_{m=0}^{p-1} \frac{N^m}{m!}$$

Per esempio, se

$$J = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix} = 2I + N$$

si ha ($p = 4$)

$$\begin{aligned} e^J &= e^2 I \cdot \sum_{m=0}^{3} \frac{N^m}{m!} \\ &= e^2 \cdot \left(I + N + \frac{1}{2}N^2 + \frac{1}{6}N^3 \right) \end{aligned}$$

Osserviamo che le matrici N^m sono sempre le stesse, qualunque sia la matrice iniziale A e dipendono solo dalla dimensione dei blocchi. Queste matrici possono essere calcolate una volta per tutte e memorizzate, e non ricalcolate ogni volta, velocizzando il calcolo.

È anche facile dare l'espressione esplicita delle matrici N^m: sia $N = J(0, p)$ un blocco di ordine p. Gli elementi non nulli sono quelli in posizione $(i, i+1)$ per $i = 1, 2, \ldots, p-1$, cioè ponendo $N = (n_{ij})$ si ha

$$n_{ij} = \delta_{i+1,j}$$

N^2 ha ancora gli elementi tutti nulli tranne quelli su una diagonale, che questa volta è data dagli elementi in posizione $(i, i+2)$ per $i = 1, 2, \ldots, p-2$.

Proseguendo in questo modo è facile vedere che N^m ha elementi non nulli solo sulla diagonale data dagli elementi in posizione $(i, i+m)$ per $i = 1, 2, \ldots, p-m$. Per esempio, ponendo $p = 4$ si hanno le matrici seguenti:

$$N = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \; N^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \; N^3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \; N^4 = 0$$

e quindi, nell'esempio precedente con $J = 2I + N$ si ha

$$e^J = e^2 \begin{pmatrix} 1 & 1 & 1/2 & 1/6 \\ 0 & 1 & 1 & 1/2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Usando la forma di Jordan, dimostriamo un'ultima proprietà dell'esponenziale.

Teorema 3.11. *Sia A una matrice quadrata (reale o complessa). Si ha*

$$\det(e^A) = e^{\operatorname{tr} A}$$

In particolare, per ogni matrice A si ha $\det(e^A) \neq 0$ e cioè e^A è invertibile.

Dimostrazione. Riducendo A in forma di Jordan, possiamo supporre che A sia triangolare superiore, con tutti gli autovalori sulla diagonale. Osserviamo che se A è reale, la sua forma di Jordan può avere elementi complessi sulla diagonale.

Le potenze A^k sono ancora triangolari superiori, con le potenze degli autovalori sulla diagonale. Dunque e^A è triangolare superiore e quindi il suo determinante è il prodotto degli elementi sulla diagonale principale. Inoltre questi elementi sono gli esponenziali degli autovalori in quanto serie il cui termine generale è della forma $\lambda^k/k!$ con λ autovalore di A.

Se allora $\lambda_1, \ldots, \lambda_n$ sono gli autovalori di A (reali o complessi) si ha:

$$\det(e^A) = e^{\lambda_1} \cdot e^{\lambda_2} \cdot \cdots \cdot e^{\lambda_n} = e^{\lambda_1 + \cdots + \lambda_n} = e^{\operatorname{tr} A}$$

perché la traccia di una matrice è la somma dei suoi autovalori. □

3.7 Esponenziale di matrici e sistemi di equazioni differenziali.

Sia A una matrice quadrata reale di ordine n e sia $X = (x_1(t), \ldots, x_n(t))$ un vettore di funzioni di una variabile reale indeterminate. Sia inoltre $X_0 = (x_0, \ldots, x_n) \in \mathbb{R}^n$. Consideriamo il problema di Cauchy

$$\begin{cases} X' = AX \\ X(0) = X_0 \end{cases}$$

Allora, come si impare nei corsi di Analisi Matematica, questo problema di Cauchy ha soluzione unica, data da

$$X(t) = e^{tA} X_0$$

È quindi importante sapere calcolare l'esponenziale di una matrice. Per una trattazione completa dei sistemi lineari a coefficienti costanti vedere, per esempio, i paragrafi 5.3 e 5.4 del libro *Differential Equations, Dynamical Systems, and Linear Algebra*, di M. Hirsh e S. Smale. Vi sono alcune copie di questo libro nella biblioteca di Matematica. Il teorema qui sopra si trova a pag. 90.

Una copia della nuova edizione del libro, intitolata *Differential Equations, Dynamical Systems, and an Introduction to Chaos*, di M. Hirsch, S, Smale, R. Devaney si trova nella biblioteca di Fisica. In questa edizione la teoria è sviluppata nel capitolo 6 e il teorema qui sopra si trova a pag. 129.

Osserviamo solo che poiché A è reale, anche la matrice e^A è reale. Se alcuni autovalori sono complessi, la sua forma di Jordan è però complessa, e cioè alcune delle funzioni $x_i(t)$ saranno a valori complessi. Queste funzioni saranno degli esponenziali complessi, che sappiamo riscrivere come combinazioni di seni e coseni. Alternativamente, leggere la sezione 6.2 di queste note per imparare a scrivere la forma reale di Jordan a partire da quella complessa.

3.8 Esercizi

Ricordiamo che la notazione $J_k(a)$ rappresenta un blocco di Jordan di autovalore a e dimensione k.

1. Calcolare l'esponenziale delle matrici:
$$A = J_4(0), \quad B = J_4(3), \quad C = J_3(i)$$
2. Calcolare l'esponenziale delle matrici:
$$A = tJ_4(0), \quad B = tJ_4(3), \quad C = tJ_3(i)$$

4 Polinomio minimo e diagonalizzazione

4.1 Le radici del polinomio minimo

Abbiamo visto nel Teorema 2.5 che il polinomio minimo e il polinomio caratteristico di una matrice hanno le stesse radici. Poiché per il teorema di Cayley-Hamilton il polinomio minimo divide il polinomio caratteristico, decomponendo i polinomi in fattori lineari con le loro radici abbiamo

$$c(t) = (t-\lambda_1)^{m_1} \cdot (t-\lambda_2)^{m_2} \cdots (t-\lambda_k)^{m_k}$$
$$m(t) = (t-\lambda_1)^{n_1} \cdot (t-\lambda_2)^{n_2} \cdots (t-\lambda_k)^{n_k}$$

dove per ogni i si ha $1 \leq n_i \leq m_i$, e $m_1 + m_2 + \cdots + m_k = n$, l'ordine della matrice A.

Sia J la forma di Jordan della matrice A. I numeri m_i e n_i hanno un significato: m_i è la molteplicità dell'autovalore λ_i e quindi è il numero di volte che λ_i compare in J. Invece n_i è la potenza minima che annulla la restrizione di J al sottospazio invariante $\ker(A-\lambda_i I)^{p_i}$ e quindi è la dimensione del più grande blocco di Jordan relativo all'autovalore λ_i. Osserviamo che in effetti $n_i = p_i$, l'esponente che compare nella dimostrazione del teorema sull'esistenza della forma di Jordan e che fornisce i sottospazi invarianti in cui decomporre lo spazio.

Conseguenza immediata di questa discussione è una nuova dimostrazione del Teorema 2.10:

Teorema 4.1. *Una matrice è diagonalizzabile se e solo se il suo polinomio minimo ha tutte radici di molteplicità* 1.

Dimostrazione. Infatti una matrice è diagonalizzabile se e solo se tutti i blocchi di Jordan hanno dimensione 1. □

Se il polinomio caratteristico ha radici distinte, allora $m_i = 1$ per ogni i e quindi in questo caso $c(t) = m(t)$ e la matrice è diagonalizzabile.

4.2 Il grado dei fattori del polinomio minimo

Vogliamo adesso dare un criterio per stabilire quando un polinomio a coefficienti complessi ha tutti i fattori di grado 1 *senza* effettuare la scomposizione. Applicando questo criterio al polinomio minimo di una matrice si avrà, usando il teorema 4.1, un criterio effettivo per decidere se una matrice è diagonalizzabile oppure no.

Sia $p(t)$ un polinomio qualunque e scriviamo la sua decomposizione in fattori irriducibili a coefficienti complessi

$$p(t) = (t - \lambda_1)^{m_1} \cdot (t - \lambda_2)^{m_2} \cdots (t - \lambda_k)^{m_k}$$

Calcoliamo la derivata del polinomio $p(t)$: la derivata di un prodotto è una somma di termini, ognuno la derivata di un fattore moltiplicato per gli altri fattori. Poiché i fattori sono delle potenze di polinomi di primo grado con coefficiente direttore 1, la derivata è il fattore stesso elevato all'esponente diminuito di 1 (e moltiplicato per l'esponente). Raccogliendo da tutti gli addendi i fattori in comune si ottiene

$$p'(t) = \prod_{i=1}^{k} (t - \lambda_i)^{m_i - 1} \cdot \sum_{i=1}^{k} \left[m_i \prod_{j \neq i} (t - \lambda_i) \right]$$

Se scomponiamo la somma in fattori, nessuno di questi fattori può essere uno dei $(t - \lambda_i)$ perché ognuno di questi divide tutti gli addendi della somma tranne uno e quindi non può dividere la somma.

Dunque il massimo comun divisore fra $p(t)$ e $p'(t)$ è

$$\mathrm{MCD}(p(t), p'(t)) = (t - \lambda_1)^{m_1 - 1} \cdot (t - \lambda_2)^{m_2 - 1} \cdots (t - \lambda_k)^{m_k - 1}$$

e si ottiene

Proposizione 4.2. *Un polinomio complesso $p(t)$ ha tutti i fattori di grado* 1 *se e solo se* $\mathrm{MCD}(p(t), p'(t)) = 1$.

Conseguenza immediata di questa proposizione e del teorema 4.1 è:

Teorema 4.3. *Sia A una matrice quadrata complessa e sia $m(t)$ il suo polinomio minimo. Allora A è diagonalizzabile se e solo se* $\mathrm{MCD}(m(t), m'(t)) = 1$.

Possiamo dare un altro criterio, usando il polinomio caratteristico invece del polinomio minimo. Sia nuovamente $p(t)$ un polinomio qualunque con decomposizione in fattori irriducibili

$$p(t) = (t - \lambda_1)^{m_1} \cdot (t - \lambda_2)^{m_2} \cdots (t - \lambda_k)^{m_k}$$

raccogliendo insieme tutti i fattori che hanno molteplicità uguale possiamo riscrivere la decomposizione come

$$p(t) = p_1(t) \cdot p_2(t)^2 \cdot \ldots \cdot p_r(t)^r$$

dove r è la massima potenza a cui compare un fattore di $p(t)$. Osserviamo anche che i polinomi $p_i(t)$ hanno tutti i fattori di primo grado e non hanno fattori in comune fra loro.

Calcolando la derivata di $p(t)$ con la regola del prodotto come in precedenza si ottiene

$$\begin{aligned} p'(t) &= \prod_{i=1}^{r} p_i(t)^{i-1} \cdot \sum_{i=1}^{r} \left[i p_i'(t) \prod_{j \neq i} p_j(t) \right] \\ &= p_2(t) \cdot p_3(t)^2 \cdot \ldots \cdot p_r(t)^{r-1} \cdot q(t) \end{aligned}$$

dove $q(t)$ è la somma di r addendi, ognuno dei quali è divisibile per tutti i fattori di $p(t)$ tranne uno e quindi non ha fattori in comune con $p(t)$. Dunque il massimo comun divisore fra $p(t)$ e $p'(t)$ è

$$d(t) = \mathrm{MCD}\,(p(t), p'(t)) = p_2(t) \cdot p_3(t)^2 \cdot \ldots \cdot p_r(t)^{r-1}$$

e quindi si ha

$$\frac{p(t)}{d(t)} = p_1(t) \cdot p_2(t) \cdot \ldots \cdot p_r(t)$$

il prodotto dei fattori di $p(t)$ tutti a primo grado. Se applichiamo questo ragionamento al polinomio caratteristico $c(t)$ possiamo, calcolando la derivata $c'(t)$, il massimo comun divisore $d(t) = \mathrm{MCD}(c(t), c'(t))$ e poi dividendo $g(t) = c(t)/d(t)$ (tutte operazioni che non richiedono la scomposizione a priori di $c(t)$ in fattori) ottenere un polinomio $g(t)$ che è il prodotto dei fattori del polinomio caratteristico tutti a primo grado. Calcoliamo ora $g(A)$. Se $g(A) \neq 0$ allora $g(t)$ non è il polinomio minimo e quindi, poiché $g(t)$ ha tutti i fattori del polinomio minimo, nel polinomio minimo uno dei fattori deve avere grado maggiore di 1 e quindi la matrice non è diagonalizzabile. Se invece $g(A) = 0$, allora $g(t)$ è divisibile per il polinomio minimo e poiché ha tutti i fattori di grado 1, deve essere il polinomio minimo. Dunque il polinomio minimo ha tutti i fattori di grado 1 e quindi la matrice è diagonalizzabile. Otteniamo perciò

Teorema 4.4. *Sia A una matrice quadrata complessa e sia $c(t)$ il suo polinomio caratteristico. Poniamo*

$$g(t) = \frac{c(t)}{\mathrm{MCD}(c(t), c'(t))}$$

Allora A è diagonalizzabile se e solo se $g(A) = 0$ (e in questo caso, $g(t)$ è il polinomio minimo di A).

4.3 Esercizi

1. Sia A una matrice con polinomio caratteristico $c(t) = (t-2)^3(t+1)^4$ e polinomio minimo $m(t) = (t-2)^2(t+1)^2$. Quali sono le possibili forme di Jordan di A?

2. Sia A una matrice di ordine 4 con due autovalori distinti. Quante (e quali) sono le possibili forme di Jordan di A.

3. Stessa domanda di prima, con A di ordine 7 e tre autovalori.

4. Ricordiamo che un endomorfismo g si dice *nilpotente* se esiste un intero $m \geq 1$ tale che g^m è l'endomorfismo nullo, e la definizione analoga si dà per una matrice quadrata A.

   Dimostrare che un endomorfismo (o una matrice quadrata) è nilpotente se e solo se ha tutti gli autovalori nulli.

5 La forma di Jordan astratta

5.1 Operatori semisemplici e nilpotenti

La forma di Jordan che abbiamo visto in modo concreto sulle matrici ha una versione più astratta relativa agli endomorfismi, cioè alle applicazioni lineari di uno spazio vettoriale V in se stesso. Ricordiamo che un endomorfismo f si dice

- *semisemplice* se esiste una base di V formata da autovettori di f. In questa base la matrice di f risulta diagonale.
- *nilpotente* se esiste un intero $m \geq 1$ tale che f^m è l'endomorfismo nullo.

Sia A una matrice e B la sua forma di Jordan. Allora $A = P^{-1}BP$, e B è della forma $B = D + N$, somma di una matrice diagonale e di una nilpotente. Scrivendo allora

$$A = (P^{-1}DP) + (P^{-1}NP)$$

vediamo che la matrice A si può scrivere come somma di una matrice diagonalizzabile e una nilpotente. Questo dice che ogni endomorfismo si può scrivere come somma di un endomorfismo semisemplice e uno nilpotente.

La decomposizione di Jordan ha l'importante proprietà di essere unica. Ci chiediamo se anche la decomposizione di un endomorfismo come semisemplice + nilpotente ha una proprietà di unicità. Notiamo che il problema ha senso perché la forma di Jordan è una tale decomposizione, e la forma di Jordan è unica, ma potrebbero esistere altre decomposizioni che non sono scritte in forma di Jordan ed essere lo stesso del tipo semisemplice + nilpotente.

Concludiamo questi preliminari con un'osservazione:

Lemma 5.1. *Un endomorfismo è nilpotente se e solo se ha tutti gli autovalori nulli.*

Dimostrazione. Sia $m(t)$ il polinomio minimo. Se gli autovalori sono tutti nulli, allora l'unica radice di $m(t)$ è 0 e quindi $m(t) = t^k$ e poiché $m(A) = A^k = 0$, l'endomorfismo è nilpotente.

Viceversa, se A è nilpotente allora esiste $n \geq 1$ tale che $A^n = 0$ e quindi il polinomio t^n è un multiplo del polinomio minimo, che è quindi $m(t) = t^k$. Questo polinomio ha come radice solo 0 e poiché gli autovalori sono radici del polinomio minimo, tutti gli autovalori sono nulli.

□

5.2 La decomposizione di Jordan astratta.

Sia V uno spazio vettoriale complesso, e $A : V \to V$ un endomorfismo. La decomposizione di Jordan è $A = A_s + A_n$, dove A_s è semisemplice, A_n è nilpotente e inoltre $A_s A_n = A_n A_s$, cioè A_s e A_n commutano (quest'ultima proprietà si verifica facilmente per le matrici che danno la forma di Jordan). Dimostriamo ora che:

Teorema 5.2. *Se $A = S+N$ si decompone come la somma di un endomorfismo semisemplice e uno nilpotente che commutano, allora la decomposizione è unica (e quindi è quella di Jordan). Inoltre S e N sono dei polinomi in A.*

Dimostrazione. Siano A_s e A_n gli endomorfismi che si ottengono dalla decomposizione di Jordan. Il punto centrale della dimostrazione è provare che A_s e A_n sono dei polinomi in A. Da questa proprietà è semplice dimostrare l'unicità, che vediamo subito. Sia infatti

$$A = A_s + A_n = S + N$$

con S semisemplice e N nilpotente e tali che S e N commutano. Allora

$$AS = (S + N)S = S^2 + NS = S^2 + SN = SA$$

e similmente per N, cioè S e N commutano con A. Poiché A_s e A_n sono dei polinomi in A, S e N commutano anche con loro. Dall'uguaglianza delle decomposizioni si ha:

$$A_s - S = N - A_n$$

A sinistra abbiamo due endomorfismi semisemplici che commutano. Sono quindi simultaneamente diagonalizzabili e quindi anche la loro differenza è diagonalizzabile, perché nella base comune di autovettori la matrice della differenza è diagonale. A destra abbiamo due endomorfismi nilpotenti che commutano e la formula del binomio di Newton mostra che la differenza è nilpotente: se gli ordini di nilpotenza sono k e l, allora $(N - A_n)^{k+l-1} = 0$.

Allora abbiamo un endomorfismo che è nilpotente, e quindi con tutti gli autovalori nulli, e anche semisemplice, e quindi simile ad una matrice diagonale: l'unica possibilità è l'endomorfismo nullo, e cioè $S = A_s$ e $N = A_n$, come voluto.

Dimostriamo ora che A_s e A_n sono polinomi in A. Basta trovare un polinomio $p(t)$ tale che $p(A) = A_s$, perché allora $q(t) = t - p(t)$ dà $q(A) = A - A_s = A_n$.

Siano $\lambda_1, \ldots, \lambda_k$ gli autovalori di A con molteplicità $m_1, \ldots, m_k$, in modo che il polinomio caratteristico di A sia

$$\det(tI - A) = (t - \lambda_1)^{m_1}(t - \lambda_2)^{m_2} \cdot \cdots \cdot (t - \lambda_k)^{m_k}$$

Dal Lemma 3.4 (e l'induzione) sappiamo che lo spazio V si decompone come somma diretta dei sottospazi

$$V_i = \ker(A - \lambda_i I)^{m_i}$$

e che A_s è l'endomorfismo semisemplice che vale $\lambda_i I$ su V_i. Basta infatti osservare che il numero p per l'autovalore λ_i usato nel Lemma 3.4 è minore o uguale a m_i e quindi $\ker(A - \lambda_i)^p = \ker(A - \lambda_i)^{m_i}$.

Consideriamo i polinomi $(t - \lambda_1)^{m_1}, \ldots, (t - \lambda_k)^{m_k}$, a cui aggiungiamo t se tutti gli autovalori sono diversi da zero. I polinomi sono relativamente primi a due a due nell'anello $\mathbb{C}[t]$ e quindi, per il teorema cinese dei resti, si ha:

Lemma 5.3. *Esiste un polinomio $p(t)$ per cui valgono simultaneamente tutte le condizioni seguenti:*

$$\begin{aligned} p(t) &= \lambda_1 + h_1(t) \cdot (t-\lambda_1)^{m_1} \\ p(t) &= \lambda_2 + h_2(t) \cdot (t-\lambda_2)^{m_2} \\ &\dots \\ p(t) &= \lambda_k + h_k(t) \cdot (t-\lambda_k)^{m_k} \\ p(t) &= h(t) \cdot t \end{aligned}$$

dove $h_1(t)$, ..., $h_2(t)$, $h(t)$ sono opportuni polinomi.

Dimostrazione. Stiamo evidentemente risolvendo il sistema di congruenze simultanee

$$\begin{aligned} p(t) &\equiv \lambda_1 \mod (t-\lambda_1)^{m_1} \\ p(t) &\equiv \lambda_2 \mod (t-\lambda_2)^{m_2} \\ &\dots \\ p(t) &\equiv \lambda_k \mod (t-\lambda_k)^{m_k} \\ p(t) &\equiv 0 \mod t \end{aligned}$$

□

Allora $p(A)$ si comporta come la moltiplicazione per λ_i sul sottospazio V_i e quindi $p(A) = A_s$, e come già detto basta porre $q(t) = t - p(t)$ per avere $q(A) = A - A_s = A_n$. Questo conclude la dimostrazione del teorema. □

6 La forma di Jordan reale

6.1 Matrici simili su $\mathbb{R}$ e su $\mathbb{C}$.

Ci occuperemo ora di estendere il teorema sulla forma di Jordan al caso di matrici reali. La prima osservazione è la seguente: se una matrice *reale* A ha tutti gli autovalori reali, allora non solo la sua forma di Jordan A_J è una matrice reale, ma anche la matrice di passaggio è reale. Infatti, tutti i ragionamenti fatti nella dimostrazione del Teorema 3.3, e in particolare la determinazione della base di vettori in cui A si scrive in forma di Jordan, non dipendono da proprietà dei numeri complessi ma solo dall'esistenza degli autovalori.

Ci possiamo allora porre la seguente domanda: se due matrici reali A e B sono simili come matrici complesse, sono anche simili su $\mathbb{R}$? Cioè, se esiste una matrice complessa P tale che $A = PBP^{-1}$, esiste una matrice reale Q tale che $A = QBQ^{-1}$? La risposta è sì, come vedremo nel seguente teorema.

Teorema 6.1. *Siano A e B due matrici reali e sia P una matrice complessa tale che $A = PBP^{-1}$. Allora esiste una matrice* reale *Q tale che $A = QBQ^{-1}$.*

Dimostrazione. Occorre dimostrare che se fra le soluzioni dell'equazione fra matrici

$$XA = BX \tag{6}$$

c'è una matrice invertibile complessa P, allora c'è anche una matrice invertibile reale Q. Poiché la (6) è un'equazione omogenea, le soluzioni complesse sono uno spazio vettoriale W su $\mathbb{C}$ con base $C_1, \ldots, C_k$. Ogni matrice C_j si può scrivere nella forma $C_j = X_j + iY_j$, dove le X_j, Y_j sono matrici reali. Poiché A e B sono matrici reali, la condizione $C_jA = BC_j$ implica che $X_jA = BX_j$ e $Y_jA = BY_j$, cioè le matrici X_j, Y_j appartengono a W, e poiché $\{X_j, Y_j\}_{j=1,\ldots,k}$ generano tutte le matrici C_j, che sono una base, generano tutto lo spazio W. Estraendo da questi generatori una base abbiamo che W ha una base formata da matrici reali $D_1, \ldots, D_k$.

Ogni matrice di W è quindi della forma

$$D = t_1D_1 + \cdots + t_kD_k$$

al variare di $t_1, \ldots, t_k \in \mathbb{C}$ e le matrici reali in W si ottengono per valori reali di $t_1, \ldots, t_k$. Poniamo $f(t_1, \ldots, t_k) = \det(t_1D_1 + \cdots + t_kD_k)$. f è un polinomio a coefficienti reali e non è il polinomio identicamente nullo perché per ipotesi esiste una matrice $P \in W$ tale che $\det(P) \neq 0$. Allora esistono certamente dei valori *reali* per $t_1, \ldots, t_k$ per cui f non si annulla, e cioè la (6) ha una soluzione reale invertibile. □

6.2 Forma di Jordan reale.

Basi di Jordan esistono sempre per matrici (endomorfismi) complessi, ma non sempre sui numeri reali. Un semplice esempio è

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

cioè la matrice di una rotazione di 90 gradi nel piano. Poiché non ci sono autovettori reali, non ci può essere forma di Jordan reale. Però, in un senso opportuno, questo è l'unico caso in cui non c'è forma di Jordan reale. Nei teoremi seguenti preciseremo questa affermazione.

Sia A una matrice reale. La parte della base di Jordan che corrisponde agli autovalori reali di A esiste e si costruisce nello stesso modo visto in precedenza (Teorema 3.3). Quindi dobbiamo solo occuparci degli autovalori non reali.

Per prima cosa osserviamo che il polinomio caratteristico $p(t) = \det(A - tI)$ è a coefficienti reali e perciò se $\lambda \in \mathbb{C}$ è una radice, anche il coniugato $\bar{\lambda}$ è una radice. Infatti

$$0 = p(\lambda) = \overline{p(\lambda)} = p(\bar{\lambda})$$

dove l'ultima uguaglianza è dovuta al fatto che i coefficienti di $p(t)$ sono reali. Dunque gli autovalori non reali di A si presentano a coppie.

Consideriamo la matrice reale A come una matrice a coefficienti complessi.

Teorema 6.2. *C'è una corrispondenza biunivoca fra i blocchi di Jordan di A corrispondenti agli autovalori λ e $\bar{\lambda}$.*

Dimostrazione. Siano P, Q matrici (quadrate) reali e sia $B = P + iQ$, e sia $z = x + iy$ dove x, y sono vettori reali. Allora

$$Bz = 0 \iff \bar{B}\bar{z} = 0$$

Infatti, separando le parti reali e immaginarie si ha che

$$Bz = (P + iQ)(x + iy) = (Px - Qx) + i(Qx + Py) = 0$$
$$\bar{B}\bar{z} = (P - iQ)(x - iy) = (Px - Qx) - i(Qx + Py) = 0$$

dicono entrambe $(Px - Qx) = (Qx + Py) = 0$ e sono quindi equivalenti.

Poniamo ora $B = (A - \lambda I)$ e quindi $\bar{B} = \overline{(A - \lambda I)} = (A - \bar{\lambda} I)$, poiché A è reale. Dunque l'applicazione $z \mapsto \bar{z}$ dà un isomorfismo (reale) fra $\ker(A - \lambda I)^n$ e $\ker(A - \bar{\lambda} I)^n$ e quindi i blocchi di Jordan per λ e per $\bar{\lambda}$ sono lo stesso numero e delle stesse dimensioni. □

Sia $J_k^*(\lambda)$ la matrice reale $2k \times 2k$ che si ottiene dal blocco di Jordan $J_k(\lambda)$ sostituendo ognuno dei suoi elementi $a + ib$ con la matrice $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$. Per esempio, se

$$J_3(2 + 3i) = \left(\begin{array}{c|c|c} 2+3i & 1 & 0 \\ \hline 0 & 2+3i & 1 \\ \hline 0 & 0 & 2+3i \end{array}\right),$$

allora

$$J_3^*(2 + 3i) = \left(\begin{array}{cc|cc|cc} 2 & 3 & 1 & 0 & 0 & 0 \\ -3 & 2 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 2 & 3 & 1 & 0 \\ 0 & 0 & -3 & 2 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 0 & -3 & 2 \end{array}\right)$$

Teorema 6.3. *Sia A una matrice reale. Allora esiste una base reale rispetto alla quale A è in forma diagonale a blocchi, con blocchi di Jordan $J_{m_1}(t_1)$, ..., $J_{m_r}(t_r)$ per gli autovalori reali t_i e blocchi $J_{n_1}^*(\lambda_1)$, ..., $J_{n_s}^*(\lambda_s)$ per gli autovalori non reali λ_i e $\bar{\lambda}_i$.*

Dimostrazione. Se λ è un autovalore di A, allora per il teorema precedente anche $\bar{\lambda}$ è un autovalore di A e a ogni blocco di Jordan $J_n(\lambda)$ corrisponde un blocco di Jordan $J_n(\bar{\lambda})$. Inoltre, sempre per il teorema precedente, se $e_1, \ldots, e_n$ è una base di Jordan per $J_n(\lambda)$, allora $\bar{e}_1, \ldots, \bar{e}_n$ è una base di Jordan per $J_n(\bar{\lambda})$.

Poniamo $e_k = x_k + iy_k$, dove gli x_k, y_k sono vettori reali. Allora i vettori $x_1, y_1, \ldots, x_n, y_n$ generano un sottospazio di dimensione $2n$ e sono quindi linearmente indipendenti (su $\mathbb{C}$, e quindi anche su $\mathbb{R}$).

Calcoliamo come diventa la matrice di A nella base $\{x_1, y_1, \ldots, x_n, y_n\}$. Si ha

$$Ae_1 = \lambda e_1 = (a + ib)(x_1 + iy_1) = (ax_1 - by_1) + i(bx_1 + ay_1)$$

e inoltre

$$Ae_1 = Ax_1 + iAy_1$$

Poiché A, x_1 e y_1 sono reali, deve essere

$$Ax_1 = ax_1 - by_1$$
$$Ay_1 = bx_1 + ay_1$$

e quindi le prime due colonne di A sono quelle di $J_n^*(\lambda)$. Per $k \geq 2$ si ha

$$Ae_k = \lambda e_k + e_{k-1} = (ax_k - by_k + x_{k-1}) + i(bx_k + ay_k + y_{k-1})$$

e di nuovo

$$Ae_k = Ax_k + iAy_k$$

Separando parte reale e immaginaria come prima, si ha che anche tutte le altre colonne di A sono quelle di $J_n^*(\lambda)$. □