Note del corso di Geometria / Geometria 1 M-Z
Corsi di Laurea in Fisica, IADA e Matematica
anno accademico 2021/22

Prof.ssa Emilia Mezzetti
mezzette@units.it

20 gennaio 2022

Indice

Capitolo 1

Preliminari di algebra

1.1 Operazioni su insiemi

Definizione 1.1.1 (Prodotto cartesiano)**.** Dati due insiemi A, B, il loro **prodotto cartesiano**, indicato con $A \times B$, è l'insieme delle coppie ordinate (a, b) con $a \in A$ e $b \in B$, cioè $A \times B = \{(a, b) \mid a \in A, b \in B\}$.

Esempio 1.1.2. *1. Se $A = \{1, 2, 3\}$, $B = \{a, b\}$, allora*

$$A \times B = \{(1, a), (1, b), (2, a), (2, b), (3, a), (3, b)\}.$$

2. Sia $\mathbb{R}$ l'insieme dei numeri reali, allora

$$\mathbb{R} \times \mathbb{R} = \{(x, y) \mid x \in \mathbb{R}, y \in \mathbb{R}\} = \mathbb{R}^2.$$

Notiamo che le coppie sono ordinate, dunque per esempio $(1, 2) \neq (2, 1)$.

Definizione 1.1.3 (Operazione interna)**.** Sia S un insieme non vuoto. Un'**operazione interna in S o legge di composizione interna in S** è un'applicazione

$$S \times S \to S$$

avente dominio $S \times S$ e codominio S. Per esempio una tale operazione può essere denotata con uno dei simboli $*$, oppure $+$, o $\cdot$, o con altri simboli; nel primo caso essa associa ad una coppia (a, b) un elemento di S, denotato $a * b$.

Notare che il termine "applicazione" è sinomino di "funzione". Un altro termine usato a volte con lo stesso significato è "mappa".

Esempio 1.1.4. *Esempi di operazione interne.*

(i) $+ : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$, *la somma è un'operazione interna in* $\mathbb{Z}$*;*

$(a, b) \to a + b$*; per esempio*

$(1, 2) \to 1 + 2 = 3.$

(ii) $\cdot : \mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$, *il prodotto è un'operazione interna in* $\mathbb{Q}$*;*

$(x, y) \to x \cdot y$*; per esempio*

$(1, 2) \to 1 \cdot 2 = 2$.

(iii) Sia $X \neq \emptyset$ *un insieme non vuoto e sia* F *l'insieme delle applicazioni di* X *in* X*, cioè aventi* X *sia come dominio sia come codominio. Si noti che* F *non è vuoto in quanto contiene almeno l'applicazione identica* $id_X : X \to X$*, tale che* $id_X(x) = x$ *per ogni* $x \in X$.

$\circ : F \times F \to F$*, la composizione è un'operazione interna in* F*;*

$(f, g) \to f \circ g$*, dove* $f \circ g$ *è l'applicazione tale che* $(f \circ g)(x) = f(g(x))$ *per ogni* $x \in X$.

Gli esempi (i) e (ii) sono esempi di operazioni numeriche.

1.2 Gruppi

Definizione 1.2.1 (Gruppo)**.** Sia G un insieme e $*$ sia un'operazione in G. La coppia $(G, *)$ è detta un **gruppo** se valgono le seguenti proprietà:

(i) *Proprietà associativa*: per ogni $a, b, c \in G$, si ha $a * (b * c) = (a * b) * c$;

(ii) *Esistenza dell'elemento neutro*: esiste $e \in G$ tale che, per ogni $a \in G$, si ha $e * a = a * e = a$; e è detto elemento neutro di G;

(iii) *Esistenza dei simmetrici, o reciproci*: per ogni $a \in G$ esiste $a' \in G$ tale che $a * a' = e = a' * a$. a' è detto reciproco di a.

Se l'operazione è indicata additivamente, ossia con il simbolo $+$, l'elemento neutro è detto "zero" e indicato 0, mentre il reciproco di a è detto opposto di a e indicato $-a$. Se l'operazione è indicata moltiplicativamente, ossia con il simbolo $\cdot$ o $\times$, l'elemento neutro è detto "uno" o unità di G e indicato 1 o 1_G, mentre il reciproco di a è detto inverso di a e indicato a^{-1}.

Definizione 1.2.2 (Gruppo abeliano)**.** Il gruppo $(G, *)$ è detto **gruppo abeliano**, o commutativo, se vale la *proprietà commutativa*, cioè per ogni $a, b \in G$ vale $a * b = b * a$.

Esempio 1.2.3.

1. $(\mathbb{Z}, +)$ *è un gruppo abeliano.*

2. $(\mathbb{Z}, \cdot)$ *non è un gruppo: la proprietà associativa è verificata, e l'*1 *esiste, però alcuni elementi non hanno l'inverso in* $\mathbb{Z}$*, si dice che "non sono invertibili" in* $\mathbb{Z}$*. Per esempio* 0 *non ha inverso, e anche* $2 \cdot z \neq 1$ *per ogni* $z \in \mathbb{Z}$*, quindi* 2 *non è invertibile in* $\mathbb{Z}$.

3. $(\mathbb{Q}, +)$ *è un gruppo abeliano.*

4. $(\mathbb{Q} \setminus \{0\}, \cdot)$ *è un gruppo abeliano.*

   Infatti, osserviamo innanzitutto che il prodotto è un'operazione interna in $\mathbb{Q} \setminus \{0\}$*, perchè il prodotto di due numeri razionali non nulli è non nullo. Poi: vale la proprietà associativa, l'elemento neutro è l'1, e per ogni* $q \in \mathbb{Q} \setminus \{0\}$ *esiste* $q^{-1} = \frac{1}{q}$ *tale che* $q \cdot \frac{1}{q} = \frac{1}{q} \cdot q = 1$.

5. *Sia* $X \neq \emptyset$ *un insieme e sia* $I(X) = \{f : X \to X \mid f \text{ biiettiva}\}$ *l'insieme delle applicazioni biunivoche di* X *in sè.*

   $(I(X), \circ)$ *è un gruppo. Infatti:*

   (i) se $f, g : X \to X$ *sono biiettive, anche* $f \circ g$ *lo è, dunque la composizione è un'operazione interna in* $I(X)$*;*

   (ii) la composizione di funzioni è associativa: $(f \circ g) \circ h = f \circ (g \circ h)$*. Infatti per ogni* $x \in X$ *si ha* $((f \circ g) \circ h)(x) = f(g(h(x))) = (f \circ (g \circ h))(x)$.

   (iii) l'applicazione identica $id_X : x \to x$*, per ogni* $x \in X$*, è l'elemento neutro di* $I(X)$*;*

   (iv) ricordiamo che un'applicazione è biiettiva se e solo se esiste l'applicazione inversa $f^{-1} : X \to X$*, tale che* $f(x) = y$ *se e solo se* $f^{-1}(y) = x$*. Infatti* f *è suriettiva e iniettiva, se e solo se, preso comunque un elemento* $y \in X$*, esiste ed è unico* $x \in X$ *tale che* $f(x) = y$*. L'applicazione* f^{-1} *è l'elemento inverso di* f *rispetto all'operazione* $\circ$.

Osserviamo che tutti i gruppi "numerici" sono abeliani. Invece il gruppo $I(X)$ non è abeliano se X ha almeno tre elementi.

Per esempio, sia $X = \{1, 2, 3\}$. Definiamo $f : X \to X$ ponendo

$$f(1) = 2, f(2) = 3, f(3) = 1,$$

e $g : X \to X$ ponendo

$$g(1) = 1, g(2) = 3, g(3) = 2.$$

Chiaramente $f \circ g \neq g \circ f$, in quanto esiste almeno un elemento $x \in X$ tale che $(f \circ g)(x) \neq (g \circ f)(x)$.

Osserviamo che in questo caso $I(X)$ ha sei elementi, corrispondenti alle permutazioni dell'insieme X; si veda il Capitolo 11.

Esercizi 1.

1. Costruire un esempio analogo al precedente per X insieme di n elementi, con $n \geq 3$ qualunque.

2. Se X ha n elementi, quanti elementi ha $I(X)$?

Proposizione 1.2.4. *Sia* $(G, *)$ *un gruppo.*

1. *L'elemento neutro in* G *è unico.*

2. *Ogni elemento* $g \in G$ *ha un unico reciproco.*

Dimostrazione. 1. Siano e, e' entrambi elementi neutri di G, ossia elementi di G tali che, per ogni $g \in G$, si ha $e * g = g * e = g$ e $e' * g = g * e' = g$. Allora $e * e' = e'$ perchè e è neutro, ma anche $e * e' = e$ perchè e' è neutro. Dunque $e = e'$.

2. Supponiamo che g', g'' siano entrambi reciproci di g. Allora si ha $g * g' = g' * g = e$ e anche $g * g'' = g'' * g = e$. Quindi

$$g' = g' * e = g' * (g * g'') = \text{per la proprietà associativa} = (g' * g) * g'' = e * g'' = g''$$

In conclusione si ha $g' = g''$. □

1.3 Relazioni d'equivalenza

Per questa sezione si vedano anche le note del corso propedeutico (Prof. Del Santo).

Una **relazione** in un insieme X è una proprietà che una coppia ordinata di elementi di X può verificare o meno. Per esempio la relazione "$<$" "minore" ha senso negli insiemi numerici $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$; la relazione "$||$" "parallelo" ha senso nell'insieme delle rette del piano, o dei piani dello spazio.

In maniera più formale, una relazione in X è un sottinsieme R del prodotto cartesiano $X \times X$. In tal caso si dirà che x è in relazione R con y se la coppia ordinata $(x, y) \in R$. Si scrive anche xRy.

Per esempio la relazione $<$ in $\mathbb{Z}$ corrisponde al sottinsieme di $\mathbb{Z} \times \mathbb{Z}$: $\{(x, y) \mid x < y\}$. Analogamente la relazione $\leq$ corrisponde al sottinsieme di $\mathbb{Z} \times \mathbb{Z}$: $\{(x, y) \mid x \leq y\}$. La relazione di parallelismo nell'insieme delle rette del piano corrisponde alle coppie di rette (r, r') tali che r, r' sono distinte e parallele oppure sono uguali.

Simboli spesso usati per denotare relazioni sono $\equiv$, $\sim$, $\simeq$, $\cong$, ecc. Un altro esempio di relazione, in $\mathbb{R}$, è il seguente: $x \sim y$ se e solo se $x^2 = y$.

Noi saremo interessati a un tipo particolare di relazioni dette relazioni d'equivalenza.

Definizione 1.3.1 (**Relazione d'equivalenza**). Sia X un insieme e $\sim$ una relazione in X. Si dice che $\sim$ è una **relazione d'equivalenza** se valgono le tre proprietà:

1. riflessiva: per ogni $x \in X$ $x \sim x$;
2. simmetrica: se $x \sim y$ allora $y \sim x$;
3. transitiva: se $x \sim y$ e $y \sim z$ allora $x \sim z$.

Esempio 1.3.2.

1. L'uguaglianza è una relazione d'equivalenza in qualunque insieme X.
2. "Essere congruenti" è una relazione d'equivalenza nell'insieme dei triangoli del piano.
3. $\leq$, $<$ non sono relazioni d'equivalenza.

Il prossimo è un esempio fondamentale. Denotiamo con $\mathbb{N}$ l'insieme dei numeri naturali: $\mathbb{N} = \{0, 1, 2, 3 \ldots\}$.

Definizione 1.3.3 (**Congruenza modulo n**). Si fissi un naturale $n \in \mathbb{N}$. La relazione di congruenza modulo n è la relazione in $\mathbb{Z}$ così definita:

$$x \equiv y \pmod{n} \text{ se e solo se esiste } k \in \mathbb{Z} \text{ tale che } x - y = kn.$$

Si scrive anche $x \equiv_n y$. Si legge "x è congruo a y modulo n.

Proposizione 1.3.4. *La relazione di congruenza modulo n è una relazione d'equivalenza in $\mathbb{Z}$.*

Dimostrazione. 1. $x - x = 0x$ per ogni $x \in \mathbb{Z}$.

2. Se $x \equiv_n y$, si ha $x - y = kn$ per un opportno $k \in \mathbb{Z}$. Allora $y - x = (-k)n$.

3. Se $x \equiv_n y$ e $y \equiv_n z$, esistono $k, h \in \mathbb{Z}$ tali che $x - y = kn$, $y - z = hn$; ma allora $x - z = (x - y) + (y - z) = kn + hn = (k + h)n$, il che prova che $x \equiv_n z$. □

D'ora in poi supporremo sempre $n \geq 2$. La seguente osservazione è importante.

Proposizione 1.3.5. *$x \equiv_n y$ se e solo se x e y hanno lo stesso resto nella divisione per n.*

Dimostrazione. Infatti se x e y hanno lo stesso resto nella divisione per n, si ha: $x = qn + r$, $y = q'n + r$, dove $0 \leq r \leq n - 1$. Ma allora $x - y = (qn + r) - (q'n + r) = (q - q')n$ e perciò $x \equiv_n y$.

Viceversa se $x \equiv_n y$, si ha $x = y + kn$. Se r è il resto della divisione di y per n, vale la relazione $y = qn + r$ con $0 \leq r \leq n - 1$; perciò si ha $x = (qn + r) + kn = (q + k)n + r$, dunque r è anche il resto della divisione di x per n. □

Definizione 1.3.6 (**Classi d'equivalenza e insieme quoziente**)**.** Sia X un insieme in cui è definita una relazione d'equivalenza $\sim$, sia $x \in X$. La **classe d'equivalenza** di x è il sottinsieme di X formato dagli elementi equivalenti a x:

$$[x] = \{y \in X \mid y \sim x\}.$$

Tale insieme si denota anche $[x]_\sim$.

L'insieme delle classi d'equivalenza è detto **insieme quoziente** di X rispetto alla relazione $\sim$ e si indica $X/\sim$.

L'insieme quoziente è un sottinsieme delle insieme delle parti di X, $\mathcal{P}(X)$. Osserviamo che $x \in [x]$ per la proprietà riflessiva. Quindi nessun elemento dell'insieme quoziente $X/\sim$ è l'insieme vuoto $\emptyset$. Inoltre le classi d'equivalenza ricoprono X, ossia X è l'unione delle classi d'equivalenza $[x]$, al variare di $x \in X$.

Definizione 1.3.7 (**Partizione**)**.** Una **partizione** di un insieme X è un sottinsieme Π dell'insieme delle parti di X che gode delle proprietà:

1. nessun elemento di Π è vuoto;
2. l'unione degli insiemi di Π è uguale a X;
3. se $S, T \in \Pi$, e $S \neq T$ allora $S \cap T = \emptyset$.

Dunque due elementi di una partizione Pi o sono disgiunti o sono uguali.

Proposizione 1.3.8. *L'insieme quoziente $X/\sim$ di una relazione d'equivalenza in X è una partizione di X.*

Dimostrazione. Le prime due proprietà sono già state osservate. Per provare la terza, consideriamo due classi d'equivalenza $[x], [y]$ tali che $[x] \cap [y] \neq \emptyset$. Allora esiste $z \in [x] \cap [y]$, cioè $z \sim x$ e $z \sim y$. Per le proprietà simmetrica e transitiva segue che $x \sim y$. Proviamo che di conseguenza $[x] = [y]$. Infatti, se $u \in [x]$, allora $u \sim x$, ma $x \sim y$, dunque per la proprietà transitiva $u \sim y$ e segue che $u \in [y]$. Abbiamo così provato che $[x] \subset [y]$. L'inclusione opposta è analoga. □

Esempio 1.3.9.

1. L'insieme quoziente $\mathbb{Z}/\!\equiv_n$ si denota $\mathbb{Z}_n$. $\mathbb{Z}_n$ ha n elementi, uno per ciascuno degli n resti della divisione per n: $0, 1, 2, \ldots, n-1$. Infatti se x ha resto r nella divisione per n, $x = qn + r$ dunque $x \equiv_n r$. Gli elementi di $\mathbb{Z}_n$ si denotano anche $[r]_n$ o $\bar{r}$. Dunque $\mathbb{Z}_n = \{\bar{0}, \bar{1}, \ldots, \overline{n-1}\}$.

Un insieme S si dice finito se esiste un numero naturale n tale che S è in biiezione con l'insieme $\{1, 2, 3, \ldots, n-1, n\}$. Dunque $\mathbb{Z}_n$ è un insieme finito con n elementi.

1.4 Operazioni in $\mathbb{Z}_n$

Sia $n \geq 2$. Nell'insieme $\mathbb{Z}_n$ si possono definire due operazioni, di somma e di prodotto, **indotte** dalle operazioni in $\mathbb{Z}$.

Siano $\bar{x}, \bar{y} \in \mathbb{Z}_n$. Definiamo

$$\bar{x} + \bar{y} = \overline{x + y},$$

$$\bar{x} \cdot \bar{y} = \overline{x \cdot y}.$$

Il prodotto si denota anche semplicemente $\bar{x}\bar{y}$. Queste operazioni di somma e prodotto sono **ben definite**, in quanto non dipendono dai particolari rappresentanti scelti per le due classi. Infatti, sia $\bar{x} = \bar{x}'$ e $\bar{y} = \bar{y}'$. Allora si ha $x' = x + kn, y' = y + hn$, per $k, h \in \mathbb{Z}$ opportuni. Quindi $(x + y) - (x' + y') = (x + y) - (x + kn + y + hn) = -(k + h)n$, da cui segue che $x + y \equiv_n x' + y'$.

Analogamente $xy - x'y' = xy - (x + kn)(y + hn) = -(xh + yk + khn)n$ e perciò $xy \equiv_n x'y'$.

Dalle proprietà della somma in $\mathbb{Z}$ seguono facilmente le proprietà della somma in $\mathbb{Z}_n$:

1. proprietà associativa: $(\bar{x} + \bar{y}) + \bar{z} = \bar{x} + (\bar{y} + \bar{z})$;
2. la classe $\bar{0}$ è l'elemento neutro della somma;
3. $\overline{-x} = -\bar{x}$;
4. proprietà commutativa: $\bar{x} + \bar{y} = \bar{y} + \bar{x}$. Ne segue

Proposizione 1.4.1. *$(\mathbb{Z}_n, +)$ è un gruppo abeliano.*

Analogamente, dalle proprietà del prodotto in $\mathbb{Z}$ segue che valgono le seguenti proprietà del prodotto in $\mathbb{Z}_n$:

1. proprietà associativa: $(\bar{x}\bar{y})\bar{z} = \bar{x}(\bar{y}\bar{z})$;
2. $\bar{1}$ è l'unità del prodotto;
3. proprietà commutativa: $\bar{x}\bar{y} = \bar{y}\bar{x}$;
4. proprietà distributiva: $(\bar{x} + \bar{y})\bar{z} = \bar{x}z + \bar{y}z$.

1.5 Campi

Definizione 1.5.1 (Campo). Sia K un insieme dotato di due operazioni, chiamate somma e prodotto e denotate $+$ e $\cdot$. La terna $(K, +, \cdot)$ si dice un **campo** se valgono le seguenti proprietà:

1. K è un gruppo abeliano rispetto alla somma;
2. proprietà associativa del prodotto;
3. esiste elemento unità;
4. ogni elemento **non nullo** di K ammette inverso;
5. proprietà commutativa del prodotto;
6. proprietà distributiva del prodotto rispetto alla somma: per ogni $a, b, c \in K$ si ha: $(a + b) \cdot c = ac + bc$.

Esempio 1.5.2.
1. Campi numerici: $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$, $(\mathbb{C}, +, \cdot)$,
2. $(\mathbb{Z}, +, \cdot)$ non è un campo perchè soltanto 1 e -1 hanno inverso.

Proposizione 1.5.3 (Proprietà generali dei campi)**.** *1. Per ogni $a \in K$ $0 \cdot a = 0$;*
2. Legge di annullamento del prodotto. Se $a \cdot b = 0$, allora $a = 0$ oppure $b = 0$;
3. Sia -1 l'opposto di 1 e $a \in K$. Allora $(-1) \cdot a = -a$.

Dimostrazione. 1. Usando la proprietà che 0 è elemento neutro per la somma e la proprietà distributiva si ottiene:

$$0 \cdot a = (0 + 0)a = 0 \cdot a + 0 \cdot a.$$

Sommando $-(0 \cdot a)$ a ambo i membri, si ottiene $0 \cdot a = 0$.

2. Sia $a \cdot b = 0$. Se $a = 0$ abbiamo finito, sia dunque $a \neq 0$. Allora esiste a^{-1}. Moltiplicando ambo i membri a sinistra per a^{-1} otteniamo

$$a^{-1} \cdot (a \cdot b) = (a^{-1} \cdot a) \cdot b = 1 \cdot b = b;$$

ma $a^{-1} \cdot (a \cdot b) = a^{-1} \cdot 0 = 0$ per il punto precedente, dunque $b = 0$.

3. $(-1) \cdot a + a = (-1) \cdot a + 1 \cdot a =$ proprietà distributiva $= ((-1) + 1) \cdot a = 0 \cdot a = 0$. Analogamente $a + (-1) \cdot a = 0$. $\square$

La legge di annullamento del prodotto garantisce che $K \setminus \{0\}$ è chiuso rispetto al prodotto. Si pu o anche esprimere dicendo che in K non vi sono divisori dello zero. Dunque le condizioni $2 - 5$ della definizione di campo si possono riassumere dicendo che $(K \setminus \{0\}, \cdot)$ è un gruppo abeliano.

D'ora in poi lavorando in un campo K useremo spesso le notazioni compatte:
$a - b = a + (-b)$
$ab = a \cdot b$
$\frac{a}{b} = a/b = ab^{-1}$: quest'ultima notazione ha senso perchè il prodotto è commutativo.

Un insieme dotato di due operazioni, che sia un gruppo abeliano rispetto alla somma, ma verificante solo la proprietà associativa per il prodotto e la proprietà distributiva (ma non necessariamente la 3., la 4. e la 5.) è detto *anello*. Se il prodotto è commutativo, è detto *anello commutativo*; se in più esiste l'unità del prodotto, è detto anello commutativo con unità. Per esempio $\mathbb{Z}$ è un anello commutativo con unità.

Un insieme verificante tutti gli assiomi di campo, eccetto la proprietà commutativa del prodotto, è detto *corpo*. Un esempio importante è il corpo dei quaternioni.

Vogliamo ora determinare per quali n $\mathbb{Z}_n$ è un campo. A tale scopo consideriamo la tabella di moltiplicazione di $\mathbb{Z}_n \setminus \{0\}$ per $n = 2, 3, 4, 5$. Per semplicità di scrittura indicheremo gli elementi di $\mathbb{Z}_n$ omettendo il segno sopra.

$n = 2$

$\cdot$	1
1	1

$n = 3$

$\cdot$	1	2
1	1	2
2	2	1

$n = 4$

$\cdot$	1	2	3
1	1	2	3
2	2	0	2
3	3	2	1

$n = 5$

$\cdot$	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1

Dalle tabelle segue che $\mathbb{Z}_4$ non è un campo, perchè $\bar{2}$ non è invertibile, mentre $\mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_5$ lo sono. In effetti, vale il seguente teorema.

Teorema 1.5.4. *Sia $n \geq 2$. Allora $\mathbb{Z}_n$ è un campo se e solo se n è un numero primo.*

Dimostrazione. Supponiamo dapprima che n non sia primo, e dimostriamo che $\mathbb{Z}_n$ non è un campo. Infatti, se n non è primo, esistono due interi a, b con $1 < a, b < n$ tali che $n = ab$. Passando alle classi di equivalenza nel quoziente $\mathbb{Z}_n$ si ottiene $\bar{n} = \bar{0} = \bar{a}\bar{b}$, che contraddice la Proposizione 1.5.3, punto 2, in quanto $\bar{a} \neq 0$ e $\bar{b} \neq 0$: $\bar{a}$ e $\bar{b}$ sono divisori dello zero.

Supponiamo ora che n sia primo e vogliamo dimostrare che $\mathbb{Z}_n$ è un campo.

Useremo le due seguenti proprietà.

1. Siano p un numero primo e $a, b \in \mathbb{Z}$. Se $p|ab$, allora o $p|a$ o $p|b$ (il segno $|$ significa "divide"). Tale proprietà segue immediatamente dal Teorema fondamentale dell'aritmetica, ossia dall'esistenza e unicità della scomposizione in fattori primi.

2. *Principio della piccionaia.* Se X è un insieme **finito** e $f : X \to X$ è un'applicazione iniettiva, allora f è anche suriettiva, e quindi è una biiezione. Infatti, se f è iniettiva, f stabilisce una biiezione fra X e $f(X)$, dunque pure $f(X)$ è finito e ha lo stesso numero di elementi di X. Essendo $f(X) \subset X$ segue che $f(X) = X$.

Fissiamo dunque $\bar{a} \in \mathbb{Z}_n$, con n primo. Supponiamo $\bar{a} \neq 0$. Vogliamo dimostrare che $\bar{a}$ è invertibile. Consideriamo l'applicazione $\varphi : \mathbb{Z}_n \to \mathbb{Z}_n$ definita da $\varphi(\bar{x}) = \bar{a}\bar{x}$: φ è la moltiplicazione per $\bar{a}$.

Osserviamo dapprima che φ è iniettiva. Infatti, se $\varphi(\bar{x}) = \varphi(\bar{y})$ ciò significa che $\bar{a}\bar{x} = \bar{a}\bar{y}$. Per definizione del prodotto in $\mathbb{Z}_n$, allora $\overline{ax} = \overline{ay}$, e quindi $ax \equiv ay \pmod{n}$. Perciò n divide $ax - ay = a(x - y)$. Dalla proprietà 1. segue che o $n|a$ o $n|x - y$. La prima è impossibile perchè $\bar{a} \neq 0$ per ipotesi, dunque $n|x - y$, ossia $\bar{x} = \bar{y}$; abbiamo così provato che φ è iniettiva.

Dunque per il Principio della piccionaia φ è anche suriettiva. Allora l'immagine di $\mathbb{Z}_n$ in φ, $\varphi(\mathbb{Z}_n)$ è tutto $\mathbb{Z}_n$. Quindi per ogni elemento $\bar{z}$ di $\mathbb{Z}_n$ esiste un $\bar{y} \in \mathbb{Z}_n$ tale che $\bar{z} = \varphi(\bar{y}) = \bar{a}\bar{y}$. In particolare se si prende $\bar{1} \in \mathbb{Z}_n$ esiste un $\bar{y}$ tale che $\bar{1} = \bar{a}\bar{y}$: questo $\bar{y}$ è l'inverso di $\bar{a}$ in $\bar{Z}_n$.

$\square$

Dunque per ogni primo p, esiste il campo finito $\mathbb{Z}_p$ con p elementi.
Il principio della piccionaia è anche chiamato principio dei cassetti. Lo si può formulare dicendo che una piccionaia con n caselle può contenere al massimo n piccioni, se non se ne vogliono mettere due nella stessa casella. Oppure: se m piccioni sono distribuiti in n caselle con $m > n$, in qualche casella ci devono stare almeno due piccioni.

Esercizi 2.

1. Sia $n \in \mathbb{N}$ un naturale non primo. Sia $1 < x < n$. Dimostrare che $\bar{x} \in \mathbb{Z}_n$ è invertibile se e solo se x è primo con n, cioè il massimo comun divisore di x e n è uguale a 1. (Suggerimento: usare l'algoritmo euclideo della divisione, il massimo comun divisore di x e n può essere espresso nella forma $ax + bn$, con opportuni $a, b \in \mathbb{Z}$).

2. In $\mathbb{R}^2$ si definiscano le seguenti operazioni:

   somma : $(x, y) + (x', y') = (x + x', y + y')$;

   prodotto : $(x, y)(x', y') = (xx' - yy', xy' + yx')$.

   Verificare che $\mathbb{R}^2$ con tali operazioni è un campo.

   Questo è un modo per introdurre il campo dei numeri complessi $\mathbb{C}$.

Capitolo 2

Spazi vettoriali

D'ora in poi K denoterà un campo fissato. Per esempio K può essere $\mathbb{Q}, \mathbb{R}, \mathbb{C}$, o $\mathbb{Z}_p$ con p un numero primo.

Definizione 2.0.1 (Operazione esterna). Un'operazione esterna con operatori in K su un insieme V è un'applicazione $K \times V \to V$. Data una coppia (λ, v) con $\lambda \in K, v \in V$, il corrispondente è indicato semplicemente λv.

In questo capitolo definiremo gli spazi vettoriali sul campo K, si tratta di insiemi dotati di due operazioni, una somma interna e un'operazione esterna con operatori in K, detta prodotto, verificanti certe proprietà. Un'operazione esterna su un insieme V con operatori in K è un'applicazione $K \times V \to V$.

Ma prima vediamo alcuni esempi, che si possono interpretare come "prototipi" di spazio vettoriale.

2.1 Primi esempi di spazio vettoriale

Esempio 2.1.1.

a) $K^n = \underbrace{K \times \cdots \times K}_{n}$, il prodotto cartesiano di n copie di K.

   $K^n = \{x = (x_1, \ldots, x_n) \mid x_i \in K \ \forall i = 1, \ldots, n\}$. Per $n = 1$ si ritrova K. La somma interna e il prodotto esterno con operatori in K sono definiti membro a membro come segue.

$$\begin{aligned} + : K^n \times K^n &\to K^n \text{ tale che} \\ (x, y) &\to x + y = (x_1 + y_1, \ldots, x_n + y_n) \end{aligned}$$

$$\begin{aligned} \cdot : K \times K^n &\to K^n \text{ tale che} \\ (\lambda, x) &\to \lambda \cdot x = (\lambda x_1, \ldots, \lambda x_n) \end{aligned}$$

   Gli elementi di K sono detti scalari. Osserviamo che lo stesso simbolo $+$ si usa per denotare sia la somma in K sia la somma in K^n.

b) Fissiamo due numeri naturali m, n: $M(m \times n, K)$ denota l'insieme delle **matrici a m righe e n colonne** con elementi (o entrate) in K. Per dare una tale matrice bisogna dare un elemento di K^{mn}, ossia una mn-upla di elementi di K; questi vanno scritti suddividendoli in m righe (orizzontali) e n colonne (verticali) e numerati con un doppio indice, il primo indica la riga e varia da 1 a m e il secondo la colonna e varia da 1 a n:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

$M(m \times n, K)$ è in biiezione con K^{mn}, cambia solo la scrittura. Anche la somma di matrici e il prodotto esterno con operatori in K (come quelle in K^{mn}) sono definiti membro a membro. Date due matrici A, B e uno scalare $\lambda \in K$, si pone

$$A + B = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + bm1 & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{pmatrix},$$

$$\lambda A = \begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \dots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \dots & \lambda a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda a_{m1} & \lambda a_{m2} & \dots & \lambda a_{mn} \end{pmatrix}.$$

Per indicare la matrice A come sopra, si usa anche la notazione compatta $A = (a_{ij})_{\substack{1 \le i \le m \\ 1 \le j \le n}}$

c) Sia $\mathbb{C}$ il campo complesso; si definisce un'operazione "esterna" con operatori in $\mathbb{R}$ semplicemente restringendo il prodotto interno in $\mathbb{C}$:

$$\begin{aligned} \cdot : \mathbb{R} \times \mathbb{C} &\to \mathbb{C} \\ (\lambda, a + ib) &\to \lambda \cdot (a + ib) = \lambda a + i\lambda b \end{aligned}$$

d) Vettori geometrici del piano: vettori applicati e vettori liberi.

Un **vettore applicato** è un segmento orientato, ha punto iniziale e punto finale; può essere pensato come una coppia ordinata (A, B) di punti del piano, dove A è il punto iniziale, o di applicazione, e B il punto finale. Ha una lunghezza, se è stata fissata un'unità di misura, e se la lunghezza è $\neq 0$ ha anche direzione e verso.

Un **vettore libero, o vettore geometrico, o semplicemente vettore** è una classe d'equivalenza di vettori applicati per la relazione di equipollenza, secondo cui due vettori applicati sono equipollenti se hanno le stessa direzione, la stessa lunghezza e lo stesso verso, ossia stanno su rette parallele e muovendo la retta dell'uno parallelamente a se stessa è possibile sovrapporlo all'altro. Un vettore libero avente come rappresentante la coppia (A, B) si denota $\overrightarrow{AB}$ o $B - A$.

Ogni vettore ha uno e un solo rappresentante applicato in A, comunque si fissi il punto A. Questa osservazione permette di definire la **somma** di due vettori: se $v = \overrightarrow{AB}$ e $w = \overrightarrow{BC}$, si pone $v + w = \overrightarrow{AC}$. Se $v = \overrightarrow{AB}$ e $w = \overrightarrow{AB'}$, allora risulta $v + w = \overrightarrow{AD}$, dove D è il quarto vertice del parallelogramma di lati AB e AB'. La somma di vettori verifica la proprietà associativa e commutativa. Il vettore nullo è $\overrightarrow{AA}$. L'opposto di $\overrightarrow{AB}$ è $\overrightarrow{BA}$.

Si definisce il prodotto di un numero reale $\lambda \in \mathbb{R}$ per un vettore v, λv: ha la direzione di v, lunghezza pari a $|\lambda|$ per la lunghezza di v e verso concorde o discorde con v a seconda che $\lambda > 0$ o $\lambda < 0$.

L'insieme dei vettori liberi del piano con queste due operazoni è uno spazio vettoriale reale. In maniera analoga si possono definire i vettori dello spazio tridimensionale.

L'algebra lineare nasce da questo esempio e dallo studio dei sistemi lineari di equazioni, che si vedrà in seguito.

2.2 K-spazi vettoriali

Definizione 2.2.1 (K-spazio vettoriale)**.** Sia K un campo. Un insieme non vuoto V è uno spazio vettoriale su K, o K-spazio vettoriale, se in V sono date due operazioni:

- un'operazione interna detta somma,
- un'operazione esterna con operatori in K detta prodotto,

per cui valgono i seguenti assiomi:

(V1) V è un gruppo abeliano rispetto alla somma; lo zero è detto vettore nullo e indicato con 0_V o semplicemente 0; l'opposto di un elemento $v \in V$ è denotato $-v$;

(V2) le due operazioni sono legate dalle seguenti quattro proprietà:

1. $\forall\ \lambda, \mu \in K, v \in V$ si ha $(\lambda + \mu)v = \lambda v + \mu v$;
2. $\forall\ \lambda \in K, v, w \in V$ si ha $\lambda(v + w) = \lambda v + \lambda w$;
3. $\forall\ \lambda, \mu \in K, v \in V$ si ha $\lambda(\mu v) = (\lambda\mu)v$;
4. $\forall\ v \in V$ vale $1 \cdot v = v$, dove 1 è l'unità di K.

Gli elementi di V sono detti vettori, quelli di K scalari. I quattro esempi precedenti sono tutti spazi vettoriali. Nell'esempio a) il vettore nullo è $0 = (0, \ldots, 0)$, $-(x_1, \ldots, x_n) = (-x_1, \ldots, -x_n)$. L'esempio b) è simile: lo zero è la matrice nulla avente in tutte le posizioni lo 0 di K. Gli esempi c) e d) sono entrambi spazi vettoriali su $\mathbb{R}$.

Proposizione 2.2.2 (Proprietà degli spazi vettoriali)**.** *In ogni spazio vettoriale valgono le seguenti proprietà.*

(i) $0 \cdot v = 0\ \forall v \in V$*;*

(ii) $\lambda \cdot 0 = 0\ \forall \lambda \in K$*;*

(iii) Se $\lambda v = 0$, allora o $\lambda = 0$ o $v = 0$;

(iv) $(-1)v = -v$.

Dimostrazione. (i) $0 \cdot v = (0+0)v = 0v + 0v$, perciò $0v = 0$;

(ii) $\lambda \cdot 0 = \lambda(0+0) = \lambda \cdot 0 + \lambda \cdot 0$, perciò $\lambda \cdot 0 = 0$;

(iii) Sia $\lambda v = 0$; se $\lambda \neq 0$, esiste $\lambda^{-1} \in K$ tale che $\lambda\lambda^{-1} = \lambda^{-1}\lambda = 1$. Allora $v = 1 \cdot v = (\lambda^{-1}\lambda)v = \lambda^{-1}(\lambda v) = \lambda^{-1} \cdot 0 = 0$ per il punto (i).

(iv) $v + (-1)v = 1 \cdot v + (-1)v = (1-1)v = 0 \cdot v = 0$.

□

Vediamo ora altri esempi importanti, iniziamo con quello dei polinomi.

Esempio 2.2.3. L'insieme dei polinomi a coefficienti in un campo K nella indeterminata t è l'insieme denotato con $K[t]$ delle espressioni del tipo $a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n$, dove $a_0, \ldots, a_n$ sono elementi di K detti coefficienti del polinomio, e $n \geq 0$ è un numero intero. Dare un polinomio equivale a dare la successione dei coefficienti $a_0, \ldots, a_n$. Il grado di un polinomio è il massimo n tale che $a_n \neq 0$. Se tutti i coefficienti sono nulli si ha il polinomio nullo, il cui grado non è definito. Si noti che un polinomio non è una funzione. Osserviamo che $K[t]$ contiene K, come insieme dei polinomi di grado 0 più il polinomio nullo.

I polinomi si sommano e si moltiplicano per elementi di K in maniera naturale e costituiscono un K-spazio vettoriale.

Esempio 2.2.4. Siano K un campo e S un insieme arbitrario. Consideriamo l'insieme delle applicazioni di dominio S e codominio K:

$$\mathcal{F}(S, K) = \{f : S \to K\}.$$

In questo insieme introduciamo due operazioni definite punto per punto. Se $f, g \in \mathcal{F}(S, K)$ si definisce la loro somma $f + g$ come l'applicazione $S \to K$ che manda un elemento $s \in S$ in $(f+g)(s) := f(s) + g(s)$. Analogamente, si definisce il prodotto λf di uno scalare λ per f, ponendo $(\lambda f)(s) = \lambda f(s)$. Si hanno così in $\mathcal{F}(S, K)$ una somma interna e un prodotto esterno con operatori in K. Elemento neutro per la somma è l'applicazione nulla 0 tale che $0(s) = 0$ per ogni $s \in S$. Gli assiomi di spazio vettoriale si verificano facilmente sfruttando le proprietà di campo di K.

2.3 Sottospazi vettoriali

Sia V un K-spazio vettoriale. Sia $W \subset V$ un sottinsieme di V.

Definizione 2.3.1 (Sottospazio vettoriale). Si dice che W è un **sottospazio vettoriale** di V se

1. $W \neq \emptyset$;
2. se $w, w' \in W$ allora $w + w' \in W$: si dice che W è chiuso rispetto alla somma;

3. se $w \in W$ e $\lambda \in K$, allora $\lambda w \in W$: W è chiuso rispetto al prodotto esterno.

A volte si ometterà l'aggettivo vettoriale e si parlerà semplicemente di sottospazi di V. Attenzione però che in seguito introdurremo anche una seconda definizone, quella di sottospazio affine di uno spazio vettoriale.

Osservazione 1. Se W è un sottospazio vettoriale di V, allora il vettore nullo appartiene a W. Infatti: W non è vuoto, dunque prendiamo un vettore $w \in W$, ma allora $0 \cdot w = 0 \in W$. Analogamente anche $-w = (-1)w \in W$.

Osservazione 2. $W = \{0\}$, l'insieme costituito dal solo vettore nullo è un sottospazio, detto sottospazio nullo, in qualunque spazio vettoriale. Invece tutto lo spazio vettoriale V è sottospazio vettoriale di se stesso, detto sottospazio improprio.

Vediamo esempi di sottinsiemi dello spazio vettoriale $\mathbb{R}^2$ che sono sottospazi vettoriali e altri che non lo sono.

Esempio 2.3.2. Sia $V = \mathbb{R}^2$.

1. $W_1 = \{(x_1, x_2) \in \mathbb{R}^2 \mid 2x_1 + 3x_2 = 4\}$ non è sottospazio, ad esempio perchè non contiene il vettore nullo di $\mathbb{R}^2$, che è la coppia $(0, 0)$.

2. $W_2 = \{(x_1, x_2) \in \mathbb{R}^2 \mid 2x_1 + 3x_2 = 0\}$ è sottospazio: è una retta passante per l'origine. Analogamente risulta un sottospazio l'insieme delle soluzioni di una qualunque equazione del tipo $ax_1 + bx_2 = 0$, con $a, b \in \mathbb{R}$.

3. $W_3 = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1\}$ non è sottospazio, non è chiuso rispetto al prodotto esterno.

4. $W_4 = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 \geq 0, x_2 \geq 0\}$ non è sottospazio, non contiene gli opposti dei suoi elementi non nulli.

Vediamo ora un esempio nell'$\mathbb{R}$-spazio vettoriale $\mathbb{R}[t]$.

Esempio 2.3.3. Sia $\mathbb{R}[t]_d$ l'insieme dei polinomi di grado d, con $d \geq 1$ fissato: non è sottospazio vettoriale perchè non è chiuso rispetto alla somma. Per esempio, sia $d = 3$ e consideriamo $f(t) = 1 + 2t - t^2 + t^3$ e $g(t) = 5 + t - t^3$; $f(t) + g(t) = 6 + 3t + t^2$ ha grado 2. Invece $\mathbb{R}[t]_{\leq d}$, l'insieme dei poliomi di grado minore o uguale a d risulta sottospazio vettoriale.

Osserviamo che le operazioni di somma e di prodotto in V si possono restringere a un sottospazio W, perchè W è chiuso rispetto a somma e prodotto e contiene il vettore nullo. Rispetto a tali operazioni W risulta anch'esso essere un K-spazio vettoriale.

2.4 Intersezione di sottospazi vettoriali e sottospazio generato

Proposizione 2.4.1. *Ogni intersezione di sottospazi vettoriali di un K-spazio vettoriale V è un sottospazio vettoriale di V.*

Dimostrazione. Sia $\{W_i\}_{i\in I}$ una famiglia di sottospazi di V, indiciata su un insieme d'indici I. Sia W la loro intersezione: $W = \bigcap_{i\in I} W_i$. Chiaramente il vettore nullo 0 appartiene a W perchè appartiene a ogni sottospazio W_i. Siano $u, w \in W$: ciò significa che $u, w \in W_i$ per ogni $i \in I$; ma ogni W_i è sottospazio vettoriale di V dunque $u + w \in W_i$ per ogni $i \in I$. Quindi concludiamo che $u + w \in W$. Analogamente, se $u \in W$ e $\lambda \in K$, siccome $u \in W_i$ per ogni $i \in I$, si ha $\lambda u \in W_i$ per ogni i e quindi $\lambda u \in W$. □

Esempio 2.4.2. In $V = K[t]$ consideriamo la famiglia di sottospazi $W_i = K[t]_{\leq i}$, con I l'insieme degli interi ≥ 0. Si ha $\bigcap_{i\in I} W_i = K$. In questo caso i sottospazi considerati formano una catena $W_0 \subset W_1 \cdots \subset W_n \subset \cdots$ e perciò l'intersezione è il primo elemento della catena.

Osserviamo che un'unione di sottospazi vettoriali non è in generale un sottospazio. Per esempio se consideriamo due rette distinte passanti per l'origine in $\mathbb{R}^2$ come nell'esempio 2.3.2, la loro unione non è un sottospazio, in quanto non è chiusa rispetto alla somma.

Osserviamo che, se W è un sottospazio di V, $w_1, w_2 \in W$ e $\lambda, \mu \in K$, allora $\lambda w_1 + \mu w_2 \in W$. Infatti λw_1 e μw_2 appartengono a W per il punto 3. della definizione, e quindi anche $\lambda w_1 + \mu w_2 \in W$ per il punto 2. Un vettore della forma $\lambda w_1 + \mu w_2$ è detto *combinazione lineare* di w_1 e w_2. Vediamo ora che vale anche il viceversa.

Proposizione 2.4.3. *Sia $W \neq \emptyset$, $W \subset V$ un sottinsieme di uno spazio vettoriale V. Supponiamo che ogni combinazione lineare di due elementi di W appartenga ancora a W, allora W è un sottospazio vettoriale.*

Dimostrazione. Per ipotesi, per ogni scelta di vettori $w_1, w_2 \in W$ e scalari $\lambda, \mu \in K$ si ha $\lambda w_1 + \mu w_2 \in W$. Se si prende $\lambda = \mu = 1$ si ottiene che $w_1 + w_2 \in W$, quindi W è chiuso rispetto alla somma, mentre se si prende $\lambda = 0$ o $\mu = 0$ si ottiene la chiusura rispetto al prodotto esterno. □

Dato un qualunque sottinsieme S di uno spazio vettoriale V, si può considerare il più piccolo sottospazio contenente S, che è l'intersezione di tutti i sottospazi di V che contengono S. E` detto **sottospazio generato da** S e denotato $L(S)$ oppure $\langle S\rangle$. Per caratterizzare i suoi elementi avremo bisogno della nozione di combinazione lineare.

2.5 Combinazioni lineari

Definizione 2.5.1 (Combinazione lineare). Siano dati vettori $v_1, \dots, v_n$ di uno spazio vettoriale V. Una loro **combinazione lineare** è un vettore della forma $\lambda_1 v_1 + \cdots + \lambda_n v_n$, con $\lambda_1, \dots, \lambda_n$ elementi di K, detti coefficienti della combinazione lineare.

Esempio 2.5.2.

1. Se $\lambda_1 = \cdots = \lambda_n = 0$ si ottiene il vettore nullo $0 = 0v_1 + \cdots + 0v_n$: questa è detta **combinazione lineare banale**.

2. Se $\lambda_i = 1$ per un certo indice i e tutti gli altri coefficienti sono nulli, si ottiene il vettore v_i.

3. Se $\lambda_1 = \lambda_2 = 1$ e tutti gli altri coefficienti sono nulli, si ottiene $v_1 + v_2$.

4. Se $n = 1$, le combinazioni lineari dell'unico vettore v sono del tipo λv, al variare di λ in K: sono detti multipli di v o vettori proporzionali a v.

Proposizione 2.5.3. *Siano $v_1, \ldots, v_n \in V$. Il sottospazio generato da $S = \{v_1, \ldots, v_n\}$ è l'insieme di tutte le combinazioni lineari di $v_1, \ldots, v_n$. Lo si denota con il simbolo $\langle v_1, \ldots, v_n \rangle$ o $L(v_1, \ldots, v_n)$. E` anche detto chiusura lineare di S.*

Dimostrazione. Dimostriamo innanzitutto che l'insieme delle combinazioni lineari di $v_1, \ldots, v_n$ è un sottospazio vettoriale W di V. Certamente contiene il vettore nullo. La somma di due combinazioni lineari è una combinazione lineare: $(\lambda_1 v_1 + \cdots + \lambda_n v_n) + (\mu_1 v_1 + \cdots + \mu_n v_n) =$ usiamo la proprietà associativa e commutativa della somma e la 1. degli spazi vettoriali $= (\lambda_1 + \mu_1)v_1 + \cdots + (\lambda_n + \mu_n)v_n$, e dunque è anche questa una combinazione lineare di $v_1, \ldots . v_n$. Infine un multiplo di una combinazione lineare è una combinazione lineare: $\alpha(\lambda_1 v_1 + \cdots + \lambda_n v_n) = (\alpha\lambda_1)v_1 + \cdots + (\alpha\lambda_n)v_n$. Anche qui abbiamo usato gli assiomi di spazio vettoriale. Osserviamo poi che W contiene S, perchè ogni elemento v_i è combinazione lineare di $v_1, \ldots, v_n$. Infine, se U è un sottospazio che contiene $v_1, \ldots, v_n$, contiene anche ogni loro combinazione lineare, quindi contiene W. Perciò W è il più piccolo sottospazio che contiene $v_1, \ldots, v_n$. □

Esempio 2.5.4. Il sottospazio generato da un vettore v è $\langle v \rangle = \{\lambda v \mid \lambda \in K\}$, l'insieme dei multipli di v. Il sottospazio generato dal vettore nullo è il sottospazio nullo $\{0\}$.

Osservazione 3. Con dimostrazione simile alla precedente, si ottiene che, se S è un insieme infinito, il sottospazio $L(S)$ generato da S è l'insieme delle **combinazioni lineari finite** di elementi di S. In altre parole si considerano tutte le possibili combinazioni lineari di un numero finito di vettori appartenenti a S.

Il prossimo esempio è fondamentale.

Esempio 2.5.5. Sia $V = K^n$, il $K-$spazio vettoriale i cui elementi sono le n-uple ordinate di elementi di K. Introduciamo la notazione:

$$e_1 = (1, 0, \ldots, 0),\ e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, 0, \ldots, 0, 1).$$

Calcoliamo una generica combinazione lineare di $e_1, \ldots, e_n$:

$$\lambda_1 e_1 + \cdots + \lambda_n e_n = \lambda_1(1, 0, \ldots, 0) + \lambda_2(0, 1, 0 \ldots) + \cdots + \lambda_n(0, 0, \ldots, 0, 1) = (\lambda_1, \ldots, \lambda_n).$$

Di conseguenza $K^n = \langle e_1, \ldots, e_n \rangle$, perchè ogni vettore di K^n si può esprimere come una loro combinazione lineare. Si dice che $e_1, \ldots, e_n$ generano K^n o sono un sistema di generatori di K^n.

Esempio 2.5.6. Sia $V = K[t]$. Consideriamo l'insieme infinito

$$S = \{1, t, t^2, \ldots, t^n, \ldots\}.$$

Le combinazioni lineari finite di elementi di S sono tutti e soli i polinomi, quindi $K[t] = L(S)$. Le potenze di t sono un sistema di generatori di $K[t]$.

Esempio 2.5.7. Consideriamo in $\mathbb{R}^2$ i vettori $v_1 = (1, 4), v_2 = (2, -2)$. Una loro combinazione lineare è un vettore della forma

$$\lambda_1 v_1 + \lambda_2 v_2 = \lambda_1(1, 4) + \lambda_2(2, -2) = (\lambda_1 + 2\lambda_2, 4\lambda_1 - 2\lambda_2),$$

dove $\lambda_1, \lambda_2 \in \mathbb{R}$.

2.6 Dipendenza e indipendenza lineare

Definizione 2.6.1 (Vettori linearmente indipendenti). Siano $v_1, \dots, v_n \in V$. Sono detti **linearmente indipendenti** se ogni loro combinazione lineare nulla è banale. In altre parole: da $\lambda_1 v_1 + \ldots + \lambda_n v_n = 0$ segue che $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$. Una combinazione lineare nulla deve avere tutti i coefficienti nulli.

Altrimenti i vettori $v_1, \dots, v_n$ sono detti linearmente dipendenti: esistono $\lambda_1, \dots, \lambda_n \in K$, **non tutti nulli** tali che $\lambda_1 v_1 + \ldots + \lambda_n v_n = 0$.

Osserviamo che un singolo vettore v è linearmente dipendente se esiste $\lambda \neq 0$ tale che $\lambda v = 0$. Per la Proposizione 2.2.2 (iii) questo si può verificare solo se $v = 0$.

Proposizione 2.6.2. *Siano $v_1, \dots, v_n$ vettori di V: $v_1, \dots, v_n$ sono linearmente dipendenti se e solo se almeno uno di essi è combinazione lineare dei rimanenti.*

Dimostrazione. Supponiamo che $v_1, \dots, v_n$ siano linearmente dipendenti e che $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$ sia una loro combinazione lineare nulla non banale. Allora esiste almeno un coefficiente $\lambda_i \neq 0$. Quindi esiste il suo inverso in K: λ_i^{-1}. Possiamo scrivere $\lambda_i v_i = -\lambda_1 v_1 - \cdots - \lambda_{i-1} v_{i-1} - \lambda_{i+1} v_{i+1} - \cdots - \lambda_n v_n$. Moltiplichiamo ora a sinistra per λ_i^{-1} e otteniamo: $\lambda_i^{-1}(\lambda_i v_i) = v_i = -\lambda_1 \lambda_i^{-1} v_1 - \cdots - \lambda_{i-1} \lambda_i^{-1} v_{i-1} - \lambda_{i+1} \lambda_i^{-1} v_{i+1} - \cdots - \lambda_n \lambda_i^{-1} v_n$.

Viceversa se $v_1 = \alpha_2 v_2 + \cdots + \alpha_n v_n$, possiamo scrivere $1 \cdot v_1 - \alpha_2 v_2 - \cdots - \alpha_n v_n = 0$: siccome il coefficiente di v_1 è uguale a 1, è diverso da 0, abbiamo così ottenuto una combinazione lineare nulla ma non banale di $v_1, \dots, v_n$. Analogo ragionamento se al posto di v_1 abbiamo un qualunque altro vettore v_i. □

Corollario 2.6.3. *Due vettori v_1, v_2 sono linearmente dipendenti se e solo se uno è combinazione lineare, cioè multiplo, dell'altro. In tal caso i due vettori si dicono proporzionali.*

Esempio 2.6.4. Nello spazio vettoriale $\mathbb{Q}^3$ i vettori $v_1 = (1, 2, 3), v_2 = (2, 4, 5)$ sono linearmente indipendenti; mentre $w_1 = (0, 1, 0), w_2 = (0, -1, 0)$ sono linearmente dipendenti.

Osservazione 4.

1. Dato un vettore v, i vettori $\{v, 2v\}$, o $\{v, -v\}$, o $\{v, \lambda v\}$ qualunque sia λ, sono linearmente dipendenti.

2. Se sono dati vettori $v_1, \dots, v_n$, con $v_i = 0$ per qualche indice i, allora $v_1, \dots, v_n$ sono linearmente dipendenti. Infatti si ha la combinazione lineare nulla non banale $0v_1 + \cdots + 0v_{i-1} + 1v_i + 0v_{i+1} + \cdots + 0v_n = 0$.

3. Supponiamo che $v_1, \dots, v_m$ siano linearmente dipendenti. Se aggiungo altri vettori qualunque $v_{m+1}, \dots, v_n$, ottengo vettori $v_1, \dots, v_n$ ancora linearmente dipendenti. Infatti basta aggiungere a una combinazione lineare nulla non banale di $v_1, \dots, v_m$ la combinazione lineare di $v_{m+1}, \dots, v_n$ con coefficienti tutti 0.

Esempio 2.6.5. Consideriamo in $\mathbb{R}^3$ i tre vettori $v_1 = (1, 2, 3)$, $v_2 = (1, -1, 0)$, $v_3 = (0, 1, 4)$. Sono linearmente dipendenti o indipendenti? Consideriamo una loro combinazione lineare nulla $x_1 v_1 + x_2 v_2 + x_3 v_3 = 0$ e analizziamo se può essere ottenuta con coefficienti non tutti nulli o meno.

$$x_1(1, 2, 3) + x_2(1, -1, 0) + x_3(0, 1, 4) = (x_1 + x_2, 2x_1 - x_2 + x_3, 3x_1 + 4x_3) = (0, 0, 0)$$

se e solo se $(x_1, x_2, x_3))$ è una soluzione del sistema di equazioni

$$\begin{cases} x_1 + x_2 & = 0 \\ 2x_1 - x_2 + x_3 & = 0 \\ 3x_1 \quad\quad + 4x_3 & = 0. \end{cases}$$

Si tratta di un sistema lineare omogeneo di 3 equazioni nelle 3 incognite x_1, x_2, x_3. In questo caso per risolverlo si può procedere esprimendo $x_2 = -x_1$ (dalla prima equazione), e $x_3 = -3/4x_1$ (dalla terza equazione), e poi sostituire nella seconda. Si ottiene $9/4x_1 = 0$ e quindi $x_1 = x_2 = x_3 = 0$. Si conclude che i tre vettori sono linearmente indipendenti.

Se sono dati 6 vettori in $\mathbb{R}^{13}$, per capire se sono linearmente indipendenti si scrive un sistema lineare omogeneo di 13 equazioni in 6 incognite. Un capitolo successivo sarà interamente dedicato alla teoria dei sistemi lineari di equazioni.

Proposizione 2.6.6. *Se $v_1, \ldots, v_n$ sono vettori linearmente indipendenti, ogni vettore $v \in L(v_1, \ldots, v_n)$ si esprime* **in maniera unica** *come loro combinazione lineare.*

Dimostrazione. La relazione $v = \lambda_1 v_1 + \ldots + \lambda_n v_n = \mu_1 v_1 + \ldots + \mu_n v_n$, si può riscrivere $\lambda_1 v_1 + \ldots + \lambda_n v_n - \mu_1 v_1 - \ldots - \mu_n v_n = 0$, o anche $(\lambda_1 - \mu_1)v_1 + \ldots + (\lambda_n - \mu_n)v_n = 0$: questa è una combinazione lineare nulla di $v_1, \ldots, v_n$, che sono linearmente indipendenti, perciò si ha $\lambda_1 - \mu_1 = \cdots = \lambda_n - \mu_n = 0$. □

L'affermazione della proposizione precedente si può anche rovesciare, ossia se ogni vettore di $L(v_1, \ldots, v_n)$ ha un'unica espressione come combinazione lineare di $v_1, \ldots, v_n$, questi sono linearmente indipendenti. Infatti se $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0 = 0v_1 + \cdots + 0v_n$, per l'unicità si deve avere $\lambda_1 = \cdots = \lambda_n = 0$.

Concludiamo questo capitolo estendendo la definizione di lineare indipendenza a famiglie qualunque, non necessariamente finite, di vettori.

Definizione 2.6.7 (Famiglia libera). Una famiglia di vettori $\{v_i\}_{i \in I}$ è detta **libera** o **linearmente indipendente**, se lo è ogni sua sottofamiglia finita. Ciò significa che non esiste una combinazione lineare nulla non banale di alcuna sottofamiglia finita di vettori presi fra i v_i.

Per esempio, in $K[t]$ le potenze di t costituiscono una famiglia libera, per definizione di polinomio.

Esercizi 3.

1. Dimostrare l'affermazione che l'unione di due rette distinte per l'origine in $\mathbb{R}^2$ non è un sottospazio vettoriale, in quanto non è chiusa rispetto alla somma.

2. Dimostrare che, se W, W' sono sottospazi vettoriali di V e $W \cup W'$ è anch'esso sottospazio vettoriale, allora o $W \subset W'$ o $W' \subset W$.

Capitolo 3

Basi

3.1 Sistemi di generatori e basi

Sia V un K-spazio vettoriale fissato.

Definizione 3.1.1 (Sistema di generatori)**.** Una famiglia $S = \{v_i\}_{i \in I}$ di elementi di V è un **sistema di generatori** di V se $V = L(S)$, cioè V coincide con il sottospazio generato da S, il che significa che ogni elemento di V può essere espresso come combinazione lineare di un numero finito di elementi v_i (vedere l'Osservazione 3).

Definizione 3.1.2 (Spazio vettoriale finitamente generato)**.** V è detto **finitamente generato** se ammette un sistema finito di generatori $v_1, \dots, v_n$.

Definizione 3.1.3 (Base)**.** Una famiglia $\{v_i\}_{i \in I}$ di elementi di V è una **base** di V se è un sistema di generatori linearmente indipendenti.

Teorema 3.1.4. *Se $\mathcal{B} = (v_1, \dots, v_n)$ è una base finita di V, ogni vettore $v \in V$ si esprime come combinazione lineare di $v_1, \dots, v_n$ in maniera unica.*

Dimostrazione. Segue dalla Proposizione 2.6.6. □

I coefficienti $x_1, \dots, x_n$ dell'unica combinazione lineare $v = x_1v_1 + \cdots + x_nv_n$ sono detti **coordinate di v rispetto alla base $\mathcal{B}$.**

Esempio 3.1.5.

1. Nello spazio euclideo $\mathbb{R}^3$ in cui è fissato un sistema di riferimento cartesiano i tre versori comunemente indicati $\vec{i}, \vec{j}, \vec{k}$ formano una base.

2. In K^n consideriamo gli n vettori $e_1, \dots, e_n$ introdotti nell'Esempio 2.5.5. Abbiamo verificato che ogni vettore di K^n è una loro combinazione lineare: $(x_1, \dots, x_n) = x_1e_1 + \cdots + x_ne_n$. D'altra parte sono linearmente indipendenti perchè se $x_1e_1 + \cdots + x_ne_n = 0 = (0, \dots, 0)$, si deve chiaramente avere $x_1 = 0, \dots, x_n = 0$. Dunque $(e_1, \dots, e_n)$ formano una base $\mathcal{C}$, detta **base canonica o base standard** di K^n. Si parla di base canonica soltanto in K^n, non in altri spazi vettoriali. Le coordinate del vettore $(x_1, \dots, x_n)$ rispetto alla base canonica sono proprio $x_1, \dots, x_n$.

3. In $M(m \times n, K)$, spazio vettoriale delle matrici $m \times n$ a coefficienti in K, consideriamo le mn matrici E_{ij} con $i = 1, \dots, m$, $j = 1, \dots, n$, così definite: E_{ij} ha tutti gli elementi nulli,

tranne quello di indici ij che è uguale a 1. Si ha che una matrice $A = (a_{ij})_{i=1,\dots,m,\ j=1,\dots,n}$ si può scrivere $A = a_{11}E_{11} + a_{12}E_{12} + \cdots + a_{mn}E_{mn}$, e tale combinazione lineare è unica. Dunque le matrici E_{ij} formano una base di $M(m \times n, K)$. Le coordinate di A rispetto a questa base sono proprio gli elementi di A.

4. Gli elementi $1, i$ formano una base di $\mathbb{C}$ come $\mathbb{R}$-spazio vettoriale.

5. Le potenze di t $(1, t, t^2, \dots, t^n, \dots)$ formano una base infinita di $K[t]$.

6. In $\mathbb{R}^2$ i vettori $v_1 = (2, 1)$ e $v_2 = (3, 4)$ formano una base. Infatti verifichiamo intanto che generano $\mathbb{R}^2$: consideriamo un qualunque vettore $a = (a_1, a_2) \in \mathbb{R}^2$, e cerchiamo se esistono x_1, x_2 tali che $a = x_1v_1 + x_2v_2 = x_1(2, 1) + x_2(3, 4) = (2x_1 + 3x_2, x_1 + 4x_2)$. Deve valere:

$$\begin{cases} 2x_1 + 3x_2 & = a_1 \\ x_1 + 4x_2 & = a_2. \end{cases} \tag{3.1}$$

Anche questo sistema lineare di due equazioni nelle due incognite x_1, x_2 si può risolvere facilmente per sostituzione, e si trova che ha una e una sola soluzione per ogni scelta di a, e precisamente:

$$\begin{cases} x_1 & = \frac{4}{5}a_1 - \frac{3}{5}a_2 \\ x_2 & = -\frac{1}{5}a_1 + \frac{2}{5}a_2. \end{cases} \tag{3.2}$$

Quindi $\mathbb{R}^2 =< v_1, v_2 >$. Inoltre v_1, v_2 sono linearmente indipendenti perchè non sono proporzionali.

3.2 Prolungamento a una base

Teorema 3.2.1 (Lemma dello scambio)**.** *Sia V un K-spazio vettoriale. Supponiamo che $\mathcal{B} = (v_1, \dots, v_n)$ sia una sua base. Sia w un vettore non nullo di V e $w = \lambda_1v_1 + \cdots + \lambda_nv_n$ la sua espressione come combinazione lineare dei vettori di $\mathcal{B}$. Se $\lambda_k \neq 0$ per un certo indice k, $1 \leq k \leq n$, allora anche $v_1, \dots, v_{k-1}, w, v_{k+1}, \dots, v_n$ formano una base di V.*

Dimostrazione. Eventualmente riordinando i vettori, possiamo supporre che $k = 1$, ossia $\lambda_1 \neq 0$, quindi esiste λ_1^{-1}. Dobbiamo dimostrare che $w, v_2, \dots, v_n$:

1. generano V;
2. sono linearmente indipendenti.

1. Sia $v \in V$ un qualunque vettore di V: poichè $\mathcal{B}$ è una base, si ha una relazione $v = \mu_1v_1 + \cdots + \mu_nv_n$, con opportuni coefficienti $\mu_1, \dots, \mu_n \in K$. Vogliamo rimpiazzare v_1 con w. Ma dall'espressione di w come combinazione lineare di $v_1, \dots, v_n$, possiamo ricavare $v_1 = \lambda_1^{-1}w - \lambda_1^{-1}\lambda_2v_2 - \cdots - \lambda_1^{-1}\lambda_nv_n$. Allora, sostituendo, otteniamo

$$\begin{aligned} v &= \mu_1(\lambda_1^{-1}w - \lambda_1^{-1}\lambda_2v_2 - \cdots - \lambda_1^{-1}\lambda_nv_n) + \mu_2v_2 + \cdots + \mu_nv_n = \\ &= \lambda_1^{-1}\mu_1w + (\mu_2 - \lambda_1^{-1}\mu_1\lambda_2)v_2 + \cdots + (\mu_n - \lambda_1^{-1}\mu_1\lambda_n)v_n. \end{aligned}$$

Quindi v è combinazione lineare di $w, v_2, \dots, v_n$, e concludiamo che questi generano V.

2. Consideriamo una combinazione lineare nulla di $w, v_2, \dots, v_n$: $\alpha_1w + \alpha_2v_2 + \cdots + \alpha_nv_n = 0$. Sostituiamo l'espressione di w e otteniamo:

$$\alpha_1(\lambda_1v_1 + \cdots + \lambda_nv_n) + \alpha_2v_2 + \cdots + \alpha_nv_n = 0$$

e quindi

$$\alpha_1\lambda_1 v_1 + (\alpha_1\lambda_2 + \alpha_2)v_2 + \cdots + (\alpha_1\lambda_n + \alpha_n)v_n = 0.$$

Ma questa è una combinazione lineare nulla dei vettori di $\mathcal{B}$, che sono linearmente indipendenti, perciò i coefficienti sono tutti nulli:

$$\begin{cases} \alpha_1\lambda_1 & = 0 \\ \alpha_1\lambda_2 + \alpha_2 & = 0 \\ \dots & \dots \\ \alpha_1\lambda_n + \alpha_n & = 0 \end{cases}$$

Dalla prima segue $\alpha_1 = 0$ perchè $\lambda_1 \neq 0$, e quindi, sostituendo nelle altre, si ha $\alpha_2 = \cdots = \alpha_n = 0$. □

Il Teorema 3.2.1 precedente, che è interessante in sè, sarà ora usato per dimostrare il seguente teorema, molto importante, ed è perciò noto come "lemma".

Teorema 3.2.2 (del completamento o prolungamento a una base)**.** *Sia V un K-spazio vettoriale. Supponiamo che $\mathcal{B} = (v_1, \dots, v_n)$ sia una sua base. Siano $w_1, \dots, w_r$, $r \geq 0$, vettori di V linearmente indipendenti. Allora:*

1. $r \leq n$;

2. esistono $n - r$ elementi di $\mathcal{B}$ $v_{i_1}, \dots, v_{i_{n-r}}$ tali che $w_1, \dots, w_r, v_{i_1}, \dots, v_{i_{n-r}}$ sia una base di V. In altre parole è possibile completare, o prolungare, $w_1, \dots, w_r$ a una base di V, aggiungendo opportuni elementi di una base data.

Dimostrazione. Per induzione su r. Il caso $r = 0$ è chiaramente vero.

Supponiamo che valgano 1. e 2. per le famiglie di $r-1$ vettori linearmente indipendenti, e le dimostriamo per r. Siano dunque $w_1, \dots, w_r$ linearmente indipendenti, allora lo sono anche $w_1, \dots, w_{r-1}$. Per ipotesi induttiva, si ha allora che $r - 1 \leq n$ ed è possibile trovare in $\mathcal{B}$ $n - r + 1$ vettori in modo che, con $w_1, \dots, w_{r-1}$, formino una base $\mathcal{B}'$. Eventualmente rinumerando, possiamo supporre che tale base sia $(w_1, \dots, w_{r-1}, v_{r+1}, \dots, v_n)$.

Dimostriamo 1.: si ha $r \leq n + 1$; se fosse $r = n + 1$, allora $n - r + 1 = 0$, e quindi già $(w_1, \dots, w_{r-1})$ sarebbe una base di V. Perciò si avrebbe che $w_r \in< w_1, \dots, w_{r-1} >$, ma questo è assurdo perchè $w_1, \dots, w_r$ sono linearmente indipendenti (per Prop. 2.6.2). Quindi $r \leq n$.

Dimostriamo 2.: osserviamo che $w_r \neq 0$, altrimenti $w_1, \dots, w_r$ sarebbero linearmente dipendenti; w_r è combinazione lineare dei vettori della base $\mathcal{B}'$: $w_r = \lambda_1 w_1 + \cdots + \lambda_{r-1}w_{r-1} + \lambda_r v_r + \cdots + \lambda_n v_n$. Osserviamo che $\lambda_r, \dots, \lambda_n$ non possono essere tutti nulli, supponiamo che sia $\lambda_r \neq 0$; allora per il Lemma dello scambio (Teorema 3.2.1), posso sostituire in $\mathcal{B}'$ w_r al posto di v_r, e allora $(w_1, \dots, w_r, v_{r+1}, \dots, v_n)$ risulta una base di V. Analogamente se è diverso da zero un altro coefficiente fra $\lambda_r, \dots, \lambda_n$. □

Corollario 3.2.3 (Conseguenze del Teorema di prolungamento a una base).

1. Se V ha una base finita, ogni sua base è finita.

2. Due basi finite di V hanno lo stesso numero di elementi.

Dimostrazione. Per la 1. supponiamo che $\mathcal{B} = (v_1, \dots, v_n)$ sia una base di V formata da n vettori, e che $\mathcal{B}' = \{w_i\}_{i \in I}$ sia un'altra base; se l'insieme d'indici I fosse infinito, potrei prendere in $\mathcal{B}'$ $n+1$ vettori $w_{i_1}, \ldots, w_{i_{n+1}}$: sono linearmente indipendenti perchè fanno parte di una base, ma per il Teorema 3.2.2 dev'essere $n+1 \leq n$: assurdo.

Per la 2., se $(v_1, \ldots, v_n)$, $(w_1, \ldots, w_m)$ sono le due basi, interpretando i primi come vettori linearmente indipendenti e i secondi come una base, per il Teorema 3.2.2 dev'essere $n \leq m$; invertendo i ruoli delle due basi si ottiene $m \leq n$. Quindi $m = n$. □

3.3 Dimensione

Definizione 3.3.1. Sia V un K-spazio vettoriale. La **dimensione di** V è uguale a n se V ha una base di n elementi. In tal caso ogni base di V è finita e composta da n vettori. Se invece V non ha basi finite, si dice che V ha dimensione infinita.

La dimensione di V si denota $\dim V$, o $\dim_K V$ se si vuole sottolineare qual è il campo base. Nel caso di dimensione infinita si scrive $\dim V = \infty$.

Esempio 3.3.2.

1. Lo spazio vettoriale nullo ha dimensione 0.
2. $\dim K^n = n$, perchè la base canonica $\mathcal{C} = (e_1, \ldots, e_n)$ è composta da n vettori. In particolare $\dim_K K = 1$. Com'è fatta la base canonica?
3. $\dim M(m \times n, K) = mn$.
4. $\dim_{\mathbb{R}} \mathbb{C} = 2$ (ma $\dim_{\mathbb{C}} \mathbb{C} = 1$).
5. $\dim_K K[t] = \infty$.

Proposizione 3.3.3 (Estrarre una base)**.** *Sia V uno spazio vettoriale non nullo, sia $v_1, \ldots, v_n$ un suo sistema di generatori finito: i vettori $v_1, \ldots, v_n$ contengono una base.*

Dimostrazione. Induzione su n. Se $n = 1$, è vero, perchè l'unico generatore dev'essere non nullo, e quindi linearmente indipendente. Supponiamo vero l'asserto per $n-1$ e lo dimostriamo per n. Per ipotesi $v_1, \ldots, v_n$ è un sistema di generatori: se sono linearmente indipendenti abbiamo finito; se non lo sono, uno è combinazione lineare dei rimanenti, supponiamo sia v_n. Allora anche $v_1, \ldots, v_{n-1}$ è un sistema di generatori (perchè?), e si conclude sfruttando l'ipotesi induttiva. □

Corollario 3.3.4. *Ogni spazio vettoriale finitamente generato ha dimensione finita.*

Teorema 3.3.5. *Sia V uno spazio vettoriale di dimensione finita n.*

1. *Se $v_1, \ldots, v_n$ sono n vettori linearmente indipendenti, formano una base di V;*
2. *se $v_1, \ldots, v_n$ generano V, sono una base di V.*

Dimostrazione. Per la 1. usare 3.2.2 e 3.2.3, e per la 2. usare 3.3.3 e 3.2.3. □

Il precedente teorema è molto utile perchè permette di concludere che un certo insieme è una base, verificando solo che si tratta di vettori linearmente indipendenti, oppure solo di un sistema di generatori. Naturalmente bisogna sapere la dimensione dello spazio vettoriale in cui si lavora.

Usando l'Assioma della scelta, o il Lemma di Zorn, si può dimostrare che ogni spazio vettoriale ha una base, eventualmente infinita.

Esercizi 4.

1. Prolungare il vettore $(2, 0, 1)$ a una base di $\mathbb{R}^3$.

2. Determinare la dimensione di $K[t]_{\leq i}$. Dopo aver verificato che 1 e $t - 1$ sono linearmente indipendenti, prolungarli a una base di $K[t]_{\leq 3}$.

Capitolo 4

Somma di sottospazi vettoriali

4.1 Dimensione di un sottospazio

Confrontiamo la dimensione di uno spazio vettoriale V di dimensione finita con quella di un suo sottospazio W.

Proposizione 4.1.1. *Se* $\dim V$ *è finita, anche* W *ha dimensione finita e* $\dim W \leq \dim V$. *Inoltre, se* $\dim W = \dim V$ *allora* $W = V$.

Dimostrazione. Se W non avesse dimensione finita, ci sarebbe in W una famiglia infinita di vettori linearmente indipendenti, ma questi sono anche vettori linearmente indipendenti di V, il che è assurdo. Quindi $\dim W$ è finita. Prendiamo una base di W $(w_1, \ldots, w_n)$, questa è costituita da elementi linearmente indipendenti che appartengono a $W \subseteq V$, perciò per il Teorema 3.2.2 $\dim W \leq \dim V$. Se $\dim W = \dim V = n$ e $(w_1, \ldots, w_n)$ è una base di W, questi sono n elementi linearmente indipendenti di V, quindi per il Teorema 3.3.5 formano una sua base, in particolare generano V, qundi $W = V$. □

4.2 Somma di sottospazi e relazione di Grassmann

Consideriamo U, W sottospazi vettoriali di un K-spazio vettoriale V.

Definizione 4.2.1. Il sottospazio somma di U e W, denotato $U + W$ è

$$U + W = \{u + w \in V \mid u \in U, w \in W\}.$$

E` immediato verificare che $U + W$ è un sottospazio vettoriale di V (contiene il vettore nullo ed è chiuso rispetto a somma e prodotto esterno), ed è precisamente il sottospazio generato da $U \cup W$. Infatti si ha che $U \subseteq U + W$ e $W \subseteq U + W$, e se V' è un sottospazio vettoriale di V che contiene sia U sia W, poichè V' è chiuso rispetto alla somma, V' contiene anche $U + W$.

Il seguente importante teorema mette in relazione le dimensioni dei sottospazi $U, W, U + W, U \cap W$ nel caso di dimensione finita.

Teorema 4.2.2 (Relazione di Grassmann). *Sia* V *un* K*-spazio vettoriale di dimensione finita,* $U, W \subseteq V$ *suoi sottospazi. Vale la relazione:*

$$\dim U + \dim W = \dim(U + W) + \dim(U \cap W).$$

Dimostrazione. Iniziamo fissando una base del sottospazio più piccolo $U \cap W$, e prolungandola poi a basi di U e di W. Sia dunque $(v_1, \dots, v_r)$ una base di $U \cap W$, $\dim(U \cap W) = r$. Poichè $v_1, \dots, v_r$ sono vettori linearmente indipendenti che appartengono a U, possiamo prolungarli a una base di U: $\mathcal{B} = (v_1, \dots, v_r, u_{r+1}, \dots, u_n)$; analogamente possiamo prolungarli a una base di W: $\mathcal{B}' = (v_1, \dots, v_r, w_{r+1}, \dots, w_m)$, dove $n = \dim U$, $m = \dim W$.

Complessivamente abbiamo ora $r + (n - r) + (m - r) = n + m - r$ vettori di V: $v_1, \dots, v_r, u_{r+1}, \dots, u_n, w_{r+1}, \dots, w_m$. Osserviamo che tali vettori appartengono tutti a $U + W$; avremo finito se riusciremo a dimostrare che formano una base di $U + W$.

a) Dimostriamo che $v_1, \dots, v_r, u_{r+1}, \dots, u_n, w_{r+1}, \dots, w_m$ generano $U + W$. Se $u + w \in U+W$, con $u \in U$, $w \in W$, allora usando le basi $\mathcal{B}, \mathcal{B}'$ u, w si possono scrivere rispettivamente nella forma:

$$u = \lambda_1 v_1 + \dots + \lambda_r v_r + \lambda_{r+1} u_{r+1} + \dots + \lambda_n u_n,$$

$$w = \mu_1 v_1 + \dots + \mu_r v_r + \mu_{r+1} w_{r+1} + \dots + \mu_m w_m.$$

Dunque sommando e facendo i conti si ottiene:

$$u + w = (\lambda_1 + \mu_1) v_1 + \dots + (\lambda_r + \mu_r) v_r + \lambda_{r+1} u_{r+1} + \dots + \lambda_n u_n + \mu_{r+1} w_{r+1} + \dots + \mu_m w_m.$$

b) Dimostriamo che $v_1, \dots, v_r, u_{r+1}, \dots, u_n, w_{r+1}, \dots, w_m$ sono linearmente indipendenti.

Consideriamo una loro combinazione lineare nulla:

$$\alpha_1 v_1 + \dots + \alpha_r v_r + \beta_{r+1} u_{r+1} + \dots + \beta_n u_n + \gamma_{r+1} w_{r+1} + \dots + \gamma_m w_m = 0. \tag{4.1}$$

Osserviamo che il vettore $v := \alpha_1 v_1 + \dots + \alpha_r v_r + \beta_{r+1} u_{r+1} + \dots + \beta_n u_n$ appartiene a U in quanto combinazione lineare dei vettori di $\mathcal{B}$; ma si ha anche $v = -\gamma_{r+1} w_{r+1} - \dots - \gamma_m w_m$, e quindi $v \in W$ in quanto combinazione lineare dei vettori di $\mathcal{B}'$ (con alcuni coefficienti nulli). Perciò $v \in U \cap W$, e quindi si può scrivere come combinazione lineare dei vettori della sua base inizialmente fissata: $v = \delta_1 v_1 + \dots + \delta_r v_r$. Allora si ha:

$$\delta_1 v_1 + \dots + \delta_r v_r = \alpha_1 v_1 + \dots + \alpha_r v_r + \beta_{r+1} u_{r+1} + \dots + \beta_n u_n$$

e quindi

$$\delta_1 v_1 + \dots + \delta_r v_r - \alpha_1 v_1 - \dots - \alpha_r v_r - \beta_{r+1} u_{r+1} - \dots - \beta_n u_n = 0$$

ossia

$$(\delta_1 - \alpha_1) v_1 + \dots + (\delta_r - \alpha_r) v_r - \beta_{r+1} u_{r+1} - \dots - \beta_n u_n = 0;$$

essendo $\mathcal{B}$ una base se ne deduce $\delta_1 - \alpha_1 = \dots = \beta_{r+1} = \dots = \beta_n = 0$. Allora la relazione (4.1) diventa $\alpha_1 v_1 + \dots + \alpha_r v_r + \gamma_{r+1} w_{r+1} + \dots + \gamma_m w_m = 0.$, Siccome $\mathcal{B}'$ è una base si conclude che $\alpha_1 = \dots = \gamma_1 = \dots = \gamma_m = 0$.

□

Esempio 4.2.3.

In $\mathbb{R}^3$ e in $\mathbb{R}^4$ per la posizione reciproca di due sottospazi vettoriali U, W di dimensione 2 (due piani) si possono presentare i seguenti casi:

n	$\dim(U \cap W)$	$\dim(U + W)$	posizione reciproca di U e W
3 o 4	2	2	$U = W$
3 o 4	1	3	incidenti lungo una retta
4	0	4	incidenti in un punto

4.3 Somma diretta

Definizione 4.3.1. La somma di due sottospazi U, W si dice **diretta** se $U \cap W = (0)$.

Se la somma di U e W è diretta la si denota $U \oplus W$; in questo caso per la relazione di Grassmann si ha $\dim U + \dim W = \dim(U + W)$.

In particolare V è somma diretta di due suoi sottospazi U, W se $V = U + W$ e $U \cap W = (0)$. Si scrive $V = U \oplus W$. In tal caso si dice che U e W sono sottospazi **supplementari**. Allora ogni vettore $v \in V$ si può esprimere **in maniera unica** nella forma $v = u + w$, con $u \in U$, $w \in W$. Infatti se $v = u+w = u'+w'$ con $u, u' \in U$, $w, w' \in W$, si ha $u-u' = w'-w$: tale vettore appartiene sia a U sia a W, ossia a $U \cap W = (0)$, e quindi $u = u'$, $w = w'$, da cui l'unicità.

In particolare il vettore nullo si può scrivere solo nella forma $0 = 0+0$, cioè da $u+w = 0$, con $u \in U$, $w \in W$, segue $u = w = 0$.

La proprietà precedente si può anche invertire. Infatti:

Teorema 4.3.2. *Sia V un K-spazio vettoriale, U, W due suoi sottospazi. Allora $V = U \oplus W$ se e solo se ogni vettore v di V si scrive* **in maniera unica** *come $v = u + w$ con $u \in U$, $w \in W$.*

Dimostrazione. Un'implicazione è già stata dimostrata. Viceversa, supponiamo che ogni vettore v di V si scriva in maniera unica come $v = u + w$ con $u \in U$, $w \in W$. Chiaramente $V = U + W$; rimane da dimostrare che $U \cap W = (0)$: se ci fosse un vettore non nullo $u \in U \cap W$, allora si potrebbe scrivere $u = u + 0 = 0 + u$: questo contraddice l'unicità ipotizzata. □

Proposizione 4.3.3. *Se $V = U \oplus W$ è uno spazio vettoriale di dimensione finita, unendo basi di U e di W si ottiene una base di V.*

Dimostrazione. Siano $(u_1, \dots, u_r)$ una base di U, e $(w_1, \dots, w_s)$ una base di W. Poichè $U \cap W = (0)$, le due basi sono disgiunte quindi la loro unione è composta da $r + s$ elementi; inoltre $\dim V = r + s$. Ogni elemento v di V si scrive come $v = u + w$, con $u \in U$, $w \in W$; u è combinazione lineare di $u_1, \dots, u_r$ e w è combinazione lineare di $w_1, \dots, w_s$. Segue che $u_1, \dots, u_r, w_1, \dots, w_s$ generano V; dal momento che sono in numero pari alla dimensione di V formano una sua base. □

La definizione di somma di sottospazi si estende per induzione a una famiglia finita qualunque di sottospazi $\{W_i\}_{i=1,\dots,r}$ con $r \geq 2$. In tal caso si può usare una delle seguenti notazioni: $W_1 + \cdots + W_r = \sum_{i=1}^{r} W_i$. Gli elementi sono le somme di r addendi presi uno in ciascuno dei sottospazi $W_1, \ldots, W_r$. E` il sottospazio generato dall'unione $\bigcup_{i=1}^{r} W_i$. Tale somma si dice diretta, e si scrive $W_1 \oplus \cdots \oplus W_r$, oppure $\bigoplus_{i=1}^{r} W_i$, se ogni suo elemento ha un'unica espressione come somma di r elementi $w_1 \in W_1, \ldots, w_r \in W_r$. La corrispondente condizione sull'intersezione risulta essere la seguente:

$$W_i \cap (\sum_{j \neq i} W_j) = (0)$$

per ogni $i = 1, \ldots, r$. (Dimostrazione per esercizio per induzione su r.)

Esercizi 5.

1. Se $(v_1, \ldots, v_n)$ è una base di V, allora $V = < v_1 > \oplus \cdots \oplus < v_n >$.
2. Se V è uno spazio vettoriale di dimensione finita e $W \subset V$ è un sottospazio vettoriale, esiste un supplementare di W, cioè un altro sottospazio U tale che $V = U \oplus W$.

Capitolo 5

Matrici e sistemi lineari di equazioni

5.1 Spazi delle righe e delle colonne

Sia

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

una matrice $m \times n$ a coefficienti in K. Useremo la seguente notazione per le righe di A e le interpreteremo come elementi di K^n:

$a_1 = (a_{11}, \dots, a_{1n})$

$\dots$

$a_m = (a_{m1}, \dots, a_{mn})$;

e similmente per le colonne di A, che interpreteremo come elementi di K^m:

$$a^1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \cdots, a^n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

Diremo che una matrice A è quadrata se $m = n$.

Definizione 5.1.1 (Spazio delle righe e delle colonne). Spazio delle righe di A è il sottospazio vettoriale di K^n generato dalle righe di A: $< a_1, \dots, a_m >$. La sua dimensione è detta **rango per righe** di A. Spazio delle colonne di A è il sottospazio vettoriale di K^m generato dalle colonne di A: $< a^1, \dots, a^n >$. La sua dimensione è detta **rango per colonne** di A.

Chiaramente si ha che entrambi i ranghi sono minori o uguali sia di m sia di n (perchè?). Anticipiamo un teorema che sarà dimostrato successivamente.

Teorema 5.1.2. *Sia A una matrice a coefficienti in un campo K. Il suo rango per righe coincide con il suo rango per colonne.*

Esempio 5.1.3.

1. **Matrice identica o matrice identità**: per ogni $n \geq 1$ è la matrice quadrata $n \times n$

$$E_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} = (\delta_{ij})_{i,j=1,\dots,n},$$

dove i simboli δ_{ij} detti **simboli di Kronecker** valgono 1 se $i = j$ e 0 se $i \neq j$. Sia le righe sia le colonne di E_n sono i vettori $e_1, \dots, e_n$ della base canonica di K^n, dunque il rango per righe e il rango per colonne di E_n valgono n.

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

i ranghi per righe e per colonne sono uguali a 2.

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

i ranghi per righe e per colonne di B sono uguali a 1.

3.
$$A = \begin{pmatrix} 0 & 1 & 3 & 4 \\ 2 & \frac{1}{2} & 1 & -1 \\ 4 & 0 & -1 & -6 \end{pmatrix}:$$

i ranghi sono uguali a 2; consideriamo il rango per righe: siccome vi sono righe non proporzionali è ≥ 2, inoltre si vede abbastanza facilmente che la terza riga è combinazione lineare delle prime due, e precisamente $a_3 = -a_1 + 2a_2$. Dal Teorema 5.1.2 segue che anche il rango per colonne è 2: verificarlo direttamente per esercizio.

5.2 Sistemi lineari di equazioni I

Sia $A = (a_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$ una matrice $m \times n$ a entrate in K, e siano $b_1, \dots, b_m \in K$ scalari in numero uguale al numero delle righe di A. Siano $x_1, \dots, x_n$ incognite in numero pari al numero delle colonne di A.

Consideriamo il seguente sistema lineare di m equazioni nelle n incognite $x_1, \dots, x_n$:

$$\begin{cases} a_{11}x_1 + & \cdots & a_{1n}x_n & = b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1}x_1 + & \cdots & a_{mn}x_n & = b_m. \end{cases} \tag{5.1}$$

Si dice che A è la **matrice dei coefficienti** del sistema lineare (5.1). **La matrice completa** del sistema è invece la matrice $m \times (n+1)$ ottenuta aggiungendo ad A una colonna b formata dagli scalari $b_1, \dots, b_m$; la si denota $(A \mid b)$.

Una **soluzione del sistema** è un vettore $v = (v_1, \dots, v_n)$ di K^n tale che sostituendo al posto di ogni incognita x_i la coordinata v_i di v le relazioni in (5.1) siano tutte soddisfatte, ossia si abbia

$$\begin{cases} a_{11}v_1+ & \cdots & a_{1n}v_n & = b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1}v_1+ & \cdots & a_{mn}v_n & = b_m. \end{cases} \tag{5.2}$$

Un sistema lineare si dice **compatibile** se ha almeno una soluzione.

Se $b = (b_1, \dots, b_m)$ è il vettore nullo, ossia se $b_1 = \cdots = v_m = 0$ si dice che il sistema lineare è **omogeneo**. Si dice anche che

$$\begin{cases} a_{11}x_1+ & \cdots & a_{1n}x_n & = 0 \\ \vdots & & \vdots & \vdots \\ a_{m1}x_1+ & \cdots & a_{mn}x_n & = 0 \end{cases} \tag{5.3}$$

è il sistema lineare omogeneo associato al sistema (5.1).

Proposizione 5.2.1. *Un sistema omogeneo (5.3) è sempre compatibile, e l'insieme delle sue soluzioni è un sottospazio vettoriale di K^n.*

Dimostrazione. Osserviamo che il vettore nullo è una soluzione, in quanto

$$\begin{cases} a_{11}0+ & \cdots & a_{1n}0 & = 0 \\ \vdots & & \vdots & \vdots \\ a_{m1}0+ & \cdots & a_{mn}0 & = 0. \end{cases}$$

Sia W l'insieme delle soluzione del sistema (5.3); W risulta chiuso rispetto alle due operazioni di somma e di prodotto esterno, infatti se $v = (v_1, \dots, v_n)$, $w = (w_1, \dots, w_n)$ sono soluzioni, anche $(v_1, \dots, v_n) + (w_1, \dots, w_n)$ e $\lambda(v_1, \dots, v_n)$ lo sono. Dunque W è un sottospazio vettoriale di K^n. □

Un sistema lineare non omogeneo può essere compatibile o meno; se lo è può avere una o più soluzioni. Per esempio nel primo caso, in cui $m = n = 1$, cioè un'equazione in un'incognita, per l'equazione $ax = b$ si possono presentare i seguenti casi:

1. $a \neq 0$, c'è una e una sola soluzione $\frac{b}{a}$;
2. $a = 0$, $b \neq 0$, il sistema non è compatibile;
3. $a = b = 0$: ogni elemento $v \in K$ è soluzione.

Osservazione 5. Il sistema lineare (5.1) si può riscrivere nella forma

$$x_1a^1 + x_2a^2 + \cdots + x_na^n = b$$

ossia

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

Se ne deduce immediatamente la seguente prima condizione perchè un sistema sia compatibile.

Proposizione 5.2.2. *Il sistema lineare (5.1) è compatibile se e solo se il vettore b appartiene allo spazio delle colonne della matrice A, cioè è combinazione lineare delle colonne di A. In tal caso una soluzione $(v_1, \dots, v_n)$ è data dai coefficienti di una combinazione lineare di $a^1, \dots, a^n$ uguale a b. Se in più $a^1, \dots, a^n$ sono linearmente indipendenti, la soluzione è unica.*

Vediamo qualche esempio di conseguenze della Proposizione 5.2.2.

Esempio 5.2.3.
1. Consideriamo un sistema omogeneo (5.3): è sempre compatibile e ha quanto meno la soluzione nulla. Ha anche soluzioni non nulle se e solo se le colonne di A sono linearmente dipendenti. Ciò si esprime dicendo che il rango per colonne di A è strettamente minore del numero n delle colonne. Dal momento che il rango per colonne è minore o uguale al minimo fra m e n, certamente il sistema ha soluzioni non nulle se $m < n$.
2. Caso $m = n$. Se le colonne di A sono linearmente indipendenti costituiscono una base di K^n, allora **qualunque sia** b, il sistema è compatibile e ha una e una sola soluzione.

5.3 Trasformazioni elementari sulle righe di una matrice

Il nostro obiettivo è di dare un metodo per risolvere i sistemi di equazioni lineari, cioè per determinare innanzitutto se un dato sistema lineare (5.1) è compatibile, e in caso affermativo per trovarne tutte le soluzioni.

A tale scopo introdurremo delle trasformazioni che mutano un dato sistema lineare in un sistema equivalente.

Definizione 5.3.1. Due sistemi lineari di m equazioni in n incognite (stesso numero di equazioni e di incognite) si dicono **equivalenti** se hanno lo stesso insieme di soluzioni, in particolare uno è compatibile se e solo se lo è l'altro.

Vedremo una classe di trasformazioni, dette trasformazioni elementari sulle righe di una matrice, che, se applicate alle righe della matrice completa $(A \mid b)$ di un sistema lineare, lo mutano in un sistema equivalente.

Introduciamo ora quattro tipi di trasformazioni elementari. Sia

$$C = (c_{ij})_{\substack{1 \le i \le m \\ 1 \le j \le n}} = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}$$

una matrice $m \times n$.

I tipo: moltiplicazione di una riga per uno scalare $\lambda \neq 0$:

$$\begin{pmatrix} c_1 \\ \vdots \\ c_i \\ \vdots \\ c_m \end{pmatrix} \overset{I}{\to} \begin{pmatrix} c_1 \\ \vdots \\ \lambda c_i \\ \vdots \\ c_m \end{pmatrix}$$

II tipo: addizione della j-esima riga alla i-esima riga:

$$\begin{pmatrix} c_1 \\ \vdots \\ c_i \\ \vdots \\ c_j \\ \vdots \\ c_m \end{pmatrix} \xrightarrow{II} \begin{pmatrix} c_1 \\ \vdots \\ c_i + c_j \\ \vdots \\ c_j \\ \vdots \\ c_m \end{pmatrix}$$

III tipo: addizione di un multiplo della j-esima riga alla i-esima riga:

$$\begin{pmatrix} c_1 \\ \vdots \\ c_i \\ \vdots \\ c_j \\ \vdots \\ c_m \end{pmatrix} \xrightarrow{III} \begin{pmatrix} c_1 \\ \vdots \\ c_i + \lambda c_j \\ \vdots \\ c_j \\ \vdots \\ c_m \end{pmatrix}$$

IV tipo: scambio della j-esima con la i-esima riga:

$$\begin{pmatrix} c_1 \\ \vdots \\ c_i \\ \vdots \\ c_j \\ \vdots \\ c_m \end{pmatrix} \xrightarrow{IV} \begin{pmatrix} c_1 \\ \vdots \\ c_j \\ \vdots \\ c_i \\ \vdots \\ c_m \end{pmatrix}$$

Osservazione 6. Una trasformazione del III tipo si può ottenere applicando successivamente trasformazioni del I e del II tipo; precisamente si moltiplica dapprima la riga j-esima per λ (I tipo), poi si somma la j-esima alla i-esima riga (II tipo), e infine si moltiplica la riga j-esima per λ^{-1} (I tipo). Anche una trasformazione del IV tipo si può ottenere applicando trasformazioni del I e del II tipo (come?).

Proposizione 5.3.2. *Sia C una matrice. Lo spazio delle righe di C non cambia se si applicano a C trasformazioni elementari sulle righe.*

Dimostrazione. La dimostrazione si basa sulla seguente semplice osservazione:

Osservazione 7. Siano $U =< u_1, \ldots, u_s >$, $W =< w_1, \ldots, w_r >$ sottospazi finitamente generati di uno spazio vettoriale V. Allora $U = W$ se e solo se $w_1, \ldots, w_r \in U$ e $u_1, \ldots, u_s \in W$. Infatti, ricordiamo che U è il minimo sottospazio che contiene $u_1, \ldots, u_s$ e W è il minimo sottospazio che contiene $w_1, \ldots, w_r$. Dunque se $w_1, \ldots, w_r \in U$, allora anche $< w_1, \ldots, w_r > \subset U$, ossia $W \subset U$. Similmente si ha l'inclusione opposta.

Nel caso di una trasformazione del I tipo, basta dunque verificare che

$$c_i \in < c_1, \ldots, \lambda c_i, \ldots, c_m >;$$

ma questo è vero perchè $c_i = 0c_1 + \cdots + \lambda^{-1}(\lambda c_i) + \cdots + 0c_m$. Per le trasformazioni del II tipo, osserviamo che $c_i = 0c_1 + \cdots + 1(c_i + c_j) + \cdots - c_j \cdots$. Per quelle del III tipo si ha $c_i = 0c_1 + \cdots + 1(c_i + \lambda c_j) + \cdots - \lambda c_j \cdots$. Il caso delle trasformazioni del IV tipo è chiaro, perchè si è solo cambiato l'ordine in cui si considerano i generatori. □

Abbiamo ora l'applicazione ai sistemi di equazioni lineari.

Proposizione 5.3.3. *Dato il sistema lineare (5.1), consideriamo la sua matrice completa $A' = (A \mid b)$. Se si applicano ad A' trasformazioni elementari sulle righe, si ottiene la matrice completa di un sistema lineare equivalente a quello di partenza.*

Dimostrazione. Infatti con una trasformazione del I tipo si moltiplica una delle equazioni per una costante non nulla; con una del II tipo si sostituisce ad un'equazione la sua somma con una delle altre equazioni. Il III tipo è una combinazione dei tipi precedenti, e con il IV tipo si scambia l'ordine di due delle equazioni. □

5.4 Algoritmo di eliminazione di Gauss

L'algoritmo di eliminazione di Gauss fa passare, mediante trasformazioni elementari sulle righe della matrice completa, da un sistema lineare (5.1) ad un sistema ad esso equivalente che è "facile da risolvere". Precisamente, la matrice completa $(A \mid b)$ viene trasformata in una matrice "a gradini o a scala".

Definizione 5.4.1. Una **matrice a gradini** o **a scala** è una matrice della forma

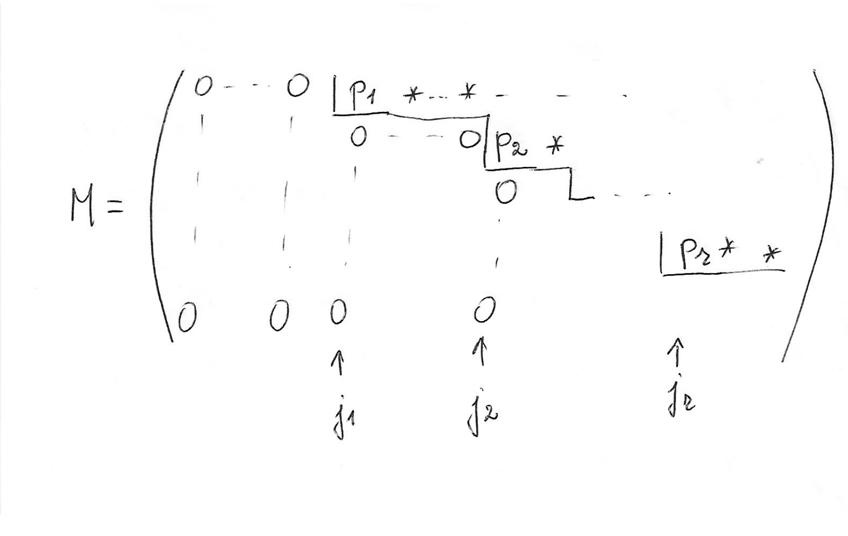

Figura 5.1: Matrice a gradini

dove $p_1, \ldots, p_r$ sono scalari diversi da 0. Sotto la "scala" è tutto 0. Gli asterischi significano che in quelle posizioni vi può essere qualunque cosa. Gli r elementi $p_1, \ldots, p_r$ si trovano nelle colonne di indici $j_1, \ldots, j_r$, e si chiamano i "pivot" della matrice M.

Il pivot p_1 si trova nella prima riga ed è il "primo" elemento non nullo della prima riga, è detto primo pivot. Sotto al primo pivot nella colonna j_1-esima è tutto 0; p_2 è il primo elemento non nullo della seconda riga, e sotto è tutto 0, e così via.

Proposizione 5.4.2. *Sia M una matrice a gradini come nella Figura 5.1. Le sue prime r righe $m_1, \dots, m_r$ formano una base dello spazio delle righe di M, dunque il rango per righe di M è uguale a r.*

Dimostrazione. Le righe $m_1, \dots, m_r$ sono linearmente indipendenti; infatti una loro combinazione lineare nulla $\lambda_1 m_1 + \cdots + \lambda_r m_r = 0$ ha la forma

$$(0, \dots, 0, \lambda_1 p_1, \dots, \lambda_2 p_2 + \lambda_1(*), \dots, \lambda_r p_r + \lambda_{r-1}(*) + \cdots, 0, \dots, 0) = 0;$$

essendo i pivot tutti non nulli, segue $\lambda_1 = \lambda_2 = \cdots = \lambda_r = 0$. Le righe successive $m_{r+1}, \dots, m_n$ sono tutte nulle. □

Passiamo ora alla descrizione dell'algoritmo di eliminazione di Gauss.

Algoritmo. Si parte da una matrice $M \in M(s \times n, K)$ e la si trasforma in una matrice M' a scala con trasformazioni elementari sulle righe.

- Se M è la matrice nulla abbiamo finito.
- Se M non è la matrice nulla, sia m^{j_1} la prima colonna non nulla. Scambiando eventualmente la prima riga con una riga successiva (trasformazione del IV tipo) possiamo supporre che sia $m_{1j_1} \neq 0$, e poniamo $p_1 = m_{1j_1}$.
- Per ogni $h \geq 2$ sommiamo alla riga h-esima un opportuno multiplo della I riga, in modo da annullare tutti gli elementi sotto a p_1 (trasformazione del III tipo). La prima riga rimane fissa.
- Consideriamo le righe $m_2, \dots, m_s$: se sono tutte nulle abbiamo finito. Altrimenti sia j_2 l'indice della prima colonna che contiene un elemento non nullo nella righe $m_2, \dots, m_s$. Scambiando eventualmente la seconda riga con una riga successiva (trasformazione del IV tipo) possiamo supporre che sia $m_{2j_2} \neq 0$, e poniamo $p_2 = m_{2j_2}$.
- Con trasformazioni elementari del III tipo, per ogni $h \geq 3$ sommiamo alla riga h-esima un opportuno multiplo della seconda riga, in modo da annullare tutti gli elementi sotto a p_2.
- Proseguiamo in questo modo, finchè arriviamo o ad avere l'ultimo pivot nell'ultima riga, oppure le ultime righe tutte nulle.

Esempio 5.4.3. Nell'esempio della Figura 5.2 sono mostrate le trasformazioni elementari eseguite; la matrice ha rango (per righe) uguale a 3; i pivot sono nelle colonne di indici $j_1 = 2$, $j_2 = 4$, $j_3 = 5$. Dopo la trasformazione, le righe m_1, m_2, m_3 sono linearmente indipendenti, mentre la riga m_4 è nulla.

Un sistema lineare è detto a gradini se lo è la sua matrice completa. Un sistema lineare a gradini si risolve con il metodo di "sostituzione all'indietro". Lo illustriamo con un esempio.

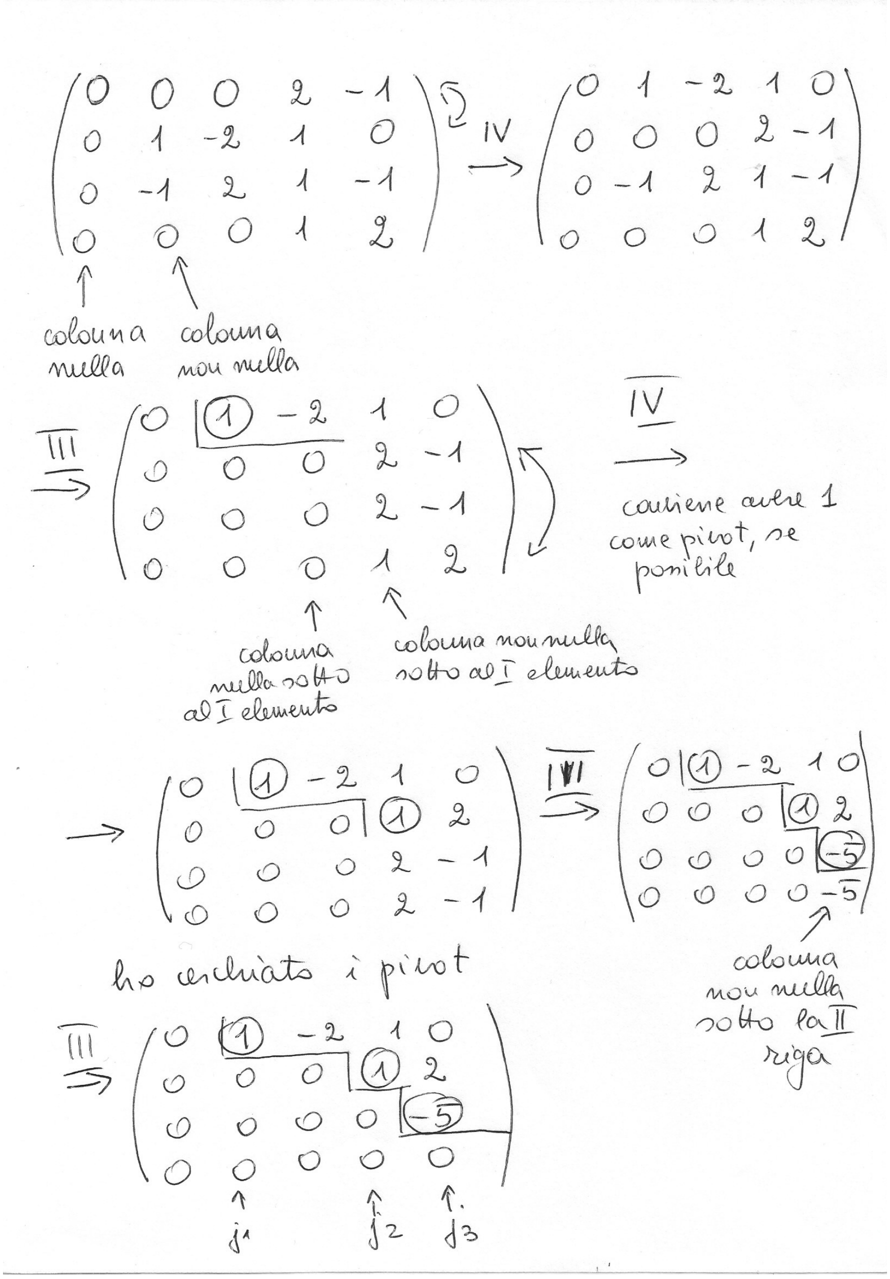

Figura 5.2: Esempio di applicazione dell'algoritmo di eliminazione di Gauss

Esempio 5.4.4. Consideriamo il sistema lineare

$$\begin{cases} x_1 + 2x_2 + x_3 = -1 \\ 2x_1 + x_2 - x_3 = 1 \\ -x_1 - x_2 + 2x_3 = 1 \end{cases} \tag{5.4}$$

Consideriamo $A' = (A \mid b)$ la sua matrice completa e le applichiamo l'algoritmo di eliminazione di Gauss:

$$\left(\begin{array}{rrr|r} 1 & 2 & 1 & -1 \\ 2 & 1 & -1 & 1 \\ -1 & -1 & 2 & 1 \end{array}\right) \xrightarrow{III} \left(\begin{array}{rrr|r} 1 & 2 & 1 & -1 \\ 0 & -3 & -3 & 3 \\ 0 & 1 & 3 & 0 \end{array}\right) \xrightarrow{I}$$

$$\rightarrow \left(\begin{array}{rrr|r} 1 & 2 & 1 & -1 \\ 0 & 1 & 1 & -1 \\ 0 & 1 & 3 & 0 \end{array}\right) \xrightarrow{III} \left(\begin{array}{rrr|r} \boxed{1} & 2 & 1 & -1 \\ 0 & \boxed{1} & 1 & -1 \\ 0 & 0 & \boxed{2} & 1 \end{array}\right):$$

abbiamo ottenuto una matrice a gradini con 3 pivot, dunque è di rango 3. Il sistema (5.4) è equivalente al seguente:

$$\begin{cases} x_1 & + & 2x_2 & + & x_3 & = & -1 \\ & & x_2 & + & x_3 & = & -1 \\ & & & & 2x_3 & = & 1 \end{cases} \tag{5.5}$$

Dall'ultima equazione ricaviamo $x_3 = \frac{1}{2}$; sostituiamo il valore di x_3 trovato nella seconda e otteniamo $x_2 = -1 - \frac{1}{2} = -\frac{3}{2}$, e infine sostituiamo nella prima i valori di x_2 e di x_3 e otteniamo $x_1 = 3 - \frac{1}{2} - 1 = \frac{3}{2}$. Quindi il sistema ha l'unica soluzione $(\frac{3}{2}, -\frac{3}{2}, \frac{1}{2})$. In effetti, le tre colonne della matrice dei coefficienti del sistema (5.5) sono chiaramente linearmente indipendenti, quindi il sistema ha una e una sola soluzione qualunque sia la colonna dei termini noti (Esempio 5.2.3, 2.). Osserviamo che trasformando A' in una matrice a gradini, si è trasformata in forma a gradini anche A.

5.5 Soluzioni di un sistema lineare compatibile

D'ora in poi per indicare il sistema lineare (5.1) useremo spesso la notazione compatta $AX = b$, dove con X intendiamo la n-upla delle incognite $x_1, \ldots, x_n$ scritte in colonna: $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. Similmente, se v è una soluzione, scriveremo $Av = b$. Il sistema omogeneo associato è $AX = 0$: abbiamo osservato nella Proposizione 5.2.1 che è sicuramente compatibile e che l'insieme delle sue soluzioni è un sottospazio vettoriale di K^n, che denoteremo con W.

Proposizione 5.5.1. *Supponiamo che il sistema lineare $AX = b$ sia compatibile e denotiamo con $S \subset K^n$ l'insieme delle sue soluzioni.*

1. Se $b \neq 0$, cioè se il sistema lineare (5.1) non è omogeneo, allora S non è un sottospazio vettoriale di K^n.

2. Se $\bar{v} \in K^n$ è una soluzione particolare del sistema, si ha che

$$S = \bar{v} + W = \{\bar{v} + w \mid w \in W\},$$

dove W è lo spazio delle soluzioni del sistema omogeneo associato.

Dimostrazione. 1. Se il sistema non è omogeneo, il vettore nullo 0 non è una soluzione, quindi S non è sottospazio vettoriale.

2. Se w è una soluzione del sistema omogeneo associato, $Aw = 0$; per ipotesi $A\bar{v} = b$, dunque si ha $A(\bar{v} + w) = b$. Viceversa, se $A\bar{v} = b$ e $Av = b$, si ha $A(v - \bar{v}) = 0$ e quindi $v - \bar{v} \in W$, da cui segue che $v \in \bar{v} + W$. □

Osservazione 8. Essendo W un sottospazio vettoriale di K^n si può considerare lo **spazio vettoriale quoziente** K^n/W introdotto nel Foglio 2 di esercizi (Esercizio 2): è il quoziente rispetto alla relazione d'equivalenza secondo cui $v \sim v'$ se e solo se $v - v' \in W$, ed è un K-spazio vettoriale rispetto alle operazioni indotte dalle operazioni di K^n. Allora $\bar{v} + W$ risulta essere la classe d'equivalenza del vettore $\bar{v}$ in K^n/W.

Si ha anche la seguente definizione di sottospazio affine di uno spazio vettoriale.

Definizione 5.5.2. Sia V un K-spazio vettoriale. Un **sottospazio affine** di V è un sottinsieme di V del tipo $S = v + W$, dove $v \in V$ e W è un sottospazio vettoriale detto **giacitura di** S. Si definisce la dimensione di S, $\dim S$, come la dimensione della sua giacitura: $\dim S = \dim W$.

Poichè $0 \in W$, allora $v \in S$: si dice che S è un sottospazio affine passante per v. Osserviamo che se $u \in S = v + W$ è un qualunque vettore di S, allora si ha anche $S = u + W = v + W$. Infatti sia $u = v + w$, con $w \in W$; per ogni $w' \in W$, $u + w' = v + (w + w') \in v + W$; viceversa $v + w' = (u - w) + w' = u + (w' - w) \in u + W$.

Esempio 5.5.3. Se $W = (0)$ è il sottospazio nullo, i sottospazi affini di giacitura W sono i singoli vettori di V.

Se $\dim W = 1$, i sottospazi di dimensione 1 di giacitura W sono le rette “parallele” a W, o di direzione W.

E così via, se $\dim W = 2$ si ottengono i piani paralleli a W.

Con questo linguaggio possiamo riformulare come segue la Proposizione 5.5.1.

Corollario 5.5.4. *L'insieme delle soluzioni di un sistema lineare compatibile è un sottospazio affine avente come giacitura lo spazio delle soluzioni del sistema omogeneo associato, e passante per una soluzione particolare del sistema.*

5.6 Risoluzione dei sistemi omogenei

Vediamo ora come trovare la dimensione e una base dello spazio W delle soluzioni di un sistema lineare omogeneo.

Iniziamo con un esempio.

Esempio 5.6.1. Un sistema lineare omogeneo di 5 equazioni in 7 incognite $x_1, \ldots, x_7$ a coefficienti in $\mathbb{R}$.

$$\left\{\begin{array}{ccccccccccccc} x_2 & & & + & 2x_4 & - & x_5 & - & 4x_6 & & & = & 0 \\ & & x_3 & - & x_4 & - & x_5 & + & 2x_6 & + & x_7 & = & 0 \\ x_2 & & & + & 2x_4 & + & x_5 & - & 2x_6 & & & = & 0 \\ & & x_3 & - & x_4 & & & + & 2x_6 & - & x_7 & = & 0 \\ x_2 & + & x_3 & + & x_4 & & & & & + & x_7 & = & 0 \end{array}\right. \tag{5.6}$$

Essendo il sistema omogeneo, la matrice completa ha soltanto una colonna nulla a destra della matrice dei coefficienti A; tale colonna rimane invariata per trasformazioni elementari; si lavora soltanto con la matrice dei coefficienti A.

$$\begin{pmatrix} 0 & \boxed{1} & 0 & 2 & -1 & -4 & 0 \\ 0 & 0 & 1 & -1 & -1 & 2 & 1 \\ 0 & 1 & 0 & 2 & 1 & -2 & 0 \\ 0 & 0 & 1 & -1 & 0 & 2 & -1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{III} \begin{pmatrix} 0 & \boxed{1} & 0 & 2 & -1 & -4 & 0 \\ 0 & 0 & \boxed{1} & -1 & -1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & 1 & -1 & 0 & 2 & -1 \\ 0 & 0 & 1 & -1 & 1 & 4 & 1 \end{pmatrix} \xrightarrow{III}$$

$$\rightarrow \begin{pmatrix} 0 & \boxed{1} & 0 & 2 & -1 & -4 & 0 \\ 0 & 0 & \boxed{1} & -1 & -1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 & 2 & 2 & 0 \end{pmatrix} \xrightarrow{I} \begin{pmatrix} 0 & \boxed{1} & 0 & 2 & -1 & -4 & 0 \\ 0 & 0 & \boxed{1} & -1 & -1 & 2 & 1 \\ 0 & 0 & 0 & 0 & \boxed{1} & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \xrightarrow{III}$$

$$\rightarrow \begin{pmatrix} 0 & \boxed{1} & 0 & 2 & -1 & -4 & 0 \\ 0 & 0 & \boxed{1} & -1 & -1 & 2 & 1 \\ 0 & 0 & 0 & 0 & \boxed{1} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \boxed{1} & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Siamo arrivati a una matrice a gradini con $j_1 = 2$, $j_2 = 3$, $j_3 = 5$, $j_4 = 6$, $r = 4$.

Il sistema a gradini corrispondente, equivalente a (5.7), è il seguente.

$$\begin{cases} \boxed{x_2} + 2x_4 - x_5 - 4x_6 = 0 \\ \boxed{x_3} - x_4 - x_5 + 2x_6 + x_7 = 0 \\ \boxed{x_5} + x_6 = 0 \\ \boxed{x_6} + 2x_7 = 0 \\ 0 = 0 \end{cases} \tag{5.7}$$

Notiamo che l'ultima equazione non è significativa e non occorre scriverla. Dall'ultima equazione non banale ricaviamo x_6 in funzione di x_7, su cui non vi sono condizioni: $x_6 = -2x_7$, poi sostituiamo nella terza: $x_5 = -x_6 = 2x_7$; su x_4 non vi sono condizioni, sostituiamo nella seconda per ricavare x_3: $x_3 = x_4 + 2x_7 + 4x_7 - x_7 = x_4 + 5x_7$; infine dalla prima ricaviamo $x_2 = -2x_4 + x_5 + 4x_6 = -2x_4 + 2x_7 - 8x_7 = -2x_4 - 6x_7$. Abbiamo ricavato i valori delle incognite corrispondenti ai pivot, x_2, x_3, x_5, x_6 in funzione delle altre, che acquistano il significato di parametri e sono dette **parametri liberi.** In particolare su x_1 non vi sono condizioni, è un parametro libero. Le incognite che abbiamo calcolato in funzione delle altre sono quelle che stanno nelle colonne dei pivot, pertanto sono in numero di r; i parametri liberi sono quindi $n - r$, e corrispondono alle colonne in cui non ci sono pivot.

La soluzione generale del sistema si ottiene dando ai parametri liberi tutti i possibili valori in $\mathbb{R}$, ed è pertanto della forma $(\mathbf{t_1}, -2t_4 - 6t_7, t_4 + 5t_7, \mathbf{t_4}, 2t_7, -2t_7, \mathbf{t_7})$, che può essere riscritta come $t_1(1,0,0,0,0,0,0) + t_4(0,-2,1,1,0,0,0) + t_7(0,-6,5,0,2,-2,1)$, al variare di t_1, t_4, t_7 in $\mathbb{R}$. Introduciamo i tre vettori di $\mathbb{R}^7$ $w_1 = (1,0,0,0,0,0,0)$, $w_2 =$

$(0, -2, 1, 1, 0, 0, 0)$, $w_3 = (0, -6, 5, 0, 2, -2, 1)$. Osserviamo che w_1 si ottiene per $t_1 = 1$, $t_4 = t_7 = 0$, w_2 si ottiene per $t_1 = t_7 = 0$ e $t_4 = 1$, e w_3 si ottiene per $t_1 = t_4 = 0$ e $t_7 = 1$. Un vettore di $\mathbb{R}^7$ è soluzione del sistema se e solo se è una combinazione lineare di w_1, w_2, w_3: $t_1 w_1 + t_4 w_2 + t_7 w_3$. Quindi lo spazio delle soluzioni W è generato da w_1, w_2, w_3, inoltre w_1, w_2, w_3 sono linearmente indipendenti; infatti una loro combinazione lineare nulla $t_1 w_1 + t_4 w_2 + t_7 w_3 = 0$ deve avere $t_1 = t_4 = t_7 = 0$, in quanto t_1, t_4, t_7 sono alcune delle sue coordinate.

Abbiamo dunque trovato una base (w_1, w_2, w_3) dello spazio W delle soluzioni, e la sua dimensione è $3 = 7 - 4$, ossia il numero delle incognite meno il numero dei pivot, o anche il numero delle incognite meno il rango per righe della matrice dei coefficienti.

Questo esempio si generalizza al caso di un qualunque sistema lineare omogeneo e ci permette di ottenere direttamente il seguente risultato generale.

Teorema 5.6.2. *Sia $AX = 0$ un sistema lineare omogeneo in n incognite a coefficienti in un campo K. Sia $W \subset K^n$ il suo spazio delle soluzioni. Si ha:*

1. $\dim W = n - r$, *dove r è il rango (per righe) della matrice A;*

2. *sia B una matrice a gradini ottenuta da A con trasformazioni elementari sulle righe; con il metodo di sostituzione all'indietro applicato al sistema $BX = 0$, equivalente a quello di partenza, si trova una soluzione generale del sistema, espressa in funzione di $n - r$ parametri liberi corrispondenti alle incognite non contenute nel colonne dei pivot;*

3. *assegnando agli $n-r$ parametri liberi i valori* $(1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, 0, \dots, 1)$ *si trova una base di W.*

5.7 Teorema di Rouché - Capelli e risoluzione dei sistemi lineari

Teorema 5.7.1 (Teorema di Rouché-Capelli). *Sia $AX = b$ un sistema lineare a coefficienti in un campo K. Il sistema è compatibile se e solo se il rango per righe della matrice dei coefficienti A coincide con quello della matrice completa $(A \mid b)$.*

Dimostrazione. Con trasformazioni elementari sulle righe di $(A \mid b)$, trasformiamo il sistema $AX = b$ in un sistema equivalente $\bar{A}X = \bar{b}$, dove $\bar{A}$ è una matrice a gradini e $(\bar{A} \mid \bar{b})$ è della seguente forma (per l'equivalenza vedere la Proposizione 5.3.3):

$$\left(\begin{array}{cccccc|c}
\cdots & \boxed{\overline{a_{1j_1}}} & * & * & * & * & \overline{b_1} \\
 & & \boxed{\overline{a_{2j_2}}} & * & * & * & \overline{b_2} \\
 & & & \ddots & & & \cdots \\
 & & & & \boxed{\overline{a_{rj_r}}} & * & \overline{b_r} \\
\cdots & 0 & 0 & \cdots & 0 & 0 & \overline{b_{r+1}} \\
 & & & \ddots & & & \cdots \\
\cdots & 0 & 0 & \cdots & 0 & 0 & \overline{b_m}
\end{array}\right)$$

Chiaramente il sistema è compatibile se e solo se $\overline{b_{r+1}} = \cdots = \overline{b_m} = 0$. Sia r il rango di $\bar{A}$: r differisce dal rango di $(\bar{A} \mid \bar{b})$ se e solo se nell'ultima colonna ci sono più di r elementi non nulli, cioè se c'è un pivot di $(\bar{A} \mid \bar{b})$ nella riga $(r+1)$-esima, e ciò equivale a dire che l'equazione $(r+1)$-esima non è compatibile. Osserviamo che i due ranghi possono differire al più di uno. □

Ricapitolando, sia A una matrice $m \times n$ a entrate in K, denotiamo con $\mathrm{rg}(A)$ il suo rango per righe.

- Il sistema lineare $AX = b$ è compatibile se e solo se $\mathrm{rg}(A) = \mathrm{rg}(A \mid b)$;
- in particolare, se il sistema è omogeneo, esso è compatibile e l'insieme delle soluzioni è un sottospazio vettoriale di K^n di dimensione $n - \mathrm{rg}(A)$;
- se il sistema non è omogeneo ed è compatibile, l'insieme delle soluzioni è un sottospazio affine di K^n del tipo $S = \bar{v} + W$, dove W è lo spazio delle soluzioni del sistema omogeneo associato $AX = 0$; inoltre $\dim S = \dim W = n - r$ dove $r = \mathrm{rg}(A) = \mathrm{rg}(A \mid\ b)$;
- per risolvere il sistema, si riduce $(A \mid b)$ a gradini; dopo aver verificato che i ranghi di A e di $(A \mid b)$ siano uguali, si procede con il metodo di sostituzione all'indietro; questo fornisce una soluzione particolare e una base di W.

Illustriamo l'ultimo punto con il seguente esempio.

Esempio 5.7.2.

$$\begin{cases} x_1 - 2x_2 + x_3 = 1 \\ x_1 - 2x_2 - x_4 = 2 \\ x_3 + x_4 = -1 \end{cases} \tag{5.8}$$

$$\left(\begin{array}{cccc|c} \boxed{1} & -2 & 1 & 0 & 1 \\ 1 & -2 & 0 & -1 & 2 \\ 0 & 0 & 1 & 1 & -1 \end{array}\right) \xrightarrow{III} \left(\begin{array}{cccc|c} \boxed{1} & -2 & 1 & 0 & 1 \\ 0 & 0 & \boxed{-1} & -1 & 1 \\ 0 & 0 & 1 & 1 & -1 \end{array}\right) \xrightarrow{I}$$

$$\to \left(\begin{array}{cccc|c} \boxed{1} & -2 & 1 & 0 & 1 \\ 0 & 0 & \boxed{-1} & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Il sistema è compatibile, $r = 2$, $j_1 = 1$, $j_2 = 3$. Il sistema a gradini è

$$\begin{cases} \boxed{x_1} - 2x_2 + x_3 = 1 \\ \boxed{-x_3} - x_4 = 1 \end{cases}$$

(non abbiamo scritto l'equazione identica $0 = 0$). Ci sono due parametri liberi t_2, t_4 relativi alle colonne senza pivot. La soluzione generale è

$$(x_1, x_2, x_3, x_4) = (2t_2 + t_4 + 2, t_2, -t_4 - 1, t_4) = t_2(2, 1, 0, 0) + t_4(1, 0, -1, 1) + (2, 0, -1, 0).$$

Poniamo $w = (2, 1, 0, 0)$, $w' = (1, 0, -1, 1)$, $\bar{v} = (2, 0, -1, 0)$. $W =< w, w' >$ è lo spazio delle soluzioni del sistema omogeneo associato, $\bar{v}$ è una soluzione particolare del sistema (5.8), ottenuta per $t_2 = t_4 = 0$.

Capitolo 6

Applicazioni lineari

6.1 Prodotto di matrici

Premettiamo la definizione e le prime proprietà dell'operazione di prodotto righe per colonne di matrici, che sarà importante nella descrizione delle applicazioni lineari.

Definizione 6.1.1. Siano $A = (a_{ij})_{\substack{1 \le i \le m \\ 1 \le j \le n}}$ e $B = (b_{jk})_{\substack{1 \le j \le n \\ 1 \le k \le p}}$, due matrici a coefficienti in un campo K, la prima di ordine $m \times n$ e la seconda di ordine $n \times p$. Il loro **prodotto righe per colonne** è la seguente matrice $m \times p$:

$$C = AB = (c_{ik})_{\substack{1 \le i \le m \\ 1 \le k \le p}}$$

dove

$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk} = \sum_{j=1}^{n} a_{ij}b_{jk}.$$

Abbiamo così definito per ogni terna di interi positivi m, n, p un'applicazione

$$M(m \times n, K) \times M(n \times p, K) \to M(m \times p, K)$$

tale che

$$(A, B) \to C = AB.$$

In particolare, se $m = n$ abbiamo definito un prodotto interno in $M(n \times n, K)$.

Esempio 6.1.2. 1.

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

Il loro prodotto è una matrice 2×2:

$$AB = \begin{pmatrix} 3 & 7 \\ 1 & 7 \end{pmatrix}.$$

Ha senso anche il prodotto BA, che risulta essere la seguente matrice 3×3:

$$BA = \begin{pmatrix} 0 & -5 & -1 \\ 0 & 5 & 1 \\ 2 & 9 & 5 \end{pmatrix}.$$

Se consideriamo invece una matrice 2×2 C, esiste solo CA ma non AC.

2. A matrice riga $1 \times n$: $A = (a_{11}, \ldots, a_{1n})$, B matrice colonna $n \times 1$: $B = \begin{pmatrix} b_{11} \\ \cdots \\ b_{n1} \end{pmatrix}$, allora $AB = (a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1n}b_{n1})$, matrice 1×1.

3. A matrice colonna $n \times 1$: $A = \begin{pmatrix} a_{11} \\ \cdots \\ a_{n1} \end{pmatrix}$, B matrice riga $1 \times m$: $B = (b_{11}, \ldots, b_{1m})$, allora AB è una matrice $n \times m$:

$$\begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & \cdots & a_{11}b_{1m} \\ a_{21}b_{11} & a_{21}b_{12} & \cdots & a_{21}b_{1m} \\ \vdots & & & \vdots \\ a_{n1}b_{11} & a_{n1}b_{12} & \cdots & a_{n1}b_{1m} \end{pmatrix},$$

4. Se A è una matrice $m \times n$, ed E_n è la matrice identica $n \times n$, si ha $AE_n = A$; analogamente si ha $E_mA = A$ dove E_m è la matrice identica di ordine m. In particolare se A è una matrice $n \times n$ $AE_n = E_nA = A$. Dunque E_n è elemento neutro per il prodotto in $M(n \times n, K)$.

5. Il prodotto righe per colonne di matrici non gode della proprietà commutativa, come mostra il seguente esempio. Siano

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

Allora

$$AB = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}, \; BA = \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}.$$

6. Se A è una matrice $m \times n$, b è una matrice $m \times 1$ e X denota una colonna di indeterminate $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, ha senso scrivere $AX = b$, ossia per esteso

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix},$$

e si ottiene così la (5.1). Questo spiega la notazione usata per indicare i sistemi lineari di equazioni.

7.

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} = \text{per la formula di addizione} =$$

$$\begin{pmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{pmatrix}$$

Definizione 6.1.3. Data una matrice $m \times n$ $A = (a_{ij})$, la **trasposta di** A, denotata con ${}^tA$, è la matrice $n \times m$ ottenuta "scambiando" le righe e le colonne di A, ossia l'elemento di indici ij di ${}^tA$ è a_{ji}.

Proposizione 6.1.4. *Il prodotto righe per colonne di matrici verifica le seguenti proprietà, per ogni scelta di matrici A, B, A', B' con il numero appropriato di righe e colonne, e di uno scalare λ :*

a) ${}^t(AB) = {}^tB\,{}^tA$;

b) $A(B + B') = AB + AB'$,
$(A + A')B = AB + A'B$;

c) $(AB)C = A(BC)$;

d) $(\lambda A)B = \lambda(AB) = A(\lambda B)$.

Dimostrazione. Le verifiche seguono direttamente dalla definizione. Per esempio, per la a), osserviamo che l'elemento di indici ih di $C = AB$ è

$$c_{ih} = \sum_{j=1}^{n} a_{ij} b_{jh}$$

mentre l'elemento di indici hi di $C' = {}^tB\,{}^tA$ è

$$c'_{hi} = {}^t(b^h)\,{}^t(a_i) = \sum_{j=1}^{n} b_{jh} a_{ij} \text{ che coincide con } c_{ih}.$$

□

6.2 Applicazioni lineari

Siano V, V' spazi vettoriali su uno stesso campo K.

Definizione 6.2.1. Un'applicazione $f : V \to V'$ è detta **applicazione lineare** o omomorfismo se

1. $f(v_1 + v_2) = f(v_1) + f(v_2)$, per ogni $v_1, v_2 \in V$;
2. $f(\lambda v) = \lambda f(v)$ per ogni $\lambda \in K$, $v \in V$.

La condizione 1. si esprime dicendo che f conserva la somma, o è compatibile con la somma, o anche è additiva; la condizione 2. che f conserva il prodotto esterno o è compatibile con il prodotto esterno, o ancora che è omogenea. Si dice anche che un'applicazione lineare è un'applicazione che conserva la struttura di spazio vettoriale.

Osservazione 9. Un'applicazione $f : V \to V'$ è lineare se e solo se per ogni scelta di $\lambda_1, \lambda_2 \in K$, $v_1, v_2 \in V$ si ha $f(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 f(v_1) + \lambda_2 f(v_2)$, ossia f conserva le combinazioni lineari. (Verifica per esercizio.)

Proposizione 6.2.2. *Sia $f : V \to V'$ un'applicazione lineare. Denotiamo con 0_V il vettore nullo di V e $0_{V'}$ il vettore nullo di V'. Si ha*

(i) $f(0_V) = 0_{V'}$;

(ii) $f(-v) = -f(v)$ *per ogni* $v \in V$.

Dimostrazione. (i) $f(0_V) = f(0 \cdot 0_V) =$ per la 2. $= 0 f(0_V) = 0_{V'}$, dove 0 denota lo zero di K.

(ii) $f(-v) = f((-1)(v)) =$ per la 2. $= (-1)f(v) = -f(v)$.

□

Come immediata conseguenza abbiamo che se $f(0_V) \neq 0_{V'}$ allora f non può essere lineare. In particolare un'applicazione lineare è costante se e solo se $f(v) = 0_{V'}$ per ogni $v \in V$: tale applicazione è detta **applicazione lineare nulla** ed esiste per ogni coppia di K-spazi vettoriali.

Esempio 6.2.3.

1. Consideriamo $\mathbb{R}$ con la sua struttura naturale di $\mathbb{R}$-spazio vettoriale. Sia $a \in \mathbb{R}$ un numero reale fissato. Consideriamo l'applicazione di moltiplicazione per a:

$$f_a : \mathbb{R} \xrightarrow{\cdot a} \mathbb{R}, \text{ tale che } f_a(x) = ax \text{ per ogni } x \in \mathbb{R}.$$

Si ha $f(\lambda x + \mu y) = a(\lambda x + \mu y) = \lambda(ax) + \mu(ay) = \lambda f_a(x) + \mu f_a(y)$, per ogni $x, y \in \mathbb{R}$, pertanto f_a è un'applicazione lineare.

2. L'applicazione $g : \mathbb{R} \to \mathbb{R}$ definita da $g(x) = 2x + 1$ non è lineare, in quanto $g(0) \neq 0$.

3. L'applicazione $h : \mathbb{R} \to \mathbb{R}$ definita da $h(x) = x^2$ non è lineare; infatti non è additiva: $h(1 + 1) = h(2) = 4 \neq h(1) + h(1) = 1 + 1 = 2$.

4. Sia $f : \mathbb{R}^2 \to \mathbb{R}^3$ definita da

$$f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3x_1 + 2x_2 \\ x_1 \\ 4x_1 + x_2 \end{pmatrix}.$$

E` facile verificare che f è lineare. Osserviamo che f si può rappresentare anche in forma matriciale mediante un prodotto righe per colonne di matrici come segue:

$$f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 1 & 0 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

5. L'applicazione $\varphi : \mathbb{R}^2 \to \mathbb{R}^3$ definita da

$$\varphi \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3x_1 + 1 \\ x_1 + x_2 \\ 4x_1 \end{pmatrix}$$

non è lineare, infatti $\varphi \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.

Gli esempi 1. e 4. fanno parte di una classe fondamentale di applicazioni lineari **associate a matrici** che ora definiremo.

6.3 L'applicazione lineare $L(A)$.

Definizione 6.3.1. Sia $A \in M(m \times n, K)$ una matrice $m \times n$ a entrate in K. Definiamo l'applicazione lineare associata alla matrice A

$$L(A) : K^n \to K^m$$

ponendo

$$L(A) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_j a_{1j}x_j \\ \vdots \\ \sum_j a_{mj}x_j \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix}.$$

Denotando con X il vettore colonna $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, scriveremo anche $L(A)(X) = AX$.

La linearità dell'applicazione $L(A)$ segue dalla Proposizione 6.1.4:

$$L(A)\Big(\lambda \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \mu \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}\Big) \overset{\text{per definizione}}{=} A\Big(\lambda \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \mu \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}\Big) =$$

$$\overset{\text{per 6.1.4 b)}}{=} A\Big(\lambda \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\Big) + A\Big(\mu \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}\Big) \overset{\text{per 6.1.4 d)}}{=} \lambda A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \mu A \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$\overset{\text{per definizione}}{=} \lambda L(A) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \mu L(A) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Osservazione 10. Vediamo chi sono i corrispondenti dei vettori della base canonica di K^n nell'applicazione lineare $L(A)$:

$$L(A)(e_1) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} = a^1 :$$

è la prima colonna di A. Similmente $L(A)(e_i) = a^i$, la colonna i-esima di A, per ogni indice i. Quindi i corrispondenti dei vettori della base canonica sono le colonne della matrice A.

6.4 Proprietà delle applicazioni lineari

Proposizione 6.4.1. *Sia $f : V \to V'$ un'applicazione lineare.*

(i) Se $W \subset V$ è un sottospazio vettoriale, allora $f(W) \subset V'$ è un sottospazio vettoriale;

(ii) se $W' \subset V'$ è un sottospazio vettoriale, allora $f^{-1}(W') \subset V$ è un sottospazio vettoriale.

Dimostrazione. (i) Per esercizio.

(ii) Siano $v_1, v_2 \in f^{-1}(W')$, dobbiamo verificare che $\lambda_1 v_1 + \lambda_2 v_2 \in f^{-1}(W')$, cioè che $f(\lambda_1 v_1 + \lambda_2 v_2) \in W'$. Ma $f(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 f(v_1) + \lambda_2 f(v_2)$ per la linearità di f. Per ipotesi $f(v_1) \in W'$ e $f(v_2) \in W'$; poichè W' è sottospazio di V' segue che anche $\lambda_1 f(v_1) + \lambda_2 f(v_2) \in W'$; ciò conclude la dimostrazione.

□

Dunque immagini e controimmagini di sottospazi vettoriali in un'applicazione lineare sono sottospazi vettoriali.

Corollario 6.4.2. *Se $f : V \to V'$ è lineare, l'immagine di f, $f(V)$, è un sottospazio vettoriale di V', detto* **sottospazio immagine** *di f. Lo si denota anche* $\operatorname{Im} f$.

La controimmagine del sottospazio nullo di V', $f^{-1}(0)$, è un sottospazio di V, detto **nucleo di** *f. Lo si denota* $\ker f$. *Tale notazione viene dal termine tedesco kern, che significa nucleo, nocciolo; analogamente in inglese kernel.*

Per definizione di applicazione suriettiva, l'applicazione lineare $f : V \to V'$ è suriettiva se e solo se $\operatorname{Im} f = V'$, cioè se l'immagine di f coincide con il codominio V'.

Il nucleo di f

$$\ker f = f^{-1}(0) = \{v \in V \mid f(v) = 0\}$$

dà invece informazioni sull'iniettività di f. Si ha:

Teorema 6.4.3. *Sia $f : V \to V'$ un'applicazione lineare di K-spazi vettoriali; f è iniettiva se e solo se* $\ker f = (0)$, *il sottospazio nullo di V.*

Dimostrazione. Ricordiamo che, essendo f lineare, $f(0_V) = 0_{V'}$. Se f è iniettiva, 0_V è l'unico vettore la cui immagine in f è $0_{V'}$, e ciò significa che $\ker f$, che è uguale a $f^{-1}(0_{V'})$, contiene solo il vettore nullo.

Viceversa supponiamo che $\ker f = (0)$. Siano $v_1, v_2 \in V$ due vettori tali che $f(v_1) = f(v_2)$; dobbiamo dimostrare che $v_1 = v_2$. Da $f(v_1) = f(v_2)$ segue $f(v_1) - f(v_2) = 0$; ma f è lineare, perciò $0 = f(v_1) - f(v_2) = f(v_1 - v_2)$. Ne segue che $v_1 - v_2 \in \ker f$, che è (0) per ipotesi, dunque $v_1 - v_2 = 0$, ossia $v_1 = v_2$. Concludiamo che f è iniettiva. $\square$

Il Teorema 6.4.3 è un criterio molto utile per dimostrare l'iniettività di un'applicazione lineare.

6.5 Spazi Hom

Siano V, V' K-spazi vettoriali. Abbiamo visto che c'è sempre almeno un'applicazione lineare da V a V', l'applicazione nulla. Useremo la notazione $\mathrm{Hom}(V, V')$ per indicare l'insieme non vuoto delle applicazioni lineari di V in V':

$$\mathrm{Hom}(V, V') = \{f : V \to V' \mid f \text{ è lineare}\}.$$

La notazione deriva dal termine "omomorfismo" che è un sinonimo di applicazione lineare.

Introduciamo in $\mathrm{Hom}(V, V')$ due operazioni, di somma e di prodotto esterno con operatori in K, definite punto per punto. Con queste operazioni $\mathrm{Hom}(V, V')$ acquista struttura di K-spazio vettoriale. Date due applicazioni lineari $f, g \in \mathrm{Hom}(V, V')$, e dato $\lambda \in K$, definiamo:

- $f + g$ è l'applicazione di V in V' che tale che, se $v \in V$ è un vettore del dominio, $(f + g)(v) = f(v) + g(v)$;
- λf è l'applicazione di V in V' che tale che, se $v \in V$ è un vettore del dominio, $(\lambda f)(v) = \lambda f(v)$.

Proposizione 6.5.1. $\mathrm{Hom}(V, V')$ *è un* $K-$*spazio vettoriale.*

Dimostrazione. Dobbiamo innanzitutto dimostrare che la somma e il prodotto per uno scalare sono operazioni interne in $Hom(V, V')$, ossia che, se f, g sono lineari e $\lambda \in K$, allora $f + g$ e λf sono lineari. Si ha: $(f + g)(av + bw) \overset{\text{def}}{=} f(av + bw) + g(av + bw) \overset{\text{linearità}}{=} af(v) + bf(w) + ag(v) + bg(w) = a(f(v) + g(v)) + b(f(w) + g(w)) \overset{\text{def}}{=} a(f + g)(v) + b(f + g)(w)$. Similmente $(\lambda f)(av + bw) = a(\lambda f)(v) + b(\lambda f)(w)$ (esercizio).

Osserviamo poi che l'applicazione nulla è elemento neutro per la somma in $\mathrm{Hom}(V, V')$, e che l'opposto di f è l'applicazione lineare $-f = (-1)f$ tale che $(-f)(v) = -f(v)$, per ogni $v \in V$. Si dimostrano poi in maniera simile gli altri assiomi di spazio vettoriale. $\square$

Un caso particolare molto importante è lo spazio vettoriale duale.

Definizione 6.5.2. Sia V un $K-$spazio vettoriale. Lo **spazio vettoriale duale** di V, denotato V^*, è $\mathrm{Hom}(V, K) = \{\varphi : V \to K \mid \varphi \text{ lineare}\}$. Gli elementi di V^* sono chiamati **forme lineari**, o funzionali, su V.

Introduciamo la terminologia usata per denotare vari tipi di applicazioni lineari. Sia $f : V \to V'$ un'applicazione lineare. Diremo che f è un

- epimorfismo se è suriettiva; si dice anche che f è un'applicazione lineare di V **su** V';
- monomorfismo se è iniettiva;
- isomorfismo se è sia iniettiva sia suriettiva;
- endomorfismo se $V = V'$;
- automorfismo se $V = V'$ e f è un isomorfismo.

Nel seguito avremo modo di studiare in maniera approfondita la nozione di endomorfismo.

Vediamo ora una proprietà significativa degli isomorfismi. Sia $f : V \to V'$ un isomorfismo. Ricordiamo che, essendo f biiettiva, esiste l'applicazione inversa $f^{-1} : V' \to V$ tale che, preso $v' \in V'$, si ha: $f^{-1}(v') = v$ se e solo se $f(v) = v'$.

Proposizione 6.5.3. *Se f è un isomorfismo, anche f^{-1} è lineare, e dunque anche f^{-1} è un isomorfismo.*

Dimostrazione. Per semplificare la scrittura poniamo $f^{-1} = g$. Dobbiamo verificare che, se $v', w' \in V'$, $\lambda, \mu \in K$, si ha $g(\lambda v' + \mu w') = \lambda g(v') + \mu g(w')$. Poichè f è biiettiva, esistono unici $v \in V$ e $w \in V$ tali che $f(v) = v'$ e $f(w) = w'$, e dunque $g(v') = v$, $g(w') = w$. Allora

$$g(\lambda v' + \mu w') = g(\lambda f(v) + \mu f(w)) \overset{f \text{ lineare}}{=} g(f(\lambda v + \mu w)) \overset{g=f^{-1}}{=} \lambda v + \mu w = \lambda g(v') + \mu g(w').$$

$\square$

In un isomorfismo due spazi vettoriali si corrispondono mediante una biiezione che conserva anche la loro struttura; se esiste un isomorfismo di V su V', si dice che V e V' sono **isomorfi** e si scrive $V \simeq V'$. Spesso spazi isomorfi vengono addirittura identificati tramite l'isomorfismo f, cioè vengono trattati come se fossero lo stesso spazio.

Capitolo 7

Teorema della dimensione per applicazioni lineari e sue conseguenze

7.1 Teorema della dimensione

Teorema 7.1.1. *Sia $f : V \to V'$ un'applicazione lineare fra $K-$spazi vettoriali, supponiamo che V abbia dimensione finita. Allora anche* $\operatorname{Im} f$ *ha dimensione finita e si ha*

$$\dim V = \dim(\ker f) + \dim(\operatorname{Im} f). \tag{7.1}$$

Dimostrazione. Poichè V ha dimensione finita, chiamiamola n, anche il suo sottospazio $\ker f$ ha dimensione finita, chiamiamola k. Prendiamo dunque una base di $\ker f$: $(u_1, \dots, u_k)$, e completiamola a una base $\mathcal{B}$ di V: $(u_1, \dots, u_k, v_{k+1}, \dots, v_n)$. Consideriamo i vettori di V' $f(u_1), \dots, f(u_k), f(v_{k+1}), \dots, f(v_n)$. Osserviamo che $f(u_1) = \dots = f(u_k) = 0$ in quanto $u_1, \dots, u_k \in \ker f$. Avremo finito se potremo dimostrare che i rimanenti $f(v_{k+1}), \dots, f(v_n)$ formano una base di $\operatorname{Im} f$.

a) $f(v_{k+1}), \dots, f(v_n)$ sono un sistema di generatori di $\operatorname{Im} f$.

Sia v' un qualunque elemento di $\operatorname{Im} f$: dunque esiste $v \in V$ tale che $v' = f(v)$; v è combinazione lineare dei vettori di $\mathcal{B}$ della forma $v = \lambda_1 u_1 + \dots + \lambda_k u_k + \lambda_{k+1} v_{k+1} + \dots + \lambda_n v_n$. Allora

$$\begin{aligned} v' = f(v) &= f(\lambda_1 u_1 + \dots + \lambda_k u_k + \lambda_{k+1} v_{k+1} + \dots + \lambda_n v_n) \overset{\text{linearità}}{=} \\ &= \lambda_1 f(u_1) + \dots + \lambda_k f(u_k) + \lambda_{k+1} f(v_{k+1}) + \dots + \lambda_n f(v_n) = \\ &= \lambda_{k+1} f(v_{k+1}) + \dots + \lambda_n f(v_n) \end{aligned}$$

poichè nei primi k termini $f(u_1) = \dots = f(u_k) = 0$.

b) $f(v_{k+1}), \dots, f(v_n)$ sono linearmente indipendenti.

Se $0 = \mu_{k+1} f(v_{k+1}) + \dots + \mu_n f(v_n) \overset{\text{linearità}}{=} f(\mu_{k+1} v_{k+1} + \dots + \mu_n v_n)$, si ha che $\mu_{k+1} v_{k+1} + \dots + \mu_n v_n \in \ker f$, e dunque è una combinazione lineare dei vettori $u_1, \dots, u_k$: $\mu_{k+1} v_{k+1} + \dots + \mu_n v_n = \mu_1 u_1 + \dots + \mu_k u_k$ per opportuni scalari $\mu_1, \dots, \mu_k$. Allora

$$\mu_1 u_1 + \dots + \mu_k u_k - \mu_{k+1} v_{k+1} - \dots - \mu_n v_n = 0,$$

da cui segue che $\mu_1 = \cdots = \mu_k = \cdots = \mu_n = 0$, in quanto $u_1, \ldots, u_k, v_{k+1}, \ldots, v_n$ formano la base $\mathcal{B}$ e sono perciò linearmente indipendenti. In particolare $\mu_{k+1} = \cdots = \mu_n = 0$ □

Definizione 7.1.2. La dimensione di $\operatorname{Im} f$ è detta **rango di** f, e la denoteremo $\operatorname{rg}(f)$.

La seguente utile proposizione estende quanto abbiamo dimostrato nella parte a) del Teorema 7.1.1:

Proposizione 7.1.3. *Sia* $\{v_i\}_{i\in I}$ *un arbitrario sistema di generatori di uno spazio vettoriale* V*, non necessariamente finitamente generato. Sia* $f : V \to V'$ *un'applicazione lineare. Allora* $\operatorname{Im} f$ *è generata dai vettori* $\{f(v_i)\}_{i\in I}$*.*

Dimostrazione. Sia $v' \in \operatorname{Im} f$, allora esiste $v \in V$ tale che $v' = f(v)$; e v può essere scritto come combinazione lineare di una sottofamiglia finita dei vettori $\{v_i\}$: $v = c_1 v_{i_1} + \cdots + c_k v_{i_k}$. Allora $v' = f(c_1 v_{i_1} + \cdots + c_k v_{i_k})$, che, per la linearità di f, è uguale a $c_1 f(v_{i_1}) + \cdots + c_k f(v_{i_k})$, e dunque appartiene al sottospazio generato dai vettori $f(v_i)$, per $i \in I$. Viceversa, sempre per la linearità di f, ogni combinazione lineare di (un numero finito di) vettori $f(v_i)$ appartiene all'immagine di f. □

7.2 Conseguenze del Teorema della dimensione

7.2.1 Applicazioni lineari iniettive e suriettive

Se $f : V \to V'$ è un'applicazione lineare **iniettiva** tra spazi vettoriali di dimensione finita, il suo nucleo $\ker f$ è il sottospazio nullo del dominio. Il Teorema 7.1.1 allora ci dice che $\dim V = \dim \operatorname{Im} f$; ma $\operatorname{Im} f$ è un sottospazio di V', quindi $\dim \operatorname{Im} f \leq \dim V'$, e otteniamo che $\dim V \leq \dim V'$.

Se invece f è suriettiva, abbiamo $\dim \operatorname{Im} f = \dim V'$, e quindi dalla (7.1) otteniamo $\dim V - \dim V' = \dim \ker f \geq 0$, e si ha dunque $\dim V \geq \dim V'$.

La prossima Proposizione analizza il caso in cui V e V' hanno la stessa dimensione.

Proposizione 7.2.1. *Sia* $f : V \to V'$ *un'applicazione lineare fra due spazi vettoriali della stessa dimensione finita:* $\dim V = \dim V'$*. Allora* f *è iniettiva se e solo se è suriettiva se e solo se è un isomorfismo.*

Dimostrazione. Infatti: per il Teorema 6.4.3 f è iniettiva se e solo $\ker f = (0)$ se e solo $\dim \ker f = 0$. Ma per il Teorema 7.1.1 si ha la relazione $\dim V = \dim(\ker f) + \dim(\operatorname{Im} f)$. Quindi f è iniettiva se e solo se $\dim V = \dim(\operatorname{Im} f)$. D'altra parte f è suriettiva se e solo se $\operatorname{Im} f = V'$ se e solo se $\dim \operatorname{Im} f = \dim V' \overset{\text{per ipotesi}}{=} \dim V$. Questo conclude la dimostrazione. □

Se f è un'applicazione lineare iniettiva, allora $\dim V = \dim \operatorname{Im} f$; possiamo restringere il codominio di f e considerare f come un'applicazione da V a $\operatorname{Im} f$. Ora la nuova applicazione $f : V \to \operatorname{Im} f$ verifica le ipotesi della Proposizione 7.2.1, quindi è un isomorfismo. In particolare V è isomorfo a $\operatorname{Im} f$: $V \simeq \operatorname{Im} f$.

7.2.2 Dimensione di V/W

Sia ora $W \subset V$ un sottospazio vettoriale di uno spazio di dimensione finita V. Consideriamo lo spazio vettoriale quoziente V/W. Definiamo un'applicazione $\pi : V \to V/W$ ponendo $\pi(v) = [v] = v + W$; chiaramente π è suriettiva.

Verifichiamo che π è lineare: $\pi(\lambda v + \mu w) = [\lambda v + \mu w] = \lambda[v] + \mu[w] = \lambda\pi(v) + \mu\pi(w)$: abbiamo usato il fatto che le operazioni in V/W sono indotte dalle operazioni in V. L'applicazione π è detta **epimorfismo canonico**.

Osserviamo che $\ker \pi = \{v \in V \mid \pi(v) = [v] = 0 \text{ in } V/W\}$. Ma lo zero di V/W è $[0_V]$, quindi $v \in \ker \pi$ se e solo se v è equivalente a 0_V, cioè $v \in W$. Quindi $\ker \pi = W$. Quindi il Teorema della dimensione ci dà: $\dim V = \dim W + \dim \operatorname{Im} \pi$; ma π è suriettiva, dunque $\dim \operatorname{Im} \pi = \dim V/W$. Possiamo dunque concludere che $\dim V/W = \dim V - \dim W$.

7.2.3 Rango per righe e rango per colonne coincidono

Sia $A \in M(m \times n, K)$. Consideriamo l'applicazione lineare $L(A) : K^n \to K^m$ introdotta nella Sezione 6.3. Il Teorema della dimensione ci dà:

$$\dim K^n = n = \dim \ker L(A) + \dim \operatorname{Im} L(A). \tag{7.2}$$

Analizziamo $\ker L(A)$ e $\operatorname{Im} L(A)$.

$$\ker L(A) = \{x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n \mid Ax = 0\} =: W;$$

W è lo spazio delle soluzioni del sistema lineare omogeneo avente A come matrice dei coefficienti.

D'altra parte, per la Proposizione 7.1.3, $\operatorname{Im} L(A)$ è il sottospazio di K^m generato dalle immagini in $L(A)$ dei vettori della base canonica $\mathcal{C}$ di K^n, che per l'Osservazione 10 sono le colonne di A:

$$\begin{aligned} &L(A)(e_1) = a^1, \quad \text{la prima colonna di} A \\ &\vdots \\ &L(A)(e_n) = a^n, \quad \text{l'ultima colonna di } A \end{aligned}$$

Dunque $\operatorname{Im} L(A) =< a^1, \dots, a^n >$ è lo spazio delle colonne della matrice A.

Quindi il Teorema della dimensione, e in particolare la relazione (7.2), ci dice che

$$n = \dim W + \text{ rango per colonne di } A.$$

D'altra parte abbiamo visto nel Teorema 5.6.2 che $\dim W = n -$ rango per righe di A.

La discussione precedente ci permette di concludere che i due ranghi sono uguali, enunciando il seguente Teorema.

Teorema 7.2.2. *Sia $A \in M(m \times n, K)$ una matrice. Il rango per righe di A coincide con il rango per colonne di A, ossia lo spazio delle righe e le spazio per colonne di A hanno la stessa dimensione.*

Quindi d'ora in poi parleremo semplicemente di **rango di una matrice** A, e lo denoteremo con il simbolo $\operatorname{rg} A$.

7.2.4 Sistemi lineari di equazioni II

Sia $Ax = 0$ un sistema lineare omogeneo di m equazioni in n incognite scritto in forma matriciale, dove $A \in M(m \times n, K)$. Come già osservato, lo spazio W delle soluzioni di tale sistema non è altro che il nucleo $\ker L(A)$ dell'applicazione lineare $L(A) : K^n \to K^m$ associata alla matrice dei coefficienti A del sistema. Il Teorema della dimensione 7.1.1 dunque ci permette di ritrovare la prima parte del Teorema 5.6.2, che afferma che $\dim W = n - r$, dove r è il rango di A, ossia la dimensione dell'immagine di $L(A)$.

Se abbiamo invece un sistema lineare non omogeneo $Ax = b$, con $b \in K^m$, osserviamo che è compatibile precisamente quando $b \in \operatorname{Im} L(A)$. In tal caso l'insieme delle sue soluzioni è $L(A)^{-1}(b)$. Se z è una soluzione particolare, $Az = b$, come visto nel Capitolo 5.5.1 ogni altra soluzione è del tipo $z + u$ con $u \in W$; dunque lo spazio affine delle soluzioni $L(A)^{-1}(b)$ può esser interpretato come un elemento dello spazio vettoriale quoziente $K^n / \ker L(A)$.

Capitolo 8

Teorema di determinazione di un'applicazione lineare e sue conseguenze

8.1 Teorema di determinazione di un'applicazione lineare

Teorema 8.1.1. *Siano V, W due $K-$spazi vettoriali, supponiamo che V abbia dimensione finita n. Sia $\mathcal{B} = (v_1, \ldots, v_n)$ una base di V e siano $w_1, \ldots, w_n$ vettori di W del tutto arbitrari. Allora esiste una ed una sola applicazione* **lineare** *$f : V \to W$ tale che $f(v_1) = w_1, \ldots, f(v_n) = w_n$.*

Vedremo che l'ipotesi che V abbia dimensione finita potrà essere rimossa; facciamo questa ipotesi solo per semplificare l'esposizione.

Si tratta di un teorema di esistenza e unicità: come si fa spesso per dimostrare teoremi di questo tipo, la strategia consiste nel dimostrare dapprima l'unicità, supponendo che una tale applicazione lineare esista; questo consente di capire come l'applicazione cercata deve operare. Poi la dimostrazione dell'esistenza si riduce a verificare che l'applicazione trovata verifica le condizioni richieste, ossia in questo caso che è lineare e manda i vettori v_i ordinatamente nei vettori w_i.

Dimostrazione. **Unicità**: supponiamo che f esista e vediamo come opera sui vettori di V. Sia dunque $v \in V$, allora c'è una e una sola espressione di v come combinazione lineare dei vettori di $\mathcal{B}$: $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$, $\lambda_1, \ldots, \lambda_n$ sono le coordinate di v rispetto a $\mathcal{B}$ e sono univocamente determinate da v. Allora

$$f(v) = f(\lambda_1 v_1 + \cdots + \lambda_n v_n) \overset{\text{linearità}}{=} \lambda_1 f(v_1) + \cdots + \lambda_n f(v_n) = \lambda_1 w_1 + \cdots + \lambda_n w_n. \quad (8.1)$$

Dunque $f(v)$ è completamente determinato e questo prova l'unicità.

Esistenza: prendiamo la (8.1) come definizione di f, ossia definiamo $f(v)$ come $\lambda_1 w_1 + \cdots + \lambda_n w_n$ dove $\lambda_1, \ldots, \lambda_n$ sono le coordinate di v rispetto a $\mathcal{B}$. Così f è ben definita; dobbiamo dimostrare che è lineare e che $f(v_i) = w_i$ per ogni indice $i = 1, \ldots, n$.

Siano dunque v, v' due vettori di V, allora $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$, $v' = \mu_1 v_1 + \cdots + \mu_n v_n$, e $v + v' = \lambda_1 v_1 + \cdots + \lambda_n v_n + \mu_1 v_1 + \cdots + \mu_n v_n = (\lambda_1 + \mu_1)v_1 + \cdots + (\lambda_n + \mu_n)v_n$; quindi le coordinate di $v + v'$ rispetto a $\mathcal{B}$ sono $\lambda_1 + \mu_1, \dots, \lambda_n + \mu_n$. Perciò, per definizione di f, $f(v) = \lambda_1 w_1 + \dots + \lambda_n w_n$, $f(v') = \mu_1 w_1 + \dots + \mu_n w_n$, $f(v + v') = (\lambda_1 + \mu_1)w_1 + \dots + (\lambda_n + \mu_n)w_n$; ma questo è uguale a $\lambda_1 w_1 + \dots + \lambda_n w_n + \mu_1 w_1 + \dots + \mu_n w_n = f(v) + f(v')$, il che prova l'additività di f. Analogamente proviamo l'omogeneità di f: se $c \in K$, $cv = c(\lambda_1 v_1 + \dots + \lambda_n v_n) = (c\lambda_1)v_1 + \dots + (c\lambda_n)v_n$, quindi le coordinate di cv rispetto a $\mathcal{B}$ sono $c\lambda_1, \dots, c\lambda_n$ e $f(cv) = (c\lambda_1)w_1 + \dots + (c\lambda_n)w_n = c(\lambda_1 w_1 + \dots + \lambda_n w_n) = cf(v)$. Quindi f è lineare.

Infine osserviamo che per ogni i $v_i = 0v_1 + \dots + 1v_i + \dots + 0v_n$, ossia le sue coordinate rispetto a $\mathcal{B}$ sono $(0, \dots, 0, 1, 0, \dots, 0)$; quindi per definizione di f si ha $f(v_i) = 0w_1 + \dots + 1w_i + \dots + 0w_n = w_i$. Questo conclude la dimostrazione dell'esistenza.

$\square$

Osservazione 11. Il Teorema 8.1.1 di determinazione di un'applicazione lineare vale anche se la dimensione di V non è finita. Precisamente, se $\{v_i\}_{i \in I}$ è una base di V, e $\{w_i\}_{i \in I}$ sono una famiglia di vettori di W indiciati sullo stesso insieme d'indici I, esiste una e una sola applicazione lineare $f : V \to W$ tale che $f(v_i) = w_i$ per ogni indice $i \in I$. La dimostrazione è del tutto analoga alla precedente. Consiglio di scriverla dettagliatamente per esercizio.

8.2 Spazi vettoriali della stessa dimensione

Iniziamo questa sezione con una semplice proposizione.

Proposizione 8.2.1. *Ogni composizione di applicazioni lineari è lineare. Precisamente: siano $f : V \to W$, $g : W \to U$ due applicazione lineari di K spazi vettoriali. Allora $g \circ f : V \to U$ è lineare.*

Dimostrazione. $(g \circ f)(\lambda v + \mu v') = g(f(\lambda v + \mu v') \overset{\text{linearità di } f}{=} g(\lambda f(v) + \mu f(v')) \overset{\text{linearità di } g}{=} \lambda g(f(v)) + \mu g(f(v')) = \lambda(g \circ f)(v) + \mu(g \circ f)(v')$. $\square$

Supponiamo ora che V, W siano K-spazi vettoriali della **stessa dimensione finita** n. Siano $\mathcal{B} = (v_1, \dots, v_n)$, $\mathcal{B}' = (w_1, \dots, w_n)$ basi rispettivamente di V e di W. Allora, per il Teorema 8.1.1, esiste una e una sola $f : V \to W$ lineare ed esiste una e una sola $g : W \to V$ lineare, tali che $f(v_i) = w_i$ e $g(w_i) = v_i$ per ogni $i = 1, \dots, n$. Possiamo allora considerare le applicazioni lineari composte: $g \circ f : V \to V$ e $f \circ g : W \to W$. Osserviamo che per ogni i

$$(g \circ f)(v_i) \overset{\text{definizione di } f}{=} g(w_i) \overset{\text{definizione di } g}{=} v_i;$$

allora, detta id_V l'applicazione identica di V, si ha $(g \circ f)(v_i) = \mathrm{id}_V(v_i)$ per ogni vettore v_i della base $\mathcal{B}$. Per la parte "unicità" del Teorema 8.1.1, visto che $g \circ f$ e id_V sono applicazioni lineari che operano allo stesso modo sui vettori di una base del dominio, possiamo concludere che $g \circ f = \mathrm{id}_V$. Nello stesso modo si ottiene che $f \circ g = \mathrm{id}_W$.

Abbiamo perciò ottenuto che f e g sono applicazioni lineari una inversa dell'altra, e dunque sono isomorfismi.

Da ciò si deduce il seguente importante fatto.

Teorema 8.2.2. *Se V, W sono K-spazi vettoriali con* $\dim V = \dim W = n$, *allora V e W sono isomorfi.*

Dimostrazione. Per costruire un isomorfismo di V in W basta fissare due basi, una in V e l'altra in W, e procedere come sopra. $\square$

In particolare:

Corollario 8.2.3. *Ogni K-spazio vettoriale di dimensione n è isomorfo a K^n.*

Gli isomorfi che abbiamo costruito sono tutti **non canonici** in quanto dipendono dalla scelta di basi nel dominio e nel codominio.

Sia $\dim V = n$ e fissiamo una sua base $\mathcal{B} = (v_1, \dots, v_n)$. Consideriamo in K^n la base canonica $\mathcal{C} = (e_1, \dots, e_n)$. Denoteremo $\kappa_{\mathcal{B}} : V \to K^n$ l'unico isomorfismo tale che $\kappa_B(v_i) = e_i$ per ogni i. Come opera $\kappa_{\mathcal{B}}$ su un generico vettore $v \in V$? Scriviamo v come combinazione lineare dei vettori di $\mathcal{B}$: $v = a_1v_1 + \dots + a_nv_n$, $a_1, \dots, a_n$ sono le coordinate di v rispetto a $\mathcal{B}$. Allora per definizione di $\kappa_{\mathcal{B}}$ si ha: $\kappa_B(v) = \kappa_{\mathcal{B}}(a_1v_1 + \dots + a_nv_n) = a_1e_1 + \dots + a_ne_n = (a_1, \dots, a_n)$: $\kappa_{\mathcal{B}}$ **associa a v la n-upla delle sue coordinate rispetto a $\mathcal{B}$.** L'isomorfismo inverso $\kappa_{\mathcal{B}}^{-1} : K^n \to V$ associa invece a una n-upla $(a_1, \dots, a_n)$ il vettore $a_1v_1 + \dots + a_nv_n$, la combinazione lineare dei vettori di $\mathcal{B}$ con coefficienti $a_1, \dots, a_n$.

Esercizi 6. Siano V, W due $K-$spazi vettoriali, $\mathcal{B} = (v_1, \dots, v_n)$ una base di V, $w_1, \dots, w_n$ vettori di W, $f : V \to W$ l'unica applicazione lineare tale che $f(v_1) = w_1, \dots, f(v_n) = w_n$ come dimostrato nel Teorema 8.1.1. Allora:

- f è suriettiva se e solo se $w_1, \dots, w_n$ sono un sistema di generatori di W;
- f è iniettiva se e solo se $w_1, \dots, w_n$ sono linearmente indipendenti.

Dimostrazione per esercizio.

8.3 Esempi

1. $\mathbb{C}$-endomorfismi di $\mathbb{C}$.

Sia $\mathbb{C}$ il campo dei numeri complessi. Abbiamo osservato che $\mathbb{C}$ può essere pensato sia come $\mathbb{C}-$ sia come $\mathbb{R}-$ spazio vettoriale. Nel primo caso la sua dimensione è 1, nel secondo è 2, infatti nell'Esempio 3.1.5, 4. abbiamo osservato che $(1, i)$ è una base di $\mathbb{C}$ come $\mathbb{R}$-spazio vettoriale. Dunque, su $\mathbb{R}$, $\mathbb{C} \simeq \mathbb{R}^2$. Nell'isomorfismo $\kappa_B : \mathbb{C} \to \mathbb{R}^2$ associato a questa base il numero complesso $a + bi$ corrisponde alla coppia (a, b). In altre parole questo isomorfismo permette di "identificare" $\mathbb{C}$ con i punti del piano reale, detto in questo caso "piano di Argand–Gauss".

A un numero complesso $z = a + bi$ possiamo associare il suo modulo $\rho = | z | = \sqrt{a^2 + b^2}$, che è un numero reale ≥ 0. Se $| z | = 1$, si ha $a^2 + b^2 = 1$, e quindi si può scrivere

$$\begin{cases} a = & \cos\theta \\ b = & \sin\theta \end{cases}$$

dove θ è determinato a meno di multipli interi di 2π. Quindi $z = \cos\theta + \sin\theta i$. Il punto corrispondente a z in $\mathbb{R}^2$ sta sulla circonferenza unitaria. Se $z \neq 0$ è un qualunque numero complesso, lo si può scrivere $z = \rho(z/\rho) = \rho z'$ dove z' ha modulo 1. E quindi $z = \rho(\cos\theta + \sin\theta i)$: questa è la cosiddetta notazione trigonometrica dei numeri complessi.

Vogliamo ora dare un'interpretazione geometrica ai $\mathbb{C}$-endomorfismi di $\mathbb{C}$, cioè alle applicazioni $\mathbb{C}$-lineari $\mathbb{C} \to \mathbb{C}$. In maniera analoga a quanto visto nell'Esempio 6.2.3 1., un tale endomorfismo è del tipo $\cdot z : \mathbb{C} \to \mathbb{C}$, la moltiplicazione per un numero complesso z. Supponiamo che $|\, z \,| = 1$ e scriviamo $z = \cos\theta + \sin\theta i$. Si ha:

$$\cdot z : a + bi \to (\cos\theta + \sin\theta i)(a + bi) = (a\cos\theta - b\sin\theta) + (a\sin\theta + b\cos\theta)i.$$

Identifichiamo ora $\mathbb{C}$ con $\mathbb{R}^2$:

$$\cdot z : \begin{pmatrix} a \\ b \end{pmatrix} \to \begin{pmatrix} a\cos\theta - b\sin\theta \\ a\sin\theta + b\cos\theta \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

dunque $\cdot z = L(A)$, dove $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$: oltre a essere $\mathbb{C}$-lineare è anche $\mathbb{R}$-lineare. Se scriviamo anche $a + bi$ in forma trigonometrica, come $\rho(\cos\alpha + \sin\alpha i)$, otteniamo che

$$L(A) \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \rho\cos\alpha \\ \rho\sin\alpha \end{pmatrix} = \rho \begin{pmatrix} \cos(\alpha + \theta) \\ \sin(\alpha + \theta) \end{pmatrix}.$$

Concludiamo dunque che $\cdot z = L(A)$ è la rotazione di angolo θ.

2. **Lo spazio vettoriale duale.**

Sia V un K-spazio vettoriale di dimensione finita n e $\mathcal{B} = (v_1, \ldots, v_n)$ una sua base. Consideriamo lo spazio vettoriale duale $V^* = \{f : V \to K \mid f \text{ lineare}\}$ i cui elementi sono le **forme lineari** su V. Per ogni $i = 1, \ldots, n$ definiamo la forma lineare $v_i^* : V \to K$ ponendo $v_i^*(v_j) = \delta_{ij}$, il simbolo di Kronecker, ossia

$$v_i^*(v_j) = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

Per il Teorema di determinazione di un'applicazione lineare 8.1.1, v_i^* lineare risulta ben definita. Se $v = x_1v_1 + \cdots + x_nv_n$, si ha $v_i^*(v) = x_1\delta_{i1} + \cdots + x_n\delta_{ni} = x_i$: è la i-esima coordinata di v rispetto a $\mathcal{B}$. Per questo motivo le funzioni $v_1^*, \ldots, v_n^*$ sono dette **funzioni coordinate rispetto alla base $\mathcal{B}$**.

Teorema 8.3.1 (Base duale)**.** *Le forme lineari $v_1^*, \ldots, v_n^*$ formano una base di V^*, detta* **base duale di $\mathcal{B}$** *e denotata $\mathcal{B}^*$.*

Dimostrazione. Proviamo che $v_1^*, \ldots, v_n^*$ sono linearmente indipendenti. Supponiamo che $\lambda_1v_1^* + \cdots + \lambda_nv_n^* = 0$: ciò significa che per ogni $v \in V$ si ha $(\lambda_1v_1^* + \cdots + \lambda_nv_n^*)(v) = 0$. Questo vale in particolare per ogni vettore di $\mathcal{B}$, dunque per ogni i si ha

$$0 = (\lambda_1v_1^* + \cdots + \lambda_nv_n^*)(v_i) = \text{usando la definizione delle operazioni in } V^*$$

$$= \lambda_1v_1^*(v_i) + \cdots + \lambda_nv_n^*(v_i) = \lambda_1\delta_{1i} + \cdots + \lambda_n\delta_{ni} = \lambda_i.$$

Quindi $\lambda_1 = \cdots = \lambda_n = 0$.

Proviamo ora che $v_1^*, \ldots, v_n^*$ generano V^*. Sia $f : V \to K$ una forma lineare; il nostro scopo è trovare degli scalari $c_1, \ldots, c_n$ tali che $f = c_1v_1^* + \cdots + c_nv_n^*$. Cerchiamo di capire quali possono essere. Se esistono scalari per cui una tale uguaglianza vale, le due applicazioni

f e $c_1v_1^*+\cdots+c_nv_n^*$ devono coincidere su ogni vettore di V, e in particolare sui vettori di $\mathcal{B}$, cioè dev'essere $f(v_1) = (c_1v_1^*+\cdots+c_nv_n^*)(v_1) = c_1, \ldots, f(v_n) = (c_1v_1^*+\cdots+c_nv_n^*)(v_n) = c_n$. Quindi l'unica possibilità è che $c_1 = f(v_1), \ldots, c_n = f(v_n)$, ossia che $f = f(v_1)v_1^* + \cdots + f(v_n)v_n^*$. E in effetti questa uguaglianza di forme lineari è vera per il Teorema 8.1.1, poichè, per ogni i, $f(v_i) = (f(v_1)v_1^* + \cdots + f(v_n)v_n^*)(v_i)$: assumono lo stesso valore sui vettori di $\mathcal{B}$. □

Osservazione 12. Abbiamo dunque dimostrato che $f = f(v_1)v_1^* + \cdots + f(v_n)v_n^*$. Osserviamo che vale anche un'uguaglianza "duale": per ogni vettore $v \in V$ si ha $v = v_1^*(v)v_1 + \cdots + v_n^*(v)v_n$. In altre parole, data la base $\mathcal{B}$ di V e la base duale $\mathcal{B}^*$ di V^*, le coordinate di f rispetto a $\mathcal{B}^*$ sono $f(v_1), \ldots, f(v_n)$, i valori assunti da f sui vettori di $\mathcal{B}$, mentre le coordinate di v rispetto a $\mathcal{B}$ sono $v_1^*(v), \ldots v_n^*(v)$, i valori assunti su v dai "vettori" di $\mathcal{B}^*$.

Capitolo 9

Matrice di un'applicazione lineare rispetto a basi fissate

9.1 Associare a un'applicazione lineare una matrice

Nella Sezione 6.3 abbiamo visto come ad ogni matrice $A \in M(m \times n, K)$ si può associare in maniera naturale un'applicazione lineare $L(A) : K^n \to K^m$. Ora vogliamo associare una matrice a entrate in K ad ogni applicazione lineare fra K-spazi vettoriali di dimensione finita. Per farlo dobbiamo preliminarmente fissare delle basi del dominio e del codominio, in modo da lavorare "in coordinate."

Sia $f : V \to W$ un'applicazione lineare fra gli spazi vettoriali V, W di dimensioni rispettivamente n e m. Fissiamo basi $\mathcal{B} = (v_1, \ldots, v_n)$ di V e $\mathcal{B}' = (w_1, \ldots, w_m)$ di W. Consideriamo gli n vettori corrispondenti in f ai vettori di $\mathcal{B}$: $f(v_1), \ldots, f(v_n) \in W$; ognuno si scrive in maniera unica come combinazione lineare di $w_1, \ldots, w_m$, nella forma

$$\begin{aligned} f(v_1) &= a_{11}w_1 + a_{21}w_2 + \cdots + a_{m1}w_m \\ f(v_2) &= a_{12}w_1 + a_{22}w_2 + \cdots + a_{m2}w_m \\ &\vdots \\ f(v_n) &= a_{1n}w_1 + a_{2n}w_2 + \cdots + a_{mn}w_m \end{aligned} \tag{9.1}$$

cioè $f(v_j) = \sum_{i=1}^m a_{ij}w_i$ per ogni $j = 1, \ldots, n$.

Definizione 9.1.1. La **matrice di f rispetto alle basi** $\mathcal{B}, \mathcal{B}'$ è la seguente matrice $m \times n$:

$$M_{\mathcal{B}}^{\mathcal{B}'}(f) = (a_{ij})_{\substack{1 \le i \le m \\ 1 \le j \le n}} = \begin{array}{c} \begin{array}{cccc} f(v_1) & f(v_2) & \ldots & f(v_n) \\ \downarrow & \downarrow & & \downarrow \end{array} \\ \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \end{array}.$$

Nelle colonne si trovano le coordinate rispetto a $\mathcal{B}'$ delle immagini dei vettori di $\mathcal{B}$.

Allora se $v = x_1v_1 + \cdots + x_nv_n \in V$, per la linearità di f si ha

$$\begin{aligned} f(v) &= x_1f(v_1) + \cdots + x_nf(v_n) = \\ &= x_1(a_{11}w_1 + \cdots + a_{m1}w_m) + \cdots + x_n(a_{1n}w_1 + \cdots + a_{mn}w_m) = \\ &= (a_{11}x_1 + \cdots + a_{1n}x_n)w_1 + \cdots + (a_{m1}x_1 + \cdots + a_{mn}x_n)w_m. \end{aligned}$$

Perciò, se chiamiamo $y_1, \ldots, y_m$ le coordinate di $f(v)$ rispetto a $\mathcal{B}'$, ossia se $f(x_1v_1 + \cdots + x_nv_n) = y_1w_1 + \cdots + y_mw_m$, abbiamo

$$\begin{cases} y_1 &= a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots & \vdots \\ y_m &= a_{m1}x_1 + \cdots + a_{mn}x_n \end{cases}$$

In forma compatta possiamo scrivere $Y = AX$, dove $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$ è la colonna delle coordinate di $f(v)$ rispetto a $\mathcal{B}'$, $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ è la colonna delle coordinate di v rispetto a $\mathcal{B}$, $A = M_{\mathcal{B}}^{\mathcal{B}'}(f)$ è la matrice di f rispetto alle basi $\mathcal{B}, \mathcal{B}'$.

Si può riassumere la situazione scrivendo un "diagramma commutativo"

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ \downarrow{\scriptstyle \kappa_{\mathcal{B}}} & & \downarrow{\scriptstyle \kappa_{\mathcal{B}'}} \\ K^n & \xrightarrow{L(A)} & K^m \end{array} \quad \text{dove} \quad \begin{array}{ccc} v & \xrightarrow{f} & f(v) \\ \downarrow{\scriptstyle \kappa_{\mathcal{B}}} & & \downarrow{\scriptstyle \kappa_{\mathcal{B}'}} \\ (x_1, \ldots, x_n) & \xrightarrow{L(A)} & (y_1, \ldots, y_m) \end{array} \tag{9.2}$$

Diagramma commutativo significa che $\kappa_{\mathcal{B}'} \circ f = L(A) \circ \kappa_{\mathcal{B}}$: posso passare da V a K^m applicando prima f poi $\kappa_{\mathcal{B}'}$, oppure applicando prima $\kappa_{\mathcal{B}}$ poi $L(A)$, e il risultato è lo stesso.

Quando si introducono delle basi, si può lavorare con le coordinate anzichè con i vettori astratti, e ci si riduce a una situazione analoga a quella che si ha negli spazi vettoriali numerici K^n. **In coordinate** ogni applicazione lineare f si rappresenta mediante una matrice A, come l'applicazione $L(A)$.

9.2 Esempi

1. La matrice associata all'applicazione lineare $L(A) : K^n \to K^m$ rispetto alle basi canoniche del dominio e del codominio è proprio A. Infatti $L(A)(e_i) = a^i$, la colonna i-esima di A. Se si cambiano basi, cambia la matrice.

2. Consideriamo la seguente matrice 2×2 a coefficienti in $\mathbb{Q}$: $A = \begin{pmatrix} 1 & 3 \\ -2 & 5 \end{pmatrix}$, e l'applicazione lineare $L(A) : \mathbb{Q}^2 \to \mathbb{Q}^2$. Consideriamo poi le seguenti basi di $\mathbb{Q}^2$ (perchè sono basi?): $\mathcal{B} = (v_1, v_2) = (\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \end{pmatrix})$, $\mathcal{B}' = (w_1, w_2) = (\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix})$. Vogliamo determinare $M_{\mathcal{B}}^{\mathcal{B}'}(L(A))$.

Dobbiamo calcolare $L(A)(v_1), L(A)(v_2)$ e poi esprimerli in funzione della base $\mathcal{B}'$.

$$L(A)(v_1) = \begin{pmatrix} 1 & 3 \\ -2 & 5 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}, \ L(A)(v_2) = \begin{pmatrix} 1 & 3 \\ -2 & 5 \end{pmatrix}\begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 12 \end{pmatrix}.$$

$$\begin{pmatrix} 4 \\ 3 \end{pmatrix} = -\begin{pmatrix} 0 \\ 1 \end{pmatrix} + 2\begin{pmatrix} 2 \\ 2 \end{pmatrix} = -w_1 + 2w_2; \ \begin{pmatrix} 5 \\ 12 \end{pmatrix} = 7\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{5}{2}\begin{pmatrix} 2 \\ 2 \end{pmatrix} = 7w_1 + \frac{5}{2}w_2.$$

Perciò $M_{\mathcal{B}}^{\mathcal{B}'}(L(A)) = \begin{pmatrix} -1 & 7 \\ 2 & \frac{5}{2} \end{pmatrix}$.

Vedremo in seguito (Teorema 10.3.1) un altro modo per trovare questa matrice.

3. L'applicazione $D : \mathbb{R}[t] \to \mathbb{R}[t]$ tale che $D(f(t)) = f'(t)$, la derivata del polinomio f, è lineare (perchè?). Osserviamo che non ha senso parlare di matrice di D perchè $\mathbb{R}[t]$ ha dimensione infinita. Per farlo, ci possiamo restringere a suoi sottospazi di dimensione finita. Consideriamo dunque gli $\mathbb{R}$-spazi vettoriali di dimensione finita $V = \mathbb{R}[t]_{\leq 2}$ con base $\mathcal{B} = (1, t, t^2)$, $W = \mathbb{R}[t]_{\leq 3}$ con base $\mathcal{B}' = (1, t, t^2, t^3)$. Definiamo un'applicazione $T : V \to W$ ponendo $T(f(t)) = t^2 f'(t+1)$. Verificare per esercizio che anche T è lineare. Calcoliamo $M_{\mathcal{B}}^{\mathcal{B}'}(T)$:

$$T(1) = 0, \ T(t) = t^2, \ T(t^2) = 2t^2 + 2t^3, \ \text{perciò} \ M_{\mathcal{B}}^{\mathcal{B}'}(T) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{pmatrix}.$$

Allora se $f(t) = a_0 + a_1 t + a_2 t^2$, le coordinate di $T(f)$ rispetto a $\mathcal{B}'$ sono

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{pmatrix}\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ a_1 + 2a_2 \\ 2a_2 \end{pmatrix};$$

quindi $T(f(t)) = (a_1 + 2a_2)t^2 + 2a_2 t^3$. Questo può essere verificato anche direttamente, usando le proprietà della derivata.

9.3 Forma canonica

Il prossimo teorema ci dice che, purchè le basi del dominio e del codominio siano scelte in maniera opportuna, la matrice di un'applicazione lineare può assumere una forma particolarmente semplice, detta "forma canonica".

Teorema 9.3.1. *Sia $f : V \to W$ un'applicazione lineare fra K-spazi vettoriali di dimensioni finite m, n, sia $r = \operatorname{rg} f = \dim \operatorname{Im} f$ il rango di f. Esistono basi $\mathcal{B}$ di V e $\mathcal{B}'$ di W tali che $M_{\mathcal{B}}^{\mathcal{B}'}(f)$ è una matrice a blocchi della forma*

$$\begin{array}{c c} & \begin{array}{ccc} r & & n-r \end{array} \\ \begin{array}{c} r \\ \\ m-r \end{array} & \begin{pmatrix} E_r & | & 0 \\ --- & | & --- \\ 0 & | & 0 \end{pmatrix} \end{array} \tag{9.3}$$

dove E_r è la matrice identica $r \times r$, e gli zeri denotano delle matrici nulle rispettivamente di tipo $r \times (n-r)$, $(m-r) \times r$, $(m-r) \times (n-r)$.

Dimostrazione. Cerchiamo di capire quali proprietà devono avere le due basi perchè la matrice di f rispetto a tali basi abbia la forma voluta.

Se la matrice $M_{\mathcal{B}}^{\mathcal{B}'}(f)$ ha la forma (9.3) rispetto a basi $\mathcal{B} = (v_1, \dots, v_n)$, $\mathcal{B}' = (w_1, \dots, w_m)$ significa che

$$\begin{aligned}
f(v_1) &= w_1 && = 1w_1 + 0w_2 + \cdots + 0w_m \\
&\vdots \\
f(v_r) &= w_r && = 0w_1 + \cdots + 1w_r + \cdots + 0w_m \\
f(v_{r+1}) &= 0 \\
&\vdots \\
f(v_n) &= 0,
\end{aligned}$$

perciò $v_{r+1}, \dots, v_n \in \ker f$, e sono linearmente indipendenti perchè fanno parte della base $\mathcal{B}$ di V. Siccome $r = \operatorname{rg} f$ e $\dim V = n$, per il Teorema della dimensione 7.1.1 $\dim \ker f = n - r$, quindi $v_{r+1}, \dots, v_n$ **formano una base di** $\ker f$. D'altra parte **i vettori** $w_1, \dots, w_r$ **generano** $\operatorname{Im} f$, perchè sono le immagini non nulle dei vettori di una base del dominio (Proposizione 7.1.3) e quindi ne formano una base, perchè sono in numero di r. Ora abbiamo gli elementi necessari per costruire le basi volute.

Partiamo da una base di $\ker f$: $v_{r+1}, \dots, v_n$, e la completiamo a una base $\mathcal{B} = (v_1, \dots, v_n)$ di V. Allora $f(v_1), \dots, f(v_n)$ generano $\operatorname{Im} f$. Siccome $f(v_{r+1}) = \cdots = f(v_n) = 0$, non contribuiscono a generare $\operatorname{Im} f$, allora gli r vettori $f(v_1), \dots, f(v_r)$ sono un sistema di r generatori, e quindi una base di $\operatorname{Im} f$. Pongo $w_1 = f(v_1), \dots, w_r = f(v_r)$, e poi li completo a una base $\mathcal{B}'$ di W aggiungendo opportuni vettori $w_{r+1}, \dots, w_m$. Rispetto a $\mathcal{B}$ e $\mathcal{B}'$ la matrice di f ha la forma (9.3). □

Esempio 9.3.2. *Consideriamo la matrice* $A = \begin{pmatrix} 1 & 2 & -1 & 1 \\ 1 & 1 & 1 & 2 \\ -1 & -2 & 1 & 0 \end{pmatrix}$. *Trovare basi $\mathcal{B}$ di $\mathbb{R}^4$ e $\mathcal{B}'$ di $\mathbb{R}^3$ tali che $M_{\mathcal{B}}^{\mathcal{B}'}(L(A))$ sia in forma canonica.*

Calcoliamo facilmente con il metodo di eliminazione di Gauss che $\operatorname{rg}(A) = 3$ *e che una base di* $\ker L(A)$ *è costituita dall'unico vettore $v = (-3, 2, 1, 0)$. Usando il teorema dello scambio, costruiamo una base di $\mathbb{R}^4$ contenente v, per esempio $\mathcal{B} = (e_1, e_2, e_4, v)$. Allora $L(A)(e_1) = w_1 = (1, 1, -1)$, $L(A)(e_2) = w_2 = (2, 1, -2)$, $L(A)(e_4) = w_3 = (1, 2, 0)$, e $\mathcal{B}' = (w_1, w_2, w_3)$ è la base di $\mathbb{R}^3$ voluta. La matrice in forma canonica è* $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$.

9.4 Rango di matrici e applicazioni lineari

Teorema 9.4.1. *Sia $f : V \to W$ un'applicazione lineare fra spazi vettoriali di dimensione finita, siano $\mathcal{B}, \mathcal{B}'$ basi di V e di W rispettivamente. Sia $A = M_{\mathcal{B}}^{\mathcal{B}'}(f)$. Allora* $\operatorname{rg} f = \operatorname{rg} A = \operatorname{rg} L(A)$.

Dimostrazione. Consideriamo il seguente diagramma commutativo dove $\kappa_{\mathcal{B}}$ e $\kappa_{\mathcal{B}'}$ sono gli isomorfismi introdotti nella Sezione 8.2:

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ \downarrow{\scriptstyle \kappa_{\mathcal{B}}} & & \downarrow{\scriptstyle \kappa_{\mathcal{B}'}} \\ K^n & \xrightarrow{L(A)} & K^m \end{array}.$$

Osserviamo che $\operatorname{Im} L(A) = \kappa_{\mathcal{B}'}(\operatorname{Im} f)$; infatti se $y \in \operatorname{Im} L(A)$ esiste $x \in K^n$ tale che $y = L(A)(x)$. Essendo $\kappa_{\mathcal{B}}$ un isomorfismo, $x = \kappa_{\mathcal{B}}(v)$ per un (unico) $v \in V$. Quindi $y = L(A)(\kappa_{\mathcal{B}}(v)) = \kappa_{\mathcal{B}'}(f(v))$ per la commutatività del diagramma, e quindi $\operatorname{Im} L(A) \subset \kappa_{\mathcal{B}'}(\operatorname{Im} f)$. Viceversa se $y \in \kappa_{\mathcal{B}'}(\operatorname{Im} f)$, esiste $v \in V$ tale che $y = \kappa_{\mathcal{B}'}(f(v)) = L(A)(\kappa_{\mathcal{B}}(v))$, il che prova l'inclusione opposta. Poichè $\kappa_{\mathcal{B}'}$ è un isomorfismo, $\operatorname{Im} f \simeq \kappa_{\mathcal{B}'}(\operatorname{Im} f) = \operatorname{Im} L(A)$ e perciò $\dim \operatorname{Im} L(A) = \dim \operatorname{Im} f$ ossia $\operatorname{rg} L(A) = \operatorname{rg} f$; infine ricordiamo che $\operatorname{rg} L(A) = \operatorname{rg} A$ (Sezione 7.2.3). $\square$

Otteniamo la seguente conseguenza:

Corollario 9.4.2. *Matrici che rappresentano la stessa applicazione lineare rispetto a basi diverse hanno lo stesso rango.*

9.5 Isomorfismo fra applicazioni lineari e matrici

Fissiamo ora due K-spazi vettoriali V, W di dimensioni finite n, m e due loro basi $\mathcal{B} = (v_1, \dots, v_n), \mathcal{B}' = (w_1, \dots, w_m)$. Associando a ogni applicazione lineare di V in W la sua matrice rispetto a tali basi possiamo costruire un'applicazione.

Teorema 9.5.1. *L'applicazione*

$$\begin{array}{rccl} M_{\mathcal{B}}^{\mathcal{B}'}: & \operatorname{Hom}(V,W) & \to & M(m \times n, K) \\ & f & \to & M_{\mathcal{B}}^{\mathcal{B}'}(f) \end{array}$$

è un isomorfismo di K-spazi vettoriali.

La dimostrazione si basa sul teorema di determinazione di un'applicazione lineare.

Dimostrazione.

1. $M_{\mathcal{B}}^{\mathcal{B}'}$ è lineare: date applicazioni lineari $f, g : V \to W$ e scalari λ, μ si tratta di verificare che $M_{\mathcal{B}}^{\mathcal{B}'}(\lambda f + \mu g) = \lambda M_{\mathcal{B}}^{\mathcal{B}'}(f) + \mu M_{\mathcal{B}}^{\mathcal{B}'}(g)$. Per la Definizione 9.1.1 di matrice di un'applicazione lineare, basta verificare che $(\lambda f + \mu g)(v_i) = \lambda f(v_i) + \mu g(v_i)$ per ogni indice $i = 1, \dots, n$; ma questo segue dalla definizione delle operazioni in $\operatorname{Hom}(V, W)$.

2. $M_{\mathcal{B}}^{\mathcal{B}'}$ è iniettiva: siccome abbiamo appena verificato che l'applicazione è lineare, basta verificare che $\ker M_{\mathcal{B}}^{\mathcal{B}'}$ è nullo. Sia dunque $f : V \to W$ un'applicazione lineare tale che $M_{\mathcal{B}}^{\mathcal{B}'}(f) = 0$, la matrice nulla. Allora le sue colonne sono tutte uguali al vettore nullo di K^m, e quindi $f(v_1) = \dots = f(v_n) = 0_W$. Ma anche l'applicazione nulla fa corrispondere a ogni vettore di $\mathcal{B}$ il vettore 0_W; per la parte di unicità del teorema 8.1.1, segue che $f = 0$.

3. $M_{\mathcal{B}}^{\mathcal{B}'}$ è suriettiva: prendiamo una qualunque matrice $A \in M(m \times n, K)$; vogliamo costruire un'applicazione lineare f tale che $M_{\mathcal{B}}^{\mathcal{B}'}(f) = A$. Consideriamo le n colonne di A: $a^1, \ldots, a^n$, e gli n vettori di W che hanno come coordinate rispetto a $\mathcal{B}'$ tali colonne: $u_1 := \kappa_{\mathcal{B}'}^{-1}(a^1) = a_{11}w_1 + \cdots + a_{m1}w_m, \ldots, u_n := \kappa_{\mathcal{B}'}^{-1}(a^n) = a_{1n}w_1 + \cdots + a_{mn}w_m$. Definiamo allora f come l'unica applicazione lineare di V in W tale che $f(v_1) = u_1, \ldots, f(v_n) = u_n$. Si ha allora $A = M_{\mathcal{B}}^{\mathcal{B}'}(f)$.

□

Corollario 9.5.2. $\dim \mathrm{Hom}(V, W) = \dim M(m \times n, K) = mn = \dim V \dim W$.

Si ritrova in particolare che $\dim V^* = \dim \mathrm{Hom}(V, K) = \dim V$ (Sezione 8.3, 2.)

9.6 Matrice di una composta di applicazioni lineari

Consideriamo la seguente situazione:

$$\begin{array}{lllll} V & \xrightarrow{f} & W & \xrightarrow{g} & U & \text{applicazioni lineari} \\ m & & n & & p & \text{dimensioni finite} \\ \mathcal{B} & & \mathcal{B}' & & \mathcal{B}'' & \text{basi} \end{array}$$

Allora la matrice dell'applicazione lineare composta è il prodotto righe per colonne delle matrici di f e di g nell'ordine appropriato:

$$\begin{array}{cccl} M_{\mathcal{B}}^{\mathcal{B}''}(g \circ f) = & M_{\mathcal{B}'}^{\mathcal{B}''}(g) & M_{\mathcal{B}}^{\mathcal{B}'}(f) & \text{prodotto righe per colonne} \\ p \times m & p \times n & n \times m & \end{array}$$

Facciamo la verifica. Siano $\mathcal{B} = (v_1, \ldots, v_n)$, $\mathcal{B}' = (v_1', \ldots, v_m')$, $\mathcal{B}'' = (v_1'', \ldots, v_p'')$, $A = M_{\mathcal{B}}^{\mathcal{B}'}(f)$, $B = M_{\mathcal{B}'}^{\mathcal{B}''}(g)$.

Per ogni $k = 1, \ldots, n$, $j = 1, \ldots, m$ si ha:

$$f(v_k) = a_{1k}v_1' + \cdots + a_{mk}v_m' = \sum_{j=1}^{m} a_{jk}v_j',$$

$$g(v_j') = b_{1j}v_1'' + \cdots + b_{pj}v_p'' = \sum_{i=1}^{p} b_{ij}v_i''.$$

Allora

$$\begin{aligned} g(f(v_k)) = g(\sum_{j=1}^{m} a_{jk}v_j') &\overset{\text{linearità}}{=} \sum_{j=1}^{m} a_{jk}g(v_j') = \\ = \sum_{j=1}^{m} a_{jk} \sum_{i=1}^{p} b_{ij}v_i'' &\overset{\text{prop.distrib.}}{=} \sum_{i=1}^{p}(\sum_{j=1}^{m} b_{ij}a_{jk})v_i'' : \end{aligned}$$

il coefficiente di v_i'' è proprio l'elemento di BA di posto ik.

Il caso degli endomorfismi

Se $f : V \to V$ è un endomorfismo di V, $\dim V = n$, si può prendere la stessa base $\mathcal{B}$ nel dominio e nel codominio; in tal caso si scrive $M_{\mathcal{B}}(f)$ anzichè $M^{\mathcal{B}}_{\mathcal{B}}(f)$.

L'isomorfismo $M_{\mathcal{B}} : \mathrm{Hom}(V, V) \to M(n \times n, K)$ gode allora della ulteriore proprietà $M_{\mathcal{B}}(g \circ f) = M_{\mathcal{B}}(g)M_{\mathcal{B}}(f)$, cioè conserva, oltre alla somma e al prodotto esterno, anche l'ulteriore operazione di "prodotto" interno, più precisamente manda la composizione di applicazioni in $\mathrm{Hom}(V, V)$ nel prodotto righe per colonne delle loro matrici in $M(n \times n, K)$. In questo caso si dice che $M_{\mathcal{B}}$ è un **isomorfismo di K-algebre**.

9.7 Applicazioni lineari di K^n in K^m

Come ulteriore applicazione della teoria sviluppata, concludiamo questo capitolo dando la caratterizzazione delle applicazioni lineari $K^n \to K^m$: sono tutte e sole del tipo $L(A)$.

Siano $\mathcal{C}$, $\mathcal{C}'$ le basi canoniche di K^n e di K^m rispettivamente. Abbiamo allora l'isomorfismo

$$\begin{array}{rccll} M^{\mathcal{C}'}_{\mathcal{C}} : & \mathrm{Hom}(K^n, K^m) & \to & M(m \times n, K) & \\ & f & \to & M^{\mathcal{C}'}_{\mathcal{C}}(f) & \\ & L(A) & \to & A & \qquad\qquad \text{Sezione 9.2, 1.} \end{array}$$

L'isomorfismo inverso è allora $L : M(m \times n, K) \to \mathrm{Hom}(K^n, K^m)$ che manda $A \to L(A)$. Ciò significa che **ogni applicazione lineare $K^n \to K^m$ può essere espressa nella forma $L(A)$ per un'unica matrice A.**

Nel caso particolare $m = n$, supponiamo che $f : K^n \to K^n$ corrisponda nell'isomorfismo $M_{\mathcal{C}}$ a una matrice $A = M_{\mathcal{C}}(f)$ e $g : K^n \to K^n$ a una matrice $B = M_{\mathcal{C}}(g)$, dunque $f = L(A)$, $g = L(B)$. Allora $g \circ f$ corrisponde a $M_{\mathcal{C}}(g \circ f) = M_{\mathcal{C}}(g)M_{\mathcal{C}}(f) = BA$. Quindi si ha $L(BA) = L(B) \circ L(A)$.

Capitolo 10

Matrici invertibili e cambiamento di base

10.1 Matrici invertibili

Definizione 10.1.1. Sia A una matrice quadrata $n \times n$ a entrate in K. A è detta **invertibile** se esiste $A' \in M(n \times n, K)$ tale che $AA' = A'A = E_n = (\delta_{ij})_{i,j=1,\dots,n}$. In tal caso A' è l'inversa di A e la si denota A^{-1}.

Denoteremo $\mathrm{GL}(n, K) \subset M(n \times n, K)$ il sottinsieme delle matrici invertibili.

Proposizione 10.1.2. *$GL(n, K)$ è un gruppo rispetto al prodotto righe per colonne, detto gruppo lineare generale.*

Dimostrazione.

1. $\mathrm{GL}(n, K)$ è chiuso rispetto al prodotto: se A, B sono matrici invertibili $n \times n$, anche AB è invertibile, e si ha $(AB)^{-1} = B^{-1}A^{-1}$. Infatti $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AE_nA^{-1} = AA^{-1} = E_n$.

2. E_n, che è l'elemento neutro rispetto al prodotto, è invertibile, e coincide con la sua inversa.

3. Se A è invertibile, anche la sua inversa A^{-1} è invertibile, e precisamente $(A^{-1})^{-1} = A$. $\square$

Osservazione 13. $\mathrm{GL}(n, K)$ non è un sottospazio vettoriale di $M(n \times n, K)$. Per esempio $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$. Quest'ultima è somma di matrici invertibili, infatti la prima è la matrice identica, e anche la seconda coincide con la propria inversa. Ma la somma non è invertibile. Infatti moltiplicata per qualunque altra matrice non può dare la matrice identica: $\begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 2c & 2d \end{pmatrix}$.

Osservazione 14. La trasposta di una matrice invertibile è invertibile e si ha $({}^tA)^{-1} = {}^t(A^{-1})$.

Teorema 10.1.3. *Sia $f : V \to W$ un'applicazione lineare fra spazi vettoriali della stessa dimensione finita n; siano $\mathcal{B}, \mathcal{B}'$ basi di V e W. Allora f è un isomorfismo se e solo se $M_{\mathcal{B}}^{\mathcal{B}'}(f)$ è una matrice invertibile; la sua inversa è proprio $M_{\mathcal{B}'}^{\mathcal{B}}(f^{-1})$.*

Dimostrazione. Se f è un isomorfismo esiste l'isomorfismo inverso f' tale che $f \circ f' = \mathrm{id}_W$ e $f' \circ f = \mathrm{id}_V$. Poniamo $A = M_{\mathcal{B}}^{\mathcal{B}'}(f)$, $A' = M_{\mathcal{B}'}^{\mathcal{B}}(f')$. Allora

$$AA' = M_{\mathcal{B}}^{\mathcal{B}'}(f) M_{\mathcal{B}'}^{\mathcal{B}}(f') \overset{9.6}{=} M_{\mathcal{B}'}^{\mathcal{B}'}(f \circ f') = M_{\mathcal{B}'}^{\mathcal{B}'}(\mathrm{id}_W) = E_n,$$

$$A'A = M_{\mathcal{B}'}^{\mathcal{B}}(f') M_{\mathcal{B}}^{\mathcal{B}'}(f) \overset{9.6}{=} M_{\mathcal{B}}^{\mathcal{B}}(f' \circ f) = M_{\mathcal{B}}^{\mathcal{B}}(\mathrm{id}_V) = E_n.$$

Dunque $A' = A^{-1}$.

Viceversa, se $A = M_{\mathcal{B}}^{\mathcal{B}'}(f)$ è invertibile, allora $L(A)$ risulta un isomorfismo con isomorfismo inverso $L(A^{-1})$: infatti $AA^{-1} = E_n$, quindi $L(AA^{-1}) \overset{9.7}{=} L(A) \circ L(A^{-1}) = L(E_n) = \mathrm{id}_{K^n}$; analogamente $L(A^{-1}) \circ L(A) = \mathrm{id}_{K^n}$, quindi $L(A^{-1})$ è l'applicazione inversa di $L(A)$. Consideriamo infine il diagramma commutativo

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ \downarrow{\scriptstyle \kappa_{\mathcal{B}}} & & \downarrow{\scriptstyle \kappa_{\mathcal{B}'}} \\ K^n & \xrightarrow{L(A)} & K^m \end{array} :$$

poichè $f = \kappa_{\mathcal{B}'}^{-1} \circ L(A) \circ \kappa_{\mathcal{B}}$, ricordando che composizione di isomorfismi è isomorfismo, concludiamo che anche f è un isomorfismo. □

Osservazione 15. $L(A)$ è un isomorfismo se e solo se A è invertibile.

Corollario 10.1.4. *Sia A una matrice quadrata $n \times n$. Le seguenti condizioni sono equivalenti:*

1. *A è invertibile;*
2. $\mathrm{rg}(A)$ *è massimo pari a n;*
3. *le n colonne di A sono linearmente indipendenti;*
4. *le n righe di A sono linearmente indipendenti;*
5. *le n colonne di A generano lo spazio delle colonne;*
6. *le n righe di A generano lo spazio delle righe.*

Dimostrazione. Dal Teorema 10.1.3 segue: A è invertibile se e solo se $L(A)$ è un isomorfismo, se e solo se $L(A)$ è suriettiva (perchè endomorfismo di uno spazio vettoriale di dimensione finita), se e solo se le n colonne di A generano il codominio K^n, se e solo se il rango di A è massimo n. Ricordando che rango per righe e per colonne coincidono (Teorema 7.2.3) e la Proposizione 7.2.1, otteniamo l'equivalenza con le altre condizioni. □

10.2 Cambiamento di base

Sia V un K-spazio vettoriale di dimensione n. Supponiamo che $\mathcal{A} = (v_1, \ldots, v_n)$, $\mathcal{B} = (w_1, \ldots, w_n)$ siano due basi di V. Ci chiediamo come si passa dalle coordinate rispetto ad $\mathcal{A}$ alle coordinate rispetto a $\mathcal{B}$ e viceversa. Consideriamo il diagramma commutativo

$$\begin{tikzcd} V \arrow[r, "\mathrm{id}_V"] \arrow[d, "\kappa_{\mathcal{A}}"] & V \arrow[d, "\kappa_{\mathcal{B}}"] \\ K^n \arrow[r] & K^n \end{tikzcd} \qquad \begin{tikzcd} v \arrow[r, "\mathrm{id}_V"] \arrow[d, "\kappa_{\mathcal{A}}"] & v \arrow[d, "\kappa_{\mathcal{B}}"] \\ (x_1, \ldots, x_n) \arrow[r] & (y_1, \ldots, y_n) \end{tikzcd}$$

dove per ogni vettore $v \in V$ si ha $v = x_1 v_1 + \cdots + x_n v_n = y_1 w_1 + \ldots + y_n w_n$.

Siccome $\mathrm{Hom}(K^n, K^n) \simeq M(n \times n, K)$, l'applicazione lineare "in basso" $K^n \to K^n$, che manda le coordinate di v rispetto ad $\mathcal{A}$ $(x_1, \ldots, x_n)$ in quelle rispetto a $\mathcal{B}$ $(y_1, \ldots, y_n)$, è del tipo $L(M)$, dove $M = M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V)$ e $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = M \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

Definizione 10.2.1. Tale matrice $M = M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V)$ è detta **matrice di passaggio da $\mathcal{A}$ a $\mathcal{B}$ o matrice del cambiamento di base.**

$$M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V) = M = \begin{array}{c} \begin{array}{cccc} \mathrm{id}_V(v_1) & \mathrm{id}_V(v_2) & \ldots & \mathrm{id}_V(v_n) \\ \downarrow & \downarrow & & \downarrow \end{array} \\ \begin{pmatrix} m_{11} & m_{12} & \ldots & m_{1n} \\ m_{21} & m_{22} & \ldots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \ldots & m_{nn} \end{pmatrix} \end{array}.$$

La prima colonna contiene le coordinate di v_1 rispetto a $\mathcal{B}$, la seconda le coordinate di v_2 rispetto a $\mathcal{B}$, ecc., cioè nelle colonne di M sono contenute le coordinate dei vettori di $\mathcal{A}$ rispetto a $\mathcal{B}$: $v_1 = m_{11} w_1 + \cdots + m_{n1} w_n, \ldots, v_n = m_{1n} w_1 + \cdots + m_{nn} w_n$.

Osservazione 16. Siccome la matrice del cambiamento di base è la matrice che rappresenta id_V, che è ovviamente un isomorfismo, rispetto a una scelta di basi, il rango di $M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V)$ è uguale a n, cioè la matrice è invertibile (per il Teorema 10.1.3). La sua inversa è proprio $M^{\mathcal{A}}_{\mathcal{B}}(\mathrm{id}_V)$: ciò segue dal Teorema 10.1.3. Lo si può vedere anche direttamente guardando il seguente diagramma commutativo:

$$\begin{tikzcd} V \arrow[r, "\mathrm{id}_V"] \arrow[d, "\kappa_{\mathcal{A}}"] & V \arrow[r, "\mathrm{id}_V"] \arrow[d, "\kappa_{\mathcal{B}}"] & V \arrow[d, "\kappa_{\mathcal{A}}"] \\ K^n \arrow[r, "L(M)"] & K^n \arrow[r, "L(M')"] & K^n \end{tikzcd}$$

in cui $M = M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V)$, $M' = M^{\mathcal{A}}_{\mathcal{B}}(\mathrm{id}_V)$. Si ha chiaramente che la composta delle due applicazioni in basso è l'applicazione identica di K^n, dunque $\mathrm{id}_{K^n} = L(M') \circ L(M) =$

$L(M'M)$ per quanto visto nella Sezione 9.7. Concludiamo che $M'M = E_n$. Analogamente si prova che anche $MM' = E_n$. Dunque $M' = M^{-1}$.

Osserviamo che le colonne di M', ossia di $M_{\mathcal{A}}^{\mathcal{B}}(\mathrm{id}_V)^{-1}$, contengono le coordinate rispetto ad $\mathcal{A}$ dei vettori della "nuova" base $\mathcal{B}$.

Osservazione 17. Dall'isomorfismo $\kappa_{\mathcal{B}} : V \to K^n$ segue che vettori $v_1, \ldots, v_n$ sono linearmente indipendenti se e solo se sono linearmente indipendenti i vettori $\kappa_{\mathcal{B}}(v_1), \ldots, \kappa_{\mathcal{B}}(v_n)$ delle loro coordinate rispetto a $\mathcal{B}$.

Perciò, data una base $\mathcal{B}$ di V, **ogni matrice invertibile M può essere interpretata come matrice di un cambiamento di base** da $\mathcal{B}$ a una nuova base. I vettori della nuova base sono quelli le cui coordinate, rispetto a $\mathcal{B}$, sono contenute nelle colonne di M^{-1}.

10.3 Matrici di un'applicazione lineare rispetto a basi diverse

Sia $f : V \to W$ un'applicazione lineare. Supponiamo che siano date due basi di V: $\mathcal{A}, \mathcal{A}'$ e due basi di W: $\mathcal{B}, \mathcal{B}'$. In questa sezione vogliamo trovare la relazione che c'è fra $M_{\mathcal{A}}^{\mathcal{B}}(f)$ e $M_{\mathcal{A}'}^{\mathcal{B}'}(f)$. Otterremo:

Teorema 10.3.1.

$$M_{\mathcal{A}'}^{\mathcal{B}'}(f) = M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{id}_W) M_{\mathcal{A}}^{\mathcal{B}}(f) M_{\mathcal{A}'}^{\mathcal{A}}(\mathrm{id}_V).$$

Dimostrazione. La relazione si ottiene facilmente considerando i diagrammi commutativi del tipo (9.2) per le applicazioni $f, \mathrm{id}_V, \mathrm{id}_W$, relativi alle basi considerate:

$$\begin{array}{ccccccc} V & \xrightarrow{\mathrm{id}_V} & V & \xrightarrow{f} & W & \xrightarrow{\mathrm{id}_W} & W \\ \downarrow{\scriptstyle \kappa_{\mathcal{A}'}} & & \downarrow{\scriptstyle \kappa_{\mathcal{A}}} & & \downarrow{\scriptstyle \kappa_{\mathcal{B}}} & & \downarrow{\scriptstyle \kappa_{\mathcal{B}'}} \\ K^n & \xrightarrow{L(T)} & K^n & \xrightarrow{L(M)} & K^m & \xrightarrow{L(S)} & K^m \end{array} \tag{10.1}$$

dove $M = M_{\mathcal{A}}^{\mathcal{B}}(f)$, $S = M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{id}_W)$ è la matrice di passaggio da $\mathcal{B}$ a $\mathcal{B}'$ basi di W, $T = M_{\mathcal{A}'}^{\mathcal{A}}(\mathrm{id}_V)$ è la matrice di passaggio da $\mathcal{A}'$ ad $\mathcal{A}$ basi di V. Allora, interpretando f come la composizione della tre applicazioni della prima riga, vediamo che la matrice di f rispetto ad $\mathcal{A}'$ e $\mathcal{B}'$ è quella che corrisponde alla composizione $L(S) \circ L(M) \circ L(T)$, ed è quindi SMT (Sezione 9.7). $\square$

In particolare, dal momento che S e T sono matrici di cambiamenti di base, abbiamo ottenuto che **due matrici che rappresentano l'applicazione lineare f rispetto a basi diverse differiscono per il prodotto a destra e a sinistra per matrici invertibili.**

Come conseguenza del Teorema 9.3.1 - Forma canonica - otteniamo allora:

Corollario 10.3.2. *Data una matrice $M \in M(m \times n, K)$, esistono matrici invertibili S $m \times m$, e T $n \times n$ tali che*

$$SMT = \begin{array}{c} \\ r \\ \\ m-r \end{array} \begin{array}{c} \begin{array}{ccc} r & \quad & n-r \end{array} \\ \left(\begin{array}{c|c} E_r & 0 \\ --- & --- \\ 0 & 0 \end{array} \right) \end{array} \tag{10.2}$$

dove $r = \mathrm{rg}(M)$.

Dimostrazione. Consideriamo $L(M) : K^n \to K^m$: M è la sua matrice rispetto alle basi canoniche. Ma esistono basi di K^n e K^m, $\mathcal{A}$ e $\mathcal{B}$, rispetto alle quali la matrice di $L(M)$ è

$$\begin{array}{cc} & \begin{array}{ccc} r & & n-r \end{array} \\ \begin{array}{c} r \\ \\ m-r \end{array} & \begin{pmatrix} E_r & | & 0 \\ --- & | & --- \\ 0 & | & 0 \end{pmatrix} \end{array} \tag{10.3}$$

Siano $T = M^{\mathcal{C}}_{\mathcal{A}}(\mathrm{id}_{K^n})$, $S = M^{\mathcal{B}}_{\mathcal{C}'}(\mathrm{id}_{K^m})$. Per il Teorema 10.3.1, SMT coincide con la matrice (10.3). □

Esempio 10.3.3. Consideriamo l'applicazione lineare $f : V \to W$ con $V = \mathbb{R}[t]_{\leq 2}$, $W = \mathbb{R}[t]_{\leq 3}$, definita da $f(p(t)) = t^2 p'(t+1)$. Se $\mathcal{B} = (1, t, t^2)$, $\mathcal{B}'(1, t, t^2, t^3)$, allora

$$M^{\mathcal{B}'}_{\mathcal{B}}(f) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{pmatrix}.$$

Ora consideriamo $\mathcal{A} = (1, t-1, (t-1)^2)$ in V e $\mathcal{A}' = (1, t-2, (t-2)^2, (t-2)^3)$ in W.

$$M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V) = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} :$$

ha rango 3 e questo conferma che $\mathcal{A}$ è una base di V.

$$M^{\mathcal{B}'}_{\mathcal{A}'}(\mathrm{id}_W) = \begin{pmatrix} 1 & -2 & 4 & -8 \\ 0 & 1 & -4 & 12 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

anche questa matrice ha rango massimo 4, dunque è invertibile, il che conferma che anche $\mathcal{A}'$ è una base. Applicando il Teorema 10.3.1, otteniamo che

$$M^{\mathcal{A}'}_{\mathcal{A}}(f) = M^{\mathcal{B}'}_{\mathcal{A}'}(\mathrm{id}_W)^{-1} M^{\mathcal{B}'}_{\mathcal{B}}(f) M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V).$$

Nella Sezione 10.5 vedremo come calcolare $M^{\mathcal{B}'}_{\mathcal{A}'}(\mathrm{id}_W)^{-1}$.

10.4 Il caso di un endomorfismo

Sia $f : V \to V$ un endomorfismo. Fissiamo la stessa base $\mathcal{A}$ di V in dominio e codominio e consideriamo la matrice $M_{\mathcal{A}}(f)$; poi prendiamo un'altra base $\mathcal{B}$ di V e consideriamo $M_{\mathcal{B}}(f)$. Dal Teorema 10.3.1 otteniamo

$$M_{\mathcal{B}}(f) = M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V) M_{\mathcal{A}}(f) M^{\mathcal{A}}_{\mathcal{B}}(\mathrm{id}_V),$$

ma ora le due matrici "ai lati" $M^{\mathcal{A}}_{\mathcal{B}}(\mathrm{id}_V)$ e $M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V)$ sono matrici una inversa dell'altra, perchè rappresentano cambiamenti di base opposti.

Posta $S = M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V)$, la matrice di passaggio da $\mathcal{A}$ a $\mathcal{B}$, abbiamo perciò $M_{\mathcal{B}}(f) = SM_{\mathcal{A}}(f)S^{-1}$.

Notiamo che, se abbiamo le coordinate dei vettori di $\mathcal{B}$ rispetto ad $\mathcal{A}$, possiamo scrivere direttamente la matrice $S^{-1} = M^{\mathcal{A}}_{\mathcal{B}}(\mathrm{id}_V)$.

Definizione 10.4.1 (Relazione di similitudine di matrici)**.** Due matrici quadrate A, B, $n \times n$ a coefficienti in K, si dicono **simili** se esiste $S \in \mathrm{GL}(n, K)$ tale che $A = SBS^{-1}$.

La similitudine è una relazione d'equivalenza in $M(n \times n, K)$. Infatti

- $A = E_n A E_n^{-1}$;
- se $A = SBS^{-1}$, allora, moltiplicando a sinistra per S^{-1} e a destra per S otteniamo $B = S^{-1}AS$.
- se $A = SBS^{-1}$ e $B = MCM^{-1}$, allora $A = S(MCM^{-1})S^{-1} = (SM)C(SM)^{-1}$.

Dal momento che ogni matrice invertibile si può interpretare come la matrice di un cambiamento di base, si ha che due matrici sono simili se e solo se rappresentano lo stesso endomorfismo rispetto a basi diverse. In particolare **matrici simili hanno lo stesso rango**.

Un problema molto importante e non banale è quello di trovare un rappresentante "semplice" o "canonico" per la classe di similitudine di una matrice data, ossia una sua "forma canonica". In altre parole, si vuole descrivere l'insieme quoziente. Tratteremo il problema della diagonalizzabilità di una matrice quadrata data M, ossia di saper dire se M è simile a una matrice diagonale. Vedremo che non sempre la risposta è affermativa. In alcuni casi in cui la risposta è negativa, vedremo che è possibile considerare un'altra forma canonica, detta di Jordan.

10.5 Algoritmo per il calcolo dell'inversa di una matrice invertibile

Iniziamo provando un risultato preliminare.

Proposizione 10.5.1. *Sia A una matrice $n \times n$ tale che esiste una matrice $n \times n$ B per cui $BA = E_n$, oppure esiste una matrice C per cui $AC = E_n$. Allora A è invertibile. Inoltre nel primo caso $A^{-1} = B$ e nel secondo caso $A^{-1} = C$.*

Dimostrazione. Se $BA = E_n$, si ha $\mathrm{id}_{K^n} = L(E_n) = L(BA) = L(B) \circ L(A)$ (Sezione 9.7). Ma allora $L(A)$ è iniettiva: infatti se $L(A)(x) = 0$, si ha anche $L(B)(L(A)(x)) = L(B)(0) = 0$, e dunque $x \in \ker(L(B) \circ L(A)) = \ker(\mathrm{id}_{K^n}) = (0)$. Essendo $L(A)$ un endomorfismo iniettivo di K^n che ha dimensione finita, $L(A)$ è un isomorfismo, e perciò A è invertibile. Ora da $BA = E_n$ segue $B = (BA)A^{-1} = E_n A^{-1} = A^{-1}$ ossia $B = A^{-1}$.

Se invece $AC = E_n$, $L(AC) = L(A) \circ L(C) = \mathrm{id}_{K^n}$. Allora $L(A)$ è suriettiva: se $y \in K^n$, si ha $y = L(A)(L(C)(y))$ e dunque $y \in \operatorname{Im} L(A)$. Come nel caso precedente allora segue che $L(A)$ è un isomorfismo e A è invertibile. Inoltre da $AC = E_n$ segue che $A^{-1}(AC) = A^{-1}$ e dunque $C = A^{-1}$. $\square$

Questo risultato permette di cercare l'inversa di una matrice A imponendo a una matrice incognita X soltanto una delle due condizioni $AX = E_n$ o $XA = E_n$.

Sia dunque A una matrice $n \times n$ invertibile. Vogliamo trovare la sua inversa. Chiamiamo $X = (x_{ij})_{i,j=1,\dots,n}$ una matrice $n \times n$ di incognite: vogliamo risolvere l'equazione matriciale $AX = E_n$.

Risolvere questa equazione matriciale è equivalente a risolvere gli n sistemi lineari:

$$\begin{aligned} AX^1 &= e_1 \\ AX^2 &= e_2 \\ &\vdots \\ AX^n &= e_n \end{aligned} \tag{10.4}$$

dove $X^1, X^2, \dots, X^n$ sono le colonne di X. Dobbiamo dunque risolvere n sistemi lineari con la stessa matrice dei coefficienti A, le cui colonne dei termini noti sono $e_1, \dots, e_n$. Poichè A è invertibile ognuno di tali sistemi ha una e una sola soluzione. Per risolverli possiamo usare l'algoritmo di Gauss, riducendo a gradini le matrici complete degli n sistemi, che sono $(A \mid e_i)$, per ogni $i = 1, \dots, n$. Le trasformazioni elementari da eseguire sono le stesse per tutti i sistemi, quindi possiamo eseguirle per tutti i sistemi contemporaneamente; partiamo dunque dalla matrice ottenuta aggiungendo ad A come colonne tutti gli n vettori della base canonica di K^n, ossia la matrice identica E_n, e operiamo con trasformazioni elementari in modo da arrivare a una matrice $(B \mid C)$ dove B è una matrice $n \times n$ a gradini con n pivot, dunque una matrice triangolare superiore:

$$\begin{pmatrix} a_{11} & \dots & a_{1n} & | & 1 & 0 & \dots & 0 \\ a_{21} & \dots & a_{2n} & | & 0 & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & | & \vdots & \ddots & \vdots & \\ a_{n1} & \dots & a_{nn} & | & 0 & 0 & \dots & 1 \end{pmatrix} = (A \mid E_n) \ \to (B \mid C). \tag{10.5}$$

La matrice $(B \mid C)$ ci dà n sistemi lineari a gradini equivalenti ai primi n. Ora continuiamo operando di nuovo con trasformazioni elementari sulle righe ma a partire dal basso, per mandare a zero tutti gli elementi sopra la diagonale principale di B.

$$\begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} & | & c_{11} & \dots & c_{1n} \\ 0 & b_{22} & \dots & b_{2n} & | & c_{21} & \dots & c_{2n} \\ 0 & 0 & \ddots & \dots & | & \dots & \ddots & \dots \\ 0 & 0 & \dots & b_{nn} & | & c_{n1} & \dots & c_{nn} \end{pmatrix} \to$$

Essendo $b_{nn} \neq 0$ possiamo usare l'ultima riga $(b_n \mid c_n)$ per mandare a zero gli elementi dell'ultima colonna di B sopra a b_{nn}, cioè alla riga i-esima $(b_i \mid c_i)$ sostituiamo $(b_i - \frac{b_{in}}{b_{nn}} b_n \mid c_i - \frac{b_{in}}{b_{nn}} c_n)$:

$$\to \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1,n-1} & 0 & | & c'_{11} & \dots & c'_{1n} \\ 0 & b_{22} & \dots & b_{2,n-1} & 0 & | & c'_{21} & \dots & c'_{2n} \\ 0 & 0 & \ddots & \dots & 0 & | & \dots & \ddots & \dots \\ 0 & 0 & \dots & b_{n-1,n-1} & 0 & | & c'_{n-1,n} & \dots & \dots \\ 0 & 0 & \dots & 0 & b_{nn} & | & c_{n1} & \dots & c_{nn} \end{pmatrix} \to$$

L'ultima riga di $(B \mid C)$ e tutto il resto di B sono rimasti invariati, mentre C è cambiata; ora usiamo la penultima riga per mandare a zero gli elementi sopra a $b_{n-1,n-1}$, e così via finchè arriviamo a una matrice del tipo:

$$\rightarrow \begin{pmatrix} b_{11} & 0 & \dots & \dots & 0 & | & d_{11} & \dots & d_{1n} \\ 0 & b_{22} & 0 & \dots & 0 & | & d_{12} & \dots & d_{2n} \\ 0 & 0 & \dots & \dots & 0 & | & \dots & \ddots & \dots \\ 0 & 0 & \dots & 0 & b_{nn} & | & d_{n1} & \dots & d_{nn} \end{pmatrix}:$$

una matrice diagonale con tutti gli elementi della diagonale non nulli, seguita da una matrice D. Ora con trasformazioni elementari del I tipo trasformiamo la matrice diagonale di sinistra in E_n, e otteniamo una matrice $(E_n \mid A')$: A' risulta essere proprio A^{-1}, l'inversa di A. Infatti i sistemi lineari (10.4) sono stati trasformati nei sistemi equivalenti:

$$\begin{array}{rl} X^1 = & a'^1 \\ X^2 = & a'^2 \\ \vdots & \\ X^n = & a'^n \end{array} . \tag{10.6}$$

Abbiamo dunque descritto un algoritmo per la costruzione dell'inversa di una matrice invertibile.

Osserviamo che, se a priori non sappiamo se una matrice A è invertibile o meno, possiamo comunque eseguire i primi passi dell'algoritmo fino alla determinazione dei pivot di A, e così rispondere a questa prima domanda.

Esempio 10.5.2. Sia $A = \begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & -1 \\ 0 & 1 & 2 \end{pmatrix}$. Eseguiamo l'algoritmo.

$$\begin{pmatrix} \boxed{-1} & 1 & 0 & | & 1 & 0 & 0 \\ 1 & -1 & -1 & | & 0 & 1 & 0 \\ 0 & 1 & 2 & | & 0 & 0 & 1 \end{pmatrix} \overset{I}{\rightarrow} \begin{pmatrix} \boxed{-1} & 1 & 0 & | & 1 & 0 & 0 \\ 0 & 0 & -1 & | & 1 & 1 & 0 \\ 0 & 1 & 2 & | & 0 & 0 & 1 \end{pmatrix} \overset{IV}{\rightarrow} \begin{pmatrix} \boxed{-1} & 1 & 0 & | & 1 & 0 & 0 \\ 0 & \boxed{1} & 2 & | & 0 & 0 & 1 \\ 0 & 0 & \boxed{-1} & | & 1 & 1 & 0 \end{pmatrix} \rightarrow$$

Siamo arrivati a determinare i pivot di A e abbiamo così verificato che A ha rango 3 ed è perciò invertibile; ora proseguiamo con l'algoritmo per arrivare ad avere una matrice diagonale nella prima metà.

$$\overset{III}{\rightarrow} \begin{pmatrix} -1 & 1 & 0 & | & 1 & 0 & 0 \\ 0 & 1 & 0 & | & 2 & 2 & 1 \\ 0 & 0 & -1 & | & 1 & 1 & 0 \end{pmatrix} \overset{III}{\rightarrow} \begin{pmatrix} -1 & 0 & 0 & | & -1 & -2 & -1 \\ 0 & 1 & 0 & | & 2 & 2 & 1 \\ 0 & 0 & -1 & | & 1 & 1 & 0 \end{pmatrix} \overset{I}{\rightarrow} \begin{pmatrix} 1 & 0 & 0 & | & 1 & 2 & 1 \\ 0 & 1 & 0 & | & 2 & 2 & 1 \\ 0 & 0 & 1 & | & -1 & -1 & 0 \end{pmatrix}.$$

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

È sempre consigliabile a questo punto fare la verifica che

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

Esempio 10.5.3 (Esempio 10.3.3 - continua)**.** Calcoliamo l'inversa di

$M^{\mathcal{B}'}_{\mathcal{A}'}(\mathrm{id}_W) = \begin{pmatrix} 1 & -2 & 4 & -8 \\ 0 & 1 & -4 & 12 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, che è già triangolare superiore con tutti i pivot uguali a 1.

$$\left(\begin{array}{cccc|cccc} 1 & -2 & 4 & -8 & 1 & 0 & 0 & 0 \\ 0 & 1 & -4 & 12 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -6 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right) \overset{III}{\to} \left(\begin{array}{cccc|cccc} 1 & -2 & 4 & 0 & 1 & 0 & 0 & 8 \\ 0 & 1 & -4 & 0 & 0 & 1 & 0 & -12 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right) \overset{III}{\to}$$

$$\to \left(\begin{array}{cccc|cccc} 1 & -2 & 0 & 0 & 1 & 0 & -4 & -16 \\ 0 & 1 & 0 & 0 & 0 & 1 & 4 & 12 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right) \overset{III}{\to} \left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 2 & 4 & 8 \\ 0 & 1 & 0 & 0 & 0 & 1 & 4 & 12 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right).$$

Abbiamo trovata l'inversa di $M^{\mathcal{B}'}_{\mathcal{A}'}(\mathrm{id}_W)$. La matrice cercata è

$$M^{\mathcal{A}'}_{\mathcal{A}}(f) = M^{\mathcal{B}'}_{\mathcal{A}'}(\mathrm{id}_W)^{-1} M^{\mathcal{B}'}_{\mathcal{B}}(f) M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V) =$$

$$= \begin{pmatrix} 1 & 2 & 4 & 8 \\ 0 & 1 & 4 & 12 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} =$$

$$= \begin{pmatrix} 0 & 4 & 24 \\ 0 & 4 & 32 \\ 0 & 1 & 14 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 4 & 16 \\ 0 & 4 & 24 \\ 0 & 1 & 12 \\ 0 & 0 & 2 \end{pmatrix}.$$

Capitolo 11

Determinante

In questo capitolo introdurremo la nozione di determinante, su uno spazio vettoriale di dimensione finita o per matrici. Premettiamo alcune proprietà dei gruppi di permutazioni che ci saranno utili per dare le definizioni.

11.1 Gruppi di permutazioni

Sia $I_n = \{1, 2, \ldots, n\}$ l'insieme dei primi n numeri naturali.

Definizione 11.1.1. Una **permutazione** dell'insieme $\{1, 2, \ldots, n\}$ è una biiezione

$$\begin{array}{rll} \sigma: & I_n \to & I_n \\ & 1 \to & \sigma(1) \\ & 2 \to & \sigma(2) \\ & \vdots & \\ & n \to & \sigma(n) \end{array}$$

Per esempio, per $n = 3$, $I_3 = \{1, 2, 3\}$ ha sei permutazioni:

$$1\ 2\ 3,\ \ 1\ 3\ 2,\ \ 2\ 1\ 3,\ \ 2\ 3\ 1,\ \ 3\ 1\ 2,\ \ 3\ 2\ 1.$$

Denoteremo $\mathcal{S}_n$ l'insieme delle permutazioni di I_n: ha $n!$ elementi, in quanto ci sono n scelte per $\sigma(1)$, $n-1$ scelte per $\sigma(2)$, ecc.

Osserviamo che $\mathcal{S}_n$ è un gruppo rispetto alla composizione di applicazioni (Esempio 1.2.3), e che non è abeliano per $n \geq 3$. È detto **gruppo simmetrico su n oggetti**.

Esempio 11.1.2.

- Per $n = 1$ $\mathcal{S}_1 = \{*\}$ si riduce a un elemento;
- Per $n = 2$ $\mathcal{S}_2$ ha due elementi: $1\ 2,\ 2\ 1$; è isomorfo a $\mathbb{Z}_2$.

Introduciamo la seguente notazione per indicare come opera una permutazione σ:

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix}$$

Per esempio

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \ \tau = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}; \ \sigma \circ \tau = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \ \tau \circ \sigma = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix},$$

che conferma che $\mathcal{S}_3$ non è abeliano.

11.2 Cicli

Definizione 11.2.1. Un **ciclo o permutazione ciclica** di lunghezza k in $\mathcal{S}_n$ è una permutazione che manda successivamente $n_1 \to n_2 \to \cdots \to n_k \to n_1$, dove $n_1, n_2, \ldots, n_k$ sono k elementi distinti di I_n, mentre gli altri elementi di I_n vengono lasciati fissi. Lo si denota $(n_1\ n_2\ \ldots n_k)$.

Vediamo alcuni esempi.

Esempio 11.2.2. 1. Sia $\tau \in \mathcal{S}_7$ la permutazione

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 5 & 6 & 2 & 1 & 7 & 3 & 4 \end{pmatrix}$$

Osserviamo che $1 \to 5 \to 7 \to 4 \to 1$, e $2 \to 6 \to 3 \to 2$: il primo è un **ciclo o permutazione ciclica di lunghezza** 4 e il secondo un ciclo di lunghezza 3. Si usa la notazione $\tau = (1\ 5\ 7\ 4)(2\ 6\ 3)$.

2. Sia $\sigma \in \mathcal{S}_7$ la seguente permutazione:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 5 & 3 & 1 & 2 & 7 & 6 & 4 \end{pmatrix}$$

Osserviamo che $1 \to 5 \to 7 \to 4 \to 2 \to 3 \to 1$: questo è un **ciclo di lunghezza** 6, invece 6 rimane fisso in σ, è un ciclo di lunghezza 1. Si scrive $\sigma = (1\ 5\ 7\ 4\ 2\ 3)$, ossia l'elemento che rimane fisso non si scrive.

3.
$$\sigma' = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 1 & 5 & 4 & 7 & 3 & 6 \end{pmatrix}$$

Questa volta $1 \to 2 \to 1$: ho un ciclo di lunghezza 2 $(1\ 2)$; $3 \to 5 \to 7 \to 6 \to 3$: un ciclo di lunghezza 4 $(3\ 5\ 7\ 6)$; $4 \to 4$: ciclo di lunghezza 1. La permutazione data σ' si indica $(1\ 2)(3\ 5\ 7\ 6)$: è prodotto di due cicli disgiunti di lunghezza 2 e 4: sono permutabili.

Per la composizione di permutazioni si usa una notazione diversa rispetto a quella usata di solito per la composizione di applicazioni: anzichè scrivere per esempio $\sigma \circ \tau$ si usa $\tau\sigma$, per intendere che si applica prima τ e poi σ. Si parla di **prodotto di permutazioni**.

Per esempio: $(1\ 3\ 5)(3\ 1\ 7\ 4)(6\ 7\ 2)$ è un prodotto di tre cicli non disgiunti; è la permutazione

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 6 & 5 & 3 & 2 & 7 & 4 \end{pmatrix} = (2\ 6\ 7\ 4\ 3\ 5):$$

è un 6-ciclo.

Definizione 11.2.3. Un 2-ciclo è chiamato **trasposizione**. È del tipo $(n_1\ n_2)$, ossia n_1 e n_2 vengono scambiati e tutti gli altri elementi rimangono fissi.

Osservazione 18.
1. Osserviamo che $(3\ 5\ 1) = (5\ 1\ 3) = (1\ 3\ 5)$, ossia un ciclo può essere "iniziato" in una posizione qualunque.
2. Due cicli disgiunti commutano.

Proposizione 11.2.4. *1. Ogni permutazione è prodotto di cicli disgiunti.*

2. Ogni ciclo è prodotto di trasposizioni, in generale non disgiunte.

3. Ogni permutazione è prodotto di trasposizioni, in generale non disgiunte.

Dimostrazione. La 1. segue subito dalla definizione di ciclo.

Per dimostrare la 2. si ragiona per induzione sulla lunghezza $k \geq 2$ del ciclo: se $k = 2$ il ciclo è una trasposizione; supponiamo vera l'affermazione per cicli di lunghezza $k-1$, e sia $\sigma = (n_1\ n_2\ \dots\ n_k)$ un ciclo di lunghezza k. L'osservazione cruciale è che σ si può scrivere come prodotto di un $(k-1)$-ciclo τ e di una trasposizione ρ, nella forma $\sigma = \tau\rho = (n_1\ n_2\ \dots\ n_{k-1})(n_k\ n_1)$. Per vederlo meglio scriviamo per esteso come operano queste permutazioni, gli elementi non indicati rimangono fissi:

$$\tau\rho = \begin{pmatrix} n_1 & n_2 & \dots & n_{k-2} & n_{k-1} & n_k \\ n_2 & n_3 & \dots & n_{k-1} & n_1 & n_k \end{pmatrix} \begin{pmatrix} n_1 & n_k \\ n_k & n_1 \end{pmatrix}.$$

Quindi concludiamo usando l'ipotesi induttiva.

La 3. segue dalle precedenti. □

11.3 Segno di una permutazione e gruppo alternante

L'espressione di una permutazione come prodotto di trasposizioni non è unica. Introdurremo ora la nozione di inversione, che permetterà di associare a ogni permutazione un segno.

Definizione 11.3.1. Sia $\sigma \in \mathcal{S}_n$ una permutazione. **Un'inversione di** σ è una coppia di indici $i < j$ compresi fra 1 e n tali che $\sigma(i) > \sigma(j)$.

Per esempio il ciclo $\sigma = (1\ 2)$ ha una inversione, in quanto $1 < 2$ ma $\sigma(1) = 2 > 1 = \sigma(2)$. Il ciclo $\tau = (1\ 2\ 3)$ ha due inversioni: $\tau = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$; le coppie da considerare sono 3 e precisamente:

$1 < 2$ e $\sigma(1) = 2 < 3 = \sigma(2); 1 < 3$ ma $\sigma(1) = 2 > 1 = \sigma(3); 2 < 3$ ma $\sigma(2) = 3 > 1 = \sigma(3)$.

Definizione 11.3.2. Sia $\sigma \in \mathcal{S}_n$ una permutazione. **Il segno di** σ è $\operatorname{sgn}\sigma = (-1)^a$, dove a è il numero di inversioni di σ, ossia

$$\operatorname{sgn}\sigma = \begin{cases} 1 & \text{se } a \text{ è pari;} \\ -1 & \text{se } a \text{ è dispari.} \end{cases}$$

Una permutazione è detta pari se ha segno 1, dispari se ha segno -1. Le permutazioni pari sono quelle che presentano un numero pari di inversioni, mentre le permutazioni dispari presentano un numero dispari di inversioni.

Proposizione 11.3.3. *Ogni trasposizione ha segno -1, cioè è dispari.*

Dimostrazione. Sia $\sigma = (n_1\ n_2)$ una trasposizione. Scriviamo σ per esteso:

$$\begin{pmatrix} 1 & 2 & \dots & n_1-1 & n_1 & n_1+1 & \dots & n_2-1 & n_2 & n_2+1 & \dots & n \\ 1 & 2 & \dots & n_1-1 & n_2 & n_1+1 & \dots & n_2-1 & n_1 & n_2+1 & \dots & n \end{pmatrix}.$$

Se $i < j$ è un'inversione di σ, notiamo subito che nessuno fra i e j può essere compreso fra 1 e $n_1 - 1$ nè fra $n_2 + 1$ e n. Invece per ogni n_k con $n_1 < n_k < n_2$, n_k compare in due inversioni: $\sigma(n_1) = n_2 > \sigma(n_k) = n_k$; $\sigma(n_k) = n_k > \sigma(n_2) = n_1$. Infine $n_1 < n_2$ ma $\sigma(n_1) = n_2 > \sigma(n_2) = n_1$. Quindi complessivamente c'è un numero dispari di inversioni e $\operatorname{sgn}\sigma = -1$. □

Il prossimo teorema viene enunciato senza dimostrazione.

Teorema 11.3.4. *Date due permutazioni $\sigma, \tau \in \mathcal{S}_n$, si ha $\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)$. Quindi l'applicazione $\operatorname{sgn} : \mathcal{S}_n \to \{-1, 1\}$, dove $\{-1, 1\}$ è il gruppo moltiplicativo degli elementi invertibili di $\mathbb{Z}$, conserva il prodotto, ed è dunque un omomorfismo di gruppi.*

La dimostrazione segue da un lemma che afferma che $\operatorname{sgn}\sigma = \prod_{i<j} \frac{\sigma(j)-\sigma(i)}{j-i}$.

Corollario 11.3.5. *1. Se $\sigma = \tau_1 \cdots \tau_k$ è un prodotto di k trasposizioni, allora $\operatorname{sgn}\sigma = \prod_{i=1}^{k} \operatorname{sgn}(\tau_i) = (-1)^k$.*

2. Se σ è un k-ciclo, allora $\operatorname{sgn}\sigma = (-1)^{k-1}$.

3. Le permutazioni σ e σ^{-1} hanno lo stesso segno.

Dimostrazione. La 1. segue immediatamente dalla Proposizione 11.3.3 e dal Teorema 11.3.4. Per dimostrare la 2. procediamo per induzione su $k \geq 2$. Se $k = 2$, un 2-ciclo è una trasposizione e ha dunque segno -1 per la Proposizione 11.3.3. Supponiamo vera la tesi per i $(k-1)$-cicli e consideriamo un k-ciclo $\sigma = (n_1, \dots, n_k)$. Come nella dimostrazione della Proposizione 11.2.4, σ si scrive come prodotto di un $(k-1)$-ciclo τ e di una trasposizione ρ. Dunque per ipotesi induttiva $\operatorname{sgn}(\sigma) = (-1)^{k-2}(-1) = (-1)^{k-1}$. Per la 3., notiamo che $\operatorname{sgn}(\sigma\sigma^{-1}) = \operatorname{sgn}(\mathrm{id}_{I_n}) = 1 = \operatorname{sgn}(\sigma)\operatorname{sgn}(\sigma^{-1})$. □

Per esempio, $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 3 & 2 & 1 \end{pmatrix} = (1\ 5)(2\ 4)$ è prodotto di due trasposizioni, dunque $\operatorname{sgn}(\sigma) = (-1)^2 = 1$.

Il Corollario 11.3.5 permette di trovare facilmente il segno di una permutazione: la si decompone come prodotto di cicli disgiunti e poi si applica la parte 2 del Corollario 11.3.5.

Definizione 11.3.6. Sia $\mathcal{A}_n = \{\sigma \in \mathcal{S}_n \mid \sigma \text{ è pari}\}$: è un sottogruppo di $\mathcal{S}_n$, detto **sottogruppo alternante** di $\mathcal{S}_n$, o semplicemente **gruppo alternante**.

Infatti dal Teorema 11.3.4 segue che $\mathcal{A}_n$ è chiuso rispetto al prodotto e che se σ è pari anche σ^{-1} lo è. Inoltre ovviamente la permutazione identica è pari.

Invece l'insieme delle permutazioni dispari non è chiuso rispetto al prodotto: il prodotto di due permutazioni dispari è pari, mentre il prodotto di una pari e una dispari è dispari.

Osservazione 19. Se τ è una permutazione dispari fissata, l'insieme $\tau\mathcal{A}_n = \{\tau\sigma \mid \sigma \in \mathcal{A}_n\}$ coincide con l'insieme di **tutte** le permutazioni dispari. Infatti se α è una qualunque permutazione dispari, la si può scrivere nella forma $\alpha = \tau(\tau^{-1}\alpha)$, dove chiaramente $\tau^{-1}\alpha$ è pari. Osserviamo che di conseguenza l'insieme delle permutazioni dispari è in biiezione con $\mathcal{A}_n$, tramite l'applicazione $\mathcal{A}_n \to \tau\mathcal{A}_n$, tale che $\sigma \to \tau\sigma$.

Quindi $\mathcal{S}_n = \mathcal{A}_n \cup \tau\mathcal{A}_n$ e i due sottinsiemi sono fra loro in biiezione, dunque hanno lo stesso numero di elementi $\frac{n!}{2}$.

Esempio 11.3.7.

$$\mathcal{S}_3 = \{\mathrm{id}_{I_3}, (1\ 2), (1\ 3), (2\ 3), (1\ 2\ 3), (1\ 3\ 2)\} \supset \mathcal{A}_3 = \{\mathrm{id}_{I_3}, (1\ 2\ 3), (1\ 3\ 2)\} :$$

$\mathcal{A}_3$ è un gruppo abeliano con 3 elementi.

Elenchiamo ora i 24 elementi di $\mathcal{S}_4$. Le permutazioni dispari sono le trasposizioni e i 4-cicli:

$$(1\ 2), (1\ 3), (1\ 4), (2\ 3), (2\ 4), (3\ 4), (1\ 2\ 3\ 4), (1\ 2\ 4\ 3), (1\ 3\ 2\ 4), (1\ 3\ 4\ 2), (1\ 4\ 2\ 3)(1\ 4\ 3\ 2).$$

Scriviamo ora gli elementi di $\mathcal{A}_4$:

$$\mathrm{id}_{I_4}, (1\,2\,3), (1\,3\,2), (1\,2\,4), (1\,4\,2), (1\,3\,4), (1\,4\,3), (2\,3\,4), (2\,4\,3), (1\,2)(3\,4), (1\,3)(2\,4), (1\,4)(2\,3);$$

vi sono dunque, oltre alla permutazione identica, i 3-cicli e i prodotti di 2 2-cicli disgiunti.

11.4 Funzioni determinante

Iniziamo con un esempio. Consideriamo una matrice quadrata di ordine 2: $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$; per definizione il determinante di A è $\det(A) = a_{11}a_{22} - a_{12}a_{21}$. Come si è visto nel Foglio 2, Es. 1, $\det(A) \neq 0$ se e solo se le righe di A sono linearmente indipendenti, che equivale a dire che il rango di A è 2 o che A è invertibile. Possiamo interpretare det come un'applicazione che associa alla coppia delle righe di una matrice A lo scalare $\det(A)$, ossia come una funzione di due "variabili", che sono vettori di K^2. Questa funzione det è lineare rispetto a ogni riga ed è nulla se e solo se le righe sono linearmente dipendenti.

Generalizzeremo questo esempio, definendo che cosa si intende per una funzione determinante su uno spazio vettoriale V.

Definizione 11.4.1. Sia V un K-spazio vettoriale di dimensione n. Fissiamo un intero $k \geq 1$. Un'applicazione

$$D : V \times \cdots \times V = V^k \to K$$

è detta **multilineare o k-lineare** se è lineare in ogni argomento, cioè se per ogni $i = 1, \ldots, k$ si ha

$$D(v_1, \ldots, v_{i-1}, \lambda v_i + \mu v_i', v_{i+1} \ldots, v_k) = \lambda D(v_1, \ldots, v_i, \ldots, v_k) + \mu D(v_1, \ldots, v_i', \ldots, v_k).$$

D è detta **alternante** se $D(v_1, \ldots, v_k) = 0$ ogni volta che due degli argomenti sono uguali, ossia esistono indici $i \neq j$ tali che $v_i = v_j$.

Definizione 11.4.2. Una funzione determinante su V, K-spazio vettoriale di dimensione n, è un'applicazione multilineare alternante $D : V^n \to K$: in questo caso il dominio è la potenza n-esima di V dove $n = \dim V$.

Proposizione 11.4.3. *Sia D una funzione determinante su V. Allora:*

1. *D è* **antisimmetrica***: $D(v_1, \dots, v_i, \dots, v_j, \dots, v_n) = -D(v_1, \dots, v_j, \dots, v_i \dots, v_n)$: se scambiamo due degli argomenti la funzione cambia segno;*

2. *se $\sigma \in \mathcal{S}_n$, allora $D(v_{\sigma(1)}, \dots, v_{\sigma(n)}) = \operatorname{sgn} \sigma \; D(v_1, \dots, v_n)$;*

3. *se $v_1, \dots, v_n$ sono linearmente dipendenti, allora $D(v_1, \dots, v_n) = 0$. In particolare se uno degli argomenti $v_i = 0$, allora $D(v_1, \dots, v_n) = 0$.*

Dimostrazione. 1. Consideriamo $D(v_1, \dots, v_i + v_j, \dots, v_i + v_j, \dots, v_n) = 0$ perchè D è alternante. Per la multilinearità questo coincide con

$$D(\dots, v_i, \dots, v_i, \dots) + D(\dots, v_i, \dots, v_j, \dots) + D(\dots, v_j, \dots, v_i, \dots) + D(\dots, v_j \dots, v_j, \dots) =$$

$$\text{per l'alternanza} = D(v_1, \dots, v_i, \dots, v_j, \dots, v_n) + D(v_1, \dots, v_j, \dots, v_i, \dots, v_n) = 0.$$

2. Se σ è una trasposizione ha segno -1 e la proprietà segue dal punto 1. Altrimenti applichiamo la Proposizione 11.2.4 e il Corollario 11.3.5: se σ è un prodotto di k trasposizioni il suo segno è $(-1)^k$; ogni volta che operiamo con una trasposizione sugli argomenti, D cambia segno, dunque alla fine avremo la tesi.

3. Se $v_1, \dots, v_n$ sono linearmente dipendenti, uno di loro è combinazione lineare dei rimanenti (Proposizione 2.6.2). Supponiamo per semplicità che sia v_n: $v_n = \lambda_1 v_1 + \dots + \lambda_{n-1} v_{n-1}$. Allora

$$\begin{aligned} D(v_1, \dots, v_n) &= D(v_1, \dots, v_{n-1}, \lambda_1 v_1 + \dots + \lambda_{n-1} v_{n-1}) = \\ &= \sum_{i=1}^{n-1} \lambda_i D(v_1, \dots, v_{n-1}, v_i) = \\ &= \sum_{i=1}^{n-1} \lambda_i 0 = 0 \end{aligned}$$

perchè D è alternante. □

Esempio 11.4.4. Un esempio banale di funzione determinante è la funzione identicamente nulla: $D(v_1, \dots, v_n) = 0$ per ogni scelta di $(v_1, \dots, v_n) \in V^n$.

11.5 Formula di Leibniz per i determinanti

Fissiamo una funzione determinante su V e sia $\mathcal{B} = (v_1, \dots, v_n)$ una base di V. Per n vettori $w_1, \dots, w_n \in V$, vogliamo esprimere $D(w_1, \dots, w_n)$ in termini delle loro coordinate rispetto a $\mathcal{B}$. Scriviamo dunque ogni w_i come combinazione lineare di $v_1, \dots, v_n$ nella forma

$w_i = \sum_{j_i=1}^n x_{ij_i} v_{j_i}$. Abbiamo bisogno di usare indici diversi per ogni vettore perchè ora li faremo variare tutti n contemporaneamente:

$$\begin{aligned} D(w_1,\dots,w_n) &= D(\sum_{j_1=1}^n x_{1j_1}v_{j_1}, \sum_{j_2=1}^n x_{2j_2}v_{j_2},\dots,\sum_{j_n=1}^n x_{nj_n}v_{j_n}) = \text{multilinearità I argomento} \\ &= \sum_{j_1=1}^n x_{1j_1} D(v_{j_1}, \sum_{j_2=1}^n x_{2j_2}v_{j_2},\dots,\sum_{j_n=1}^n x_{nj_n}v_{j_n}) = \text{multilin. per gli altri} \\ &= \sum_{j_1=1}^n \sum_{j_2=1}^n \cdots \sum_{j_n=1}^n x_{1j_1}x_{2j_2}\cdots x_{nj_n} D(v_{j_1},\dots,v_{j_n}). \end{aligned}$$

Ma $D(v_{j_1},\dots,v_{j_n})$ può essere $\neq 0$ soltanto se gli indici $j_1,\dots,j_n$ sono tutti diversi, cioè se formano una permutazione di $1,\dots,n$. Quindi possiamo riscrivere così:

$$\begin{aligned} D(w_1,\dots,w_n) &= \sum_{\sigma\in\mathcal{S}_n} x_{1\sigma(1)}\dots x_{n\sigma(n)} D(v_{\sigma(1)},\dots,v_{\sigma(n)}) = \\ &= \sum_{\sigma\in\mathcal{S}_n} (\operatorname{sgn}\sigma) x_{1\sigma(1)}\dots x_{n\sigma(n)} D(v_1,\dots,v_n). \end{aligned} \tag{11.1}$$

L'espressione (11.1) è nota come **formula di Leibniz per il determinante**. È una somma con $n!$ addendi quindi è utilizzabile solo per n piccolo. Come conseguenza della formula di Leibniz, abbiamo il seguente corollario.

Corollario 11.5.1. *Sia D una funzione determinante non nulla su V. Allora $(v_1,\dots,v_n)$ è una base di V se e solo se $D(v_1,\dots,v_n)\neq 0$.*

Dimostrazione. Se $v_1,\dots,v_n$ non formano una base, sono linearmente dipendenti, dunque $D(v_1,\dots,v_n)=0$ per la Proposizione 11.4.3. Viceversa: essendo $D\neq 0$ esistono n vettori $w_1,\dots,w_n$ tali che $D(w_1,\dots,w_n)\neq 0$. Allora se $v_1,\dots,v_n$ formano una base, applicando la (11.1) otteniamo che dev'essere $D(v_1,\dots,v_n)\neq 0$. □

Il seguente teorema, molto importante, ci dà l'unicità della funzione determinante a meno di costanti.

Teorema 11.5.2. *Sia V un K-spazio vettoriale di dimensione n, $(v_1,\dots,v_n)$ una base di V. Allora esiste una e una sola funzione determinante su V, $D: V^n\to K$, tale che $D(v_1,\dots,v_n)=1$. Ogni altra funzione determinante è del tipo λD, con $\lambda\in K$.*

Dimostrazione. Come si fa usualmente, dimostriamo prima l'unicità e poi l'esistenza. Supponendo che esista una funzione determinante D tale che $D(v_1,\dots,v_n)=1$, la formula di Leibniz (11.1) ci dà che, presi n vettori $w_i=\sum_j x_{ij}v_j$, dev'essere

$$D(w_1,\dots,w_n) = \sum_{\sigma\in\mathcal{S}_n} (\operatorname{sgn}\sigma) x_{1\sigma(1)}\dots x_{n\sigma(n)}, \tag{11.2}$$

e perciò abbiamo l'unicità. Dobbiamo dunque dimostrare che, prendendo la (11.2) come definizione di D, otteniamo una funzione multilineare alternante tale che $D(v_1,\dots,v_n)=1$.

Per l'ultima affermazione osserviamo che $v_i = \sum_j \delta_{ij} v_j$ e quindi dalla (11.2) otteniamo $D(v_1, \dots, v_n) = \sum_{\sigma \in \mathcal{S}_n} (\operatorname{sgn} \sigma) \delta_{1\sigma(1)} \dots \delta_{n\sigma(n)} = 1$, perchè rimane un unico addendo corrispondente alla permutazione identica.

Verifichiamo che D è multilineare:

$$\begin{aligned} D(w_1, \dots, \lambda w_i + \mu w_i', \dots, w_n) &= \sum_{\sigma \in \mathcal{S}_n} (\operatorname{sgn} \sigma) x_{1\sigma(1)} \cdots (\lambda x_{i\sigma(i)} + \mu x'_{i\sigma(i)}) \dots x_{n\sigma(n)} = \\ &= \lambda D(w_1, \dots, w_i, \dots w_n) + \mu D(w_1, \dots, w_i', \dots w_n). \end{aligned}$$

Verifichiamo infine che D è alternante; consideriamo vettori $w_1, \dots, w_n$ con $w_i = w_j$ dove i, j sono due indici diversi. Ciò significa che $x_{ik} = x_{jk}$ per ogni indice k. Sia $\tau = (i\ j)$ la trasposizione che scambia proprio gli indici i e j. Allora $\mathcal{S}_n = \mathcal{A}_n \cup \tau \mathcal{A}_n$ (Osservazione 19). Quindi

$$\begin{aligned} D(w_1, \dots, w_i, \dots, w_j, \dots, w_n) &= \sum_{\sigma \in \mathcal{S}_n} x_{1\sigma(1)} \cdots x_{n\sigma(n)} = \\ &= \sum_{\sigma \in \mathcal{A}_n} x_{1\sigma(1)} \cdots x_{n\sigma(n)} - \sum_{\sigma \in \mathcal{A}_n} x_{1\sigma(\tau(1))} \cdots x_{n\sigma(\tau(n))}. \end{aligned}$$

Ma

$$\begin{aligned} x_{1\sigma(\tau(1))} \cdots x_{i\sigma(\tau(i))} \cdots x_{j\sigma(\tau(j))} \cdots x_{n\sigma(\tau(n))} &= \text{per definizione di } \tau \\ = x_{1\sigma(1)} \cdots x_{i\sigma(j)} \cdots x_{j\sigma(i)} \cdots x_{n\sigma(n)} &= \text{per l'ipotesi che } w_i = w_j \\ = x_{1\sigma(1)} \cdots x_{j\sigma(j)} \cdots x_{i\sigma(i)} \cdots x_{n\sigma(n)}, & \end{aligned}$$

quindi la somma è nulla. □

11.6 Determinante di una matrice

Consideriamo ora il caso in cui $V = K^n$ in cui è fissata la base canonica $\mathcal{C} = (e_1, \dots, e_n)$. Il determinante standard, denotato det, è l'unica funzione determinante

$$\det : K^n \times \cdots \times K^n = (K^n)^n \to K$$

tale che $\det(e_1, \dots, e_n) = 1$. In questo caso il dominio $(K^n)^n$ si interpreta come spazio delle matrici quadrate di ordine n, ossia se $(w_1, \dots, w_n) \in K^n \times \cdots \times K^n$ le n-uple $w_1, \dots, w_n$ sono pensate come le n **righe** di una matrice $n \times n$. Quindi $\det(E_n) = \det(e_1, \dots, e_n) = 1$: è il determinante della matrice identica E_n.

Se $A = (a_{ij})_{i,j=1,\dots,n}$ è una matrice $n \times n$, useremo la notazione

$$\det(A) = \det(\underbrace{a_1, \dots, a_n}_{\text{righe di } A}) = \mid A \mid .$$

Esempio 11.6.1 (Determinante delle matrici di ordine 2 e 3)**.**
Se A è una matrice di ordine 2, $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, per la regola di Leibniz si ha:

$$\det(A) = \det((a_{11}, a_{12}), (a_{21}, a_{22})) = \underbrace{a_{11}a_{22}}_{\sigma = \mathrm{id}_{I_2}} - \underbrace{a_{12}a_{21}}_{\sigma = (1\ 2)} .$$

Se A è una matrice di ordine 3, $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$, lo sviluppo del suo determinante secondo la regola di Leibniz contiene 6 addendi, corrispondenti alle 6 permutazioni di $\mathcal{S}_3$ (Esempio 11.3.7):

$$\det(A) = a_{11}a_{22}a_{33} - \underbrace{a_{11}a_{23}a_{32}}_{\sigma=(2\ 3)} + \underbrace{a_{12}a_{23}a_{31}}_{\sigma=(1\ 2\ 3)} - \underbrace{a_{12}a_{21}a_{33}}_{\sigma=(1\ 2)} + \underbrace{a_{13}a_{21}a_{32}}_{\sigma=(1\ 3\ 2)} - \underbrace{a_{13}a_{22}a_{31}}_{\sigma=(1\ 3)}.$$

La **regola di Sarrus** è un trucco per scrivere velocemente il determinante di una matrice 3×3 (*solo matrici* 3×3). Riscriviamo la matrice A, ripetendo poi le prime due colonne ed evidenziando tre diagonali:

$$\left(\begin{array}{ccc|cc} \color{red}{a_{11}} & \color{green}{a_{12}} & \color{blue}{a_{13}} & a_{11} & a_{12} \\ a_{21} & \color{red}{a_{22}} & \color{green}{a_{23}} & \color{blue}{a_{21}} & a_{22} \\ a_{31} & a_{32} & \color{red}{a_{33}} & \color{green}{a_{31}} & \color{blue}{a_{32}} \end{array}\right)$$

I primi tre addendi di $\det(A)$ sono i tre prodotti degli elementi sulle tre diagonali evidenziate: $a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}$. Ora dobbiamo invece sottrarre i tre prodotti degli elementi contenuti nelle tre diagonali secondarie evidenziate di seguito:

$$\left(\begin{array}{ccc|cc} a_{11} & a_{12} & \color{red}{a_{13}} & \color{green}{a_{11}} & \color{blue}{a_{12}} \\ a_{21} & \color{red}{a_{22}} & \color{green}{a_{23}} & \color{blue}{a_{21}} & a_{22} \\ \color{red}{a_{31}} & \color{green}{a_{32}} & \color{blue}{a_{33}} & a_{31} & a_{32} \end{array}\right)$$

Si ottiene: $-a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}$. Confrontando con l'Esempio 11.6.1, si vede che quello che si trova è proprio il determinante della matrice considerata.

Proposizione 11.6.2. *Sia A una matrice quadrata $n \times n$, e sia ${}^tA$ la sua trasposta. Allora $\det(A) = \det({}^tA)$.*

Dimostrazione. Sia $A = (a_{ij})_{i,j=1,\ldots,n}$, ${}^tA = (a'_{ij})_{i,j=1,\ldots,n}$ con $a'_{ij} = a_{ji}$ per ogni coppia di indici. Si ha:

$$\begin{aligned} \det({}^tA) &= \sum_{\sigma \in \mathcal{S}_n} (\operatorname{sgn}\sigma) a'_{1\sigma(1)} \cdots a'_{n\sigma(n)} = \\ &= \sum_{\sigma} (\operatorname{sgn}\sigma) a_{\sigma(1)1} \cdots a_{\sigma(n)n} = \text{cambiando eventualmente l'ordine} \\ &= \sum_{\sigma} (\operatorname{sgn}\sigma) a_{1\sigma^{-1}(1)} \cdots a_{n\sigma^{-1}(n)} = \\ &= \det(A) \quad \text{perchè } \sigma^{-1} \text{ descrive tutto } \mathcal{S}_n. \end{aligned}$$

□

Osservazione 20. Come conseguenza osserviamo che il determinante è una funzione multilineare alternante anche sulle colonne di A.

11.7 Comportamento di un determinante per trasformazioni elementari

Vediamo ora come cambia il determinante di una matrice A se si opera con trasformazioni elementari sulle righe. Scriviamo la matrice mettendo in evidenza le sue righe.

I tipo: $A = \begin{pmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{pmatrix} \xrightarrow{I} \begin{pmatrix} a_1 \\ \vdots \\ \lambda a_i \\ \vdots \\ a_n \end{pmatrix}$; per la multilinearità $\det \begin{pmatrix} a_1 \\ \vdots \\ \lambda a_i \\ \vdots \\ a_n \end{pmatrix} = \lambda \det \begin{pmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{pmatrix}$: **il determinante viene moltiplicato per λ**;

II tipo: $A = \begin{pmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{pmatrix} \xrightarrow{II} \begin{pmatrix} a_1 \\ \vdots \\ a_i + a_j \\ \vdots \\ a_n \end{pmatrix}$; di nuovo per la multilinearità $det \begin{pmatrix} a_1 \\ \vdots \\ a_i + a_j \\ \vdots \\ a_n \end{pmatrix} = \det \begin{pmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{pmatrix} + \det \begin{pmatrix} a_1 \\ \vdots \\ a_j \\ \vdots \\ a_n \end{pmatrix}$, ma il secondo determinante contiene due righe uguali, ed è perciò nullo; **una trasformazione elementare del II tipo lascia il determinante invariato**;

III tipo: $A = \begin{pmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{pmatrix} \xrightarrow{III} \begin{pmatrix} a_1 \\ \vdots \\ a_i + \lambda a_j \\ \vdots \\ a_n \end{pmatrix}$; per la multilinearità $det \begin{pmatrix} a_1 \\ \vdots \\ a_i + \lambda a_j \\ \vdots \\ a_n \end{pmatrix} = \det \begin{pmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{pmatrix} + \lambda \det \begin{pmatrix} a_1 \\ \vdots \\ a_j \\ \vdots \\ a_n \end{pmatrix}$, quindi anche **una trasformazione elementare del III tipo lascia il determinante invariato**;

IV tipo: due righe della matrice vengono scambiate, perciò **in una trasformazione elementare del IV tipo il determinante cambia segno**.

Segue subito il seguente Corollario.

Corollario 11.7.1. *Se A è una matrice quadrata di ordine n e λ è uno scalare, allora $det(\lambda A) = \lambda^n det(A)$. In particolare $det(-A) = (-1)^n det(A)$.*

Quanto abbiamo appena osservato riguardo alla trasformazioni elementari potrà essere applicato al calcolo di determinanti, alla luce della seguente proposizione.

Proposizione 11.7.2. *Sia A una matrice triangolare superiore, $A = (a_{ij})_{i,j=1,\dots,n}$ con $a_{ij} = 0$ per ogni coppia di indici tali che $i > j$:* $\begin{pmatrix} a_{11} & * & * & \cdots & * \\ 0 & a_{22} & * & \cdots & * \\ \vdots & & \ddots & & * \\ 0 & 0 & 0 & \cdots & a_{nn} \end{pmatrix}$. *Allora $det(A) = a_{11}a_{22}\cdots a_{nn}$: è il prodotto degli elementi che stanno sulla diagonale principale.*

Dimostrazione. $\det(A) = \sum_\sigma (\operatorname{sgn}\sigma) a_{1\sigma(1)} \dots a_{n\sigma(n)}$; osserviamo che se $\sigma \neq \mathrm{id}_{I_n}$ c'è almeno un indice i tale che $i > \sigma(i)$.

Per l'ipotesi che A sia triangolare superiore, questo implica che nel prodotto $a_{1\sigma(1)} \dots a_{n\sigma(n)}$ c'è almeno un fattore nullo, e quindi in $\det(A)$ rimane solo l'addendo corrispondente a id_{In}. □

La stessa proprietà vale anche per le matrici triangolari inferiori (con lo stesso ragionamento).

Allora per calcolare il determinante di una matrice A, possiamo trasformarla in una matrice a gradini con trasformazioni elementari, tenendo conto del fatto che quelle del II e III tipo lasciano il determinante invariato, quelle del IV tipo fanno cambiar segno al determinante, quelle del I tipo lo moltiplicano per uno scalare. Il risultato è una matrice triangolare superiore.

Esempio 11.7.3. Presentiamo un esempio di una matrice 4×4; lo sviluppo del suo determinante secondo la regola di Leibniz contiene $4! = 24$ addendi; con l'algoritmo appena esposto il calcolo è molto più rapido.

$$\begin{vmatrix} 0 & 3 & 1 & 2 \\ 5 & 1 & 3 & 3 \\ 0 & 2 & 0 & 0 \\ 1 & 3 & 1 & 1 \end{vmatrix} = \text{scambio due righe} = -\begin{vmatrix} \boxed{1} & 3 & 1 & 1 \\ 5 & 1 & 3 & 3 \\ 0 & 2 & 0 & 0 \\ 0 & 3 & 1 & 2 \end{vmatrix} = \text{III tipo} = -\begin{vmatrix} \boxed{1} & 3 & 1 & 1 \\ 0 & -14 & -2 & -2 \\ 0 & 2 & 0 & 0 \\ 0 & 3 & 1 & 2 \end{vmatrix} =$$

$$\text{IV e I tipo} = 2\begin{vmatrix} \boxed{1} & 3 & 1 & 1 \\ 0 & \boxed{1} & 0 & 0 \\ 0 & -14 & -2 & -2 \\ 0 & 3 & 1 & 2 \end{vmatrix} = \text{III tipo} = 2\begin{vmatrix} \boxed{1} & 3 & 1 & 1 \\ 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & -2 & -2 \\ 0 & 0 & 1 & 2 \end{vmatrix} = \text{I tipo} =$$

$$= -4\begin{vmatrix} \boxed{1} & 3 & 1 & 1 \\ 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & \boxed{1} & 1 \\ 0 & 0 & 1 & 2 \end{vmatrix} = \text{III tipo} = -4\begin{vmatrix} \boxed{1} & 3 & 1 & 1 \\ 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & \boxed{1} & 1 \\ 0 & 0 & 0 & \boxed{1} \end{vmatrix} = -4.$$

Abbiamo visto nel Corollario 11.5.1 che una funzione determinante non nulla, definita su V, non si annulla su vettori $v_1, \dots, v_n$ se e solo se questi formano una base. Ne segue:

Proposizione 11.7.4. *Sia A una matrice quadrata. Allora $det(A) = \begin{vmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{vmatrix} \neq 0$ se e solo se le righe di A formano una base di K^n, se e solo se A è invertibile.*

Proposizione 11.7.5. *Sia $A = \begin{pmatrix} B & \mid & * \\ --- & \mid & --- \\ 0 & \mid & C \end{pmatrix}$ una matrice a blocchi, con B, C matrici quadrate. Allora $det(A) = det(B)det(C)$.*

Dimostrazione. Lo si dimostra con trasformazioni elementari, trasformando B e C in matrici triangolari superiori senza moltiplicare nessuna riga per uno scalare, ma operando solo con trasformazioni elementari di II, III e IV tipo. Prima si passa da B a B' triangolare con trasformazioni elementari che comprendono k scambi di righe, lavorando solo sulle righe di B; poi si passa da C a C' triangolare, con trasformazioni elementari sulle righe di C, con l scambi di righe: $A \to \begin{pmatrix} B' & \mid & * \\ --- & \mid & --- \\ 0 & \mid & C' \end{pmatrix} = A'$. Allora $\det(A') =$ triangolare $= \det(B')\det(C') = (-1)^k\det(B)(-1)^l\det(C) = (-1)^{k+l}\det(A)$. Si conclude che $\det(A) = \det(B)\det(C)$. □

11.8 Teorema di Binet

La dimostrazione di questo importante teorema, di enunciato molto semplice, si basa sulla teoria dei determinanti che abbiamo costruito.

Teorema 11.8.1. *Siano A, B matrici quadrate $n \times n$ su K. Allora $det(AB) = det(A)det(B)$. In particolare l'applicazione $det : GL(n, K) \to K \setminus \{0\}$ è un omomorfismo di gruppi (moltiplicativi).*

Dimostrazione. Consideriamo dapprima il caso particolare in cui $\det(B) = 0$: allora le righe di B sono linearmente dipendenti, quindi il sistema lineare omogeneo $BX = 0$ ha soluzioni non nulle: esiste $v \neq 0$, $v \in K^n$, tale che $Bv = 0$. Allora anche $(AB)v = A(Bv) = A0 = 0$: ciò significa che anche il sistema lineare omogeneo $(AB)X = 0$ ha una soluzione non nulla, dunque AB ha rango $< n$, e quindi $\det(AB) = 0$.

Sia ora $\det(B) \neq 0$. Consideriamo l'applicazione

$$f_B : M(n \times n, K) \to K$$

definita da $f_B(A) = \frac{\det(AB)}{\det(B)}$. Vogliamo dimostrare che $f_B(A) = \det(A)$. Per il Teorema 11.5.2 basta dimostrare che:

1. f_B è multilineare e alternante nelle righe di A;
2. $f_B(E_n) = 1$.

La 2. è immediata: $f_B(E_n) = \frac{\det(E_nB)}{\det(B)} = 1$. Per dimostrare la 1. consideriamo una matrice A con $a_j = \lambda a'_j + \mu a''_j$; allora al posto d'indici jk di AB abbiamo $a_j b^k = (\lambda a'_j + \mu a''_j) b^k = \lambda(a'_j b^k) + \mu(a''_j b^k)$; perciò la riga j-esima di AB è una combinazione lineare di due righe. Ma il determinante è lineare nella j-esima componente, perciò $\det(AB) = \lambda \det(A'B) + \mu(A''B)$, dove A' è ottenuta da A sostituendo la riga a_j con a'_j, e similmente A''. Quindi $f_B(A) = \lambda f_B(A') + \mu f_B(A'')$. Inoltre: se A ha due righe uguali, anche AB ha due righe uguali, e quindi $\det(AB) = 0$, che implica $f_B(AB) = 0$.

Concludiamo così che $\det(A) = f_B(A)$. $\square$

Corollario 11.8.2. *Sia A una matrice quadrata invertibile. Allora $det(A^{-1}) = (detA)^{-1}$.*

Dimostrazione. $AA^{-1} = E_n$; allora per il Teorema di Binet 11.8.1, $\det(AA^{-1}) = 1 = \det(A)\det(A^{-1})$. $\square$

Corollario 11.8.3. *Se A, B sono matrici quadrate simili, allora $det(A) = det(B)$.*

Dimostrazione. Da $B = S^{-1}AS$ segue $\det(B) = \det(S^{-1}AS) = \det(S^{-1})\det(A)\det(S) =$ per il Corollario precedente $= \det(S)^{-1}\det(A)\det(S) = \det(A)$. $\square$

Il Corollario 11.8.3 permette di dare la definizione di determinante di un endomorfismo.

Definizione 11.8.4. Sia $f : V \to V$ un endomorfismo di un K-spazio vettoriale V di dimensione finita. Si definisce il **determinante di** f ponendo $\det(f) = \det M_{\mathcal{B}}(f)$, qualunque sia la base $\mathcal{B}$ di V considerata.

La definizione è ben posta perchè se si prende un'altra base $\mathcal{B}'$ di V, le matrici $M_{\mathcal{B}}(f)$ e $M_{\mathcal{B}'}$ di f sono simili (Sezione 10.4) e hanno perciò lo stesso determinante. Con questa definizione, abbiamo che f è un isomorfismo di V in sè se e solo se $\det(f) \neq 0$.

11.9 Matrice aggiunta

Sia $A = (a_{ij})_{i,j=1,\dots,n}$ una matrice $n \times n$. Per ogni i, j denotiamo A'_{ij} la seguente matrice $n \times n$:

$$\begin{array}{cc}
 & \begin{array}{ccccccccc} & & & & \overset{j}{\downarrow} & & & & \end{array} \\
\begin{array}{c} \\ \\ \\ i \to \\ \\ \\ \\ \end{array} &
\begin{pmatrix}
a_{11} & \dots & a_{1,j-1} & | & 0 & | & a_{1,j+1} & \dots & a_{1n} \\
\vdots & & & | & \vdots & | & & & \\
a_{i-1,1} & \dots & a_{i-1,j-1} & | & 0 & | & a_{i-1,j+1} & \dots & a_{i-1,n} \\
0 & & 0 & & 1 & & 0 & \dots & 0 \\
a_{i+1,1} & \dots & a_{i+1,j-1} & | & 0 & | & a_{i+1,j+1} & \dots & a_{i+1,n} \\
\vdots & & & | & \vdots & | & & & \\
a_{n1} & \dots & a_{n,j-1} & | & 0 & | & a_{n,j+1} & \dots & a_{nn}
\end{pmatrix}
\end{array} \tag{11.3}$$

Abbiamo sostituito all'i-esima riga e_j e alla j-esima colonna e_i. Osserviamo che, se procediamo con trasformazioni elementari del III tipo su A'_{ij} come segue: alla I riga sommiamo la riga i-esima moltiplicata per a_{1j}, alla II riga sommiamo la riga i-esima moltiplicata per a_{2j}, e così via, lasciando invariata soltanto la riga i-esima, otteniamo la matrice

$$\begin{pmatrix} a_1 \\ \vdots \\ a_{i-1} \\ e_j \\ a_{i+1} \\ \vdots \\ a_n \end{pmatrix}.$$

Con queste trasformazioni elementari il determinante non è cambiato.

Definizione 11.9.1. La matrice $\widetilde{A} = (\widetilde{a}_{ij})_{ij=1,\ldots,n}$, dove $\widetilde{a}_{ij} :=\mid A'_{ji} \mid$, è detta **matrice aggiunta di** A.

Notiamo che gli indici sono scambiati, dunque $\widetilde{A} = {}^t(\mid A'_{ij} \mid)_{i,j=1,\ldots,n}$.

Teorema 11.9.2. *Sia A una matrice quadrata. Si ha $\widetilde{A}A = A\widetilde{A} = det(A)E_n$.*

Dimostrazione. Denotiamo con b_{ik} l'elemento di $A\widetilde{A}$ di indici ik. Si ha:

$$b_{ik} = \sum_{j=1}^{n} a_{ij}\widetilde{a}_{jk} = \sum_{j=1}^{n} a_{ij} \mid A'_{kj} \mid = \sum_{j=1}^{n} a_{ij} \det \begin{pmatrix} a_1 \\ \vdots \\ a_{k-1} \\ e_j \\ a_{k+1} \\ \vdots \\ a_n \end{pmatrix} = \det \begin{pmatrix} a_1 \\ \vdots \\ a_{k-1} \\ \sum_{j=1}^{n} a_{ij}e_j \\ a_{k+1} \\ \vdots \\ a_n \end{pmatrix}.$$

L'ultima uguaglianza segue dalla linearità del determinante rispetto al k-esimo argomento. Ma $\sum_j a_{ij}e_j = a_i$: è la riga i-esima di A che ora va nella posizione k-esima. Quindi si trova:

$$b_{ik} = \det \begin{pmatrix} a_1 \\ \vdots \\ a_{k-1} \\ a_i \\ a_{k+1} \\ \vdots \\ a_n \end{pmatrix} = \begin{cases} 0 & \text{se } i \neq k \\ \det(A) & \text{se } i = k \end{cases} = \det(A)\delta_{ik}.$$

Perciò $A\widetilde{A} = \det(A)E_n$. Per calcolare $\widetilde{A}A$ si lavora in maniera analoga sulle colonne, ricordando che il determinante è multilineare alternante anche sulle colonne della matrice (Osservazione 20). □

Corollario 11.9.3 (Formula per l'inversa di una matrice)**.** *Se A è una matrice quadrata invertibile, si ha*

$$A^{-1} = \frac{1}{det(A)}\widetilde{A}.$$

Notiamo che si tratta di una formula che dà direttamente l'espressione dell'inversa di una matrice, da distinguersi dall'algoritmo della Sezione 10.5.

Esempio 11.9.4. Sia A una matrice 2×2. Se

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

$$\widetilde{a}_{11} = \begin{vmatrix} 1 & 0 \\ 0 & a_{22} \end{vmatrix} = a_{22},\ \widetilde{a}_{22} = \begin{vmatrix} a_{11} & 0 \\ 0 & 1 \end{vmatrix} = a_{11},\ \widetilde{a}_{21} = \begin{vmatrix} 0 & 1 \\ a_{21} & 0 \end{vmatrix} = -a_{21},\ \widetilde{a}_{12} = -a_{12}.$$

Dunque

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

11.10 Sviluppo di Laplace di un determinante

Data una matrice quadrata A $n \times n$, denotiamo con A_{ij} la sua sottomatrice $(n-1) \times (n-1)$ ottenuta cancellando la riga i-esima e la colonna j-esima di A. A_{ij} è detta sottomatrice complementare dell'elemento a_{ij} di A e il suo determinante **minore complementare di** a_{ij}.

Confrontiamo il determinante della matrice A_{ij} con quello della matrice A'_{ij} definita nella Sezione 11.9.

Proposizione 11.10.1. *Si ha* $det(A'_{ij}) = \widetilde{a}_{ji} = (-1)^{i+j} det(A_{ij}$.

Dimostrazione. Infatti da A'_{ij} con $i-1$ scambi di righe e $j-1$ scambi di colonne passiamo alla matrice a blocchi

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & A_{ij} & \\ 0 & & & \end{pmatrix}$$

che ha lo stesso determinante di A_{ij}. Otteniamo quindi l'espressione voluta perchè $(i-1)+(j-1)$ ha la stessa parità di $i+j$. □

Definizione 11.10.2. L'elemento $(-1)^{i+j}\det(A_{ij})$ è detto **complemento algebrico o cofattore dell'elemento** a_{ij} di A. La matrice aggiunta è detta anche **matrice cofattore** di A.

Teorema 11.10.3 (I Teorema di Laplace per lo sviluppo di un determinante)**.**

Sia $A = (a_{ij})_{i,j=1,\dots,n}$ *una matrice* $n \times n$.

1. Sia i un indice di riga fissato. Allora

$$det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} det(A_{ij}),$$

ogni addendo è prodotto dell'elemento a_{ij} *per il suo complemento algebrico. Questo è detto sviluppo di* $det(A)$ *seconda la riga i-esima.*

2. Sia j un indice di colonna fissato. Allora

$$det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij}\, det(A_{ij}),$$

anche qui ogni addendo è prodotto dell'elemento a_{ij} per il suo complemento algebrico. Questo è detto sviluppo di $det(A)$ seconda la colonna j-esima.

Dimostrazione. Sappiamo che $A\widetilde{A} = (b_{ij})_{i,j} = \det(A)E_n$. Allora per ogni i

$$b_{ii} = \det(A) = \sum_{j=1}^{n} a_{ij}\widetilde{a}_{ji} = \sum_{j} a_{ij} \mid A'_{ij} \mid = \sum_{j} (-1)^{i+j} a_{ij} \mid A_{ij} \mid .$$

Se si considera $\widetilde{A}A$ si ottiene l'analoga espressione per le colonne. □

Esempio 11.10.4. 1. Consideriamo il seguente determinante:

$$\begin{vmatrix} 0 & 1 & 2 \\ 3 & 2 & 1 \\ 1 & 1 & 0 \end{vmatrix} = \text{sviluppiamo secondo la terza riga e otteniamo} =$$

$$= (-1)^{3+1} \cdot 1 \cdot \begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} + (-1)^{3+2} \cdot 1 \cdot \begin{vmatrix} 0 & 2 \\ 3 & 1 \end{vmatrix} + 0 = -3 + 6 = 3.$$

2. $\begin{vmatrix} 4 & 0 & 5 & 6 \\ 7 & 0 & 8 & 2 \\ 1 & 3 & 2 & 1 \\ 0 & 0 & -1 & 0 \end{vmatrix}$: lo sviluppiamo secondo la seconda colonna $= -3 \begin{vmatrix} 4 & 5 & 6 \\ 7 & 8 & 2 \\ 0 & -1 & 0 \end{vmatrix} =$

sviluppiamo secondo la terza riga: $= -3 \begin{vmatrix} 4 & 6 \\ 7 & 2 \end{vmatrix} = -3(8 - 42) = 102.$

3. Risoluzione dei sistemi lineari omogenei di due equazioni in tre incognite.
Consideriamo il sistema lineare

$$\begin{cases} ax + by + cz &= 0 \\ a'x + b'y + c'z &= 0 \end{cases} \quad \text{avente come matrice} \quad \begin{pmatrix} a & b & c \\ a' & b' & c' \end{pmatrix}.$$

Tale sistema ha la seguente soluzione $\left(\begin{vmatrix} b & c \\ b' & c' \end{vmatrix}, -\begin{vmatrix} a & c \\ a' & c' \end{vmatrix}, \begin{vmatrix} a & b \\ a' & b' \end{vmatrix} \right)$.

Per vederlo, considerare lo sviluppo di Laplace secondo la prima riga dei seguenti determinanti con due righe uguali:

$$\begin{vmatrix} a & b & c \\ a & b & c \\ a' & b' & c' \end{vmatrix} = 0, \quad \begin{vmatrix} a' & b' & c' \\ a & b & c \\ a' & b' & c' \end{vmatrix} = 0.$$

Una proprietà analoga vale per i sistemi lineari di $n-1$ equazioni in n incognite.

Osservazione 21. Come conseguenza del Teorema di Laplace 11.10.3, si ha che, se si fissano due indici di riga diversi $i \neq k$, allora

$$\sum_{j=1}^{n}(-1)^{i+j}a_{ij}\det(A_{kj}) = 0, \tag{11.4}$$

ossia se si sommano i prodotti degli elementi di una riga per i complementi algebrici degli elementi di un'altra riga, si ottiene 0. In effetti la (11.4) si può pensare come lo sviluppo di Laplace del determinante di una matrice con due righe uguali, cioè la matrice ottenuta sostituendo in A alla riga k-esima la riga i-esima. Questo è talvolta citato come **II Teorema di Laplace.**

La proprietà analoga vale anche per le colonne.

11.11 Teorema di Cramer

Consideriamo un sistema lineare di n equazioni in n incognite la cui matrice dei coefficienti A ha rango massimo n. Equivalentemente A è invertibile, o ha determinante non nullo; si dice anche che A è non singolare o non degenere. In tal caso il sistema ha una e una sola soluzione qualunque sia la colonna dei termini noti (Sezione 5.7). Il Teorema di Cramer fornisce una formula esplicita per l'unica soluzione.

Teorema 11.11.1. *Sia A una matrice $n \times n$ invertibile a coefficienti in K, siano $a^1, \dots, a^n$ le sue colonne e sia $Ax = b$ un sistema lineare con $b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \in K^n$. Allora l'unica soluzione del sistema è la n-upla $(y_1, \dots, y_n)$ di coordinate*

$$y_j = \frac{det(a^1, \cdots, a^{j-1}, b,\ a^{j+1}, \cdots, a^n)}{det(A)}$$

per ogni $j = 1, \dots, n$: è un quoziente di due determinanti, a denominatore quello di A, a numeratore quello della matrice ottenuta sostituendo alla colonna j-esima di A la colonna dei termini noti.

Dimostrazione. Sia $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ la soluzione. Allora $Ay = b$; ciò equivale a

$$y_1a^1 + y_2a^2 + \cdots + y_na^n = b.$$

Perciò $\det(a^1, \cdots, a^{j-1}, b,\ a^{j+1}, \cdots, a^n) = \det(a^1, \cdots, a^{j-1},\ y_1a^1 + y_2a^2 + \cdots + y_na^n,\ a^{j+1}, \cdots, a^n) =$ per la multilinearità $= \sum_{i=1}^{n} y_i\ \det(a^1, \cdots, a^{j-1}, a^i,\ a^{j+1}, \cdots, a^n) = y_j\det(A)$. Essendo $\det(A) \neq 0$, possiamo dividere per $\det(A)$ e otteniamo la tesi. $\square$

Osserviamo che l'unica soluzione è anche esprimibile come $y = A^{-1}b$.

11.12 Rango e minori di una matrice

Definizione 11.12.1. I minori di una matrice sono i determinanti delle sue sottomatrici quadrate. Si dice che un minore ha ordine s se è il determinante di una sottomatrice $s \times s$.

Vale la seguente caratterizzazione del rango di una matrice.

Proposizione 11.12.2. *Sia $A \in M(m \times n, K)$ una matrice $m \times n$. Allora A ha rango r se e solo se A ha almeno un minore non nullo di ordine r e ogni minore di A di ordine $s > r$ è nullo. Equivalentemente, il rango di A è il massimo ordine di un minore non nullo di A.*

Dimostrazione. Se A ha rango r, A contiene r righe linearmente indipendenti; possiamo supporre che siano $a_1, \cdots, a_r$. Allora la sottomatrice $A' = \begin{pmatrix} a_1 \\ \vdots \\ a_r \end{pmatrix}$ formata dalle prime r righe di A è una matrice $r \times n$ di rango r. Quindi A' ha r colonne linearmente indipendenti: questo ci dà una sottomatrice $r \times r$ invertibile di A. Dato che, per ipotesi, se $s > r$, s righe o colonne di A comunque prese sono linearmente dipendenti, ogni sottomatrice quadrata di ordine s ha righe e colonne linearmente dipendenti.

Viceversa, supponiamo che la sottomatrice formata dalla prime r righe e colonne di A sia invertibile; allora le prime r righe e colonne di A sono linearmente indipendenti, quindi $\operatorname{rg} A \geq r$. Ma $\operatorname{rg} A$ non può essere $> r$ altrimenti per la precedente implicazione potrei trovare un minore non nullo di ordine $> r$. □

Si può dimostrare il seguente teorema di Kronecker, o degli orlati, che permette di semplificare il procedimento espresso dal Teorema 11.12.2. In una matrice A, considerata una sua sottomatrice quadrata B di ordine p con determinante diverso da zero, si definiscono **orlati di** B tutte le sottomatrici quadrate di ordine $p+1$, ottenute aggiungendo a B (parte di) una riga e una colonna di A. Il Teorema afferma che, se tutti gli orlati di B hanno determinante nullo, allora $\operatorname{rg} A = p$.

11.13 Il determinante come area o volume

Siano $v_1, \ldots, v_n$ vettori non nulli di $\mathbb{R}^n$.

Definizione 11.13.1. Il parallelepipedo generato da $v_1, \ldots, v_n$ è il sottinsieme di $\mathbb{R}^n$:

$$P(v_1, \ldots, v_n) := \{v = \lambda_1 v_1 + \ldots \lambda_n v_n \mid 0 \leq \lambda_i \leq 1,\ i = 1, \ldots, n\}.$$

Per esempio, per $n = 1$ troviamo $P(v) = \{\lambda v \mid 0 \leq \lambda \leq 1\} \subset \mathbb{R}$: è il segmento corrispondente al vettore v. Per $n = 2$, troviamo il parallelogramma $P(v_1, v_2)$ avente i due vettori v_1, v_2 come lati. Se v_1, v_2 sono linearmente dipendenti, il parallelogramma è detto degenere, è il caso di due vettori paralleli.

Proposizione 11.13.2. *Siano $v_1, v_2 \in \mathbb{R}^2$ vettori linearmente indipendenti. L'area di $P(v_1, v_2)$ è il valore assoluto del determinante dei due vettori: $|\ det \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}\ |$.*

Cenno di dimostrazione. Consideriamo il parallelogramma di lati v_1, v_2 come in figura. L'area è il prodotto "base $\times$ altezza". Modifichiamo ora v_1 senza cambiare tale area, sostituiamo cioè a v_1 un vettore $v_1' = v_1 + \lambda v_2$, dove λ è scelto in modo che $v_1 + \lambda v_2$ sia parallelo a e_1, primo vettore della base canonica, quindi $v_1' = ae_1 = (a, 0)$. Osserviamo che passando da $P(v_1, v_2)$ a $P(v_1', v_2)$ l'area non cambia perchè non cambia l'altezza perpendicolare a v_2.

Ripetiamo ora sostituendo a v_2 un vettore $v_2' = \mu v_1' + v_2$ con μ scelto in modo che v_2' risulti parallelo a e_2, quindi $v_2' = be_2 = (0, b)$. Questa volta non cambia l'altezza perpendicolare a v_1', dunque l'area è invariata.

Abbiamo ottenuto $v_1' = ae_1 = (a, 0)$, $v_2' = be_2 = (0, b)$, con $a, b \in \mathbb{R}$; sono i lati di un rettangolo di area $\mid a \mid\mid b \mid$. D'altra parte

$$\det \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \det \begin{pmatrix} v_1 + \lambda v_2 \\ v_2 \end{pmatrix} = \det \begin{pmatrix} v_1' \\ v_2 + \mu v_1' \end{pmatrix} = \det \begin{pmatrix} v_1' \\ v_2' \end{pmatrix} = \det \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} = ab.$$

$\square$

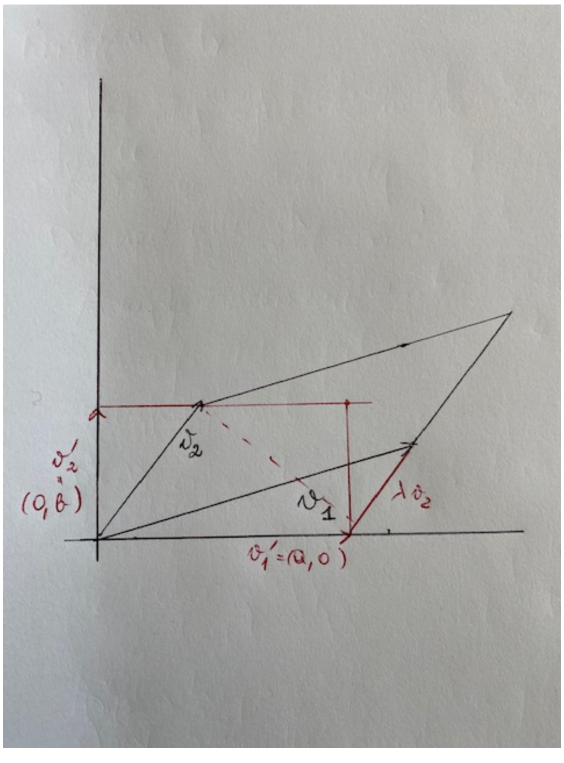

Figura 11.1: Parallelogramma di lati v_1, v_2 e rettangolo di lati v_1', v_2'

Questo ragionamento si estende a dimensione n qualunque: si definisce per induzione il volume del parallelepipedo $P(v_1, \dots, v_n)$ come valore assoluto del determinante $\det(v_1, \dots, v_n)$: è il prodotto del volume n-dimensionale di una base per la relativa altezza.

Se $f : \mathbb{R}^2 \to \mathbb{R}^2$ è un endomorfismo del tipo $f = L(A)$, dove A una matrice 2×2, e $v_1, v_2 \in \mathbb{R}^2$ sono due vettori, si dimostra con un conto diretto che $\det(f(v_1), f(v_2)) = \det(A)\det(v_1, v_2)$. Dunque il determinante di A, che coincide con il determinante di f, misura il rapporto fra le aree dei due parallelogrammi, di lati v_1, v_2 e $f(v_1), f(v_2)$.

Capitolo 12

Diagonalizzazione

Ci occuperemo ora del problema di trovare un rappresentante "naturale" nella classe di similitudine di una data matrice A di $M(n \times n, K)$. La prima domanda che ci poniamo è quando la classe di A contiene una matrice diagonale. Come vedremo la risposta dipende strettamente dal campo K su cui si lavora.

Fissiamo dunque un campo K .

12.1 Autovalori, autovettori e autospazi

Definizione 12.1.1 (Autovalori e autovettori)**.**

1. Sia V un K-spazio vettoriale e sia $f : V \to V$ un endomorfismo di V. Uno scalare $\lambda \in K$ è un **autovalore di** f se esiste un vettore $v \in V$ **non nullo** tale che $f(v) = \lambda v$. In tal caso v è detto **autovettore di** f **di autovalore** λ**.**

2. Sia $A \in M(n \times n, K)$ una matrice quadrata. Uno scalare λ è un **autovalore di** A se esiste un vettore $x \in K^n$ **non nullo** tale che $Ax = \lambda x$. In tal caso x è detto **autovettore di** A **di autovalore** λ**.**

Chiaramente dire che λ, v sono rispettivamente un autovalore e un autovettore di A equivale a dire che lo sono dell'endomorfismo $L(A) : K^n \to K^n$.

Notiamo che **gli autovettori sono vettori non nulli**, altrimenti la nozione di autovalore sarebbe banale. Infatti, per ogni $\lambda \in K$, indicato con 0_V il vettore nullo di V, si ha $f(0_V) = 0_V = \lambda 0_V$.

Osservazione 22. Sia V uno spazio vettoriale di dimensione finita. λ è un autovalore di f se e solo se λ è un autovalore della matrice $M = M_{\mathcal{B}}(f)$, per ogni base $\mathcal{B}$ di V. Infatti se $x_1, \ldots, x_n$ sono le coordinate di v rispetto a $\mathcal{B}$, $f(v) = \lambda v$ se e solo se $MX = \lambda X$, dove $X = {}^t(x_1, \ldots, x_n)$.

Esempio 12.1.2. Sia $A = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$. Allora $Ax = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 0 \end{pmatrix}, \lambda x = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \end{pmatrix}$.

Dunque $Ax = \lambda x$ se e solo se $\begin{cases} 2x_1 = \lambda x_1 \\ 0 = \lambda x_2 \end{cases}$, se e solo se $\begin{cases} (2-\lambda)x_1 = 0 \\ \lambda x_2 = 0 \end{cases}$. Questo accade, con $x = (x_1, x_2) \neq 0$, se e solo se $\lambda = 2$ per ogni coppia del tipo $(x_1, 0)$, oppure $\lambda = 0$ per ogni coppia del tipo $(0, x_2)$: $\lambda_1 = 2$ e $\lambda_2 = 0$ sono gli autovalori di A.

Per $\lambda_1 = 2$, i relativi autovettori sono tutti i vettori non nulli con $x_2 = 0$; formano il complementare del vettore nullo nel sottospazio $< (1,0) >=< e_1 >\subset K^2$.

Se $\lambda_2 = 0$, i relativi autovettori sono tutti i vettori non nulli con $x_1 = 0$; formano il complementare del vettore nullo nel sottospazio $< (0,1) >=< e_2 >\subset K^2$.

In altre parole, la restrizione di $L(A)$ alla retta $x_2 = 0$ è la moltiplicazione per 2; mentre la restrizione alla retta $x_1 = 0$ è la mappa nulla.

Definizione 12.1.3 (Autospazi). Sia $f : V \to V$ un endomorfismo, $\lambda \in K$. L'insieme $\mathrm{Aut}(\lambda) = \{v \in V \mid f(v) = \lambda v\} = \{0_V\} \cup \{\text{autovettori di } \lambda\}$ è detto **autospazio di** λ.

Analoga definizione nel caso di una matrice $A \in M(n \times n, K)$. L'autospazio di λ è $\mathrm{Aut}(\lambda) = \{x \in K^n \mid Ax = \lambda x\} = \{0_{K^n}\} \cup \{\text{autovettori di } \lambda\}$.

Osserviamo che $\mathrm{Aut}(\lambda)$ è un sottospazio vettoriale di V, risp. di K^n, e $\mathrm{Aut}(\lambda) \neq (0)$ se e solo se λ è autovalore. Infatti $0 \in \mathrm{Aut}(\lambda)$; inoltre se $v, w \in \mathrm{Aut}(\lambda)$ si ha $f(v) = \lambda v$, $f(w) = \lambda w$, dunque, presi due scalari $a, b \in K$, $f(av+bw) =$ linearità di $f = af(v)+bf(w) = a\lambda v + b\lambda w = \lambda(av + bw)$, perciò $av + bw \in \mathrm{Aut}(\lambda)$.

Definizione 12.1.4. La **molteplicità geometrica** dello scalare λ, denotata $m_g(\lambda)$, è la dimensione del relativo autospazio: $m_g(\lambda) = \dim \mathrm{Aut}(\lambda)$.

Quindi λ è un autovalore se e solo se $m_g(\lambda) \geq 1$.

Proposizione 12.1.5. *Siano $\lambda \neq \mu$ due autovalori distinti di f. Allora* $\mathrm{Aut}(\lambda) \cap \mathrm{Aut}(\mu) = (0)$.

Dimostrazione. Supponiamo per assurdo che $v \in \mathrm{Aut}(\lambda) \cap \mathrm{Aut}(\mu)$ sia un vettore non nullo; dunque v è autovettore sia di λ sia di μ. Allora $f(v) = \lambda v = \mu v$, quindi $\lambda v - \mu v = 0 = (\lambda - \mu)v$. Siccome v è non nullo, ciò implica $\lambda - \mu = 0$, il che è assurdo. □

Questa osservazione si generalizza nel seguente teorema.

Teorema 12.1.6. *Autovettori relativi ad autovalori distinti sono linearmente indipendenti. Ossia, se $\lambda_1, \ldots, \lambda_k$ sono autovalori di f, con $\lambda_i \neq \lambda_j$ per ogni $i \neq j$, e se $v_1, \ldots, v_k$ sono autovettori relativi rispettivamente a $\lambda_1, \ldots, \lambda_k$, allora $v_1, \ldots, v_k$ sono linearmente indipendenti.*

Dimostrazione. Per induzione su k. Se $k = 1$, il teorema è vero perchè ogni autovettore è non nullo.

Sia $k \geq 2$ e supponiamo che la tesi del teorema sia vera per $k-1$. Consideriamo una combinazione lineare nulla di $v_1, \ldots, v_k$:

$$\mu_1 v_1 + \cdots + \mu_k v_k = 0. \tag{12.1}$$

Allora $f(\mu_1 v_1 + \cdots + \mu_k v_k) = f(0) = 0 =$ per la linearità di $f = \mu_1 f(v_1) + \cdots + \mu_k f(v_k) \overset{\text{autovettori}}{=} \mu_1(\lambda_1 v_1) + \cdots + \mu_k(\lambda_k v_k) = \lambda_1(\mu_1 v_1) + \cdots + \lambda_k(\mu_k v_k)$. Per usare l'ipotesi induttiva sostituiamo in (12.1) a $\mu_k v_k$ l'espressione $-\mu_1 v_1 - \cdots - \mu_{k-1} v_{k-1}$ e otteniamo:

$$\lambda_1 \mu_1 v_1 + \cdots + \lambda_{k-1} \mu_{k-1} v_{k-1} - \lambda_k(\mu_1 v_1 + \cdots + \mu_{k-1} v_{k-1}) = 0$$

da cui

$$(\lambda_1\mu_1 - \lambda_k\mu_1)v_1 + \cdots + (\lambda_{k-1}\mu_{k-1} - \lambda_k\mu_{k-1})v_{k-1} = 0.$$

Per ipotesi induttiva $v_1, \ldots, v_{k-1}$ sono linearmente indipendenti, quindi otteniamo

$$\begin{cases} \lambda_1\mu_1 - \lambda_k\mu_1 = (\lambda_1 - \lambda_k)\mu_1 = 0 \\ \cdots \\ (\lambda_{k-1} - \lambda_k)\mu_{k-1} = 0 \end{cases}$$

Essendo gli autovalori tutti distinti $\lambda_1 - \lambda_k \neq 0, \ldots, \lambda_{k-1} - \lambda_k \neq 0$, e quindi $\mu_1 = \cdots = \mu_{k-1} = 0$. Allora la (12.1) diventa $\mu_k v_k = 0$, ma $v_k \neq 0$, e concludiamo che anche $\mu_k = 0$. □

Corollario 12.1.7. *Se f è un endomorfismo di V con $\dim V = n$, f non può avere più di n autovalori distinti. Analogamente, una matrice $n \times n$ non può avere più di n autovalori distinti.*

Osserviamo che $0 \in K$ può essere un autovalore di f (si veda l'esempio 12.1.2); ciò accade se e solo se esiste un vettore non nullo v tale che $f(v) = 0v = 0$, cioè se e solo se f non è iniettiva. In tal caso $\mathrm{Aut}(0) = \ker(f)$. Analogamente per una matrice A, 0 è un autovalore se e solo se $L(A)$ non è iniettiva, ossia $\ker L(A) \neq 0$ ossia $\mathrm{rg}(A) < n$. In tal caso $\mathrm{Aut}(0) = \ker L(A)$.

12.2 Ricerca degli autovalori e polinomio caratteristico

L'autospazio di uno scalare λ si può anche esprimere come nucleo di un endomorfismo. Infatti $\mathrm{Aut}(\lambda) = \ker(f - \lambda\,\mathrm{id}_V)$, dove

$$\begin{array}{rccl} f - \lambda\,\mathrm{id}_V : & V & \to & V \\ & v & \to & (f - \lambda\,\mathrm{id}_V)(v) = f(v) - \lambda v. \end{array}$$

Analogamente per le matrici: $x \in \mathrm{Aut}(\lambda)$, autospazio di una matrice A, se e solo se $x \in \ker(L(A) - \lambda\,\mathrm{id}_{K^n}) = \ker(L(A) - \lambda L(E_n)) = \ker(L(A - \lambda E_n))$. Quindi gli autovettori di λ sono le soluzioni non nulle del sistema lineare $(A - \lambda E_n)(x) = 0$.

Questa interpretazione ci dà un metodo per cercare gli autovalori di un endomorfismo o di una matrice. Se $f : V \to V$ è un endomorfismo di V, spazio vettoriale di dimensione finita, λ è un autovalore di f se e solo se $\mathrm{Aut}(\lambda) = \ker(f - \lambda\,\mathrm{id}_V)$ è non nullo, se e solo se $f - \lambda\,\mathrm{id}_V$ non è un isomorfismo, se e solo se il determinante $\det(f - \lambda\,\mathrm{id}_V) = 0$.

Se $\mathcal{B}$ è una base di V, $\det(f - \lambda\,\mathrm{id}_V) = \det(M_{\mathcal{B}}(f - \lambda\,\mathrm{id}_V)) = \det(M_{\mathcal{B}}(f) - \lambda M_{\mathcal{B}}(\mathrm{id}_V)) = \det(M_{\mathcal{B}}(f) - \lambda E_n)$. Posto $A = M_{\mathcal{B}}(f)$, concludiamo che

$$\lambda \text{ è autovalore di } f \text{ se e solo se } \det(A - \lambda E_n) = 0. \tag{12.2}$$

In particolare, se A è una matrice quadrata

$$\lambda \text{ è autovalore di } A \text{ se e solo se } \det(A - \lambda E_n) = 0. \tag{12.3}$$

Osservazione 23. Osserviamo che se A e A' sono matrici simili,

$$\det(A - \lambda E_n) = \det(A' - \lambda E_n).$$

Infatti se $A' = S^{-1}AS$, si ha $\det(S^{-1}AS - \lambda E_n) = \det(S^{-1}AS - \lambda S^{-1}E_nS) = \det(S^{-1}AS - S^{-1}(\lambda E_n)S) = \det(S^{-1}(A - \lambda E_n)S) = \det(A - \lambda E_n)$, per il Corollario 11.8.3 del Teorema di Binet.

Sia x un'indeterminata: consideriamo la matrice $A - xE_n$. Scritta per esteso è del tipo:

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} - x \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} = \begin{pmatrix} a_{11} - x & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - x & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - x \end{pmatrix} = (a_{ij} - x\delta_{ij})_{i,j=1,\dots,n}.$$

Sviluppiamo il determinante di $A - xE_n$ con la formula di Leibniz (Teorema 11.1):

$$\det(A - xE_n) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma)(a_{1\sigma(1)} - x\delta_{1\sigma(1)}) \cdots (a_{n\sigma(n)} - x\delta_{n\sigma(n)}).$$

Analizziamo i vari addendi:

- per $\sigma = \mathrm{id}_{I_n}$ abbiamo $(a_{11} - x)(a_{22} - x) \cdots (a_{nn} - x)$; sviluppando troviamo $(-1)^n x^n + (-1)^{n-1}(a_{11} + a_{22} + \cdots + a_{nn})x^{n-1} +$ termini di grado più basso;
- per ogni altra permutazione σ troviamo un prodotto di n fattori di cui al più $n - 2$ sono sulla diagonale principale di $A - xE_n$, in quanto c'è esattamente un fattore per ogni riga e uno per ogni colonna; quindi tale prodotto è un polinomio in x di grado $\leq n - 2$.

In definitiva possiamo scrivere

$$\det(A - xE_n) = (-1)^n x^n + (-1)^{n-1}(a_{11} + a_{22} + \cdots + a_{nn})x^{n-1} + \alpha_{n-2}x^{n-2} + \cdots + \alpha_1 x + \det(A); \tag{12.4}$$

l'ultimo termine è il termine noto che si ottiene ponendo $x = 0$, i coefficienti $\alpha_1, \dots, \alpha_{n-2}$ sono scalari. Abbiamo quindi ottenuto un polinomio di grado n in x, a coefficienti in K.

Definizione 12.2.1. $\det(A - xE_n) = \det(f - xE_n)$ è detto **polinomio caratteristico** di f, e denotato $p_f(x)$.

Per l'Osservazione 23, il polinomio $p_f(x)$ dipende solo da f, non dalla base scelta nè dalla relativa matrice $A = M_{\mathcal{B}}(f)$. In particolare, **dipendono soltanto da f tutti i coefficienti di $p_f(x)$, oltre al suo termine noto che è $\det(f)$**. Inoltre **dipendono solo da f anche le radici del polinomio caratteristico**, cioè gli elementi a di K tali che $p_f(a) = 0$. Il coefficiente di $(-1)^{n-1}x^{n-1}$ è $a_{11} + a_{22} + \cdots + a_{nn}$ è detto **traccia di** f; è la traccia della matrice A, cioè la somma degli elementi della diagonale principale.

Da questa discussione abbiamo ottenuto:

Proposizione 12.2.2. *Sia $f : V \to V$ un endomorfismo di un K-spazio vettoriale di dimensione finita. Gli autovalori di f sono le radici in K del polinomio caratteristico $p_f(x)$, ossia sono le soluzioni in K dell'equazione algebrica $p_f(x) = 0$.*

Se anzichè partire dall'endomorfismo f si parte da una matrice, il polinomio caratteristico di A è $p_A(x) = \det(A - xE_n)$.

Se λ è un autovalore di f, è una radice di $p_f(x)$, allora per il teorema di Ruffini $p_f(x)$ è divisibile per $x - \lambda$, cioè esiste un polinomio $q_1(x)$ di grado $n - 1$ tale che $p_f(x) = (x-\lambda)q(x)$. Ora possiamo considerare $q_1(\lambda)$: se è $\neq 0$ abbiamo finito, altrimenti applichiamo di nuovo il teorema di Ruffini e otteniamo che esiste un secondo polinomio $q_2(x)$ tale che $q_1(x) = (x - \lambda)q_2(x)$, e dunque $p_f(x) = (x - \lambda)^2 q_2(x)$. Possiamo ripetere il procedimento, al massimo per n volte, fino ad arrivare ad esprimere il polinomio caratteristico nella forma $p_f(x) = (x - \lambda)^a q(x)$ con $q(\lambda) \neq 0$. Tale numero a è detto **molteplicità algebrica dell'autovalore** λ e denotato $m_a(\lambda)$. E` il massimo esponente di $x - \lambda$ che divide $p_f(x)$.

12.3 Esempi I

Esempio 12.3.1. Riprendiamo l'Esempio 12.1.2. Il polinomio caratteristico della matrice $A = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$ è $p_A(x) = \mid A - xE_2 \mid = \begin{vmatrix} 2-x & 0 \\ 0 & -x \end{vmatrix} = -(2 - x)x = (x - 2)x$. Tale polinomio è già scomposto in due fattori della forma $x-\lambda$, da cui si vede che vi sono due autovalori uguali a 0 e 2, ciascuno con molteplicità algebrica 1. Ciò conferma quanto trovato in precedenza con un ragionamento diretto.

Esempio 12.3.2. Sia $V = \mathbb{R}^2$; consideriamo la matrice $R_\alpha := \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$ dove $\alpha \in \mathbb{R}$. L'endomorfismo di $\mathbb{R}^2$ $L(R_\alpha)$ è la rotazione di angolo α (Sezione 8.3). In tale rotazione i vettori della base canonica vengono mandati rispettivamente nei vettori $v_1 = (\cos\alpha, \sin\alpha)$, e $v_2 = (-\sin\alpha, \cos\alpha)$ (colonne di R_α). Il polinomio caratteristico è

$$p_{R_\alpha}(x) = \begin{vmatrix} \cos\alpha - x & -\sin\alpha \\ \sin\alpha & \cos\alpha - x \end{vmatrix} = x^2 - 2\cos\alpha\ x + 1.$$

Per trovare gli autovalori dobbiamo dunque risolvere l'equazione di II grado

$$x^2 - 2\cos\alpha\ x + 1 = 0. \tag{12.5}$$

Dal momento che $\frac{\Delta}{4} = \cos^2\alpha - 1 = -\sin^2\alpha$ è sempre ≤ 0, possiamo distinguere due casi:

1. $\sin^2\alpha \neq 0$: in tal caso l'equazione (12.5) non ha soluzioni reali, dunque la rotazione non ha autovalori nè autovettori; ciò si verifica se α non è un multiplo intero di π;

2. $\sin^2\alpha = 0$; in tal caso $\alpha = 2k\pi$ oppure $\alpha = \pi + 2k\pi$, $k \in \mathbb{Z}$. Nel primo caso $R_0 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = E_2$; $p_{R_0}(x) = (x - 1)^2$, dunque c'è il solo autovalore $\lambda = 1$ con $m_a(1) = 2$. $R_0 - E_2$ è la matrice nulla dunque tutto $\mathbb{R}^2$ è l'autospazio $\mathrm{Aut}(1)$, e anche $m_g(1) = 2$. Nel secondo caso $R_\pi := \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -E_2$; $p_{R_\pi}(x) = (x+1)^2$, dunque c'è il solo autovalore $\lambda = -1$ con $m_a(-1) = 2$. Anche questa volta $R_\pi + E_2 = 0$ dunque $\mathrm{Aut}(-1) = \mathbb{R}^2$ e anche $m_g(-1) = 2$.

Se cambiamo campo base e lavoriamo su $\mathbb{C}$ anzichè su $\mathbb{R}$, la matrice R_α definisce un endomorfismo denotato ancora $L(R_\alpha)$ di $\mathbb{C}^2$. Questa volta se α non è un multipo intero di π, l'equazione (12.5) ha due soluzioni complesse coniugate: ${}_1\lambda_2 = \cos\alpha \pm \sqrt{-\sin^2\alpha} = \cos\alpha \pm i\sin\alpha$. Quindi abbiamo due autovalori e due autospazi dell'endomorfismo complesso $L(R_\alpha)$.

12.4 Diagonalizzazione

Definizione 12.4.1. 1. Un endomorfismo $f : V \to V$ con $\dim V = n$, è detto **diagonalizzabile** o **semplice** se esiste una base $\mathcal{B} = (v_1, \ldots, v_n)$ di V formata da autovettori di f. Se $\lambda_1, \ldots, \lambda_n$ sono i loro autovalori, **non necessariamente distinti**, si ha $f(v_i) = \lambda_i v$ per ogni $i = 1, \ldots, n$. Allora

$$M_{\mathcal{B}}(f) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \tag{12.6}$$

è una matrice diagonale, che contiene sulla diagonale principale gli autovalori $\lambda_1, \ldots, \lambda_n$. In altre parole, esiste una base rispetto a cui la matrice di f è diagonale. Una tale base è detta **diagonalizzante**.

2. Una matrice quadrata A è **diagonalizzabile** se lo è $L(A)$, cioè se A è simile a una matrice diagonale, ossia esistono una matrice invertibile S e una matrice diagonale D tali che $D = S^{-1}AS$.

Osserviamo esplicitamente che in tal caso sulla diagonale principale di D compaiono gli autovalori di A, eventualmente con ripetizioni.

Teorema 12.4.2. *Sia $f : V \to V$ un endomorfismo,* $\dim V = n$*, e sia λ un autovalore di f. Allora $m_a(\lambda) \geq m_g(\lambda)$.*

Dimostrazione. Fissiamo una base $(v_1, \ldots, v_m)$ di $\mathrm{Aut}(\lambda)$, dove $m = m_g(\lambda)$. La prolunghiamo a una base $\mathcal{B}$ di V: $(v_1, \ldots, v_m, \ldots, v_n)$. Sia $A = M_{\mathcal{B}}(f)$. Si ha $f(v_1) = \lambda v_1, \ldots, f(v_m) = \lambda v_m$, perciò A è una matrice a blocchi della forma:

$$A = \left(\begin{array}{c|c} \lambda E_m & B \\ --- & --- \\ 0 & A' \end{array} \right)$$

dove B, A' sono matrici opportune, e A' è quadrata di ordine $n - m$. Allora il polinomio caratteristico di f è

$$p_f(x) = \mid A - xE_n \mid = \left| \begin{array}{c|c} (\lambda - x)E_m & B \\ --- & --- \\ 0 & A' - xE_{n-m} \end{array} \right| = (\lambda - x)^m \mid A' - xE_{n-m} \mid = (\lambda - x)^m g(x).$$

Dunque $m_g(\lambda) = m \leq m_a(\lambda)$, e vale la disuguaglianza stretta se e solo se $g(\lambda) = 0$. $\square$

Esempio 12.4.3. Sia $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Si ha $p_A(x) = \begin{vmatrix} -x & 1 \\ 0 & -x \end{vmatrix} = x^2$. Dunque l'unico autovalore è $\lambda = 0$ con $m_a(0) = 2$.

$\mathrm{Aut}(0) = \{(x_1, x_2) \mid Ax = 0\} = \{(x_1, x_2) \mid x_2 = 0\} =< e_1 >$: ha dimensione 1; dunque $m_g(0) < m_a(0)$.

Osserviamo che A non è diagonalizzabile, altrimenti sarebbe simile alla matrice nulla, perchè 0 è il suo unico autovalore; ma la classe di similitudine della matrice nulla chiaramente non contiene matrici non nulle.

Osservazione 24. Se $f : V \to V$ con $\dim V = n$ ha n **autovalori distinti** $\lambda_1, \dots, \lambda_n$, per il Teorema 12.1.6 n autovettori corrispondenti a tali autovalori sono linearmente indipendenti, perciò formano una base di V, dunque f **è diagonalizzabile**. In questo caso il polinomio caratteristico è della forma $p_f(x) = (-1)^n(x-\lambda_1)\cdots(x-\lambda_n) = (\lambda_1 - x)\cdots(\lambda_n - x)$, e tutte le molteplicità algebriche risultano uguali a 1. Per il Teorema 12.4.2 anche le molteplicità geometriche sono dunque uguali a 1. Gli autospazi hanno tutti dimensione 1. A risulta simile alla matrice

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \tag{12.7}$$

e a ogni matrice diagonale in cui gli elementi della diagonale principale compaiono in un ordine diverso.

Per esempio, da quanto visto nell'esempio 12.3.2, segue che sul campo complesso $\mathbb{C}$, la matrice di una rotazione

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

è simile alla matrice

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix},$$

dove $\lambda_1 = \cos\alpha + i\sin\alpha = e^{i\alpha}$, $\lambda_2 = \cos\alpha - i\sin\alpha = e^{-i\alpha}$ (abbiamo usato la notazione esponenziale per i numeri complessi).

Osservazione 25. Sia $f : V \to V$ un endomorfismo con $p_f(x) = (\lambda - x)^n$; allora si ha $m_a(\lambda) = n$ e f ha il solo autovalore λ. Dunque f è diagonalizzabile se e solo se V ha una base di autovettori tutti di autovalore λ, rispetto a cui la matrice di f è λE_n, $\mathrm{Aut}(\lambda) = V$ e $m_g(\lambda) = n$. In questo caso $f = \lambda\, \mathrm{id}_V$, ossia $f(v) = \lambda v$ per ogni vettore $v \in V$. Un tale endomorfismo è detto **omotetia di rapporto** λ. È rappresentato da λE_n rispetto a qualunque base, e in effetti se una matrice è simile a λE_n coincide con λE_n: $S^{-1}(\lambda E_n)S = \lambda(S^{-1}E_nS) = \lambda E_n$.

Enunciamo finalmente un teorema che ci dà (in due modi diversi) condizioni necessarie e sufficienti per la diagonalizzabilità di un endomorfismo.

Teorema 12.4.4. *Sia $f : V \to V$ un endomorfismo di un K-spazio vettoriale V di dimensione finita n. Le seguenti condizioni sono equivalenti:*

1) f è diagonalizzabile;

2) il polinomio caratteristico $p_f(x)$ si fattorizza in $K[x]$ come prodotto di n fattori lineari **non necessariamente distinti***: $p_f(x) = (\lambda_1 - x)(\lambda_2 - x) \cdots (\lambda_n - x)$, e per ogni autovalore λ_i si ha $m_a(\lambda_i) = m_g(\lambda_i)$;*

3) detti $\mu_1, \ldots, \mu_k$ gli autovalori distinti di f, si ha

$$V = \mathrm{Aut}(\mu_1) \oplus \mathrm{Aut}(\mu_2) \oplus \cdots \oplus \mathrm{Aut}(\mu_k). \tag{12.8}$$

Dimostrazione. 1) $\Rightarrow$ 2) Sia f diagonalizzabile, allora esiste una base $\mathcal{B}$ di autovettori di f e $M_{\mathcal{B}}(f)$ ha la forma (12.7). Quindi $p_f(x) = \mid M_{\mathcal{B}}(f) - xE_n \mid = (\lambda_1 - x)(\lambda_2 - x) \cdots (\lambda_n - x)$. Questo prova la prima asserzione. Consideriamo $m_a(\lambda_i)$: è pari al numero di volte che λ_i compare sulla diagonale principale di $M_{\mathcal{B}}(f)$, ossia al numero di vettori della base $\mathcal{B}$ che hanno λ_i come autovettore. D'altra parte $m_g(\lambda_i) = \dim \mathrm{Aut}(\lambda_i) = n - \mathrm{rg}(M_{\mathcal{B}}(f) - \lambda_i E_n)$: coincide con $m_a(\lambda_i)$.

2) $\Rightarrow$ 3) Mettiamo in evidenza nel polinomio caratteristico le molteplicità algebriche dei vari autovettori: $p_f(x) = (\mu_1 - x)^{a_1} \cdots (\mu_k - x)^{a_k}$, dove a_i è la molteplicità algebrica di μ_i. Osserviamo che, visto che $p_f(x)$ ha grado n, si ha $a_1 + a_2 + \cdots + a_k = n$. Per ipotesi, per ogni i,$a_i = m_a(\mu_i) = m_g(\mu_i)$. Perciò

$$\dim \mathrm{Aut}(\mu_1) + \cdots + \dim \mathrm{Aut}(\mu_k) = n = \dim V.$$

Ci rimane da dimostrare che la somma $\mathrm{Aut}(\mu_1) + \mathrm{Aut}(\mu_2) + \cdots + \mathrm{Aut}(\mu_k)$ è diretta, perchè se lo è ha dimensione n e quindi coincide con V. Ricordiamo che una somma di sottospazi $W_1, \ldots, W_r$ è diretta se ogni elemento di $W_1 + \cdots + W_r$ ha un'unica espressione $w_1 + \cdots + w_r$ come somma di vettori presi nei vari sottospazi considerati (vedere Sezione 4.3). Sia dunque $w_1 + \cdots + w_k = w_1' + \cdots + w_k'$, con $w_i, w_i' \in \mathrm{Aut}(\mu_i)$. Allora

$$(w_1 - w_1') + (w_2 - w_2') + \cdots + (w_k - w_k') = 0, \tag{12.9}$$

e ogni addendo $w_i - w_i' \in \mathrm{Aut}(\mu_i)$ (perchè è sottospazio). Alcuni addendi possono risultare nulli; se ce ne fossero di non nulli, la (12.9) sarebbe una combinazione lineare nulla non banale (con coefficienti uguali a 1) di autovettori relativi ad autovalori distinti: ma questo contraddice il Teorema 12.1.6, quindi in realtà gli addendi sono tutti nulli e per ogni i si ha $w_i = w_i'$, cioè l'unicità dell'espressione, e la somma è diretta.

3) $\Rightarrow$ 1) Se vale la (12.8), si può costruire una base di V facendo l'unione di basi dei sottospazi $\mathrm{Aut}(\mu_1), \ldots, \mathrm{Aut}(\mu_k)$ (Proposizione 4.3.3), si ha così una base di autovettori, quindi l'endomorfismo f è diagonalizzabile. $\square$

Esempio 12.4.5. Sia $A \in M(3 \times 3, \mathbb{R})$ la seguente matrice:

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

è diagonalizzabile? Consideriamo il suo polinomio caratteristico

$$p_A(x) = \begin{vmatrix} -x & -1 & 1 \\ -3 & -2-x & 3 \\ -2 & -2 & 3-x \end{vmatrix} = -x^3 + x^2 + x - 1.$$

È possibile fattorizzarlo in fattori lineari, infatti $-x^3 + x^2 + x - 1 = -x^2(x-1) + (x-1) = -(x^2-1)(x-1) = -(x-1)^2(x+1)$. Gli autovalori sono dunque

- $\lambda_1 = 1$ con $m_a(1) = 2$;
- $\lambda_2 = -1$ con $m_a(-1) = 1$ e quindi anche $m_g(-1) = 1$.

Dunque A è diagonalizzabile se e solo se $m_g(1) = 2$. $\mathrm{Aut}(1) = \ker(L(A - E_3))$, dove $A - E_3 = \begin{pmatrix} -1 & -1 & 1 \\ -3 & -3 & 3 \\ -2 & -2 & 2 \end{pmatrix}$: ha rango 1 perchè le sue righe sono tutte e tre proprozionali, dunque il nucleo ha dimensione 2, quindi $m_g(1) = 2$ e A è diagonalizzabile, e precisamente è simile alla matrice diagonale $D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$.

Vogliamo ora trovare una base $\mathcal{B} = (v_1, v_2, v_3)$ diagonalizzante per A, tale che $M_{\mathcal{B}}(L(A)) = D$. Dobbiamo prendere $v_1, v_2 \in \mathrm{Aut}(1)$, $v_3 \in \mathrm{Aut}(-1)$.

$\mathrm{Aut}(1)$ è lo spazio delle soluzioni dell'unica equazione $x_1 + x_2 - x_3 = 0$, corrispondente a una riga di $A - E_3$. Osserviamo che x_2, x_3 sono entrambe variabili libere, dunque una generica soluzione è del tipo $(-x_2 + x_3, x_2, x_3)$. Due soluzioni linearmente indipendenti si ottengono ponendo $(x_2, x_3) = (0, 1)$ oppure $(1, 0)$. Otteniamo così $v_1 = (1, 0, 1)$, $v_2 = (-1, 1, 0)$.

$\mathrm{Aut}(-1) = \ker(L(A+E_3))$, dove $A+E_3 = \begin{pmatrix} 1 & -1 & 1 \\ -3 & -1 & 3 \\ -2 & -2 & 4 \end{pmatrix}$. Operiamo con trasformazioni elementari per trasformarla in una matrice a gradini, verifichiamo così che ha rango 2 (come sapevamo) e troviamo una base per lo spazio delle soluzioni:

$$\mathrm{Aut}(-1) =< (1, 3, 2) >=< v_3 > .$$

Abbiamo dunque $\mathcal{B} = ((1, 0, 1), (-1, 1, 0), (1, 3, 2)) = (v_1, v_2, v_3)$.

Vogliamo infine determinare una matrice invertibile S e la sua inversa S^{-1}, tali che $D = S^{-1}AS$. S si può scrivere immediatamente perchè $S = M^{\mathcal{B}}_{\mathcal{C}}(\mathrm{id}_{\mathbb{R}^3})$ e dunque $S = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & 3 \\ 1 & 0 & 2 \end{pmatrix}$. Calcoliamo S^{-1} applicando la formula $S^{-1} = \frac{1}{\det(S)}\tilde{S}$, dove $\tilde{S}$ è l'aggiunta di S (Corollario 11.9.3). Otteniamo $\det(S) = -2$ e

$$S^{-1} = -\frac{1}{2} \begin{pmatrix} 2 & 2 & -4 \\ 3 & 1 & -3 \\ -1 & -1 & 1 \end{pmatrix}.$$

Se vogliamo fare la verifica, osserviamo che $D = S^{-1}AS$ se e solo se $SD = AS$.

Se prendiamo invece come base diagonalizzante $\mathcal{B}' = (v_3, v_1, v_2)$ troveremo anzichè D la matrice diagonale $D' = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Attenzione! Per trasformazioni elementari sulle righe di una matrice A, il polinomio caratteristico cambia, quindi non è possibile trasformare preliminarmente A in una matrice a gradini per semplificare il calcolo del polinomio caratteristico.

12.5 Esempi II

Esempio 12.5.1. Consideriamo l'applicazione $\rho : \mathbb{R}^2 \to \mathbb{R}^2$ definita da

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

Chiaramente ρ è un endomorfismo, perchè è del tipo $L(A)$, con $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. È la **riflessione rispetto all'asse** x che è detta asse della riflessione. ρ è già espressa nella forma $L(A)$ con A diagonale, e vediamo che ha due autovalori reali distinti $1, -1$. $\mathrm{Aut}(1)$ è l'asse della riflessione $< e_1 >$, mentre $\mathrm{Aut}(-1) =< e_2 >$.

Ora ruotiamo di $\alpha/2$ l'asse della riflessione, e otteniamo l'applicazione ρ_r, riflessione di asse r. Per scrivere equazioni di ρ_r, osserviamo che, se $\mathcal{B} = (v_1, v_2)$ è la base ottenuta dalla base canonica con la rotazione di $\alpha/2$, la matrice di ρ_r rispetto a $\mathcal{B}$ è $M_{\mathcal{B}}(\rho_r) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Allora $M_{\mathcal{C}}(\rho_r) = M_{\mathcal{B}}^{\mathcal{C}}(\mathrm{id}_{\mathbb{R}^2}) M_{\mathcal{B}}(\rho_r) M_{\mathcal{C}}^{\mathcal{B}}(\mathrm{id}_{\mathbb{R}^2})$, dove $M_{\mathcal{B}}^{\mathcal{C}}(\mathrm{id}_{\mathbb{R}^2}) = \begin{pmatrix} \cos\frac{\alpha}{2} & -\sin\frac{\alpha}{2} \\ \sin\frac{\alpha}{2} & \cos\frac{\alpha}{2} \end{pmatrix}$ è la matrice della rotazione di $\frac{\alpha}{2}$ e $M_{\mathcal{B}}^{\mathcal{C}}(\mathrm{id}_{\mathbb{R}^2}) = \begin{pmatrix} \cos\frac{\alpha}{2} & \sin\frac{\alpha}{2} \\ -\sin\frac{\alpha}{2} & \cos\frac{\alpha}{2} \end{pmatrix}$ è la matrice della trasformazione inversa, che è la rotazione di $-\frac{\alpha}{2}$. Eseguendo il prodotto otteniamo $M_{\mathcal{C}}(\rho_r) = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix} =: S_\alpha$. Dunque S_α è la matrice, rispetto alla base canonica, della riflessione di asse r che forma l'angolo di $\frac{\alpha}{2}$ con l'asse x. È una matrice a traccia nulla simile ad A. Il polinomio caratteristico è

$$p_{S_\alpha}(x) = \begin{vmatrix} \cos\alpha - x & \sin\alpha \\ \sin\alpha & -\cos\alpha - x \end{vmatrix} = x^2 - 1 = (x-1)(x+1).$$

Dunque per ogni α vi sono i due autovalori $1, -1$ con molteplicità algebrica e geometrica uguali a 1. Questo era aspettato perchè S_α è simile ad A (e quindi hanno lo stesso polinomio caratteristico). Dunque tutti gli endomorfismi ρ_r sono diagonalizzabili. I due autospazi sono: per $\lambda = 1$ la retta r di equazione $(\cos\alpha - 1)x + \sin\alpha y = 0$, per $\lambda = -1$ la retta di equazione $(\cos\alpha + 1)x + \sin\alpha y = 0$.

Esempio 12.5.2. Sia $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$; il suo polinomio caratteristico è $P_A(x) = \begin{vmatrix} -x & 1 \\ -1 & -x \end{vmatrix} = x^2 + 1$. A non è diagonalizzabile su $\mathbb{R}$, perchè x^2+1 non ha radici reali, ma lo è su $\mathbb{C}$. Infatti in $\mathbb{C}[x]$ $x^2+1 = (x+i)(x-i)$, quindi sul campo complesso A è simile alla matrice diagonale $\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$. $\mathrm{Aut}(i) = \ker L \begin{pmatrix} -i & 1 \\ -1 & -i \end{pmatrix}$, che ha rango 1, e ha dunque equazione $x_1 + ix_2 = 0$. Una soluzione è $(i, -1)$. Analogamente $\mathrm{Aut}(-i) =< (i, 1) >$. Una base diagonalizzante è $\mathcal{B} = ((i, -1), (i, 1))$. Si ha $\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} = S^{-1} A S$, con $S = \begin{pmatrix} i & i \\ -1 & 1 \end{pmatrix}$. Si ha $\det S = 2i$, dunque $S^{-1} = \frac{1}{2i} \begin{pmatrix} 1 & -i \\ i & i \end{pmatrix} = -\frac{1}{2} \begin{pmatrix} i & 1 \\ i & -1 \end{pmatrix}$.

Esempio 12.5.3. Sia J la matrice triangolare

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

Il suo polinomio caratteristico è $p_J(x) = (2-x)^3$, prodotto di fattori lineari in $\mathbb{R}[x]$ e in $\mathbb{Q}[x]$. L'unico autovalore è $\lambda = 2$ con $m_a(2) = 3$. L'autospazio $\mathrm{Aut}(2)$ ha dimensione $3 - \mathrm{rg}(J - 2E_3)$.

$$J - 2E_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}:$$

ha rango 2, perciò $m_g(2) = 1$, J non è diagonalizzabile su nessun campo.

12.6 Teorema fondamentale dell'Algebra

Abbiamo visto esempi di endomorfismi diagonalizzabili, ed esempi di endomorfismi che non lo sono, perchè non vale una o l'altra delle due condizioni necessarie e sufficienti del Teorema 12.4.4, punto 2. Negli esempi 12.3.2 e 12.5.2 sul campo reale, il polinomio caratteristico non si fattorizza in fattori lineari in $\mathbb{R}[x]$. Nell'esempio 12.5.3, pur fattorizzandosi $p_J(x)$ in fattori lineari su qualunque campo, non vale la condizione di uguaglianza delle due molteplicità algebrica e geometrica.

La condizione che il polinomio caratteristico $p_f(x)$ si fattorizzi in $K[x]$ come prodotto di n fattori lineari si esprime dicendo che $p_f(x)$ ha tutte le sue radici in K.

A tale proposito, enunciamo senza dimostrazione l'importante teorema:

Teorema 12.6.1 (Teorema fondamentale dell'Algebra)**.** *Sia $p(x) \in \mathbb{C}[x]$ un polinomio di grado $n \geq 1$ a coefficienti complessi. Allora $p(x)$ ha almeno una radice λ in $\mathbb{C}$.*

Dunque $p(x)$ è divisibile per $x - \lambda$, cioè $p(x) = (x - \lambda)p_1(x)$, con $p_1(x)$ di grado $n - 1$; $p_1(x)$ si trova eseguendo la divisione di polinomi. Ora, se $n - 1 \geq 1$, si può ripetere il procedimento, fino a concludere che $p(x)$ è prodotto di n fattori lineari in $\mathbb{C}[x]$, dunque $p(x)$ ha tutte le sue n radici in $\mathbb{C}$, purchè si conti ogni radice tante volte quant'è la sua molteplicità algebrica.

Tuttavia, anche se tutte le radici di qualunque polinomio $p(x)$ esistono in $\mathbb{C}$, non è sempre possibile calcolarle. Per polinomi di grado $n = 2$, esiste la ben nota formula risolutiva che esprime le soluzioni dell'equazione $p(x) = 0$ in funzione dei coefficienti di p, mediante un'estrazione di radice quadrata. Per gradi $n = 3, 4$ pure esistono della formule risolutive "per radicali", dette **formule di Cardano**, che esprimono le radici di un polinomio $p(x)$ in funzione dei coefficienti ricorrendo all'estrazione di radici terze e quarte. Invece, per gradi $n \geq 5$, è stato dimostrato che **non esiste una formula risolutiva per radicali.** Questo risultato, parzialmente dimostrato da Paolo Ruffini nei primi anni dell'800, fu ottenuto dal matematico norvegese Niels Abel nel 1826.

Vi sono tuttavia delle formule risolutive che fanno uso di altre funzioni, anzichè di radicali; nella pratica, però, dovendo risolvere equazioni specifiche, si ricorre a metodi numerici che forniscono le soluzioni approssimate, con il grado di approssimazione desiderato.

Se si lavora invece su campi $\mathbb{Z}_p$, che hanno un numero finito di elementi, le soluzioni di un'equazione si possono cercare, con un numero finito di tentativi, sostituendo via via all'incognita i vari elementi del campo.

Capitolo 13

Triangolarizzazione e forma canonica di Jordan

13.1 Matrici di Jordan

Definizione 13.1.1. Una matrice di Jordan, o un blocco di Jordan, è una matrice J $n \times n$ della forma

$$\begin{pmatrix} \lambda & 0 & \cdots & 0 & 0 \\ 1 & \lambda & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 1 & \lambda & 0 \\ 0 & 0 & \cdots & 1 & \lambda \end{pmatrix} \quad \text{oppure} \quad \begin{pmatrix} \lambda & 1 & \cdots & 0 & 0 \\ 0 & \lambda & \ddots & 0 & 0 \\ \vdots & & \ddots & \ddots & 0 \\ 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{pmatrix},$$

dove λ è un qualunque scalare. Nel primo caso J è una matrice triangolare inferiore con tutti λ sulla diagonale principale e 1 sulla diagonale immediatamente sotto, mentre nel secondo caso J è una matrice triangolare superiore con tutti λ sulla diagonale principale e tutti 1 sulla diagonale immediatamente sopra.

Questa classe di matrici generalizza il precedente esempio 12.5.3. Il nome viene dal matematico francese Camille Jordan (1838-1922).

Il polinomio caratteristico di J, in entrambi i casi, è $P_J(x) = (\lambda - x)^n$; $p_J(x)$ si fattorizza in n fattori lineari uguali, J ha un unico autovalore λ con molteplicità algebrica n. L'autospazio $\mathrm{Aut}(\lambda)$ ha dimensione pari a $n - \mathrm{rg}(A - \lambda E_n)$, dove (nel primo caso)

$$A - \lambda E_n = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ & \ddots & \ddots & 0 \\ 0 & \cdots & 1 & 0 \end{pmatrix},$$

dunque $\mathrm{rg}(A - \lambda E_n) = n-1$ e $m_g(\lambda) = 1$. Analogamente nel secondo caso. Inoltre nel primo caso $\mathrm{Aut}(\lambda) =< e_n >$ perchè ha equazioni $x_1 = \cdots = x_{n-1} = 0$, mentre nel secondo caso $\mathrm{Aut}(\lambda) =< e_1 >$ perchè ha equazioni $x_2 = \cdots = x_n = 0$. Dunque J non è diagonalizzabile, qualunque sia il campo base, ma è una matrice triangolare.

13.2 Triangolarizzazione

Definizione 13.2.1. 1. Sia $f : V \to V$ un endomorfismo di un K-spazio vettoriale, $\dim V = n$. f è trangolarizzabile (o triangolabile) se esiste una base $\mathcal{B}$ di V tale che la matrice $M_{\mathcal{B}}(f)$ sia triangolare (superiore o inferiore).

2. Una matrice $A \in M(n \times n, K)$ è triangolarizzabile se è simile a una matrice triangolare.

Teorema 13.2.2. *Un endomorfismo $f : V \to V$ è triangolarizzabile se e solo se il suo polinomio caratteristico $p_f(x) \in K[x]$ è fattorizzabile in fattori lineari, cioè ha tutte le sue radici in K.*

Dimostrazione. Un'implicazione è facile: se f è triangolarizzabile, esiste $\mathcal{B}$, base di V, tale che

$$M_{\mathcal{B}}(f) = \begin{pmatrix} \lambda_1 & a_{12} & \cdots & a_{1n} \\ 0 & \lambda_2 & \cdots & a_{2n} \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Perciò $p_f(x) = (\lambda_1 - x) \cdots (\lambda_n - x)$. Analogamente se la matrice è diagonale inferiore.

Viceversa procediamo per induzione su $n = \dim V$. Per $n = 1$ è ovvio perchè ogni matrice 1×1 è triangolare. Supponiamo vero il teorema per $n - 1$ e lo dimostriamo per n. Sia dunque f un endomorfismo di V, spazio vettoriale di dimensione n, e supponiamo che $p_f(x)$ sia prodotto di fattori lineari. Allora $p_f(x)$ ha almeno una radice $\lambda \in K$ che è dunque un autovalore di f, sia $v_1 \neq 0$ un relativo autovettore: $f(v_1) = \lambda v_1$. Prolunghiamo v_1 a una base di V: $\tilde{\mathcal{B}} = (v_1, w_2, \ldots, w_n)$. Allora $V =< v_1 > \oplus W$, dove $W =< w_2, \ldots, w_n >$, W ha dimensione $n - 1$ e $\mathcal{B}' = (w_2, \ldots, w_n)$ è una sua base. Consideriamo

$$A = M_{\tilde{\mathcal{B}}}(f) = \begin{pmatrix} \lambda & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \cdots & & & \\ 0 & a_{n2} & \cdots & a_{nn} \end{pmatrix};$$

possiamo interpretare la sottomatrice $A' = \begin{pmatrix} a_{22} & \cdots & a_{2n} \\ \cdots & & \\ a_{n2} & \cdots & a_{nn} \end{pmatrix}$ come matrice rispetto a $\mathcal{B}'$ di un endomorfismo g di W, il cui polinomio caratteristico è $|A' - xE_{n-1}|$. Ma $p_f(x) = (\lambda - x)|A' - xE_{n-1}|$, perchè A è triangolare a blocchi; per ipotesi $p_f(x)$ è prodotto di fattori lineari, quindi anche $p_g(x)$ è prodotto di fattori lineari. Possiamo quindi applicare l'ipotesi induttiva: g è triangolarizzabile, quindi esiste una base $(v_2, \ldots, v_n)$ di W rispetto a cui la matrice di g è triangolare superiore. Chiamiamo allora $\mathcal{B}$ la base di V ottenuta come unione di v_1 con $\mathcal{B}'$; la matrice $M_{\mathcal{B}}(f)$ risulta triangolare superiore.

In effetti: ogni vettore v di V ha un'unica espressione come somma di un multiplo di v_1 e di un vettore di W (perchè $V =< v_1 > \oplus W$); detta π l'applicazione $\pi : V \to W$ che manda v nel suo addendo che sta in W, g risulta uguale a $\pi \circ f$ ristretta a W. $\square$

Usando il Teorema 13.2.2 e il Teorema fondamentale dell'Algebra 12.6.1, si ottiene il seguente Corollario.

Corollario 13.2.3. *Ogni endomorfismo di un $\mathbb{C}$-spazio vettoriale di dimensione finita è triangolarizzabile. Similmente, ogni matrice quadrata di $M(n \times n, \mathbb{C})$ è simile a una matrice traingolare.*

13.3 Potenze di un endomorfismo e autospazi generalizzati

Sia $g : V \to V$ un endomorfismo. Le **potenze di** g sono gli endomorfismi $g^2 := g \circ g$, $g^3 := g \circ g \circ g$, e così via. Valgono le seguenti semplici proprietà:

Proposizione 13.3.1.
1. $\ker g \subset \ker g^2 \subset \ker g^3 \subset \dots$ *(*)*
2. se $\ker g^k = \ker g^{k+1}$ *per un certo* $k \geq 1$*, da quel momento in poi nella catena (*) sono tutte uguaglianze;*
3. Se $v \in \ker g^k$*, allora* $g(v) \in \ker g^{k-1}$*.*

Dimostrazione. 1. Se $g^k(v) = 0$ allora $g^{k+1}(v) = g(g^k(v)) = 0$. Dunque $\ker g^k \subset \ker g^{k+1}$ per ogni k.

2. Supponiamo $\ker g^k = \ker g^{k+1}$, e sia $v \in \ker g^{k+2}$. Allora $g^{k+2}(v) = 0 = g^{k+1}g(v)$, e perciò $g(v) \in \ker g^{k+1} = \ker g^k$. Dunque $g^k(g(v)) = 0$ ossia $g^{k+1}(v) = 0$. Si conclude che $\ker g^{k+2} = \ker g^{k+1}$.

3. Se $g^k(v) = 0$, allora $g^{k-1}(g(v)) = 0$, e quindi $g(v) \in \ker g^{k-1}$. □

Applicheremo questa proposizione nel caso in cui è dato un endomorfismo $f : V \to V$, di cui ci interessa trovare una forma "canonica", $\lambda \in K$ è un autovalore di f, e $g = f - \lambda\,\mathrm{id}_V$. In questo caso $\ker g = \mathrm{Aut}(\lambda)$ e ha dimensione $m_g(\lambda)$.

Vale la seguente proposizione di cui omettiamo la dimostrazione.

Proposizione 13.3.2. *Sia* $g = f - \lambda\,\mathrm{id}_V$*.*
1. Se $m_g(\lambda) = m_a(\lambda)$*, nella catena (*) della Proposizione 13.3.1 sono tutte uguaglianze.*
2. Se $m_g(\lambda) < m_a(\lambda)$*, allora* $\mathrm{Aut}(\lambda) = \ker g$ *è contenuto strettamente in* $\ker g^2$*, e* $\dim \ker g^2 \leq m_a(\lambda)$*.*
3. Nella catena (*) i sottospazi hanno tutti dimensione $\leq m_a(\lambda)$*; le inclusioni sono strette finchè si arriva a un sottospazio* $\ker g^k$ *di dimensione pari a* $m_a(\lambda)$*, e allora la catena si stabilizza.*

Poniamo $a = m_a(\lambda)$. Si usa la seguente notazione:

$$\begin{array}{ccccccccc}
\mathrm{Aut}_1(\lambda) & \subsetneq & \mathrm{Aut}_2(\lambda) & \subsetneq & \cdots & \subsetneq & \mathrm{Aut}_k(\lambda) & = \cdots = & \mathrm{Aut}_a(\lambda) \\
\| & & \| & & & & \| & & \| \\
\ker g & \subsetneq & \ker g^2 & \subsetneq & \cdots & \subsetneq & \ker g^k & = \cdots = & \ker g^a \\
\| & & & & & & & & \\
\mathrm{Aut}(\lambda) & & & & & & & &
\end{array}$$

I sottospazi che compaiono nella catena sono detti **autospazi generalizzati dell'autovalore** λ. La catena si stabilizza certamente dopo a passi ma potrebbe stabilizzarsi anche prima.

Consideriamo ora un endomorfismo $f : V \to V$ tale che $p_f(x)$ ha tutte le sue radici in K:

$$p_f(x) = (\lambda_1 - x)^{a_1}(\lambda_2 - x)^{a_2} \cdots (\lambda_p - x)^{a_p}$$

con $a_i = m_a(\lambda_i)$ per ogni indice i, dunque f è triangolarizzabile. Si ha $n = a_1 + a_2 + \cdots + a_p$. Per ogni autovalore, consideriamo il relativo autospazio generalizzato $\mathrm{Aut}_{a_i}(\lambda_i)$ di dimensione a_i.

Teorema 13.3.3. *Sia f un endomorfismo triangolarizzabile di V. Allora*

$$V = \mathrm{Aut}_{a_1}(\lambda_1) \oplus \mathrm{Aut}_{a_2}(\lambda_2) \oplus \cdots \oplus \mathrm{Aut}_{a_p}(\lambda_p).$$

Dimostrazione. Omessa. □

Questo teorema servirà a costruire una forma canonica per la matrice di f.

13.4 Forma canonica di Jordan

In questa sezione chiameremo blocco di Jordan una matrice J come nella definizione 13.1.1 triangolare inferiore. Osserviamo che se J è la matrice di un endomorfismo f rispetto a una base $\mathcal{B} = (v_1, \dots, v_n)$, si ha:

$$\begin{array}{ccccc}
f(v_1) = \lambda v_1 + v_2 & \Longleftrightarrow & (f - \lambda\,\mathrm{id}_V)(v_1) = & g(v_1) = & v_2 \\
f(v_2) = \lambda v_2 + v_3 & \Longleftrightarrow & (f - \lambda\,\mathrm{id}_V)(v_2) = & g(v_2) = & g^2(v_1) = v_3 \\
\vdots & & \vdots & & \\
f(v_{n-1}) = \lambda v_{n-1} + v_n & \Longleftrightarrow & (f - \lambda\,\mathrm{id}_V)(v_{n-1}) = & g(v_{n-1}) = & g^{n-1}(v_1) = v_n \\
f(v_n) = \lambda v_n & \Longleftrightarrow & v_n \text{ è un autovettore di } \lambda, & g(v_n) = 0. &
\end{array}$$

Questo è il significato della condizione $M_{\mathcal{B}}(f) = J$.

Il seguente teorema ci dice che, se f ha tutti gli autovalori in K, c'è una base di V rispetto a cui la matrice di f è composta da blocchi di Jordan.

Teorema 13.4.1 (Forma canonica o normale di Jordan)**.** *Sia $f : V \to V$ un endomorfismo tale che*

$$p_f(x) = (\lambda_1 - x)^{a_1}(\lambda_2 - x)^{a_2} \cdots (\lambda_p - x)^{a_p}$$

con $\dim V = n = a_1 + a_2 + \cdots + a_p$. Allora esiste una base $\mathcal{B}$ di V, detta base di Jordan, unione di basi degli autospazi generalizzati $\mathrm{Aut}_{a_i}(\lambda_i)$, $i = 1, \dots, p$, tale che

$$M_{\mathcal{B}}(f) = \begin{pmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & & 0 \\ \cdots & & \ddots & \\ 0 & 0 & & J_r \end{pmatrix},$$

con $r \geq p$, matrice diagonale a blocchi, dove ogni blocco J_i è un blocco di Jordan relativo a uno degli autovalori. Inoltre:

1. la forma canonica di Jordan è unica a meno di permutazioni dei blocchi;

2. il numero di blocchi di ogni autovalore λ_i è uguale a $m_g(\lambda_i)$;

3. preso un autovalore λ, e posto $g = f - \lambda\,\mathrm{id}_V$, la massima dimensione di un blocco di Jordan di f relativo a λ è il minimo esponente k per cui $\ker g^k$ ha dimensione $m_a(\lambda)$.

Dimostrazione. Omessa. $\square$

Se f ha un solo autovalore λ, con $m_a(\lambda) = n$, la massima dimensione di un blocco di Jordan è il minimo k tale che $\dim \ker g^k = n$. Se f corrisponde alla matrice A e g alla matrice $B = A - \lambda E_n$, allora $\dim \ker g^k = n$ se e solo se $\ker g^k = V$ se e solo se $g^k = 0$ se e solo se $B^k = 0$.

Quindi la massima dimensione di un blocco di Jordan è il minimo k tale che $B^k = 0$. In questo caso B è una matrice **nilpotente**, cioè con una potenza nulla, e il minimo k per cui ciò accade è detto **indice di nilpotenza** di B.

Capitolo 14

Prodotti scalari reali e complessi.

14.1 Definizione di prodotto scalare

La nozione di prodotto scalare in uno spazio vettoriale su $\mathbb{R}$ o su $\mathbb{C}$ sta alla base di tutte le nozioni di carattere metrico, in particolare lunghezza (norma), angolo, ortogonalità. La definizione che daremo è diversa nei due casi reale e complesso. In entrambi i casi c'è un esempio fondamentale di prodotto scalare canonico o standard rispettivamente in $\mathbb{R}^n$ e in $\mathbb{C}^n$.

Esempio 14.1.1. Il prodotto scalare standard in $\mathbb{R}^n$ è l'applicazione $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ tale che $\langle x, y \rangle = x_1 y_1 + \cdots + x_n y_n$, dove $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$. Osserviamo che si può scrivere in maniera concisa $\langle x, y \rangle = {}^t xy$, dove ora x, y denotano le relative n-uple scritte in colonna.

Il prodotto scalare standard in $\mathbb{C}^n$ è l'applicazione $\langle \cdot, \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ tale che $\langle z, w \rangle = \bar{z_1} w_1 + \cdots + \bar{z_n} w_n$, dove $z = (z_1, \ldots, z_n)$, $w = (w_1, \ldots, w_n)$. Anche questa volta si può scrivere in maniera concisa $\langle z, w \rangle = {}^t \bar{z} y$, dove z, w denotano le relative n-uple scritte in colonna, e dove $\bar{z}$ denota il complesso coniugato del numero complesso z.

Daremo ora la definizione generale di prodotto scalare nei due casi. Premettiamo un richiamo sul coniugio nel campo complesso. Si tratta dell'applicazione $J : \mathbb{C} \to \mathbb{C}$, tale che $J(z) = \bar{z}$, dove se $z = a + bi$, $\bar{z} = a - bi$. Si ha $z = \bar{z}$ se e solo se $z \in \mathbb{R}$. Valgono le proprietà: $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$, $\overline{\lambda z} = \bar{\lambda} \bar{z}$, ossia $J(\lambda z) = \bar{\lambda} J(z)$. Dunque J non è un endomorfismo lineare di $\mathbb{C}$ come $\mathbb{C}$-spazio vettoriale, infatti conserva la somma ma non il prodotto esterno; è invece un'applicazione $\mathbb{R}$-lineare. La proprietà $J(\lambda z) = \bar{\lambda} J(z)$ si esprime dicendo che J è anti-lineare.

Ricordiamo inoltre che:

- $z + \bar{z} = 2\Re(z)$, dove $\Re(z)$ denota la parte reale di z;
- $z - \bar{z} = 2i\Im(z)$, dove $\Im(z)$ denota la parte immaginaria di z;
- $z\bar{z} = (a + bi)(a - bi) = a^2 + b^2 = \mid z \mid^2$, il modulo di z al quadrato.

Definizione 14.1.2. Sia V un $\mathbb{R}$-spazio vettoriale. **Un prodotto scalare in** V è un'applicazione $b : V \times V \to \mathbb{R}$ verificante le proprietà:

1. b è bilineare, ossia lineare in ciascuno dei due argomenti;
2. b è simmetrica, ossia $b(v, v') = b(v', v)$, per ogni $v, v' \in V$;
3. b è definita positiva, ossia $b(v, v) \geq 0$ per ogni $v \in V$, e inoltre $b(v, v) = 0$ se e solo se $v = 0$.

Le tre propretà si riassumono dicendo che b è **una forma bilineare simmetrica definita positiva**.

Definizione 14.1.3. Sia V un $\mathbb{C}$-spazio vettoriale. **Un prodotto scalare in** V è un'applicazione $b : V \times V \to \mathbb{C}$ verificante le proprietà:

1. b è sesquilineare, ossia lineare nel secondo argomento ma antilineare nel primo;
2. b è hermitiana, ossia $b(w, z) = \overline{b(z, w)}$, per ogni $z, w \in V$;
3. b è definita positiva, ossia $b(z, z) \geq 0$ per ogni $z \in V$, e inoltre $b(z, z) = 0$ se e solo se $z = 0$.

Osserviamo che l'ultima condizione ha senso, perchè dalla seconda segue che $b(z, z) \in \mathbb{R}$. Le tre propretà si riassumono dicendo che b è **una forma sesquilineare hermitiana definita positiva**.

I prodotti scalari standard su $\mathbb{R}^n$ e su $\mathbb{C}^n$ soddisfano le tre condizioni richieste per essere un prodotto scalare.

Osservazione 26. Le tre proprietà che si richiedono a un prodotto scalare complesso si particolarizzano a quelle dei prodotti scalari reali quando si lavora con elementi reali, in quanto il coniugato di un numero reale x è x stesso.

Definizione 14.1.4. Uno **spazio vettoriale euclideo** è uno spazio vettoriale reale in cui è stato fissato un prodotto scalare reale. Uno **spazio vettoriale unitario** è uno spazio vettoriale complesso in cui è stato fissato un prodotto scalare complesso.

14.2 Rappresentazione matriciale

Le forme bilineari sugli spazi vettoriali su $\mathbb{R}$ e quelle sesquilineari sugli spazi vettoriali su $\mathbb{C}$ si possono rappresentare analiticamente in coordinate mediante opportune matrici, una volta fissata una base.

Sia dunque $b : V \times V \to K$ una tale forma, con $K = \mathbb{R}$ o $\mathbb{C}$, sia $\mathcal{B} = (v_1, \ldots, v_n)$ una base di V.

Definizione 14.2.1. La matrice di b rispetto a $\mathcal{B}$ è la matrice $n \times n$ $M_{\mathcal{B}}(b) = (b(v_i, v_j))_{i,j=1,\ldots,n}$: al posto d'indici i, j c'è lo scalare $b(v_i, v_j)$.

Siano v, w due vettori di V, di coordinate rispetto a $\mathcal{B}$ rispettivamente $x_1, \ldots, x_n$, e $y_1, \ldots, y_n$. Denotiamo con x la colonna delle coordinate x_i, e con y quella delle y_j. Allora si ha:

$$\text{caso reale : } b(v, w) = b(\sum_{i=1}^{n} x_i v_i, \sum_{j=1}^{n} y_j v_j) = \text{ bilinearità } = \sum_{i,j=1}^{n} x_i y_j b(v_i, v_j) = \sum_{i,j=1}^{n} x_i b(v_i, v_j) y_j,$$

caso complesso : $b(v,w) = b(\sum_{i=1}^{n} x_i v_i, \sum_{j=1}^{n} y_j v_j) =$ sesquilinearità $= \sum_{i,j=1}^{n} \overline{x_i} b(v_i, v_j) y_j$.

Allora, posto $M = M_{\mathcal{B}}(b)$, si ha nel caso reale

$$^t x M y = (x_1, \dots, x_n) M \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = (\sum_{i=1}^{n} x_i b(v_i, v_1), \dots, \sum_{i=1}^{n} x_i b(v_i, v_n)) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = b(v,w),$$

e nel caso complesso

$$^t \bar{x} M y = (\bar{x_1}, \dots, \bar{x_n}) M \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = (\sum_{i=1}^{n} \bar{x_i} b(v_i, v_1), \dots, \sum_{i=1}^{n} \bar{x_i} b(v_i, v_n)) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = b(v,w).$$

Abbiamo dunque $b(v,w) = {}^t\bar{x} M_{\mathcal{B}}(b) y$, dove x, y sono i vettori colonna delle coordinate di v, w rispetto a $\mathcal{B}$. Nel caso reale questo si particolarizza in $b(v,w) = {}^t x M_{\mathcal{B}}(b) y$.

Osserviamo che $b(v_i, v_j) = {}^t\bar{e_i} M e_j$, perchè le coordinate dei vettori di base sono date dai vettori della base canonica.

Ricordiamo che una matrice quadrata A è detta simmetrica se $A = {}^t A$, coincide con la sua trasposta.

Definizione 14.2.2. Una matrice quadrata complessa A è detta **hermitiana** se ${}^t A = \bar{A}$, cioè la trasposta di A coincide con la coniugata di A, ottenuta coniugando tutti gli elementi di A.

Sulla diagonale principale di una matrice hermitiana tutti gli elementi sono numeri reali.

Se la forma b considerata è simmetrica reale (rispettivamente hermitiana complessa), la matrice $M_{\mathcal{B}}(b)$ risulta simmetrica, cioè $M_{\mathcal{B}}(b) = {}^t M_{\mathcal{B}}(b)$ (rispettivamente hermitiana, cioè ${}^t M_{\mathcal{B}}(b) = \overline{M_{\mathcal{B}}(b)}$). Infatti $b(v_j, v_i) = b(v_i, v_j)$ (rispettivamente $b(v_j, v_i) = \overline{b(v_i, v_j)}$.

Abbiamo visto come, data una base di V, a ogni forma bilineare simmetrica (risp. sesquilineare hermitiana) si può associare una matrice. Vedremo ora che il procedimento si può invertire.

Proposizione 14.2.3. *Sia A una matrice quadrata $n \times n$.*

1. *A è simmetrica reale se e solo se l'applicazione $b : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ definita da $b(x,y) = {}^t x A y$ è una forma bilineare simmetrica.*

2. *A è hermitiana complessa se e solo se l'applicazione $b : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ definita da $b(x,y) = {}^t\bar{x} A y$ è una forma sesquilineare hermitiana.*

Dimostrazione. Un'implicazione è già stata vista. Osserviamo inoltre che il caso 2. è un caso particolare del caso 1., dunque supponiamo che A sia hermitiana e definiamo b ponendo $b(x,y) = {}^t\bar{x} A y$. Verifichiamo la linearità di b nel secondo argomento:

$$b(x, \lambda y + \mu y') = {}^t\bar{x} A(\lambda y + \mu y') = \text{ proprietà distributiva } = {}^t\bar{x} A(\lambda y) + {}^t\bar{x} A(\mu y') =$$

$$= \text{proprietà del prodotto di matrici} = \lambda({}^t\bar{x}Ay) + \mu({}^t\bar{x}Ay') = \lambda b(x,y) + \mu b(x,y').$$

Verifichiamo l'antilinearità nel primo argomento:

$$b(\lambda x + \mu x', y) = {}^t(\overline{\lambda x + \mu x'})Ay = \text{proprietà del coniugio} = {}^t(\bar{\lambda}\bar{x} + \bar{\mu}\bar{x}')Ay =$$

$$= \text{proprietà del prodotto di matrici} = \bar{\lambda}{}^t\bar{x}Ay + \bar{\mu}{}^t\bar{x}'Ay = \bar{\lambda}b(x,y) + \bar{\mu}b(x',y).$$

Verifichiamo l'hermitianità: $b(y,x) = {}^t\bar{y}Ax =$ una matrice 1×1 coincide con la sua trasposta $= {}^t({}^t\bar{y}Ax) = {}^tx{}^tA\bar{y} = \overline{{}^t\bar{x}{}^t\bar{A}y} = A$ è hermitiana $= \overline{{}^t\bar{x}Ay} = \overline{b(x,y)}$. □

Osservazione 27. Osserviamo che se $A = (a_{ij}), B = (b_{ij})$ sono matrici simmetriche reali (risp. hermitiane complesse) che definiscono la stessa forma b, ossia se ${}^txAy = {}^txBy$ per ogni $x, y \in \mathbb{R}^n$ (risp. ${}^t\bar{x}Ay = {}^t\bar{x}By$ per ogni $x, y \in \mathbb{C}^n$) allora $A = B$. Infatti se come vettori x, y prendiamo due vettori della base canonica $\mathcal{C}$ di $\mathbb{R}^n$ (risp. di $\mathbb{C}^n$), abbiamo ${}^t\bar{e_i}Ae_j = a_{ij} = {}^t\bar{e_i}Be_j = b_{ij}$.

Si ha così una biiezione fra forme bilineari simmetriche su $\mathbb{R}^n$ e matrici simmetriche reali di ordine n (risp. tra forme sesquilineari hermitiane su $\mathbb{C}^n$ e matrici hermitiane complesse di ordine n).

Una biiezione analoga si ha su uno spazio vettoriale V di dimensione n, reale o complesso, se è stata fissata una sua base.

Esempio 14.2.4. 1. Se $\mathcal{C}$ è la base canonica di $\mathbb{R}^n$ (risp. di $\mathbb{C}^n$) e b è il prodotto scalare standard, si ha $b(e_i, e_j) = \delta_{ij}$ e dunque $M_{\mathcal{C}}(b) = E_n$.

2. Sia $A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ matrice simmetrica reale di ordine 3. La forma bilineare simmetrica corrispondente rispetto a $\mathcal{C}$ su $\mathbb{R}^3$ è tale che $b(e_1, e_1) = 2$, $b(e_1, e_2) = 1 = b(e_2, e_1)$, $b(e_1, e_3) = 0 = b(e_3, e_1)$, $b(e_2, e_2) = 1$, $b(e_2, e_3) = 0 = b(e_3, e_2)$, $b(e_3, e_3) = 3$. Dunque

$$b(x,y) = \sum_{ij=1}^{3} x_i y_j b(e_i, e_j) = 2x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2 + 3x_3y_3.$$

Per vedere se b è o meno un prodotto scalare consideriamo $b(x,x)$ e vediamo se risulta sempre positivo o nullo, e se si annulla solo nella terna nulla.

$$b(x,x) = 2x_1^2 + 2x_1x_2 + x_2^2 + 3x_3^2 = x_1^2 + (x_1^2 + 2x_1x_2 + x_2^2) + 3x_3^2 = x_1^2 + (x_1 + x_2)^2 + 3x_3^2;$$

essendo una somma di quadrati è sempre ≥ 0, e si annulla solo se

$$\begin{cases} x_1 = 0 \\ x_1 + x_2 = 0 \\ x_3 = 0 \end{cases}$$

se e solo se $x_1 = x_2 = x_3 = 0$. Quindi b è un prodotto scalare.

3. Sia $S = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. La forma bilineare simmetrica corrispondente su $\mathbb{R}^3$ è

$$b(x,y) = {}^t x S y = x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2 + x_3y_3.$$

$$b(x,x) = {}^t x S x = x_1^2 + 2x_1x_2 + x_2^2 + x_3^2 = (x_1+x_2)^2 + x_3^2 \geq 0$$

per ogni $x \in \mathbb{R}^3$; inoltre $b(x,x) = 0$ se e solo se

$$\begin{cases} x_1 + x_2 = 0 \\ x_3 = 0 \end{cases}.$$

Questa volta dunque $b(x,x) = 0$ per ogni $x \in \langle(1,-1,0)\rangle$, dunque b è una forma bilineare simmetrica non definita positiva, non è un prodotto scalare.

4. Sia $B = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. B definisce $b(x,y) = x_1y_2 + x_2y_1 + x_3y_3$, $b(x,x) = 2x_1x_2 + x_3^2$: non è sempre ≥ 0. Per esempio $b(e_1,e_1) = 0$, $b(\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}) = -2$. Non è definita positiva.

5. Sia $H = \begin{pmatrix} 2 & 1-i \\ 1+i & 2 \end{pmatrix}$, matrice hermitiana 2×2. La forma sesquilineare hermitiana h associata ad H rispetto alla base canonica di $\mathbb{C}^2$ opera come segue su $z = (z_1, z_2)$, $w = (w_1, w_2)$:

$$h(z,w) = (\bar{z}_1 \; \bar{z}_2) H \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = 2\bar{z}_1w_1 + (1-i)\bar{z}_1w_2 + (1+i)\bar{z}_2w_1 + 2\bar{z}_2w_2.$$

In particolare

$$h(z,z) = 2\bar{z}_1z_1 + (1-i)\bar{z}_1z_2 + (1+i)\bar{z}_2z_1 + 2\bar{z}_2z_2 = 2|z_1|^2 + 2|z_2|^2 + 2\Re(\bar{z}_1z_2) + 2\Im(\bar{z}_1z_2).$$

Scriviamo ora $z_1 = a_1 + b_1i$, $z_2 = a_2 + b_2i$, e sostituiamo:

$$h(z,z) = 2(a_1^2+b_1^2+a_2^2+b_2^2+(a_1a_2+b_1b_2)+(a_1b_2-a_2b_1)) = (a_1+a_2)^2+(b_1+b_2)^2+(a_1+b_2)^2+(a_2-b_1)^2.$$

Chiaramente questa somma è sempre ≥ 0 e si annulla se e solo se $a_1 + a_2 = b_1 + b_2 = a_1 + b_2 = a_2 - b_1 = 0$, che equivale a $a_1 = b_1 = a_2 = b_2 = 0$ ossia a $z = (z_1, z_2) = 0$. Quindi h è definita positiva, e risulta essere un prodotto scalare su $\mathbb{C}^2$.

6. Un esempio di carattere diverso di prodotto scalare su $V = \mathbb{R}[t]$ è definito ponendo

$$\langle p, q\rangle = \int_0^1 p(t)q(t)\mathrm{d}t.$$

E\` una forma bilineare simmetrica per le proprietà dell'integrale; è definita positiva perchè

$$\langle p, p\rangle = \int_0^1 p^2(t)\mathrm{d}t \geq 0$$

per ogni polinomio $p(t)$ e si annulla solo se p è il polinomio nullo.

7. Sia $V = M(m \times n, \mathbb{R})$. Definiamo $b : V \times V \to \mathbb{R}$ ponendo $b(A, B) = \mathrm{tr}({}^tBA)$, la traccia della matrice prodotto. E` immediato verificare che b è bilineare e simmetrica; inoltre $b(A, A) = \mathrm{tr}({}^tAA)$. La matrice ${}^tAA$ al posto d'indici ii ha $(a_{1i} \dots a_{ni}) \begin{pmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{pmatrix} = a_{1i}^2 + a_{2i}^2 + \cdots + a_{ni}^2$. Dunque $b(A, A)$ è la somma dei quadrati di tutti gli elementi della matrice A, perciò b è un prodotto scalare.

14.3 Norma

Sia V un K-spazio vettoriale, con $K = \mathbb{R}$ o $K = \mathbb{C}$.

Definizione 14.3.1. Un'applicazione $N : V \to \mathbb{R}$ è una **norma** su V se:

1. $N(\lambda v) = |\lambda| N(v)$, per ogni $v \in V$, dove λ denota il valore assoluto di λ nel caso reale e il modulo di λ nel caso complesso;

2. $N(v + w) \leq N(v) + N(w)$, per ogni $v, w \in V$; questa è detta **disuguaglianza triangolare**;

3. $N(v) = 0$ se e solo se $v = 0$.

Osserviamo che dalla definizione segue subito che $N(v) \geq 0$ per ogni vettore v; infatti $N(v - v) = N(0) = 0$ per la 3.; ma $0 = N(v - v) \leq N(v) + N(-v) = N(v) + |-1|N(v) = 2N(v)$; perciò $N(v) \geq 0$.

Un'applicazione norma è spesso indicata con il simbolo $\|\cdot\|$. In tal caso le tre proprietà si scrivono $\|\lambda v\| = |\lambda|\|v\|$, $\|v + w\| \leq \|v\| + \|w\|$, $\|v\| = 0$ se e solo se $v = 0$.

Esempio 14.3.2. Esempi in $\mathbb{R}^n$.

1. Norma euclidea: $\|(x_1, \dots, x_n)\| = \sqrt{x_1^2 + \cdots + x_n^2}$. Le proprietà 1. e 2. di una norma sono immediate, la disuguaglianza triangolare sarà dimostrata in seguito.

2. Norma 1: $\|(x_1, \dots, x_n)\|_1 = |x_1| + \cdots + |x_n|$, somma dei valori assoluti. Per dimostrare la disuguaglianza traingolare si usa il fatto che in $\mathbb{R}$ $|x + y| \leq |x| + |y|$.

3. Norma ∞: $\|(x_1, \dots, x_n)\|_\infty = \max_{i=1,\dots,n}\{|x_i|\}$: massimo dei valori assoluti delle componenti.

Esempio in $\mathbb{C}^n$, norma standard: $\|(z_1, \dots, z_n)\| = \sqrt{\bar{z}_1 z_1 + \cdots + \bar{z}_n z_n}$.

Vedremo ora che a ogni prodotto scalare si può associare una norma. Denoteremo spesso un prodotto scalare con il simbolo $\langle \cdot, \cdot \rangle$ anzichè con b.

Proposizione 14.3.3. *Sia $\langle \cdot, \cdot \rangle$ un prodotto scalare su V, spazio vettoriale reale o complesso. Allora l'applicazione $\|\cdot\| : V \to \mathbb{R}$ tale che $\|v\| = \sqrt{\langle v, v \rangle}$ è una norma su V.*

Osserviamo che la definizione è ben posta perchè il prodotto scalare su V è definito positivo.

Dimostrazione. La 3. vale perchè il prodotto scalare è definito positivo. Verifichiamo la 1. nel caso complesso:

$$\|\lambda v\| = \sqrt{\langle \lambda v, \lambda v\rangle} = \sqrt{\bar{\lambda}\lambda\langle v, v\rangle} = \sqrt{|\lambda|^2\langle v, v\rangle} = |\lambda|\sqrt{\langle v, v\rangle} = |\lambda|\|v\|.$$

La dimostrazione della 2. sarà data dopo il Teorema 14.4.1. □

Si può dimostrare che non ogni funzione norma su uno spazio vettoriale proviene da un prodotto scalare, ma che, perchè lo sia, occorre che sia verificata per ogni coppia di vettori v, w anche la "Legge del parallelogramma": $\|v + w\|^2 + \|v - w\|^2 = 2\|v\|^2 + 2\|w\|^2$.

14.4 Disuguaglianza di Cauchy-Schwarz e sue conseguenze

Teorema 14.4.1 (Disuguaglianza di Cauchy-Schwarz)**.** *Sia $(V, \langle\cdot,\cdot\rangle)$ uno spazio vettoriale euclideo o unitario. Allora per ogni coppia di vettori $v, w \in V$ vale la disuguaglianza*

$$|\langle v, w\rangle| \leq \|v\|\|w\| \tag{14.1}$$

dove $|\cdot|$ denota il valore assoluto nel caso reale e il modulo nel caso complesso. Inoltre vale l'uguaglianza in (14.1) se e solo se v, w sono linearmente dipendenti.

Dimostrazione. Vediamo la dimostrazione nel caso complesso, quello reale è un caso particolare.

Se $w = 0$, $|\langle v, 0\rangle| = 0 = \|v\|\|0\|$, quindi vale la (14.1) con segno di uguaglianza.

Se $w \neq 0$, consideriamo lo scalare

$$\lambda := \frac{\overline{\langle v, w\rangle}}{\langle w, w\rangle}.$$

Si ha:

$$0 \leq \langle v - \lambda w, v - \lambda w\rangle = \langle v, v\rangle - \lambda\langle v, w\rangle - \bar{\lambda}\langle w, v\rangle + \bar{\lambda}\lambda\langle w, w\rangle =$$

sostituiamo il valore di λ

$$= \langle v, v\rangle - \frac{\overline{\langle v, w\rangle}}{\langle w, w\rangle}\langle v, w\rangle - \frac{\langle v, w\rangle}{\langle w, w\rangle}\overline{\langle v, w\rangle} + \frac{\langle v, w\rangle}{\langle w, w\rangle}\frac{\overline{\langle v, w\rangle}}{\langle w, w\rangle}\langle w, w\rangle =$$

abbiamo usato che il prodotto scalare è hermitiano e che $\langle w, w\rangle$ è reale; gli ultimi due addendi si cancellano:

$$= \|v\|^2 - \frac{|\langle v, w\rangle|^2}{\|w\|^2}.$$

Moltiplichiamo per $\|w\|^2$, che è positivo, e otteniamo

$$0 \leq \|v\|^2\|w\|^2 - |\langle v, w\rangle|^2$$

ossia

$$|\langle v, w\rangle|^2 \leq \|v\|^2\|w\|^2.$$

Estraendo le radici otteniamo

$$|\langle v, w\rangle| \leq \|v\|\|w\|,$$

che è la disuguaglianza voluta.

Se in (14.1) vale l'uguaglianza, o $w = 0$, oppure $\langle v-\lambda w, v-\lambda w\rangle = 0$, dunque $v-\lambda w = 0$, e quindi v, w sono linearmente dipendenti. Viceversa, se v, w sono linearmente dipendenti, con $w \neq 0$, esiste uno scalare λ tale che $v = \lambda w$; perciò $\langle v, w\rangle = \langle \lambda w, w\rangle = \bar{\lambda}\langle w, w\rangle = \bar{\lambda}\|w\|^2$. Perciò $|\langle v, w\rangle| = |\lambda|\|w\|^2$, e $\|v\|\|w\| = \|\lambda w\|\|w\| = |\lambda|\|w\|\|w\|$; segue l'uguaglianza in (14.1). □

Corollario 14.4.2. *La norma associata a un prodotto scalare soddisfa la disuguaglianza triangolare.*

Dimostrazione.

$$\begin{aligned}
\|v + w\|^2 &= \langle v + w, v + w\rangle = \\
&= \langle v, v\rangle + \langle v, w\rangle + \langle w, v\rangle + \langle w, w\rangle = \\
&= \|v\|^2 + \langle v, w\rangle + \overline{\langle v, w\rangle} + \|w\|^2 = \\
&= \|v\|^2 + 2\Re\langle v, w\rangle + \|w\|^2 \leq
\end{aligned}$$

(Ora usiamo il fatto che per un numero complesso z si ha $\Re(z) \leq |z|$; infatti, posto $z = a + bi$, $\Re(z) = a \leq |a| = \sqrt{a^2} \leq \sqrt{a^2 + b^2} = |z|$. Inoltre vale l'uguaglianza se e solo se $a = |a| = \sqrt{a^2 + b^2}$, cioè z è reale positivo.)

$$\leq \|v\|^2 + 2\,|\langle v, w\rangle| + \|w\|^2 \leq$$

per la disuguaglianza di Cauchy-Schwarz

$$\leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 = (\|v\| + \|w\|)^2;$$

estraendo le radici quadrate da ambo i membri della disuguaglianza si ha la tesi. □

Osservazione 28. Osserviamo che nella disuguaglianza triangolare vale l'uguaglianza se e solo se sono entrambe uguaglianze le due disuguaglianze che compaiono nella dimostrazione; questo equivale al fatto che v, w siano linearmente dipendenti e che il loro prodotto scalare $\langle v, w\rangle$ sia reale positivo, ossia vale una relazione del tipo $v = \lambda w$ con λ reale ≥ 0.

Proposizione 14.4.3 (Formule di polarizzazione)**.** *Una funzione norma può essere indotta da un unico prodotto scalare, cioè se si conosce come opera la funzione norma associata a un prodotto scalare, si può ricostruire $\langle v, w\rangle$ per ogni coppia di vettori v, w.*

Dimostrazione. Nella dimostrazione del Corollario 14.4.2, abbiamo visto che

$$\|v + w\|^2 = \|v\|^2 + 2\Re\langle v, w\rangle + \|w\|^2. \tag{14.2}$$

Nel caso reale ne segue la seguente relazione, detta formula di polarizzazione reale:

$$\langle v, w\rangle = \frac{1}{2}(\|v + w\|^2 - \|v\|^2 - \|w\|^2) \tag{14.3}$$

che esprime $\langle v, w\rangle$ utilizzando la norma dei vettori $v, w, v + w$.

Nel caso complesso, dalla (14.2) si ottiene soltanto

$$\Re\langle v, w\rangle = \frac{1}{2}(\|v + w\|^2 - \|v\|^2 - \|w\|^2); \tag{14.4}$$

per ricostruire $\langle v, w\rangle$ è quindi necessario esprimere anche $\Im\langle v, w\rangle$ in funzione delle norme di opportuni vettori. In effetti si ha:

$$\Im\langle v, w\rangle = \frac{1}{2}(\|v - iw\|^2 - \|v\|^2 - \|w\|^2). \tag{14.5}$$

La verifica è lasciata per esercizio. Le relazioni (14.4) e (14.5) costituiscono la formula di polarizzazione complessa. $\square$

14.5 Angoli e ortogonalità

Sia V uno spazio vettoriale reale euclideo. La disuguaglianza di Cauchy-Schwarz ci dice che il valore assoluto di $\langle v, w\rangle$ è limitato superiormente da $\|v\|\|w\|$, per ogni coppia di vettori v, w. Ciò significa che

$$-\|v\|\|w\| \leq \langle v, w\rangle \leq \|v\|\|w\|.$$

Se v, w sono entrambi non nulli, le norme $\|v\|, \|w\|$ sono strettamente positive, e otteniamo

$$-1 \leq \frac{\langle v, w\rangle}{\|v\|\|w\|} \leq 1.$$

Ricordiamo che la funzione coseno è continua strettamente decrescente nell'intervallo chiuso $[0, \pi]$, e assume tutti i valori compresi fra -1 e 1. Perciò per ogni $y \in [-1, 1]$ esiste uno ed un solo $x \in [0, \pi]$ tale che $y = \cos x$. Possiamo allora definire l'angolo convesso di due vettori non nulli.

Definizione 14.5.1. Siano $v, w \in V$, $v \neq 0, w \neq 0$. **L'angolo convesso** di v e w è l'unico $\alpha \in [0, \pi]$ tale che

$$\cos\alpha = \frac{\langle v, w\rangle}{\|v\|\|w\|}.$$

Osserviamo che la definizione data ha senso soltanto nel caso reale, in quanto nel caso complesso in generale $\langle v, w\rangle \in \mathbb{C}$. Tuttavia, in entrambi i casi, reale e complesso, si dà la definizione di ortogonalità.

Definizione 14.5.2. Sia V uno spazio vettoriale euclideo o unitario. Due vettori $v, w \in V$ si dicono **ortogonali** se e solo se $\langle v, w\rangle = 0$. Per indicare che v, w sono ortogonali si scrive anche $v \perp w$.

Nel caso euclideo due vettori non nulli sono ortogonali se e solo se il loro angolo convesso vale $\frac{\pi}{2}$.

Definizione 14.5.3. 1. Sia V uno spazio vettoriale euclideo o unitario, siano $U, W \subset V$ sottospazi vettoriali. U, W si dicono ortogonali se $\langle u, w\rangle = 0$ per ogni $u \in U$ e per ogni $w \in W$. Si scrive $U \perp W$.

2. Sia $S \subset V$ un sottinsieme. L'ortogonale di S è il sottospazio vettoriale di V

$$S^{\perp} = \{v \in V \mid \langle v, s \rangle = 0 \text{ per ogni } s \in S\}.$$

(Verifica che è un sottospazio per esercizio)

Se $W \subset V$ è un sottospazio vettoriale, $W^{\perp}$ è detto **complemento ortogonale di** W.

Se $w \in V$ e $W = \langle w \rangle$ è il sottospazio generato da w, verificare che si ha $w^{\perp} = W^{\perp}$.

Definizione 14.5.4. Una famiglia di vettori di V $\{v_i\}_{i \in I}$ si dice **famiglia ortogonale** se $v_i \perp v_j$ per ogni $i, j \in I$, $i \neq j$: sono a due a due ortogonali.

Un vettore di norma 1 è detto un **versore**.

Una famiglia ortogonale di vettori è detta **ortonormale** se in più $\|v_i\| = 1$ per ogni $i \in I$. Ciò si può anche esprimere dicendo che $\langle v_i, v_j \rangle = \delta_{ij}$ per ogni $i, j \in I$.

Una famiglia ortonormale è dunque una famiglia ortogonale di versori. Una **base ortonormale** di V è una base che risulta una famiglia ortonormale.

Se $v \neq 0$ è un vettore non nullo, **normalizzare** v significa passare da v al vettore

$$\frac{1}{\|v\|} v = \frac{v}{\|v\|},$$

che ha chiaramente norma 1.

Esempio 14.5.5. 1. Per il prodotto scalare standard di $\mathbb{R}^n$ e di $\mathbb{C}^n$ la base canonica è una base ortonormale.

2. In $\mathbb{R}^2$ e in $\mathbb{C}^2$ con prodotto scalare standard, i vettori $(1, 1)$, $(1, -1)$ sono ortogonali; normalizzandoli si ottiene $v_1 = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$, $v_2 = (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$: v_1, v_2 formano una base ortonormale.

Proposizione 14.5.6. *Sia $\mathcal{B} = (v_1, \dots, v_n)$ una base ortonormale di uno spazio vettoriale ortogonale o euclideo V. Allora ogni vettore $v \in V$ si scrive nella forma*

$$v = \langle v_1, v \rangle v_1 + \cdots + \langle v_n, v \rangle v_n. \tag{14.6}$$

Dimostrazione. Siano $\lambda_1, \dots, \lambda_n$ le coordinate di v rispetto a $\mathcal{B}$: $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$. Allora, per ogni $j = 1, \dots, n$ si ha:

$$\begin{aligned}
\langle v_j, v \rangle &= \langle v_j, \lambda_1 v_1 + \cdots + \lambda_n v_n \rangle = \text{per la linearità nel secondo argomento} \\
&= \lambda_1 \langle v_j, v_1 \rangle + \cdots + \lambda_n \langle v_j, v_n \rangle = \\
&= \sum_{i=1}^{n} \lambda_i \langle v_j, v_i \rangle = \\
&= \sum_{i=1}^{n} \lambda_i \delta_{ji} = \lambda_j.
\end{aligned}$$

□

La Proposizione 14.5.6 mostra che lavorare con basi ortonormali presenta notevoli vantaggi. La coordinata i-esima di $v \neq 0$ rispetto a una tale base è dunque $\langle v_i, v\rangle = \cos\theta_i \|v\|$, dove θ_i è l'angolo convesso tra v_i e v. La relazione (14.6) si può dunque riscrivere nella forma

$$v = \|v\|(\cos\theta_1 v_1 + \cdots + \cos\theta_n v_n), \tag{14.7}$$

dove il vettore in parentesi risulta un versore e precisamente il normalizzato di v.

Proposizione 14.5.7. *Sia $\{v_i\}_{i\in I}$ una famiglia ortogonale con $v_i \neq 0$ per ogni i. Allora:*

1. *i vettori v_i sono linearmente indipendenti;*
2. *$\{\frac{v_i}{\|v_i\|}\}_{i\in I}$ è una famiglia ortonormale.*

Dimostrazione. 1. Consideriamo una combinazione lineare nulla dei v_i; con ciò intendiamo una combinazione lineare nulla di una qualunque sottofamiglia finita della famiglia di partenza. Ricordiamo infatti che non ha senso considerare somme infinite. Questa combinazione lineare nulla si può scrivere nella forma $\sum_{i\in I}\lambda_i v_i = 0$ con l'avvertenza che i coefficienti λ_i sono supposti **quasi tutti nulli**, cioè soltanto un numero finito di scalari λ_i può essere diverso da 0. Allora $\langle v_j, \sum_{i\in I}\lambda_i v_i\rangle = \langle v_j, 0\rangle = 0$, e d'altra parte $\langle v_j, \sum_{i\in I}\lambda_i v_i\rangle = \sum_{i\in I}\lambda_i\langle v_j, v_i\rangle = \lambda_j$. Perciò la combinazione lineare considerata è banale.

2. Come abbiamo già osservato $\frac{v_i}{\|v_i\|}$ ha norma 1. Inoltre, se $i \neq j$ si ha

$$\langle \frac{v_i}{\|v_i\|}, \frac{v_j}{\|v_j\|}\rangle = \frac{1}{\|v_i\|\|v_j\|}\langle v_i, v_j\rangle = 0.$$

□

14.6 Proiezioni ortogonali e ortonomalizzazione

Sia V uno spazio vettoriale euclideo o unitario, sia $W \subset V$ un suo sottospazio vettoriale. Supponiamo che $(w_1, \dots, w_m)$ sia una base **ortonormale** di W.

Definizione 14.6.1. Dato un vettore $v \in V$, la sua **proiezione ortogonale su** W è il vettore $\tilde{v} := \langle w_1, v\rangle w_1 + \cdots + \langle w_m, v\rangle w_m$.

Osserviamo che:

1. $\tilde{v} \in W$ perchè è combinazione lineare di $w_1, \dots, w_m$;

2. $v - \tilde{v}$ è ortogonale a W; infatti verifichiamo che $v - \tilde{v}$ risulta ortogonale a w_i per ogni $i = 1, \dots, m$ ed è quindi ortogonale a ogni vettore di W:

$$\begin{aligned}\langle w_i, v - \tilde{v}\rangle = \langle w_i, v\rangle - \langle w_i, \sum_{j=1}^m \langle w_j, v\rangle w_j\rangle = &\text{ linearità nel secondo argomento}\\ = \langle w_i, v\rangle - \sum_{j=1}^m \langle w_j, v\rangle\langle w_i, w_j\rangle = &\\ = \langle w_i, v\rangle - \sum_{j=1}^m \langle w_j, v\rangle\delta_{ij} = 0.&\end{aligned}$$

Vedremo ora che basi ortonormali esistono in ogni spazio vettoriale euclideo o unitario di dimensione finita.

Teorema 14.6.2 (Teorema di ortonormalizzazione di Gram-Schmidt)**.** *Sia V uno spazio vettoriale euclideo o unitario di dimensione finita n; sia $W \subset V$ un suo sottospazio vettoriale di dimensione m. Allora ogni base ortonormale $(w_1, \dots, w_m)$ di W si può prolungare a una base ortonormale di V.*

Dimostrazione. Per induzione su $n - m$, la codimensione di W in V. Se $n - m = 0$, risulta $W = V$ e la tesi del teorema è automaticamente verificata.

Sia dunque $n - m > 0$ e supponiamo vera la tesi del teorema per $n - m - 1$; ciò significa che ogni base ortonormale di un sottospazio vettoriale di dimensione $m + 1$ di V si può prolungare a una base ortonormale di V.

Essendo $n - m > 0$, $W \subsetneq V$ e dunque esiste un vettore $v \in V \setminus W$. Consideriamo la sua proiezione ortogonale su W: $\tilde{v} := \langle w_1, v\rangle w_1 + \cdots + \langle w_m, v\rangle w_m$. Abbiamo che $v - \tilde{v}$ è ortogonale a W, e $v - \tilde{v} \neq 0$, altrimenti $v = \tilde{v} \in W$. Allora lo possiamo normalizzare e definire $w_{m+1} := \frac{v-\tilde{v}}{\|v-\tilde{v}\|}$. Otteniamo che $\|w_{m+1}\| = 1$ e $w_{m+1} \perp w_i$ per ogni $i = 1, \dots, m$, dunque $w_1, \dots, w_m, w_{m+1}$ formano una famiglia ortonormale di vettori. Consideriamo allora il sottospazio $W' = \langle w_1, \dots, w_m, w_{m+1}\rangle$: $w_1, \dots, w_{m+1}$ è una sua base ortonormale, e $\dim W' = m+1$. Allora $\dim V - \dim W' = n-m-1$ e possiamo applicare l'ipotesi induttiva, quindi la base ortonormale $w_1, \dots, w_{m+1}$ si può prolungare a una base ortonormale di V. Ciò conclude la dimostrazione, perchè la base ortonormale di V appena ottenuta prolunga anche la base di W $(w_1, \dots, w_m)$ fissata in partenza. □

Corollario 14.6.3. *Ogni spazio vettoriale euclideo o unitario di dimensione finita $n > 0$ possiede una base ortonormale.*

Dimostrazione. Infatti preso in V un vettore non nullo v, possiamo normalizzarlo e porre $w_1 = \frac{v}{\|v\|}$. Consideriamo ora $W = \langle w_1\rangle$ e applichiamo il Teorema 14.6.2: la sua base ortonormale di W costituita da w_1 può essere prolungata a una base ortonormale di V. □

Osservazione 29. Sia $(V, \langle\cdot,\cdot\rangle)$ uno spazio vettoriale euclideo o unitario, e $\mathcal{B} = (v_1, \dots, v_n)$ una sua base ortonormale. La matrice $M_{\mathcal{B}}(\langle,\rangle)$ del prodotto scalare rispetto a $\mathcal{B}$ risulta la matrice identica E_n, infatti al posto di indici ij contiene $\langle v_i, v_j\rangle = \delta_{ij}$. Quindi l'espressione analitica del prodotto scalare, in coordinate rispetto a una base ortonormale, coincide con quella del prodotto scalare canonico di $\mathbb{R}^n$ o di $\mathbb{C}^n$ rispettivamente.

Definizione 14.6.4. Sia V uno spazio vettoriale euclideo o unitario. V è detto **somma ortogonale** dei suoi sottospazi vettoriali $V_1, \dots, V_k$, e si scrive $V = V_1 \bigoplus V_2 \bigoplus \cdots \bigoplus V_k$, se

1. $V = V_1 + \cdots + V_k$;
2. $V_i \perp V_j$ per ogni $i \neq j$.

Osservazione 30. Osserviamo che dalla condizione 2. segue che $V_i \cap V_j = (0)$ se $i \neq j$, in quanto se $v \in V_i \cap V_j$ dev'essere $\langle v, v\rangle = 0$ e dunque $v = 0$. Non solo, ma segue anche $V_i \perp (\sum_{j\neq i} V_j)$, e perciò $V_i \cap (\sum_{j\neq i} V_j) = (0)$. Concludiamo che **ogni somma ortogonale è una somma diretta.**

Proposizione 14.6.5. *Sia V uno spazio vettoriale euclideo o unitario di dimensione finita n. Sia $W \subset V$ un suo sottospazio vettoriale. Allora $V = W \bigoplus W^\perp$. In particolare si ha $\dim V = \dim W + \dim W^\perp$.*

Dimostrazione. Sia $w_1, \ldots, w_m$ una base ortonormale di W, la prolunghiamo a una base ortonormale di V: $(w_1, \ldots, w_m, v_{m+1}, \ldots, v_n)$. Allora ogni vettore $v \in V$ si scrive

$$v = \langle w_1, v\rangle w_1 + \cdots + \langle w_m, v\rangle w_m + \langle v_{m+1}, v\rangle v_{m+1} + \cdots + \langle v_n, v\rangle v_n = \tilde{v} + v',$$

con $\tilde{v} \in W$ e $v' \in W^\perp$, il che prova che $V = W + W^\perp$; siccome $W \perp W^\perp$ la somma è ortogonale. □

Osserviamo in particolare che i vettori aggiunti alla base di partenza $v_{m+1}, \ldots, v_n$ costituiscono una base ortonormale di $W^\perp$.

Esempio 14.6.6. 1. **Ortonormalizzare una base con il metodo di Gram-Schmidt.**

Consideriamo la matrice simmetrica $S = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 1 & 0 \\ -1 & 0 & 3 \end{pmatrix}$. La forma bilineare simmetrica su $\mathbb{R}^3$ ad essa associata rispetto alla base canonica opera come segue:

$$\langle x, y\rangle = 2x_1y_1 + x_1y_2 + x_2y_1 - x_1y_3 - x_3y_1 + x_2y_2 + 3x_3y_3,$$

$$\langle x, x\rangle = 2x_1^2 + 2x_1x_2 + x_2^2 - 2x_1x_3 + 3x_3^2 = (x_1 + x_2)^2 + (x_1 - x_3)^2 + 2x_3^2,$$

ed è perciò un prodotto scalare. Vogliamo costruire una base ortonormale di $\mathbb{R}^3$ rispetto a tale prodotto scalare, a partire dalla base canonica, applicando il teorema di Gram-Schmidt 14.6.2. Partiremo da e_1, lo normalizzeremo e otterremo v_1; poi considereremo e_2 che non appartiene a $W_1 = \langle v_1\rangle$; otterremo v_2 come $\frac{e_2 - \tilde{e}_2}{\|e_2 - \tilde{e}_2\|}$, dove $\tilde{e}_2$ è la proiezione ortogonale di e_2 su W_1; infine considereremo e_3 che non appartiene a W_2, il sottospazio generato da v_1, v_2, e otterremo v_3 come $\frac{e_3 - \tilde{e}_3}{\|e_3 - \tilde{e}_3\|}$, dove $\tilde{e}_3$ è la proiezione ortogonale di e_3 su W_2.

Osserviamo dunque che $\langle e_1, e_1\rangle = 2$, dunque $\|e_1\| = \sqrt{2}$ e $v_1 = (\frac{1}{\sqrt{2}}, 0, 0)$. Notiamo che e_2 non appartiene a $\langle v_1\rangle = \langle e_1\rangle$, e che $\langle e_1, e_2\rangle = 1 \neq 0$, dunque e_1, e_2 non sono ortogonali. Costruiamo la proiezione ortogonale di e_2 sul sottospazio $W_1 = \langle v_1\rangle$:

$$\tilde{e}_2 = \langle v_1, e_2\rangle v_1 = \langle (\frac{1}{\sqrt{2}}, 0, 0), (0, 1, 0)\rangle (\frac{1}{\sqrt{2}}, 0, 0) = \frac{1}{\sqrt{2}}(\frac{1}{\sqrt{2}}, 0, 0) = (\frac{1}{2}, 0, 0).$$

Allora

$$e_2 - \tilde{e}_2 = (0, 1, 0) - (\frac{1}{2}, 0, 0) = (-\frac{1}{2}, 1, 0),$$

$$\|e_2 - \tilde{e}_2\| = \sqrt{\frac{1}{2} - 1 + 1} = \frac{1}{\sqrt{2}}, \text{ quindi}$$

$$v_2 = \sqrt{2}(-\frac{1}{2}, 1, 0) = (-\frac{\sqrt{2}}{2}, \sqrt{2}, 0).$$

Osserviamo che e_3 non appartiene al sottospazio $W_2 = \langle v_1, v_2 \rangle = \langle e_1, e_2 \rangle$. Costruiamo la proiezione ortogonale di e_3 su W_2:

$$\begin{aligned}
\tilde{e}_3 &= \langle v_1, e_3 \rangle v_1 + \langle v_2, e_3 \rangle v_2 = \\
&= \langle \big(\frac{1}{\sqrt{2}}, 0, 0\big), (0,0,1) \rangle \big(\frac{1}{\sqrt{2}}, 0, 0\big) + \langle \big(-\frac{\sqrt{2}}{2}, \sqrt{2}, 0\big), (0,0,1) \rangle \big(-\frac{\sqrt{2}}{2}, \sqrt{2}, 0\big) = \\
&= -\frac{1}{\sqrt{2}}(-\frac{\sqrt{2}}{2}, \sqrt{2}, 0) + \frac{\sqrt{2}}{2}(-\frac{\sqrt{2}}{2}, \sqrt{2}, 0) = (-1, 1, 0).
\end{aligned}$$

Dunque

$$\begin{aligned}
e_3 - \tilde{e}_3 &= (1, -1, 1), \\
\|e_3 - \tilde{e}_3\| &= \sqrt{2}, \\
v_3 &= \frac{(1, -1, 1)}{\sqrt{2}}.
\end{aligned}$$

Abbiamo così costruito una base ortonormale (v_1, v_2, v_3) con sottospazi generati $\langle v_1 \rangle = \langle e_1 \rangle$, $\langle v_1, v_2 \rangle = \langle e_1, e_2 \rangle$.

2. Consideriamo su $\mathbb{R}^3$ il prodotto scalare canonico; vogliamo costruire una base ortogonale avente come primo vettore $v_1 = (1, 1, 1)$. Osserviamo che i due vettori rimanenti devono essere ortogonali a v_1, e quindi le loro coordinate x, y, z devono verificare l'equazione $x + y + z = 0$. Una soluzione non nulla è $v_2 = (1, -1, 0)$. Il terzo vettore dev'essere ortogonale sia a v_1 sia a v_2, quindi le sue coordinate devono essere una soluzione del sistema lineare $x + y + z = x - y = 0$, la cui matrice dei coefficienti è

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

Allora, applicando l'Esempio 11.10.4, 3, abbiamo immediatamente una soluzione non nulla $v_3 = (1, 1, -2)$. Possiamo ottenere una base ortonormale normalizzando i tre vettori, e otteniamo:

$$\frac{1}{\sqrt{3}}(1,1,1), \frac{1}{\sqrt{2}}(1,-1,0), \frac{1}{\sqrt{6}}(1,1,-2).$$

3. Il procedimento seguito nel precedente Esempio 1. porta al procedimento più generale di ortonormalizzazione di una base $\mathcal{B} = (v_1, \dots, v_n)$. Poniamo

$$W_1 =< v_1 >,\ W_2 =< v_1, v_2 >, \dots, W_n =< v_1, \dots, v_n >= V.$$

- Normalizziamo v_1 e otteniamo $w_1 = \frac{v_1}{\|v_1\|}$. Osserviamo che $W_1 =< v_1 >=< w_1 >$ e w_1 costituisce una sua base ortonormale.
- $v_2 \notin W_1$, costruiamo la sua proiezione ortogonale $\tilde{v_2}$ su W_1 di cui abbiamo a disposizione una base ortonormale. Poniamo $w_2 = \frac{v_2 - \tilde{v_2}}{\|v_2 - \tilde{v_2}\|}$. Osserviamo che $W_2 =< v_1, v_2 >=< w_1, w_2 >$ e che quest'ultima è una sua base ortonormale.
- $v_3 \notin W_2$, ne facciamo la proiezione ortogonale $\tilde{v_3}$ su W_2 di cui conosciamo una base ortonormale. Poniamo w_3 la normalizzazione di $v_3 - \tilde{v_3}$, e così via.

Capitolo 15

Endomorfismi ortogonali e unitari

15.1 Definizione e prime proprietà

Gli endomorfismi di uno spazio vettoriale euclideo o unitario che **conservano** la struttura metrica sono detti rispettivamente ortogonali e unitari. La definizione precisa è la seguente.

Definizione 15.1.1. Sia V uno spazio vettoriale euclideo, rispettivamente unitario. Un endomorfismo $f : V \to V$ è detto **ortogonale**, rispettivamente **unitario**, se, per ogni coppia di vettori $v, w \in V$, si ha $\langle v, w\rangle = \langle f(v), f(w)\rangle$, ossia se f conserva il prodotto scalare.

Chiaramente se f conserva il prodotto scalare, e si considera il caso in cui $v = w$, si ottiene $\langle v, v\rangle = \langle f(v), f(v)\rangle$ per ogni $v \in V$, e perciò si ha anche $\|v\| = \|f(v)\|$: f **conserva la norma** indotta dal prodotto scalare. Questa proprietà si può rovesciare grazie alle formule di polarizzazione.

Proposizione 15.1.2. *Sia V uno spazio vettoriale euclideo, rispettiamente unitario. Sia $f : V \to V$ un endomorfismo che conserva la norma, ossia $\|v\| = \|f(v)\|$ per ogni $v \in V$. Allora f è ortogonale, rispettivamente unitario.*

Dimostrazione. Basta osservare che il prodotto scalare di due vettori qualsiasi $v, w \in V$ si può esprimere usando soltanto le norme di opportuni vettori, grazie alle formule di polarizzazione (14.3), (14.4), (14.5). □

Osservazione 31. Un endomorfismo ortogonale, o unitario, è sicuramente iniettivo. Infatti, se $v \in \ker f$, si ha $f(v) = 0$ e perciò $\|f(v)\| = 0$; ma $\|v\| = \|f(v)\|$, quindi $\|v\| = 0$, che implica $v = 0$. Perciò il nucleo di f è nullo, e f risulta iniettivo. Allora se $\dim V$ è finita, un endomorfismo ortogonale (o unitario) di V è un automorfismo di V, cioè un isomorfismo di V in sè.

Gli endomorfismi ortogonali, o unitari, sono anche detti isometrie vettoriali o isometrie lineari.

15.2 Autovalori degli endomorfismi ortogonali e unitari

Proposizione 15.2.1. *Sia V un K-spazio vettoriale euclideo, risp. unitario, dove K denota il campo reale, risp. il campo complesso. Sia $f : V \to V$ un endomorfismo ortogonale, risp. unitario. Se λ è un autovalore di f, allora il valore assoluto di λ, risp. il modulo di λ, è uguale a 1: $|\lambda| = 1$.*

Dimostrazione. Essendo λ un autovalore di f per ipotesi, esiste un vettore $v \neq 0$ in V tale che $f(v) = \lambda v$. Allora $\|v\| = \|f(v)\| = \|\lambda v\| = |\ \lambda\ |\ \|v\|$. Essendo $\|v\| \neq 0$, si ottiene $|\lambda| = 1$. □

Nel caso reale un autovalore di un endomorfismo ortogonale f può valere solo 1 o -1. Nel caso complesso un autovalore λ di un endomorfismo unitario f appartiene alla circonferenza unitaria: $\lambda = \cos\alpha + i \sin\alpha = e^{i\alpha}$.

Osserviamo che in particolare nessun autovalore è nullo, cosa che peraltro segue anche dal fatto che f è iniettivo.

Proposizione 15.2.2. *Siano $\lambda \neq \mu$ autovalori distinti di un endomorfismo ortogonale, o unitario, di V. Siano v un autovettore di λ e w un autovettore di μ. Allora $v \perp w$, ossia autovettori di autovalori distinti sono ortogonali (e non solo linearmente indipendenti).*

Dimostrazione. $f(v) = \lambda v$, $f(w) = \mu w$, perciò

$$\langle v, w\rangle = \langle f(v), f(w)\rangle = \langle \lambda v, \mu w\rangle = \bar{\lambda}\mu\langle v, w\rangle.$$

Allora $(1 - \bar{\lambda}\mu)\langle v, w\rangle = 0$. Ci sono due casi:

- $\langle v, w\rangle = 0$ e v, w sono ortogonali;
- $1 = \bar{\lambda}\mu$. Ma $|\lambda| = 1$ per la Proposizione 15.2.1, e perciò $\bar{\lambda}\lambda = 1$, e dunque $\bar{\lambda} = \frac{1}{\lambda}$. Allora $1 = \bar{\lambda}\mu = \frac{\mu}{\lambda}$, che implica $\lambda = \mu$: assurdo.

(Nel caso reale la dimostrazione può essere semplificata in quanto la relazione che si trova è $\lambda\mu = 1$ da cui $\mu = \frac{1}{\lambda}$, ma λ, μ valgono entrambi 1 o -1, quindi sono uguali. □

Più in generale:

Proposizione 15.2.3. *Se $\lambda_1, \dots, \lambda_k$ sono autovalori di f a due a due distinti, con f endomorfismo ortogonale o unitario, allora la somma dei corrispondenti autospazi è una somma ortogonale:*

$$\mathrm{Aut}(\lambda_1) \textcircled{\perp} \mathrm{Aut}(\lambda_2) \textcircled{\perp} \cdots \textcircled{\perp} \mathrm{Aut}(\lambda_k) \subset V.$$

Dimostrazione. Basta osservare che, per ogni $i = 1, \dots, k$, ogni autovettore di λ_i è ortogonale a ogni somma di autovettori degli altri autovalori. □

15.3 Matrici ortogonali e unitarie

Considereremo ora le matrici associate a endomorfismi ortogonali e unitari rispetto a basi di V. Ci interesserà il caso in cui la base considerata è ortonormale.

Definizione 15.3.1 (Matrici ortogonali e unitarie)**.**
Sia A una matrice $n \times n$ invertibile.
Se A ha entrate **reali**, A è detta **ortogonale** se $A^{-1} = {}^tA$.
Se A ha entrate **complesse**, A è detta **unitaria** se $A^{-1} = {}^t\bar{A}$.

Osserviamo che se A è ortogonale o unitaria, anche la sua trasposta ${}^tA$ lo è.

Proposizione 15.3.2. *Sia A una matrice quadrata reale, risp. complessa. Le seguenti proprietà sono equivalenti:*

1. *A è ortogonale, risp. unitaria;*
2. *le colonne di A formano una base ortonormale di $\mathbb{R}^n$, risp. di $\mathbb{C}^n$, con il prodotto scalare standard;*
3. *le righe di A formano una base ortonormale di $\mathbb{R}^n$, risp. di $\mathbb{C}^n$, con il prodotto scalare standard.*

Dimostrazione. Vediamo la dimostrazione nel caso complesso, quello reale è un caso particolare. Osserviamo che A è unitaria se e solo se ${}^t\bar{A}A = E_n = (\delta_{ij})_{i,j=1,\dots,n}$. Indicate con $a^1, \dots, a^n$ le colonne di A, ciò equivale a ${}^t\bar{a}^i a^j = \delta_{ij}$ per ogni coppia di indici i, j; ma questo è proprio $\langle \bar{a}^i, a^j \rangle$ perchè stiamo considerando il prodotto scalare standard, e dunque le colonne di A formano una base ortonormale di $\mathbb{C}^n$.

Per provare la proprietà per le righe, basta osservare che $A^t\bar{A} = E_n$ se e solo se (usando il coniugio) $\overline{A^t\bar{A}} = \bar{A}^tA = E_n$, e si prosegue poi come sopra, lavorando con la trasposta di A. $\square$

Corollario 15.3.3. *Sia $A = M^{\mathcal{B}}_{\mathcal{B}'}(\mathrm{id}_V)$ la matrice di un cambiamento di base. Se A è una matrice ortogonale, o unitaria, e se $\mathcal{B}$ è una base ortonormale, allora anche $\mathcal{B}'$ è ortonormale.*

Esempio 15.3.4. 1. Matrice della rotazione di angolo α in $\mathbb{R}^2$: $R_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$.

$$ {}^tR_\alpha R_\alpha = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. $$

2. Matrice di una riflessione in $\mathbb{R}^2$: $S_\alpha = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}$.

$$ {}^tS_\alpha S_\alpha = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix} \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. $$

Il legame tra le matrici appena introdotte e gli endomorfismi ortogonali e unitari è dato dal seguente teorema.

Teorema 15.3.5. *Sia $f : V \to V$ un endomorfismo di uno spazio vettoriale euclideo, risp. unitario, di dimensione finita. Sia $\mathcal{B} = (v_1, \dots, v_n)$ una base* **ortonormale** *di V. Allora f è ortogonale, risp. unitario, se e solo se $M_{\mathcal{B}}(f)$ è una matrice ortogonale, risp. unitaria.*

Dimostrazione. Come sempre, vediamo la dimostrazione nel caso complesso; il caso reale è un caso particolare. Siano $v, w \in V$ vettori con colonne delle coordinate rispetto a $\mathcal{B}$ rispettivamente $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$. Sia $A = M_{\mathcal{B}}(f)$. Osserviamo che, essendo $\mathcal{B}$ ortonormale, $\langle v, w \rangle = {}^t\bar{x}y$; inoltre le coordinate di $f(v)$ rispetto a $\mathcal{B}$ sono date da Ax, e quelle di $f(w)$ da Ay.

Allora: f è unitario se e solo se per ogni coppia di vettori di V si ha $\langle v, w \rangle = \langle f(v), f(w) \rangle$; passando in coordinate, questo si traduce nell'affermazione che f è unitario se e solo se per ogni $x, y \in \mathbb{C}^n$ si ha ${}^t\bar{x}y = {}^t(\overline{Ax})Ay$, o equivalentemente ${}^t\bar{x}E_n y = {}^t\bar{x}{}^t\bar{A}Ay$. Per l'Osservazione 27, ciò equivale a $E_n = {}^t\bar{A}A$, ossia A è unitaria. $\square$

Corollario 15.3.6. *1. Consideriamo $\mathbb{R}^n$ con la base canonica $\mathcal{C}$ e il prodotto scalare canonico; allora A è una matrice ortogonale $n \times n$ se e solo se $L(A)$ è un endomorfismo ortogonale.*

2. Consideriamo $\mathbb{C}^n$ con la base canonica $\mathcal{C}$ e il prodotto scalare standard: allora A è una matrice unitaria $n \times n$ se e solo se $L(A)$ è un endomorfismo unitario.

Dimostrazione. Infatti in entrambi i casi $\mathcal{C}$ è una base ortonormale per il prodotto scalare standard e $A = M_{\mathcal{C}}(L(A))$. $\square$

Quindi per l'Esempio 15.3.4 rotazioni e riflessioni sono endomorfismi ortogonali di $\mathbb{R}^2$ con prodotto scalare canonico.

L'insieme delle matrici reali $n \times n$ ortogonali si denota $O(n)$: è un sottogruppo di $GL(n, \mathbb{R})$, detto **gruppo ortogonale** di grado n.

Infatti: se $A, B \in O(n)$, si ha ${}^tAA = E_n = {}^tBB$, quindi ${}^t(AB)AB = {}^tB{}^tAAB = {}^tBE_nB = E_n$, quindi $O(n)$ è chiuso rispetto al prodotto righe per colonne. Inoltre E_n è una matrice ortogonale, perchè le sue colonne costituiscono la base canonica. Infine se A è ortogonale, anche A^{-1} lo è, perchè ${}^t(A^{-1})A^{-1} = {}^t({}^tA)A^{-1} = AA^{-1} = E_n$.

In maniera del tutto analoga l'insieme delle matrici complesse $n \times n$ unitarie costituisce un sottogruppo di $GL(n, \mathbb{C})$, denotato $U(n)$, detto **gruppo unitario** di grado n.

Proposizione 15.3.7. *Se A è una matrice ortogonale, risp. unitaria, si ha $\mid det(A) \mid= 1$ (valore assoluto, risp. modulo, del determinante).*

Dimostrazione. $\det(A^t\bar{A}) = \det(E_n) = 1$; d'altra parte, per il Teorema di Binet, $\det(A^t\bar{A}) = \det(A)\det({}^t\bar{A}) = \det(A)\det(\bar{A}) = \det(A)\overline{\det(A)} =\mid \det(A) \mid^2$. Percioò $\mid \det(A) \mid^2= 1$ e si conclude che $\mid \det(A) \mid= 1$. $\square$

Le matrici ortogonali, risp. unitarie, con determinante uguale a 1, formano un sottogruppo del gruppo ortogonale, risp. unitario:

- $SO(n) = O(n) \cap SL(n, \mathbb{R}) = \{A \text{ ortogonale con } \det(A) = 1\}$ **gruppo ortogonale speciale**;

- $SU(n) = U(n) \cap SL(n, \mathbb{C}) = \{A \text{ unitaria con } \det(A) = 1\}$ **gruppo unitario speciale**.

15.4 Forma normale per endomorfismi unitari e ortogonali

Teorema 15.4.1. *Sia V un $\mathbb{C}$-spazio vettoriale unitario di dimensione finita n, sia $f : V \to V$ un endomorfismo unitario. Allora esiste una base ortonormale $\mathcal{B}$ formata da autovettori di f. In particolare f è diagonalizzabile.*

Dimostrazione. Per induzione su n.

Se $n = 1$, V ha dimensione 1: esiste $v \neq 0$, $v \in V$; lo normalizzo, $v_1 := \frac{v}{\|v\|}$ costituisce una base ortonormale di V, ogni vettore di V gli è proporzionale, dunque in particolare v_1 è un autovettore di f.

Passo induttivo: supponiamo vero il teorema per spazi unitari di dimensione $n-1$. Sia f endomorfismo unitario di V di dimensione $n > 1$, consideriamo il polinomio caratteristico $p_f(x) \in \mathbb{C}[x]$: ha almeno una radice in $\mathbb{C}$, λ, che risulta un autovalore di f; dunque esiste $v_1 \neq 0$ tale che $f(v_1) = \lambda v_1$, e possiamo supporre, eventualmente dopo averlo normalizzato, che $\|v_1\| = 1$. Sia $W = v_1^\perp = \langle v_1 \rangle^\perp$: ha dimensione $n-1$ per la Proposizione 14.6.5. Vogliamo dimostrare che $f(W) \subset W$, perchè in tal caso $f|_W \colon W \to W$ risulta un endomorfismo unitario di W e gli possiamo applicare l'ipotesi induttiva. Sia dunque $w \in W$: allora $\langle w, v_1 \rangle = 0$. Consideriamo $f(w)$: vogliamo dimostrare che $\langle f(w), v_1 \rangle = 0$. Da $\langle w, v_1 \rangle = 0$ segue $\langle f(w), f(v_1) \rangle = 0 = \langle f(w), \lambda v_1 \rangle = \lambda \langle f(w), v_1 \rangle$. Poichè $\lambda \neq 0$ (Sezione 15.2), possiamo concludere che $\langle f(w), v_1 \rangle = 0$, e quindi $f(w) \in W$ come volevamo.

Consideriamo dunque $f|_W \colon W \to W$: W è spazio vettoriale unitario con il prodotto scalare di V ristretto ai vettori di W, e $f|_W$ risulta un endomorfismo unitario di W. Allora per ipotesi induttiva esiste una base ortonormale di W, $(v_2, \dots, v_n)$, formata da autovettori di $f|_W$: sono anche autovettori di f. Si ha che ogni $v_i \in W = v_1^\perp$, per $i = 2, \dots, n$, perciò $\langle v_1, v_i \rangle = 0$. Poichè $V = \langle v_1 \rangle \bigoplus W$, otteniamo che $(v_1, v_2, \dots, v_n)$ è una base di V, perchè è unione di basi di due addendi diretti, e risulta una base ortogonale di autovettori di f. $\square$

Il Teorema 15.4.1 ha una versione equivalente per matrici.

Teorema 15.4.2. *Sia $A \in U(n)$ una matrice complessa unitaria di ordine n. Allora esiste una matrice unitaria S di ordine n tale che $S^{-1}AS = {}^t\bar{S}AS$ sia diagonale. In altre parole A è diagonalizzabile mediante una matrice unitaria.*

Dimostrazione. Consideriamo $\mathbb{C}^n$ come spazio unitario con il prodotto scalare canonico. Allora $L(A) : \mathbb{C}^n \to \mathbb{C}^n$ è un automorfismo unitario, quindi, per il Teorema 15.4.1 , esiste una base $\mathcal{B} = (v_1, \dots, v_n)$, ortonormale per il prodotto scalare canonico, formata da autovettori di $L(A)$. Quindi $M_\mathcal{B}(L(A))$ è una matrice diagonale avente sulla diagonale principale gli autovalori di A, ciascuno tante volte quant'è la sua molteplicità algebrica. Ricordiamo che anche la base canonica $\mathcal{C}$ è ortonormale e che $A = M_\mathcal{C}(L(A))$. Abbiamo allora

$$M_\mathcal{B}(L(A)) = M_\mathcal{C}^\mathcal{B}(\mathrm{id}_{\mathbb{C}^n}) M_\mathcal{C}(L(A)) M_\mathcal{B}^\mathcal{C}(\mathrm{id}_{\mathbb{C}^n}) = S^{-1}AS$$

dove S è la matrice di passaggio da $\mathcal{B}$ a $\mathcal{C}$, avente nelle colonne le coordinate dei vettori di $\mathcal{B}$, $v_1, \ldots, v_n$, rispetto alla base canonica. Siccome entrambe le basi sono ortonormali, per il Teorema 15.3.5 la matrice S risulta unitaria. $\square$

Dalla Proposizione 15.2.3 e dal Teorema 15.4.1, segue subito il seguente Corollario.

Corollario 15.4.3. *Se f è un endomorfismo unitario di V e $\lambda_1, \ldots, \lambda_k$ sono i suoi autovalori distinti, si ha $V = \mathrm{Aut}(\lambda_1) \bigoplus \cdots \bigoplus \mathrm{Aut}(\lambda_k)$.*

Nel caso reale degli endomorfismi ortogonali o delle matrici ortogonali, non è sempre possibile diagonalizzare. Per esempio la matrice R_α di una rotazione è ortogonale (Esempio 15.3.4) ma non è diagonalizzabile se $\alpha \neq 0, \pi$ (Esempio 12.3.2). Però, se A è una matrice ortogonale 2×2, A è del tipo R_α o S_α.

Proposizione 15.4.4. *Sia $A \in O(2)$ una matrice reale ortogonale di ordine 2. Allora se $det(A) = 1$, A è della forma R_α, mentre se $det(A) = -1$, A è della forma S_α.*

Dimostrazione. Sia $A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$. A è ortogonale se e solo se ${}^tAA = E_2$, ossia se e solo se

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} a^2 + b^2 & ac + bd \\ ac + bd & c^2 + d^2 \end{pmatrix}.$$

Poichè $a^2 + b^2 = c^2 + d^2 = 1$, esistono unici angoli α, β, con $0 \leq \alpha, \beta < 2\pi$, tali che $a = \cos\alpha, b = \sin\alpha, c = \sin\beta, d = \cos\beta$. Ma $ac + bd = 0 = \cos\alpha \sin\beta + \sin\alpha \cos\beta = \sin(\alpha + \beta)$, quindi $\alpha + \beta$ è un multiplo intero di π. Quindi si verifica una delle seguenti condizioni: $\alpha + \beta = 0, \pi, 2\pi, 3\pi$.

- Se $\alpha + \beta = 0$ o 2π, $\alpha = -\beta + 2\pi$, dunque $\sin\beta = -\sin\alpha$ e $\cos\beta = \cos\alpha$, quindi $A = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$ è la matrice di una rotazione, ha determinante 1.
- Se $\alpha + \beta = \pi$ o 3π, $\alpha = -\beta + \pi$ o $\alpha = -\beta + 3\pi$, α e β sono angoli supplementari. Allora $\sin\beta = \sin\alpha$ e $\cos\beta = -\cos\alpha$, quindi $A = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}$ è la matrice di una riflessione, ha determinante -1.

$\square$

Di conseguenza, gli unici endomorfismi ortogonali di $\mathbb{R}^2$, con prodotto scalare standard, sono le rotazioni e le riflessioni, detti anche movimenti rigidi del piano. Nel caso di una rotazione, come già osservato, A non ha autovalori, tranne che se $\alpha = 0$ o π, nei quali casi si ha $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ o $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$. Nel caso di una riflessione A ha sempre gli autovalori 1 e -1, $\mathrm{Aut}(1)$, $\mathrm{Aut}(-1)$ hanno dimensione 1, dunque sono due rette ortogonali, e $\mathbb{R}^2 = \mathrm{Aut}(1) \bigoplus \mathrm{Aut}(-1)$. A è simile a $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Il prossimo teorema dice che quelli appena visti sono i "mattoni" con cui costruire la forma normale per che gli automorfismi ortogonali.

Teorema 15.4.5. *Sia V un $\mathbb{R}$-spazio vettoriale euclideo di dimensione finita n, sia $f : V \to V$ un endomorfismo ortogonale. Allora esiste una base ortonormale $\mathcal{B}$ di V tale che $M_{\mathcal{B}}(f)$ sia del tipo*

$$\begin{pmatrix} 1 &&&&&&&& \\ & \ddots &&&&&&& \\ && -1 &&& \Large 0 &&& \\ &&& \ddots &&&&& \\ &&&& -1 &&&& \\ &&&&& \cos\alpha_1 & -\sin\alpha_1 && \\ &&&&& \sin\alpha_1 & \cos\alpha_1 && \\ &&&&&&& \ddots & \\ && \Large 0 &&&&&& \begin{matrix} \cos\alpha_k & -\sin\alpha_k \\ \sin\alpha_k & \cos\alpha_k \end{matrix} \end{pmatrix} \tag{15.1}$$

È una matrice diagonale a blocchi: nella prima parte ci può essere un certo numero di 1 *e* -1 *sulla diagonale principale e* 0 *sul resto delle righe; nella seconda parte vi possono essere un certo numero di matrici di rotazione.*

Questo è il primo esempio di teorema in cui il caso complesso è più semplice del caso reale, dovuto al fatto che $\mathbb{C}$ è algebricamente chiuso. Per esempio, se $n = 3$, due esempi di forma normale sono i seguenti:

- $\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$. L'endomorfismo di $\mathbb{R}^3$ corrispondente è f tale che $v = (x, y, z) \to f(v) = (x, -y, -z)$. L'asse x è fisso, nel piano yz ho una rotazione di π; f è la rotazione di π intorno all'asse x. Il piano yx è l'autospazio di autovalore -1.
- $\begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}$. L'asse x viene riflesso intorno all'origine, il piano yz viene ruotato di un angolo α. $(x, y, z) = (x, 0, 0) + (0, y, z) \to (-x, y\cos\alpha - z\sin\alpha, y\sin\alpha + z\cos\alpha) = (-x, 0, 0) + (0, y\cos\alpha - z\sin\alpha, y\sin\alpha + z\cos\alpha)$.

Diamo soltanto lo schema della dimostrazione del teorema, senza entrare in tutti i dettagli, che comunque non sono difficili.

Dimostrazione. La dimostrazione è per induzione su $n = \dim V$. La base dell'induzione consiste nei casi $n = 1$ e $n = 2$. Se $n = 1$, una matrice ortogonale è uguale a (1) o a (-1), quindi del tipo (15.1). Se $n = 2$, fissata una base ortonormale di V, f è rappresentata da una matrice del tipo R_α o S_α. Nel primo caso abbiamo uno dei blocchi 2×2 della (15.1), nel secondo caso la matrice è diagonalizzabile con autovalori 1, -1; prendendo due versori v_1, v_2 nei relativi autospazi, otteniamo una base ortonormale rispetto a cui la matrice di f è $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Per dimostrare il passo induttivo, avremo bisogno del seguente lemma.

Lemma 15.4.6. *Sia V un $\mathbb{R}$-spazio vettoriale di dimensione finita, e sia $f : V \to V$ un endomorfismo. Allora esiste un sottospazio vettoriale $W \subset V$, con $\dim W = 1$ o 2, tale che $f(W) \subset W$; un tale sottospazio è detto invariante rispetto a f.*

Idea della dimostrazione del Lemma. Se f ha un autovalore λ, sia v un suo autovettore: $f(v) = \lambda v$. Allora $W = \langle v \rangle$ ha dimensione 1 ed è invariante per f, perchè $f(cv) = cf(v) = c\lambda v \in W$. Altrimenti, se f non ha autovalori, sia $\mathcal{B}$ una base di V e sia $A = M_{\mathcal{B}}(f)$, matrice reale $n \times n$. **Passiamo nell'ambiente complesso** considerando $L_{\mathbb{C}}(A) : \mathbb{C}^n \to \mathbb{C}^n$, $z \to Az$. Ora, poichè siamo su $\mathbb{C}$, $L_{\mathbb{C}}(A)$ ha almeno un autovalore (non reale) $\lambda \in \mathbb{C}$: λ è una radice complessa di $p_{L_{\mathbb{C}}(A)}(x) = p_A(x)$, polinomio a coefficienti reali perchè A è reale. Si provano ora i seguenti fatti:

1. anche $\bar{\lambda}$ è radice di $p_A(x)$ ed è dunque autovalore di $L_{\mathbb{C}}(A)$;
2. fissiamo v un autovettore di λ, allora il suo coniugato $\bar{v}$ è un autovettore di $\bar{\lambda}$, e $v, \bar{v}$ sono linearmente indipendenti;
3. i vettori $v + \bar{v}, i(v - \bar{v})$ sono reali, linearmente indipendenti e generano un sottospazio W' di $\mathbb{R}^n$ invariante per $L(A) : \mathbb{R}^n \to \mathbb{R}^n$, ossia $L(A)(W') \subset W'$;
4. consideriamo l'isomorfismo $\kappa_{\mathcal{B}} : V \to \mathbb{R}^n$; in questo isomorfismo W' corrisponde a un sottospazio W di V invariante per f, di dimensione 2.

Questo conclude la dimostrazione del Lemma. □

Torniamo a $f : V \to V$ ortogonale, con $\dim V = n > 2$. Per il lemma esiste un sottospazio $W \subset V$ invariante per f di dimensione 1 o 2, ossia $f(W) \subset W$. Siccome f è un isomorfismo, $\dim W = \dim f(W)$, e perciò $f(W) = W$. Da ciò si dimostra che anche $f(W^\perp) = W^\perp$.

Ora possiamo considerare le restrizioni $f\,|_W \colon W \to W$, $f\,|_{W^\perp} \colon W^\perp \to W^\perp$, sono entrambi endomorfismi ortogonali. Possiamo prendere basi ortonormali $\mathcal{B}_1$ di W come nella base dell'induzione, e $\mathcal{B}_2$ di $W^\perp$ come da ipotesi induttiva; essendo $V = W \bigoplus W^\perp$, l'unione di $\mathcal{B}_1$ e $\mathcal{B}_2$ è una base $\mathcal{B}$ ortonormale di V, e $M_{\mathcal{B}}(f)$ è una matrice a blocchi della forma

$$\begin{pmatrix} M_{\mathcal{B}_1}(f\,|_W) & | & 0 \\ ---- & | & ---- \\ 0 & | & M_{\mathcal{B}_2}(f\,|_{W^\perp}) \end{pmatrix}$$

come richiesto dal teorema. □

Corollario 15.4.7 (Corollario per le matrici ortogonali)**.** *Se A è una matrice ortogonale, allora esiste una matrice ortogonale S tale che $S^{-1}AS = {}^tSAS$ è della forma descritta dal Teorema 15.1.*

La dimostrazione è analoga a quella vista nel caso unitario.

Esempio 15.4.8 (Endomorfismi ortogonali di $\mathbb{R}^3$)**.** In questo caso il polinomio caratteristico $p_f(x) \in R[x]$ ha grado 3, dunque può avere o 3 radici reali, o una radice reale e due radici complesse coniugate. Quindi in ogni caso f ha almeno un autovalore (reale) λ_1, pari a 1 o a

-1 per la Proposizione 15.2.1. Se v_1 è un versore autovettore di λ_1, $W := v_1^\perp$ ha dimensione 2 ed è invariante: $f(W) = W$. Presa una base ortonormale $\mathcal{B}'$ di W formata da v_2, v_3, la matrice di f rispetto a $\mathcal{B} = (v_1, v_2, v_3)$ è del tipo $\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & & \\ 0 & A' & \end{pmatrix}$ dove A' rappresenta $f\mid_W$ rispetto a $\mathcal{B}'$, dunque $A' \in O(2)$. Ora distinguiamo i seguenti casi:

1. $\det(A) = 1$, $\lambda_1 = -1$: allora $\det(A') = -1$ e la forma normale è $\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$: rotazione di π intorno all'asse y;
2. $\det(A) = 1$, $\lambda_1 = 1$: allora $\det(A') = 1$ e la forma normale è $\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}$: rotazione di α intorno all'asse x;
3. $\det(A) = -1$, $\lambda_1 = 1$: allora $\det(A') = -1$ e la forma normale è $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$: riflessione rispetto al piano xy che è fisso, ed è l'autospazio di autovalore 1;
4. $\det(A) = -1$, $\lambda_1 = -1$: allora $\det(A') = 1$ e la forma normale è $\begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}$.

Nei due casi in cui $\det(A) = 1$, abbiamo una rotazione intorno a un asse, in effetti il gruppo ortogonale speciale $SO(3)$ è detto anche gruppo delle rotazioni di $\mathbb{R}^3$.

Esempio 15.4.9 (Forma normale unitaria per R_α)**.** Come già osservato, la matrice R_α è una matrice reale ortogonale, ma può essere pensata anche come una matrice complessa unitaria. Se $\alpha \neq 0$ o π, R_α non è diagonalizzabile su $\mathbb{R}$ ma lo è su $\mathbb{C}$, dove gli autovalori sono $\lambda_1 = \cos\alpha + i\sin\alpha = e^{i\alpha}$, $\lambda_2 = \cos\alpha - i\sin\alpha = e^{-i\alpha}$. Si trova $\mathrm{Aut}(\lambda_1) = \langle (1, -i) \rangle$ e $\mathrm{Aut}(\lambda_2) = \langle (1, i) \rangle$. Perciò una base ortonormale $\mathcal{B}$ di $\mathbb{C}^2$ di autovettori di R_α si ottiene normalizzando i due generatori trovati (che sono fra loro ortogonali): $v_1 = (\frac{1}{\sqrt{2}}, \frac{-i}{\sqrt{2}})$, $v_2 = (\frac{1}{\sqrt{2}}, \frac{i}{\sqrt{2}})$, e osserviamo che $v_2 = \bar{v}_1$. La matrice del cambiamento di base da $\mathcal{B}$ a $\mathcal{C}$ è perciò

$$M_{\mathcal{C}}^{\mathcal{B}}(\mathrm{id}_{\mathbb{C}^2}) = S = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{i}{\sqrt{2}} & \frac{i}{\sqrt{2}} \end{pmatrix};$$

è una matrice unitaria, perchè matrice di passaggio fra due basi ortonormali, e lo si può verificare direttamente controllando che ${}^t\bar{S}S = E_2$. Inoltre si ha

$${}^t\bar{S}R_\alpha S = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} e^{i\alpha} & 0 \\ 0 & e^{-i\alpha} \end{pmatrix}.$$

Questa è la forma normale di R_α come matrice unitaria.

15.5 Esempio di calcolo della forma normale ortogonale

Sia A la matrice ortogonale

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

Vogliamo trovare la sua forma normale ortogonale. Si ha $\det(A) = 1$, quindi $A \in SO(3)$ e rappresenta una rotazione intorno a un asse (Esempio 15.4.8). Si ha: $p_A(x) = -(x^3 - 1) = -(x-1)(x^2+x+1)$, quindi l'unico autovalore reale è $\lambda = 1$. Si calcola che $\mathrm{Aut}(1) = \langle u_1 \rangle = \langle (1,1,1) \rangle$; normalizzando il generatore u_1 si trova $v_1 = \frac{1}{\sqrt{3}}(1,1,1)$. Imponendo l'ortogonalità al vettore v_1 rispetto al prodotto scalare canonico di $\mathbb{R}^3$, si trova l'equazione $x_1+x_2+x_3 = 0$ che definisce il suo ortogonale $W = v_1^\perp$. Allora $W = \langle u_2, u_3 \rangle = \langle (-1,1,0), (-1,0,1) \rangle$. Vogliamo trovare una base ortonormale per W: possiamo procedere con Gram-Schmidt, oppure considerare il prodotto vettoriale di u_1, u_2 per trovare un vettore ortogonale a entrambi (Definizione 16.4.2):

$$u_1 \times u_2 = \begin{vmatrix} e_1 & e_2 & e_3 \\ 1 & 1 & 1 \\ -1 & 1 & 0 \end{vmatrix} = -e_1 - e_2 + 2e_3 = (-1,-1,2).$$

Questo vettore è ortogonale a u_1 e u_2 dunque appartiene a W e, con u_2, forma una sua base ortogonale. Normalizziamo: $v_2 = \frac{1}{\sqrt{2}}(-1,1,0)$, $v_3 = \frac{1}{\sqrt{6}}(-1,-1,2)$: v_2, v_3 formano una base ortonormale di W. Calcoliamo la matrice di $L(A)\mid_W$ rispetto a questa base:

$$Av_2 = (0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}),\ \ Av_3 = (\frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}})$$

Li dobbiamo esprimere come combinazione lineare di v_2 e v_3; facendo il conto otteniamo $Av_2 = -\frac{1}{2}v_2 + \frac{\sqrt{3}}{2}v_3$; $Av_3 = -\frac{\sqrt{3}}{2}v_2 - \frac{1}{2}v_3$. Quindi la forma normale trovata è:

$$A' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\frac{2\pi}{3} & -\sin\frac{2\pi}{3} \\ 0 & \sin\frac{2\pi}{3} & \cos\frac{2\pi}{3} \end{pmatrix} :$$

concludiamo che $L(A)$ è la rotazione di $\frac{2\pi}{3}$ intorno all'asse di rotazione $\mathrm{Aut}(1) = \langle (1,1,1) \rangle$.

Si ha $A' = {}^tSAS$ dove S è la matrice che contiene nelle colonne le coordinate di v_1, v_2, v_3:

$$S = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix}$$

(fare la verifica!).

Capitolo 16

Endomorfismi autoaggiunti

16.1 Definizione e prime proprietà

Definizione 16.1.1. Sia V uno spazio vettoriale euclideo o unitario. Un endomorfismo $f : V \to V$ è detto **autoaggiunto** se, per ogni coppia di vettori $v, w \in V$, si ha $\langle f(v), w\rangle = \langle\ v, f(w)\rangle$.

Osserviamo subito che, se f è un endomorfismo autoaggiunto, per ogni vettore $v \in V$ si ha $\langle f(v), v\rangle = \langle v, f(v)\rangle$. Nel caso reale questo non dice niente di nuovo perchè il prodotto scalare è simmetrico, ma nel caso complesso il prodotto scalare è hermitiano quindi $\langle f(v), v\rangle = \overline{\langle v, f(v)\rangle}$. Otteniamo dunque:

Proposizione 16.1.2. *Se $f : V \to V$ è un endomorfismo autoaggiunto di uno spazio vettoriale complesso unitario, per ogni $v \in V$ si ha che $\langle v, f(v)\rangle$ è reale.*

Vedremo ora di quali proprietà godono la matrici degli endomorfismi autoaggiunti **rispetto alle basi ortonormali**.

Proposizione 16.1.3. *Sia $\mathcal{B}$ una base ortonormale dello spazio vettoriale V euclideo o unitario di dimensione finita. Un endomorfismo f è autoaggiunto se e solo se*

- *caso reale: $M_{\mathcal{B}}(f)$ è simmetrica;*
- *caso complesso: $M_{\mathcal{B}}(f)$ è hermitiana.*

Dimostrazione. Dimostriamo il caso complesso, quello reale è un caso particolare. Sia $A = M_{\mathcal{B}}(f)$. Allora, preso un vettore v, denotato con $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ il vettore delle coordinate di v rispetto a $\mathcal{B}$, il vettore delle coordinate di $f(v)$ è Ax. Quindi f è autoaggiunto se e solo se, per ogni $v, w \in V$, $\langle f(v), w\rangle = \langle v, f(w)\rangle$ se e solo se, per ogni $x, y \in \mathbb{C}^n$, ${}^t(\overline{Ax})y = {}^t\bar{x}{}^t\bar{A}y = {}^t\bar{x}(Ay)$, perchè rispetto a una base ortonormale l'espressione analitica del prodotto scalare è quella del prodotto scalare canonico di $\mathbb{C}^n$. Come nell'Osservazione 27, prendendo per v e w due vettori della base $\mathcal{B}$ di V, ossia per x e y due vettori della base canonica di $\mathbb{C}^n$, si deduce che ${}^t\bar{A} = A$, cioè A è hermitiana. $\square$

Gli endomorfismi autoaggiunti nel caso reale sono anche chiamati simmetrici, proprio perchè lo sono le loro matrici rispetto alle basi ortonormali.

16.2 Autovalori e autovettori degli endomorfismi autoaggiunti

Proposizione 16.2.1. *Sia $f : V \to V$ un endomorfismo autoaggiunto di uno spazio vettoriale euclideo o unitario. Allora*

1. *se λ è un autovalore di f, λ è reale;*

2. *autovettori relativi ad autovalori distinti sono ortogonali.*

La 1. è ovvia nel caso euclideo. La 2. è analoga alla proprietà che vale anche per endomorfismi ortogonali e unitari (Prop. 15.2.2).

Dimostrazione. 1. Sia $v \neq 0$ un autovettore di λ, cioè $f(v) = \lambda v$. Si ha allora: $\langle f(v), v\rangle =$ f è autoaggiunto $= \langle v, f(v)\rangle = \langle v, \lambda v\rangle = \lambda\langle v, v\rangle$. D'altra parte $\langle f(v), v\rangle = \langle \lambda v, v\rangle = \bar{\lambda}\langle v, v\rangle$. Poichè $\langle v, v\rangle \neq 0$, segue che $\lambda = \bar{\lambda}$, cioè $\lambda \in \mathbb{R}$.

2. Siano $\lambda \neq \mu$ autovalori di f, v, w autovettori relativi. Allora

$$\langle f(v), w\rangle = \langle v, f(w)\rangle = \langle v, \mu w\rangle = \mu\langle v, w\rangle.$$

D'altra parte $\langle f(v), w\rangle = \langle \lambda v, w\rangle =$ λ è reale per il punto 1. $= \lambda\langle v, w\rangle$. Quindi $(\lambda - \mu)\langle v, w\rangle = 0$; ma $\lambda - \mu \neq 0$, quindi $\langle v, w\rangle = 0$ dunque v, w sono ortogonali. □

16.3 Il teorema spettrale

Il teorema spettrale risponde al problema della diagonalizzabilità degli endomorfismi autoaggiunti.

Teorema 16.3.1 (Il teorema spettrale)**.** *Sia V uno spazio vettoriale euclideo o unitario di dimensione finita n, sia $f : V \to V$ un endomorfismo autoaggiunto. Allora esiste una base ortonormale $\mathcal{B}$ formata da autovettori di f, in particolare f è diagonalizzabile.*

Dimostrazione. Tratteremo separatamente il caso complesso (che è più immediato) e il caso reale.

Caso complesso. La dimostrazione è per induzione su $n = \dim V$, e segue lo stesso schema della dimostrazione del Teorema 15.4.1. Per $n = 1$ il teorema è vero come nel Teorema 15.4.1. Sia dunque $n > 1$ e supponiamo vero il teorema per spazi unitari di dimensione $n - 1$. Siccome siamo su $\mathbb{C}$, f ha almeno un autovalore, necessariamente reale (Prop. 16.2.1). Sia v_1 un suo autovettore, che possamo supporre sia di norma 1 (altrimenti lo normalizziamo). Poniamo $W = v_1^{\perp}$. Allora $\dim W = n-1$ e $V = \langle v_1\rangle \bigoplus W$. Dimostriamo che W è invariante per f, ossia che $f(W) \subset W$: sia $w \in W$, allora $\langle v_1, w\rangle = 0$. Vogliamo dimostrare che $f(w) \in W$, quindi consideriamo

$$\langle v_1, f(w)\rangle = \ f \text{ autoaggiunto } = \langle f(v_1), w\rangle = \langle \lambda v_1, w\rangle = \ \lambda \text{ è reale} = \lambda\langle v_1, w\rangle = 0.$$

Dunque $f(w) \in W$. Allora $f\mid_W : W \to W$ è un endomorfismo; W è dotato della struttura di spazio euclideo ottenuta restringendo a W il prodotto scalare di V, e $f\mid_W$ è autoaggiunto. Poichè $\dim W = n-1$, possiamo applicare l'ipotesi induttiva e affermare che esiste una base ortonormale di W $(v_2, \dots, v_n)$ formata da autovettori di $f\mid_W$, quindi anche di f. Allora $\mathcal{B} = (v_1, \dots, v_n)$ è una base ortonormale di V ed è formata da autovettori di f.

Caso reale. Vogliamo dimostrare che f ha almeno un autovalore. Per farlo passeremo all'ambiente complesso. Fissiamo dunque una qualunque base ortonormale $\mathcal{B}$ di V; per la Proposizione 16.1.3 $A = M_{\mathcal{B}}(f)$ è una matrice simmetrica reale, in particolare è una matrice hermitiana. Il polinomio caratteristico $p_f(x) = p_A(x)$, a coefficienti reali, ha almeno una radice λ in $\mathbb{C}$. Consideriamo l'endomorfismo $L_{\mathbb{C}}(A) : \mathbb{C}^n \to \mathbb{C}^n$, tale che $z \to Az$, e osserviamo che λ è autovalore anche di $L_{\mathbb{C}}(A)$. Essendo A hermitiana, $L_{\mathbb{C}}(A)$ è un endomorfismo autoaggiunto di $\mathbb{C}^n$ con prodotto scalare canonico, allora per la Proposizione 16.2.1 λ è reale. Possiamo così concludere che f ha almeno un autovalore.

A questo punto la dimostrazione continua e si conclude come nel caso complesso. □

Osservazione 32. Dalla dimostrazione del Teorema spettrale 16.3.1 segue l'importante proprietà che, data una matrice **simmetrica reale** A, il suo polinomio caratteristico $p_A(x)$ ha tutte le sue radici in $\mathbb{R}$, ossia $p_A(x)$ si fattorizza in fattori lineari in $\mathbb{R}[x]$. A ha tutti gli autovalori reali.

Vediamo ora alcune prime conseguenze del Teorema spettrale.

Corollario 16.3.2 (Teorema spettrale per le matrici)**.**

1. Sia A una matrice simmetrica reale. Allora esiste una matrice ortogonale S tale che $S^{-1}AS = {}^tSAS$ sia una matrice diagonale.

2. Sia A una matrice complessa hermitiana. Allora esiste una matrice unitaria S tale che $S^{-1}AS = {}^t\bar{S}AS$ sia una matrice diagonale a valori reali.

Dimostrazione. Chiamiamo K il campo base, che può essere $\mathbb{R}$ o $\mathbb{C}$. Consideriamo l'endomorfismo $L(A) : K^n \to K^n$, dove K^n è dotato del prodotto scalare canonico. Ricordiamo che A è la matrice di $L(A)$ rispetto alla base canonica $\mathcal{C}$. Per il Teorema spettrale esiste una base $\mathcal{B}$ ortonormale di autovettori di $L(A)$, quindi $M_{\mathcal{B}}(L(A))$ è una matrice diagonale avente sulla diagonale principale gli autovalori di A, tutti reali, e si ha $M_{\mathcal{B}}(L(A)) = S^{-1}M_{\mathcal{B}}(L(A))S = S^{-1}AS$, dove $S = M^{\mathcal{C}}_{\mathcal{B}}(\mathrm{id}_{K^n})$. Essendo $\mathcal{B}$, $\mathcal{C}$ entrambe ortonormali, S è una matrice ortogonale, risp. unitaria. □

Corollario 16.3.3. *Sia $f : V \to V$ un endomorfismo autoaggiunto. Allora $V = \mathrm{Aut}(\lambda_1) \oplus \cdots \oplus \mathrm{Aut}(\lambda_k)$, è somma diretta ortogonale degli autospazi di f.*

Dimostrazione. Segue immediatamente dal Teorema spettrale. □

16.4 Esempi

Vedremo ora come operare in pratica per applicare il Teorema spettrale 16.3.1. Supponiamo dato un endomorfismo f autoaggiunto. Si considera il suo polinomio caratteristico $p_f(x)$, lo si fattorizza in fattori lineari

$$p_f(x) = (\lambda_1 - x)^{a_1} \dots (\lambda_k - x)^{a_k}$$

dove $\lambda_1, \dots, \lambda_k \in \mathbb{R}$, e $a_i = m_a(\lambda_i) = m_g(\lambda_i)$ per ogni i, perchè sappiamo già che f è diagonalizzabile. Poi per ogni autovalore λ_i si risolve un sistema lineare omogeneo per determinare una base di $\mathrm{Aut}(\lambda_i)$. Si ortonormalizza la base trovata, per esempio con il metodo di Gram-Schmidt 14.6.2. Infine si fa l'unione delle k basi trovate per i vari autospazi.

Esempio 16.4.1. Sia A la matrice simmetrica reale

$$A = \frac{1}{15}\begin{pmatrix} 10 & 5 & 10 \\ 5 & -14 & 2 \\ 10 & 2 & -11 \end{pmatrix}.$$

Il suo polinomio caratteristico è

$$p_A(x) = \begin{vmatrix} \frac{10}{15} - x & \frac{5}{15} & \frac{10}{15} \\ \frac{5}{15} & -\frac{14}{15} - x & \frac{2}{15} \\ \frac{10}{15} & \frac{2}{15} & -\frac{11}{15} - x \end{vmatrix} = \frac{1}{(15)^3}\begin{vmatrix} 10 - 15x & 5 & 10 \\ 5 & -14 - 15x & 2 \\ 10 & 2 & -11 - 15x \end{vmatrix} =$$

sviluppando il determinante

$$= \frac{1}{(15)^3}(-(15)^3x^3 - (15)^3x^2 + (15)^3x + (15)^3) = -x^3 - x^2 + x + 1 = -(x+1)^2(x-1).$$

A ha due autovalori $\lambda_1 = 1$ con $m_a(1) = 1$, $\lambda_2 = -1$ con $m_a(-1) = 2$.

$$\mathrm{Aut}(1) : \begin{pmatrix} -5 & 5 & 10 \\ 5 & -29 & 2 \\ 10 & 2 & -26 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0 \text{ da cui si trova } \mathrm{Aut}(1) = \langle u_1 \rangle = \langle (5,1,2) \rangle.$$

Normalizziamo u_1 e abbiamo

$$v_1 = \frac{(5,1,2)}{\sqrt{30}}.$$

$$\mathrm{Aut}(-1) : \begin{pmatrix} 25 & 5 & 10 \\ 5 & 1 & 2 \\ 10 & 2 & 4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0 \text{ da cui si trova } \mathrm{Aut}(-1) = \langle u_2, u_3 \rangle = \langle (-1,5,0), (-2,0,5) \rangle.$$

Ora possiamo procedere con Gram-Schmidt a ortonormalizzare u_2, u_3.

$$\begin{aligned}
v_2 &= \frac{u_2}{\|u_2\|} = \frac{(-1,5,0)}{\sqrt{26}}, \\
\widetilde{u_3} &= \langle v_2, u_3\rangle v_2 = \\
&= \langle \frac{(-1,5,0)}{\sqrt{26}}, (-2,0,5)\rangle \frac{(-1,5,0)}{\sqrt{26}} = \\
&= \frac{2}{\sqrt{26}}\Big(\frac{-1}{\sqrt{26}}, \frac{5}{\sqrt{26}}, 0\Big) = \\
&= \Big(-\frac{1}{13}, \frac{5}{13}, 0\Big), \\
u_3 - \widetilde{u_3} &= (-2,0,5) - \Big(-\frac{1}{13}, \frac{5}{13}, 0\Big) = \Big(-\frac{25}{13}, -\frac{5}{13}, 5\Big), \\
\|u_3 - \widetilde{u_3}\| &= \sqrt{\Big(\frac{25}{13}\Big)^2 + \Big(\frac{5}{13}\Big)^2 + 5^2} = \frac{5}{13}\sqrt{195}, \\
v_3 &= \frac{u_3 - \widetilde{u_3}}{\|u_3 - \widetilde{u_3}\|} = \frac{1}{\sqrt{195}}(-5,-1,13).
\end{aligned}$$

La matrice S è la seguente:

$$\begin{pmatrix} \frac{5}{\sqrt{30}} & -\frac{1}{\sqrt{26}} & -\frac{5}{\sqrt{195}} \\ \frac{1}{\sqrt{30}} & \frac{5}{\sqrt{26}} & -\frac{1}{\sqrt{195}} \\ \frac{2}{\sqrt{30}} & 0 & \frac{13}{\sqrt{195}} \end{pmatrix},$$

è ortogonale con le coordinate dei tre vettori v_1, v_2, v_3 come colonne.

In alternativa, anzichè applicare Gram-Schmidt, possiamo procedere così: sappiamo che $u_1 = (5,1,2)$ e $u_2 = (-1,5,0)$ sono ortogonali. Cerchiamo un vettore (x_1, x_2, x_3) ortogonale a entrambi:

$$\begin{cases} 5x_1 + x_2 + 2x_3 = 0 \\ -x_1 + 5x_2 = 0 \end{cases} \qquad \text{con matrice dei coefficienti } \begin{pmatrix} 5 & 1 & 2 \\ -1 & 5 & 0 \end{pmatrix}.$$

Una soluzione è

$$\Big(\begin{vmatrix} 1 & 2 \\ 5 & 0 \end{vmatrix}, -\begin{vmatrix} 5 & 2 \\ -1 & 0 \end{vmatrix}, \begin{vmatrix} 5 & 1 \\ -1 & 5 \end{vmatrix},\Big) = (-10,-2,26) = 2(-5,-1,13).$$

Poi normalizziamo i tre vettori e troviamo la stessa base trovata in precedenza.

Quello che abbiamo fatto nel secondo caso è il **prodotto esterno o vettoriale** dei due vettori u_1 e u_2 di $\mathbb{R}^3$.

Definizione 16.4.2. Il prodotto esterno o vettoriale (o prodotto wedge) di due vettori di $\mathbb{R}^3$ è l'applicazione $\times : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ tale che $(x,y) \to x \times y$ o $x \wedge y$ definito da

$$x \times y = \Big(\begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix}, -\begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix}, \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix},\Big)$$

che si usa anche scrivere simbolicamente come determinante

$$\begin{vmatrix} e_1 & e_2 & e_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix},$$

dove (e_1, e_2, e_3) come al solito è la base canonica di $\mathbb{R}^3$. Il determinante va pensato sviluppato con la regola di Laplace secondo la prima riga.

Il prodotto vettoriale gode delle seguenti proprietà:

1. è bilineare;

2. è antisimmetrico;

3. $x \times y = 0$ se e solo se x, y sono linearmente dipendenti;

4. $x \times y$ è ortogonale sia a x sia a y, dunque è ortogonale al piano generato dai due vettori.

Dimostrazione. Le 1,2,3 seguono dal fatto che le coordinate di $x \times y$ sono i minori di ordine 2 della matrice $\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{pmatrix}$; la 4. segue dallo sviluppo di Laplace dei determinanti nulli

$$\begin{vmatrix} x_1 & x_2 & x_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}, \begin{vmatrix} y_1 & y_2 & y_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}.$$

$\square$

Osserviamo che si ha anche $\langle x \times y, z \rangle = \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}$, ed è nullo se e solo se x, y, z sono linearmente dipendenti: è detto **prodotto misto** dei tre vettori x, y, z.

Calcoliamo la norma di $x \times y$:

$$\begin{aligned} \|x \times y\|^2 &= (x_2y_3 - x_3y_2)^2 + (x_1y_3 - x_3y_1)^2 + (x_1y_2 - x_2y_1)^2 = \\ &= (x_1^2 + x_2^2 + x_3^2)(y_1^2 + y_2^2 + y_3^2) - (x_1y_1 + x_2y_2 + x_3y_3)^2 = \\ &= \|x\|^2\|y\|^2 - \langle x, y\rangle^2 = \\ &= \|x\|^2\|y\|^2 - \|x\|^2\|y\|^2 \cos^2\alpha = \quad \alpha \text{ l'angolo convesso di } x \text{ e } y \\ &= \|x\|^2\|y\|^2(1 - \cos^2\alpha) = \\ &= \|x\|^2\|y\|^2 \sin^2\alpha. \end{aligned}$$

Perciò $\|x \times y\| = \|x\|\|y\| \sin\alpha$, perchè α è convesso, e coincide con l'area del parallelogramma di lati i vettori x e y.

16.5 Forme bilineari simmetriche e cambiamenti di base

Sia V uno spazio vettoriale reale di dimensione finita, $b : V \times V \to \mathbb{R}$ una forma bilineare simmetrica. Siano $\mathcal{B}, \mathcal{B}'$ basi di V. Vogliamo vedere quale relazione intercorre fra le matrici $A = M_{\mathcal{B}}(b)$ e $A' = M_{\mathcal{B}'}(b)$.

Proposizione 16.5.1. *Detta $S = M^{\mathcal{B}}_{\mathcal{B}'}(\mathrm{id}_V)$ la matrice del cambiamento di base, si ha $A' = {}^tSAS$.*

Dimostrazione. Siano $v, w \in V$ due vettori e denotiamo con x, risp. x' le colonne delle coordinate di v rispetto a $\mathcal{B}$, risp. $\mathcal{B}'$, e con y, risp. y', quelle di w. Allora $x = Sx'$, $y = Sy'$. Le espressioni analitiche di b nelle coordinate rispetto alle due basi sono $b(v, w) = {}^txAy = {}^tx'A'y'$. D'altra parte ${}^txAy = {}^t(Sx')A(Sy') = {}^tx'({}^tSAS)y'$, e quindi $A' = {}^tSAS$ (per l'Osservazione 27). $\square$

Definizione 16.5.2. Due matrici quadrate $n \times n$ $A, A' \in M(n \times n, K)$ sono dette **congruenti** se esiste una matrice invertibile S tale che $A' ={}^t SAS$.

Osservazione 33. La Proposizione 16.5.1 afferma che due matrici che rappresentano la stessa forma bilineare simmetrica rispetto a basi diverse sono congruenti. È facile verificare che la relazione di congruenza fra matrici quadrate dello stesso ordine è una relazione d'equivalenza. Tale relazione è diversa dalla relazione di similitudine, che intercorre invece fra matrici che rappresentano lo stesso endomorfismo rispetto a basi diverse.

Tuttavia, il Teorema spettrale per le matrici (Corollario 16.3.2) afferma che per ogni matrice simmetrica reale A, esiste una matrice **ortogonale** S che diagonalizza A, ossia ${}^tSAS = S^{-1}AS = D$, matrice diagonale; infatti ${}^tS = S^{-1}$. Quindi tramite S, che è ortogonale, A è sia simile sia congruente a una matrice diagonale, e precisamente alla matrice diagonale che ha sulla diagonale principale gli autovalori $\lambda_1, \dots, \lambda_n$ di A, presi ciascuno tante volte quant'è la sua molteplicità algebrica. In questo modo S diagonalizza sia la forma bilineare simmetrica sia l'endomorfismo $L(A)$ di $\mathbb{R}^n$ aventi A come matrice rispetto alla base canonica.

Se $\mathbb{R}^n$ è dotato della struttura di spazio euclideo con il prodotto scalare canonico, la matrice ortogonale S si può interpretare come matrice di passaggio da $\mathcal{C}$ a una base $\mathcal{B}$ ortonormale. Quindi come corollario del Teorema spettrale, data una matrice simmetrica reale A, esiste una base ortonormale $\mathcal{B}$ di $\mathbb{R}^n$ rispetto a cui la forma bilineare simmetrica b su $\mathbb{R}^n$ associata ad A è rappresentata da una matrice diagonale.

Nelle coordinate x, y rispetto a tale base $\mathcal{B}$, $b(v, w) = \lambda_1x_1y_1 + \lambda_2x_2y_2 + \cdots + \lambda_nx_ny_n$. Il problema della diagonalizzazione per forme bilineari simmetriche reali ha dunque sempre risposta positiva. Ciò viene ulteriormente precisato e migliorato nel seguente teorema.

Teorema 16.5.3 (Teorema di trasformazione ad assi principali)**.** *Consideriamo $\mathbb{R}^n$ spazio euclideo con il prodotto scalare canonico. Sia $b : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ una forma bilineare simmetrica. Sia $A = M_{\mathcal{C}}(b)$ la matrice simmetrica $n \times n$ che la rappresenta rispetto alla base canonica. Allora:*

1. esiste una base ortonormale $\mathcal{B} = (v_1, \ldots, v_n)$ formata da autovettori di A; ossia esiste $S = M_{\mathcal{C}}^{\mathcal{B}}(\mathrm{id}_{\mathbb{R}^n})$, matrice ortogonale, tale che

$$M_{\mathcal{B}}(b) = {}^tSAS = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \tag{16.1}$$

diagonale, con $\lambda_1, \ldots, \lambda_n$ gli autovalori di A, tutti reali.

2. esiste una base ortogonale $\mathcal{B}'$ di autovettori di A, tale che

$$M_{\mathcal{B}'}(b) = {}^tTAT = \begin{pmatrix} \begin{matrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & 1 \end{matrix} & 0 & 0 \\ 0 & \begin{matrix} -1 & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & -1 \end{matrix} & 0 \\ 0 & 0 & 0 \end{pmatrix} : \tag{16.2}$$

è una matrice diagonale con soli 1, -1, 0 sulla diagonale principale, e precisamente tanti 1 quanti sono gli autovalori positivi, tanti -1 quanti gli autovalori negativi.

Dimostrazione. La 1. è ancora una riformulazione del Teorema spettrale.

Per dimostrare la 2., eventualmente riordinando i vettori di $\mathcal{B}$, possiamo supporre che gli autovalori $\lambda_1, \ldots, \lambda_r$ siano > 0, gli autovalori $\lambda_{r+1}, \ldots, \lambda_{r+s}$ siano < 0, e che $\lambda_{r+s+1} = \cdots = \lambda_n = 0$. Definiamo la base $\mathcal{B}' = (v'_1, \ldots, v'_n)$ ponendo

$$v'_i = \begin{cases} \frac{v_i}{\sqrt{\lambda_i}} & \text{se } i \leq r, \text{ ossia } \lambda_i > 0, \\ \frac{v_i}{\sqrt{-\lambda_i}} & \text{se } r < i \leq r+s, \text{ ossia } \lambda_i < 0, \\ v_i & \text{se } r+s < i \leq n, \text{ ossia } \lambda_i = 0. \end{cases}$$

Allora $b(v'_i, v'_j) = 0$ se $i \neq j$, mentre

$$b(v'_i, v'_i) = \begin{cases} b(\frac{v_i}{\sqrt{\lambda_i}}, \frac{v_i}{\sqrt{\lambda_i}}) = (\frac{1}{\sqrt{\lambda_i}})^2 b(v_i, v_i) = 1 & \text{se } i \leq r, \\ b(\frac{v_i}{\sqrt{-\lambda_i}}, \frac{v_i}{\sqrt{-\lambda_i}}) = (\frac{1}{\sqrt{-\lambda_i}})^2 b(v_i, v_i) = -1 & \text{se } r < i \leq r+s, \\ b(v_i, v_i) = 0 & \text{se } r+s < i \leq n. \end{cases}$$

□

Con la notazione precedente, $M_{\mathcal{B}}(b)$ ha r autovalori positivi, s autovalori negativi, e i rimanenti $n - r - s$ autovalori uguali a 0. Nelle coordinate x', y' rispetto alla nuova base $\mathcal{B}'$, $b(v, w) = x'_1 y'_1 + x'_2 y'_2 + \cdots + x'_r y'_r - x'_{r+1} y'_{r+1} - \cdots - x'_{r+s} y'_{r+s}$. Osserviamo che la matrice $T = M_{\mathcal{C}}^{\mathcal{B}'}(\mathrm{id}_{\mathbb{R}^n})$ in generale non è ortogonale. Le rette generate dai vettori v_i, che coincidono con le rette generate dai vettori v'_i, sono dette **assi principali** per b. Il cambiamento di base effettuato per passare da $\mathcal{B}$ a $\mathcal{B}'$ non cambia le direzione dei tre assi coordinati, ma consiste semplicemente nel riscalare i vettori.

Definizione 16.5.4. Data una forma bilineare simmetrica $b : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, l'applicazione $q : \mathbb{R}^n \to \mathbb{R}$ definita da $q(x) = b(x, x)$ è detta **forma quadratica** associata a b.

Detta A la matrice simmetrica $M_{\mathcal{C}}(b)$ di b rispetto alla base canonica, $q(x) = {}^t xAx = \sum_{ij} a_{ij}x_ix_j$ è rappresentata da un polinomio omogeneo di secondo grado in $x_1, \ldots, x_n$. Per la forma quadratica, il Teorema 16.5.3 si può riformulare così:

Teorema 16.5.5. *Sia $q : \mathbb{R}^n \to \mathbb{R}$ la forma quadratica su $\mathbb{R}^n$ (dotato della struttura euclidea standard) associata a una forma bilineare simmetrica. Allora:*

1. *esiste una base ortonormale di $\mathbb{R}^n$ tale che, in coordinate rispetto a questa base, q si scrive $q(x_1, \ldots, x_n) = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2$;*

2. *esiste una base ortogonale di $\mathbb{R}^n$ tale che, in coordinate rispetto a questa base, q si scrive $q(x'_1, \ldots, x'_n) = x_1'^2 + \cdots + x_r'^2 - x_{r+1}'^2 - \cdots - x_{r+s}'^2$.*

Teorema 16.5.6 (Teorema di Sylvester - Legge d'inerzia)**.** *1. Sia V uno spazio vettoriale reale di dimensione finita n, sia $b : V \times V \to \mathbb{R}$ una forma bilineare simmetrica. Il numero di autovalori positivi, negativi e nulli di $M_{\mathcal{B}}(b)$ dipende solo da b ed è indipendente dalla base fissata.*

2. Ogni matrice simmetrica reale è congruente a una e una sola matrice diagonale della forma (16.2).

Dimostrazione. Sia $\mathcal{B} = (v_1, \ldots, v_n)$ e supponiamo che $M_{\mathcal{B}}(b)$ sia della forma (16.1) con $\lambda_1, \ldots, \lambda_r > 0$, $\lambda_{r+1}, \ldots, \lambda_{r+s} < 0$, e $\lambda_i = 0$ per $i > r + s$. Sia $W_+ = \langle v_1, \ldots, v_r \rangle$, $W_- = \langle v_{r+1}, \ldots, v_{r+s} \rangle$, $W_0 = \langle w_{r+s+1}, \ldots, v_n \rangle$. Allora:

- $b(v, v)$ è sempre > 0 su W_+, < 0 su W_-, 0 su W_0;

- $V = W_+ \oplus W_- \oplus W_0$;

- r è la massima dimensione dei sottospazi $U \subset V$ tali che $b(v, v) > 0$ per ogni $v \in U$ $v \neq 0$, ossia b è definita positiva se ristretta a U. Infatti se b è definita positiva su U, allora $U \cap (W_+ \cup W_0) = (0)$, quindi la somma $U \oplus W_- \oplus W_0$ è diretta. Allora $\dim(U \oplus W_- \oplus W_0) = \dim U + \dim W_+ \dim W_0 = \dim U + s + (n - r - s) = \dim U + n - r \leq n = \dim V$. Quindi $\dim U \leq r$.

Questa caratterizzazione di r non dipende dalla base scelta. Per s vale una caratterizzazione analoga, come massima dimensione dei sottospazi dove $b(v, v) < 0$ per ogni $v \neq 0$. Questo conclude la dimostrazione. □

Definizione 16.5.7. Con la notazione precedente, la **segnatura** di una forma bilineare simmetrica reale b, o della forma quadratica q a essa associata, o della matrice che le rappresenta rispetto a una base fissata, è la coppia (r, s), o la terna (r, s, n).

La forma bilineare simmetrica b, o la forma quadratica q, è detta:

1. **definita positiva** se ha segnatura $(n, 0)$, o $(n, 0, n)$;

2. **definita negativa** se ha segnatura $(0, n)$, o $(0, n, n)$;

3. **semidefinita positiva** se ha segnatura $(m, 0)$ o $(m, 0, n)$ con $m < n$;

4. **semidefinita negativa** se ha segnatura $(0, m)$ o $(0, m, n)$ con $m < n$;

5. **indefinita** in tutti gli altri casi.

Due matrici simmetriche reali sono dunque congruenti se e solo se hanno la stessa segnatura.

Notiamo che il segno assunto da q nei 5 casi è come segue:

1. $q(v) \geq 0$ per ogni v e $q(v) = 0$ se e solo se $v = 0$; ossia b è un prodotto scalare;

2. $q(v) \leq 0$ per ogni v e $q(v) = 0$ se e solo se $v = 0$;

3. $q(v) \geq 0$ per ogni v ma esiste $v \neq 0$ tale che $q(v) = 0$;

4. $q(v) \leq 0$ per ogni v ma esiste $v \neq 0$ tale che $q(v) = 0$;

5. esistono vettori $v, w \in V$ con $q(v) > 0, q(w) < 0$.

Concludiamo enunciando senza dimostrazione un utile criterio che permette di stabilire la segnatura di una matrice simmetrica reale solo guardando il suo polinomio caratteristico.

Teorema 16.5.8 (Criterio di Cartesio)**.** *Sia $p(x) \in \mathbb{R}[x]$ un polinomio di grado n a coefficienti reali:*

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n; \; n > 0, \; a_n \neq 0.$$

Sia m il minimo intero tale che $a_m \neq 0$. **Supponiamo che tutte le radici di p siano reali.** *Allora:*

1. *0 è radice di $p(x)$ se e solo se $m > 0$ e in tal caso m è la molteplicità algebrica di 0 come radice di $p(x)$;*

2. *il numero di radici positive di $p(x)$ (contate con la relativa molteplicità algebrica) è uguale al numero di variazioni di segno nella sequenza dei coefficienti* **non nulli** *di $p(x)$;*

3. *il numero di radici negative di $p(x)$ è pari a $n - ($numero di radici positive$) - m$.*

Il Criterio di Cartesio si applica al polinomio caratteristico di ogni matrice simmetrica reale, in quanto sappiamo che le sue radici sono tutte reali.

Corollario 16.5.9. *La forma bilineare simmetrica b associata a una matrice A di ordine n è un prodotto scalare se e solo se la sequenza dei coefficienti del polinomio caratteristico $p_A(x)$ presenta n variazioni di segno; in particolare nessun coefficiente di $p_A(x)$ può annullarsi, e si ha $m = 0$ ossia 0 non è una sua radice.*

Dimostrazione. Infatti dire che b è un prodotto scalare significa che gli autovalori di A sono tutti positivi, ossia che il polinomio caratteristico di A ha radici tutte positive. In particolare 0 non può essere una sua radice, e la sequenza dei coefficienti deve avere segni alterni. $\square$

Esempio 16.5.10. Consideriamo la matrice

$$A_c = \begin{pmatrix} 4 & 1 & -1 \\ 1 & 1 & 0 \\ -4 & 0 & c \end{pmatrix}$$

dove c è un parametro reale. Vogliamo determinare per quali valori di c A_c definisce un prodotto scalare. Il polinomio caratteristico è

$$p_{A_c}(x) = \begin{vmatrix} 4-x & 1 & -1 \\ 1 & 1-x & 0 \\ -4 & 0 & c-x \end{vmatrix} = -x^3 + (c+5)x^2 + (13-5c)x + (3c-16).$$

Il primo segno è negativo, dunque, per avere un prodotto scalare, dovranno valere le condizioni $c+5>0$, $13-5c<0$ e $3c-16>0$; queste tre condizioni equivalgono all'unica condizione $c > \frac{16}{3}$. Osserviamo che soltanto per $c = \frac{16}{3}$ il termine noto si annulla, dunque 0 è un autovalore; in tal caso ci sono due variazioni di segno, dunque due radici positive e una nulla, e la forma bilineare corrispondente è semidefinita positiva. Per tutti gli altri valori di c vi sono due variazioni di segno e 0 non è radice, dunque la terza radice è negativa, e la forma risulta indefinita.

16.6 Esempio di trasformazione ad assi principali

Sia A la seguente matrice simmetrica reale:

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

A è la matrice dell'endomorfismo autoaggiunto $L(A)$ rispetto alla base canonica di $\mathbb{R}^3$, con prodotto scalare standard. Ad A corrisponde anche la forma bilineare simmetrica b su $\mathbb{R}^3$ definita da: $b(x,y) = {}^t x A y = 2x_1y_1 - 3x_1y_3 - 3x_3y_1 - x_2y_2 + 2x_3y_3$. Si ha $A = M_{\mathcal{C}}(b)$. La forma quadratica $q : \mathbb{R}^3 \to \mathbb{R}$ è definita da $q(x) = b(x,x) = 2x_1^2 - 6x_1x_3 - x_2^2 + 2x_3^2$. Applichiamo il Teorema spettrale per diagonalizzare A mediante una matrice ortogonale. Calcoliamo il polinomio caratteristico di A per trovare i suoi autovalori:

$$p_A(x) = \begin{vmatrix} 2-x & 0 & -3 \\ 0 & -1-x & 0 \\ -3 & 0 & 2-x \end{vmatrix} = -(1+x)[(2-x)^2-9] = -(1+x)(x^2-4x-5) = -(x+1)^2(x-5).$$

Gli autovalori sono -1 e 5, con $m_a(-1) = 2$ e $m_a(5) = 1$. Allora il Teorema spettrale ci garantisce che esiste una matrice ortogonale S tale che

$$S^{-1}AS = {}^tSAS = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{pmatrix}, \tag{16.3}$$

matrice diagonale con gli autovalori di A sulla diagonale. S è la matrice di un cambiamento di base: $S = M_{\mathcal{C}}^{\mathcal{B}}(\mathrm{id}_{\mathbb{R}^3})$, dove $\mathcal{B} = (v_1, v_2, v_3)$ è fatta così: v_1, v_2 formano una base ortonormale di $\mathrm{Aut}(-1)$, e v_3 è un vettore di norma 1 che genera $\mathrm{Aut}(5)$.

Per trovare $\mathrm{Aut}(-1)$, calcoliamo $A + E_3 = \begin{pmatrix} 3 & 0 & -3 \\ 0 & 0 & 0 \\ -3 & 0 & 3 \end{pmatrix}$ di rango 1, quindi $x_1 - x_3 = 0$ è l'equazione di $\mathrm{Aut}(-1)$. Una base è formata da $(1, 0, 1), (0, 1, 0)$, è già ortogonale; la normalizziamo e troviamo $v_1 = \frac{1}{\sqrt{2}}(1, 0, 1)$, $v_2 = e_2$. Per trovare $\mathrm{Aut}(5)$, calcoliamo $A - 5E_3 = \begin{pmatrix} -3 & 0 & -3 \\ 0 & -6 & 0 \\ -3 & 0 & -3 \end{pmatrix}$ di rango 2, quindi le equazioni sono $\begin{cases} x_1 + x_3 & = 0 \\ x_2 = 0. \end{cases}$

Una base di $\mathrm{Aut}(5)$ è $(1, 0, -1)$, lo normalizziamo e otteniamo $v_3 = \frac{1}{\sqrt{2}}(1, 0, -1)$. Abbiamo così ottenuto la base ortonormale diagonalizzante $\mathcal{B} = (v_1, v_2, v_3)$. Allora per questa scelta della base $\mathcal{B}$ otteniamo

$$S = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{pmatrix}:$$

matrice ortogonale. Come abbiamo visto nella Sezione 10.2, le equazioni del cambiamento di coordinate sono $x = Sy$, cioè per esteso $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = S \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$, dove x_1, x_2, x_3 sono coordinate rispetto a $\mathcal{C}$ e y_1, y_2, y_3 rispetto a $\mathcal{B}$. Quindi equivalentemente $y = S^{-1}x = {}^tSx$, perchè S è ortogonale. Esplicitamente abbiamo:

$$\begin{cases} y_1 & = \frac{1}{\sqrt{2}}x_1 + \frac{1}{\sqrt{2}}x_3 \\ y_2 & = x_2 \\ y_3 & = \frac{1}{\sqrt{2}}x_1 - \frac{1}{\sqrt{2}}x_3. \end{cases}$$

Gli assi principali sono gli assi y_1, y_2, y_3, cioè le tre rette generate dai tre vettori di $\mathcal{B}$: l'asse y_1 è la retta $\langle v_1 \rangle$ di equazioni $y_2 = y_3 = 0$, ossia $x_2 = x_1 - x_3 = 0$, l'asse y_2 è la retta $\langle v_2 \rangle$ di equazioni $y_1 = y_3 = 0$ ossia $x_1 + x_3 = x_1 - x_3 = 0$; l'asse y_3 è la retta $\langle v_3 \rangle$ di equazioni $y_1 = y_2 = 0$, ossia $x_1 + x_3 = x_2 = 0$. La forma quadratica q allora si può scrivere:

$$q(x) = b(x, x) = {}^txAx = {}^t(Sy)A(Sy) = {}^ty({}^tSAS)y \overset{\text{per la (16.3)}}{=} -y_1^2 - y_2^2 + 5y_3^2 = q'(y),$$

quindi q è stata diagonalizzata: la sua espressione rispetto agli assi principali non contiene i termini "misti" y_iy_j con $i \neq j$.

Per il Teorema 16.5.3, possiamo fare un ulteriore cambiamento di base, senza cambiare i tre assi, ma solo riscalando, in modo che la forma quadratica sia rappresentata da una matrice diagonale con soli $1, -1$ sulla diagonale. Essendo $\lambda_1 = \lambda_2 = -1$, $\lambda_3 = 5$ avremo la base ortogonale $\mathcal{B}'$ con $v_1' = v_1$, $v_2' = v_2$, $v_3' = \frac{1}{\sqrt{5}}v_3$. Quindi $M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{id}_{\mathbb{R}^3}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{\sqrt{5}} \end{pmatrix}$. Le corrispondenti coordinate z_1, z_2, z_3 sono date da

$$y = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{\sqrt{5}} \end{pmatrix} z, \quad \text{quindi} \begin{cases} y_1 & = z_1 \\ y_2 & = z_2 \\ y_3 & = \frac{1}{\sqrt{5}} z_3 \end{cases}.$$

La forma quadratica allora si trasforma: $-y_1^2 - y_2^2 + 5y_3^2 = -z_1^2 - z_2^2 + z_3^2$. e corrisponde alla matrice diagonale $\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Osserviamo che la segnatura è la coppia $(1, 2)$, la forma è indefinita.

Possiamo dare un'interpretazione geometrica considerando l'insieme

$$X = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid q(x) = 2x_1^2 - 6x_1x_3 - x_2^2 + 2x_3^2 = 1\}.$$

È una **superficie quadrica** di cui vogliamo capire la geometria. Se passiamo alle coordinate (y_1, y_2, y_3) l'equazione della superficie diventa $-y_1^2 - y_2^2 + 5y_3^2 = 1$. Ora se tagliamo X con un piano affine di equazione $y_3 = k$, troviamo $y_1^2 + y_2^2 = 5k^2 - 1$: è l'equazione di

- una circonferenza se $5k^2 - 1 > 0$, cioè $k > \frac{\sqrt{5}}{5}$ o $k < -\frac{\sqrt{5}}{5}$;
- un punto se $k = \pm\frac{\sqrt{5}}{5}$
- vuoto se $-\frac{\sqrt{5}}{5} < k < \frac{\sqrt{5}}{5}$

Se invece tagliamo X con piani del tipo $y_1 = k$ o $y_2 = k$ troviamo delle iperboli, per esempio $5y_3^2 - y_1^2 = 1 + k^2$.

Nella trasformazione $x = Sy$, o $y = S^{-1}x = {}^tSx$ la superficie non viene "deformata" in quanto, essendo S ortogonale, si conservano le lunghezze. Quindi passare agli assi principali ci permette di capire la "forma" di X. Nel passare invece alle coordinate z_1, z_2, z_3 le lunghezze non si conservano e la superficie viene deformata.