

APPUNTI DI GEOMETRIA E CALCOLO NUMERICO

M.V.Catalisano A.Perelli

Indice

Capitolo 1. Preliminari

1.1. Insiemi	3
1.2. Applicazioni	4
1.3. Numeri complessi	5
1.4. Polinomi	8

Capitolo 2. Sistemi lineari e matrici

2.1. Sistemi lineari: algoritmo di riduzione gaussiana	10
2.2. Matrici	15
2.3. Determinante e caratteristica	19
2.4. Sistemi lineari: teorema di Rouché-Capelli	22
2.5. Fattorizzazione LU	25

Capitolo 3. Spazi vettoriali e trasformazioni lineari

3.1. Spazi vettoriali	27
3.2. Trasformazioni lineari	34
3.3. Sistemi lineari e trasformazioni lineari	39

Capitolo 4. Geometria analitica, parte I

4.1. Vettori geometrici	41
4.2. Geometria lineare nel piano	44
4.3. Coniche in forma canonica	48
4.4. Geometria lineare nello spazio	52
4.5. Quadriche in forma canonica	59

Capitolo 5. Diagonalizzazione

5.1. Trasformazioni lineari e matrici diagonalizzabili	64
5.2. Diagonalizzazione	69

Capitolo 6. Prodotto scalare e forme quadratiche

6.1. Prodotto scalare	72
6.2. Ortogonalità	74
6.3. Diagonalizzazione delle matrici simmetriche reali	75
6.4. Forme quadratiche	78

Capitolo 7. Geometria analitica, parte II	
7.1. Coniche e quadriche	83
7.2. Cenni su linee e superfici	86
Capitolo 8. Elementi di calcolo numerico	
8.1. Norme matriciali	88
8.2. Sistemi lineari: stabilità	90

Capitolo 1. PRELIMINARI

1.1. Insiemi

Insiemi

Si danno per note le nozioni intuitive di *insieme* e di *elemento* di un insieme. Useremo le seguenti notazioni e definizioni:

$A = \{\dots \text{ elementi } \dots\}$	<i>insieme</i>
$a \in A; a \notin A$	<i>a appartiene ad A; a non appartiene ad A</i>
$\exists; \nexists$	<i>esiste; non esiste</i>
\forall	<i>per ogni</i>
$: , \text{t.c., } $	<i>tale che</i>
$\Rightarrow; \Leftrightarrow$	<i>implica; se e solo se</i>
$A \supset B$	<i>A contiene B, cioè: $b \in B \Rightarrow b \in A$</i>
$A \subset B$	<i>A contenuto in B, cioè: $a \in A \Rightarrow a \in B$</i>
\emptyset	<i>insieme vuoto.</i>

Se $A \subset B$ allora A è *sottoinsieme* di B .

Unione di insiemi: $A \cup B = \{x : x \in A \text{ e/o } x \in B\}$. Osserviamo che gli elementi comuni ad A e B vengono contati in $A \cup B$ una sola volta. Ad esempio: se $A = \{1, 2, 3\}$ e $B = \{1, 4, 5\}$ allora $A \cup B = \{1, 2, 3, 4, 5\}$.

Intersezione di insiemi: $A \cap B = \{x : x \in A \text{ e } x \in B\}$. Osserviamo che se non vi sono elementi comuni ad A e B allora $A \cap B = \emptyset$.

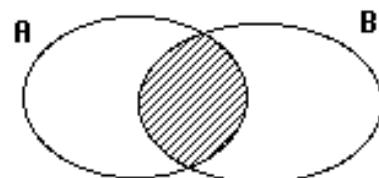
Differenza di insiemi: $A \setminus B = \{x : x \in A \text{ e } x \notin B\}$.

Complementare di B in A , con A e B tali che $A \supset B$: $C_{AB} = A \setminus B$.

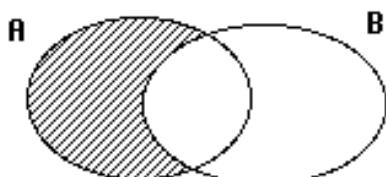
Graficamente, unione, intersezione, differenza e complementare si possono rappresentare come segue:



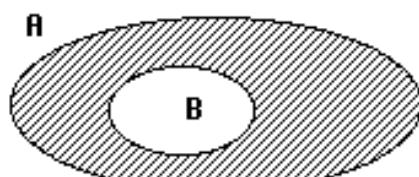
$A \cup B$



$A \cap B$



$A \setminus B$



C_{AB}

Insiemi numerici

Esempi importanti di *insiemi numerici* sono:

- \mathbb{N} numeri *naturali* = $\{0, 1, 2, 3, \dots\}$
- \mathbb{Z} numeri *interi* = $\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$
- \mathbb{Q} numeri *razionali* = $\{\frac{a}{b} : a, b \in \mathbb{Z}, b \neq 0\}$; conviene assumere che la frazione $\frac{a}{b}$ sia *ridotta*, ovvero m.c.d.(a, b)=1
- \mathbb{R} numeri *reali*; ci accontentiamo della definizione intuitiva di \mathbb{R} : i numeri reali possono essere pensati come i *punti di una retta*
- \mathbb{C} numeri *complessi*: vedi paragrafo 1.3.

Una proprietà importante di \mathbb{R} è la seguente:

$$\alpha, \beta \in \mathbb{R}, \alpha < \beta \Rightarrow \exists \gamma \in \mathbb{R} \text{ con } \alpha < \gamma < \beta.$$

In realtà esistono infiniti tali γ ; la stessa proprietà vale anche per \mathbb{Q} , ma \mathbb{R} è "più numeroso" di \mathbb{Q} , in un senso ben preciso che però non approfondiamo. Ben noti esempi di numeri in $\mathbb{R} \setminus \mathbb{Q}$ sono $\sqrt{2}$, e , π . Abbiamo

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C},$$

le inclusioni essendo strette.

Prodotto cartesiano

Dati due insiemi A e B definiamo l'operazione *prodotto cartesiano*:

$$A \times B = \{(a, b) : a \in A, b \in B\}.$$

L'idea di prodotto cartesiano nasce dal concetto di *piano cartesiano*, dato per noto. Infatti il piano cartesiano è l'insieme delle coppie (x, y) con $x, y \in \mathbb{R}$, e viene denotato con $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. L'operazione di prodotto cartesiano può essere *iterata*; in generale

$$A_1 \times A_2 \times \dots \times A_n = \{(a_1, \dots, a_n) : a_i \in A_i, i = 1, \dots, n\}.$$

Ad esempio, lo *spazio euclideo* è l'insieme delle terne (x, y, z) con $x, y, z \in \mathbb{R}$, e viene denotato con $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$. In generale scriviamo

$$\mathbb{R}^n = \mathbb{R} \times \dots \times \mathbb{R} \quad n - \text{volte}$$

ed analogamente per altri insiemi.

1.2. Applicazioni

Applicazioni

Definizione. Siano A e B due insiemi. Una *applicazione* $f : A \rightarrow B$ è una legge che ad *ogni* elemento $a \in A$ fa corrispondere un *unico* elemento $f(a) \in B$.

Useremo le seguenti notazioni e definizioni:

- A *dominio* di f
- B *codomnio* di f
- $f(a)$ *immagine* di $a \in A$
- $\text{Im } f$ *immagine* di $f = \{b \in B : b \text{ è immagine di qualche } a \in A\}$
- $f^{-1}(b)$ *controimmagine* di $b \in B = \{a \in A : f(a) = b\}$; abbiamo $f^{-1}(b) = \emptyset \Leftrightarrow b \notin \text{Im } f$
- Γ *grafico* di $f = \{(a, f(a)) \in A \times B\}$.

Esempio. $f : \mathbb{R} \rightarrow \mathbb{R}$ definita da $f(x) = x^2$ è un'applicazione, il cui grafico è una parabola.

Un'applicazione $f : A \rightarrow B$ è *iniettiva* se $a, a' \in A, a \neq a' \Rightarrow f(a) \neq f(a')$. Un'applicazione $f : A \rightarrow B$ è *surgettiva* se $\text{Im } f = B$. Un'applicazione $f : A \rightarrow B$ è *bigettiva* se è iniettiva e surgettiva. Un esempio di applicazione bigettiva è l'*applicazione identica* di un insieme A :

$$\text{id}_A : A \rightarrow A, \quad \text{id}_A(a) = a.$$

Applicazione composta e inversa

Dati tre insiemi A, B, C e due applicazioni $f : A \rightarrow B, g : B \rightarrow C$ si definisce l'*applicazione composta* $g \circ f$:

$$g \circ f : A \rightarrow C, \quad g \circ f(a) = g(f(a)).$$

Se $f : A \rightarrow B$ è bigettiva si definisce l'*applicazione inversa* f^{-1} (da non confondersi con la controimmagine):

$$f^{-1} : B \rightarrow A, \quad f^{-1}(b) = \text{quell'elemento } a \in A \text{ t.c. } f(a) = b.$$

La definizione è *ben posta* grazie alla bigettività di f . E' facile verificare che

$$f \circ f^{-1} = \text{id}_B, \quad f^{-1} \circ f = \text{id}_A.$$

1.3. Numeri complessi

Forma cartesiana

I numeri complessi nascono dall'esigenza di risolvere equazioni del tipo $x^2 + 1 = 0$, che non hanno soluzioni in \mathbb{R} . Si definisce formalmente il *numero immaginario* i , che soddisfa $i^2 = -1$. L'insieme \mathbb{C} dei *numeri complessi* è l'insieme delle espressioni formali del tipo $a + ib$ con $a, b \in \mathbb{R}$, ovvero

$$\mathbb{C} = \{a + ib : a, b \in \mathbb{R}\}.$$

I numeri complessi si denotano con la lettera z e si possono rappresentare come punti del piano cartesiano, per mezzo dell'applicazione bigettiva

$$z = a + ib \leftrightarrow (a, b). \tag{1.1}$$

In tale rappresentazione l'asse delle ascisse prende il nome di *asse reale*, quello delle ordinate di *asse immaginario* ed il piano cartesiano di *piano complesso*.

Dato $z = a + ib \in \mathbb{C}$ definiamo $a = \operatorname{Re} z$ *parte reale* e $b = \operatorname{Im} z$ *parte immaginaria* di z . Segue che, tramite l'applicazione (1.1), i numeri reali corrispondono ai numeri complessi z con $\operatorname{Im} z = 0$. In questo senso abbiamo quindi che

$$\mathbb{R} \subset \mathbb{C}.$$

La rappresentazione dei numeri complessi nella forma $z = a + ib$ si chiama *forma cartesiana* dei numeri complessi.

Introduciamo in \mathbb{C} le operazioni di somma e prodotto, in modo che siano coerenti con quelle ben note in \mathbb{R} e con il fatto che $i^2 = -1$. Dati $z = a + ib$ e $z' = a' + ib'$ definiamo

$z + z' = (a + a') + i(b + b')$	<i>somma</i>
$zz' = (aa' - bb') + i(ab' + a'b)$	<i>prodotto</i> .

Segue che:

$$0 = 0 + i0 \text{ soddisfa } 0 + z = z \text{ e } 0z = 0, \forall z \in \mathbb{C};$$

$$\text{se } \lambda \in \mathbb{R} \text{ allora } \lambda z = \lambda a + i\lambda b, \forall z \in \mathbb{C};$$

$$-z = -a - ib.$$

Definiamo il *coniugato* \bar{z} di $z \in \mathbb{C}$ come $\bar{z} = a - ib$. E' facile verificare che

$$\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2, \quad \overline{z_1 z_2} = \bar{z}_1 \bar{z}_2, \quad z \in \mathbb{R} \Leftrightarrow z = \bar{z}.$$

Se $z \neq 0$ abbiamo allora

$$\frac{1}{z} = \frac{\bar{z}}{z\bar{z}} = \frac{a - ib}{a^2 + b^2} = \frac{a}{a^2 + b^2} - i \frac{b}{a^2 + b^2}.$$

Definiamo il *modulo* $|z|$ di $z \in \mathbb{C}$ come $|z| = \sqrt{a^2 + b^2}$. Geometricamente $|z|$ rappresenta la distanza del punto z dall'origine degli assi, per il teorema di Pitagora. Inoltre

$$|z|^2 = z\bar{z}.$$

Forma polare e formula di De Moivre

Passiamo ora ad una diversa rappresentazione dei numeri complessi. Denotiamo con ρ il modulo di z , $\rho = |z|$, e con θ l'*argomento* di z , ovvero l'angolo (misurato in *radiani!*) che il segmento congiungente z con l'origine forma con l'asse reale. Tale angolo viene orientato in senso *antiorario*. E' chiaro che modulo e argomento determinano il numero complesso z . Ovviamente l'argomento è determinato a meno di multipli di 2π , ovvero i dati ρ , θ e ρ , $\theta + 2k\pi$, con $k \in \mathbb{Z}$, determinano lo stesso numero complesso. Per questo motivo si adotta la *convenzione* seguente: l'argomento θ soddisfa

$$0 \leq \theta < 2\pi.$$

La trigonometria fornisce le formule

$$\begin{cases} a = \rho \cos \theta \\ b = \rho \sin \theta \end{cases} \quad (1.2)$$

e quindi

$$z = \rho(\cos \theta + i \sin \theta),$$

detta *forma polare* dei numeri complessi. Osserviamo che le funzioni $\cos x$ e $\sin x$ sono periodiche con periodo 2π , quindi la forma polare di un numero complesso è indipendente dalla determinazione di θ scelta. Le formule inverse sono

$$\begin{cases} \rho = \sqrt{a^2 + b^2} \\ \cos \theta = \frac{a}{\rho} \\ \sin \theta = \frac{b}{\rho}, \end{cases} \quad (1.3)$$

ed è ben noto dalla trigonometria che $\cos \theta$ e $\sin \theta$ determinano θ . Le formule (1.2) e (1.3) costituiscono le *formule di passaggio* dalla forma cartesiana a quella polare di un numero complesso, e viceversa.

Il prodotto di numeri complessi è più agevole in forma polare: da ben note formule di trigonometria si deduce infatti

$$zz' = \rho\rho'(\cos(\theta + \theta') + i \sin(\theta + \theta')), \quad \frac{1}{z} = \frac{1}{\rho}(\cos \theta - i \sin \theta) \quad \text{e} \quad \bar{z} = \rho(\cos \theta - i \sin \theta).$$

La forma polare dei numeri complessi può essere espressa in modo compatto mediante la *funzione esponenziale complessa* e^z , le cui principali proprietà sono

$$e^{z+z'} = e^z e^{z'}, \quad e^{i\theta} = \cos \theta + i \sin \theta \quad \text{se} \quad \theta \in \mathbb{R} \quad (\text{formula di Eulero}).$$

Quindi

$$z = \rho(\cos \theta + i \sin \theta) = \rho e^{i\theta}$$

ed anche, per $n \in \mathbb{N}$,

$$z^n = \rho^n e^{in\theta}, \quad \bar{z} = \rho e^{-i\theta}, \quad \frac{1}{z} = \frac{1}{\rho} e^{-i\theta}.$$

Infine, si può dimostrare che l'equazione $z^n = a$, $a \in \mathbb{C}$, ha esattamente n soluzioni, tutte distinte. Tali soluzioni, dette *radici n-esime* di a , sono fornite dalla *formula di De Moivre*: posto $a = \rho e^{i\theta}$, le soluzioni z_0, \dots, z_{n-1} dell'equazione $z^n = a$ sono

$$z_k = \sqrt[n]{\rho} e^{i(\frac{\theta+2k\pi}{n})}, \quad k = 0, \dots, n-1.$$

E' interessante notare che le radici n -esime di a , una volta rappresentate sul piano complesso, dividono la circonferenza di centro l'origine e raggio $\sqrt[n]{\rho}$ in n parti uguali. Questo si verifica facilmente osservando che l'argomento di z_{k+1} differisce di $\frac{2\pi}{n}$ da quello di z_k .

1.4. Polinomi

Polinomi

Un *polinomio* è un'espressione del tipo

$$P(z) = a_n z^n + a_{n-1} z^{n-1} + \dots + a_1 z + a_0$$

dove $n \in \mathbb{N}$, e i *coefficienti* $a_n, a_{n-1}, \dots, a_1, a_0$ e la *variabile* z sono numeri complessi.

L'insieme di tutti i polinomi si denota con $\mathbb{C}[z]$; il *grado* di $P(z)$ si definisce come $\deg P = \max\{n \in \mathbb{N} : a_n \neq 0\}$ e, se $n = \deg P$, a_n è il *coefficiente direttivo* di $P(z)$. Se $\deg P = 0$ allora $P(z) = c \in \mathbb{C}$.

Vale il *principio di identità* per i polinomi:

$$P(z) = 0 \quad \forall z \in \mathbb{C} \quad \Leftrightarrow \quad a_i = 0 \quad \text{per } i = 0, \dots, n.$$

Un numero complesso α è *radice* di $P(z)$ se $P(\alpha) = 0$. Vale il seguente (difficile)

Teorema Fondamentale dell'Algebra. *Ogni polinomio $P \in \mathbb{C}[z]$ con $\deg P \geq 1$ ha almeno una radice.*

La *divisione tra polinomi* è analoga a quella tra numeri; infatti si ha

$$Q(z) \text{ divide } P(z) \Leftrightarrow \exists M(z) \text{ t.c. } P(z) = Q(z)M(z),$$

e si usa la notazione $Q(z)|P(z)$. Anche l'algoritmo di divisione è analogo a quello tra numeri; infatti dati $P, Q \in \mathbb{C}[z]$ esiste $R(z) \in \mathbb{C}[z]$ tale che

$$P(z) = Q(z)M(z) + R(z), \quad 0 \leq \deg R < \deg Q. \quad (1.4)$$

I polinomi $M(z)$ e $R(z)$, rispettivamente *quoziente* e *resto* della divisione, possono essere calcolati per mezzo del ben noto *metodo di Ruffini*.

Radici e divisibilità sono tra loro collegate; vale infatti la

Proposizione 1.1. *Siano $P \in \mathbb{C}[z]$ e $\alpha \in \mathbb{C}$. Allora $P(\alpha) = 0 \Leftrightarrow (z - \alpha)|P(z)$.*

Dimostrazione. Dalla (1.4) con $Q(z) = z - \alpha$ otteniamo, essendo $\deg(z - \alpha) = 1$,

$$P(z) = (z - \alpha)M(z) + c, \quad c \in \mathbb{C}.$$

Segue che

$$P(\alpha) = 0 \Leftrightarrow c = 0 \Leftrightarrow (z - \alpha)|P(z). \quad \square$$

La Proposizione 1.1 porta al concetto di *moltiplicità* di una radice α di $P(z)$, definita come

$$\mu_\alpha = \max\{k \in \mathbb{N} : (z - \alpha)^k | P(z)\}.$$

Teorema 1.1. *Ogni polinomio $P \in \mathbb{C}[z]$ con $\deg P(z) = n \geq 1$ ha esattamente n radici, contate con molteplicità.*

Dimostrazione. Dal Teorema Fondamentale dell'Algebra segue l'esistenza di una radice α_1 di $P(z)$. Dalla Proposizione 1.1 abbiamo

$$P(z) = (z - \alpha_1)M_1(z), \quad \deg M_1 = n - 1. \quad (1.5)$$

Se $n - 1 = 0$ il teorema è dimostrato, altrimenti riapplichiamo la stessa procedura a $M_1(z)$, ottenendo l'esistenza di una radice α_2 di $M_1(z)$, e quindi di $P(z)$, tale che

$$M_1(z) = (z - \alpha_2)M_2(z), \quad \deg M_2 = n - 2$$

e quindi dalla (1.5) abbiamo

$$P(z) = (z - \alpha_1)(z - \alpha_2)M_2(z), \quad \deg M_2 = n - 2.$$

Iterando questa procedura fino ad arrivare a $M_n(z)$ con $\deg M_n = n - n = 0$, e quindi $M_n(z) = c \in \mathbb{C}$, otteniamo che

$$P(z) = (z - \alpha_1)(z - \alpha_2) \cdots (z - \alpha_n)c, \quad (1.6)$$

ed il teorema è dimostrato. \square

Denotando con $\alpha_1, \dots, \alpha_r$ le radici distinte di $P(z)$ e con μ_1, \dots, μ_r le loro molteplicità possiamo scrivere la (1.6) nella forma

$$P(z) = c \prod_{i=1}^r (z - \alpha_i)^{\mu_i};$$

se $\deg P(z) = n$ allora chiaramente

$$n = \sum_{i=1}^r \mu_i \quad \text{e} \quad c = a_n.$$

Polinomi a coefficienti reali

Esaminiamo più in dettaglio i polinomi a coefficienti reali, il cui insieme viene denotato con $\mathbb{R}[z]$.

Proposizione 1.2. *Sia $P \in \mathbb{R}[z]$. Allora $P(\alpha) = 0 \Leftrightarrow P(\bar{\alpha}) = 0$.*

Dimostrazione. Sia $P(z) = a_n z^n + \dots + a_0$. Chiaramente

$$P(\alpha) = 0 \Leftrightarrow \overline{P(\alpha)} = 0.$$

Poichè i coefficienti di $P(z)$ sono reali abbiamo

$$\overline{P(\alpha)} = \overline{a_n \alpha^n + \dots + a_0} = \overline{a_n \alpha^n} + \dots + \overline{a_0} = \overline{a_n} \overline{\alpha}^n + \dots + \overline{a_0} = a_n \overline{\alpha}^n + \dots + a_0 = P(\bar{\alpha}),$$

e la proposizione è dimostrata. \square

Segue che se $P(z) \in \mathbb{R}[z]$ ha la radice α allora ha anche la radice $\bar{\alpha}$. Quindi le radici $\alpha \in \mathbb{C} \setminus \mathbb{R}$ di tali polinomi si possono raggruppare in coppie $\alpha, \bar{\alpha}$. Ovviamente α e $\bar{\alpha}$ hanno la stessa molteplicità. Osserviamo infine che

- i) $(z - \alpha)(z - \bar{\alpha}) = z^2 - 2z\operatorname{Re}\alpha + |\alpha|^2 \in \mathbb{R}[z]$;
- ii) se $P(z) \in \mathbb{R}[z]$ ha grado dispari, allora ha necessariamente almeno una radice $\alpha \in \mathbb{R}$.

Capitolo 2. SISTEMI LINEARI E MATRICI

2.1. Sistemi lineari: algoritmo di riduzione gaussiana

Sistemi lineari

Un *sistema lineare* \mathfrak{L} è un sistema di m equazioni lineari, ossia di grado 1, in n incognite x_1, \dots, x_n , ovvero un'espressione del tipo

$$\begin{cases} a_{11}x_1 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + \dots + a_{2n}x_n = b_2 \\ \dots \dots \dots \dots \dots \\ a_{m1}x_1 + \dots + a_{mn}x_n = b_m \end{cases}$$

con $a_{ij}, b_i \in \mathbb{R}/\mathbb{C}$ per $i = 1, \dots, m$ e $j = 1, \dots, n$. Useremo le seguenti notazioni e definizioni:

a_{ij} , $i = 1, \dots, m$ e $j = 1, \dots, n$	<i>coefficients</i>
x_j , $j = 1, \dots, n$	<i>incognites</i>
b_i , $i = 1, \dots, m$	<i>terms known</i>
$R_i : a_{i1}x_1 + \dots + a_{in}x_n = b_i$	<i>row i-th</i>
$a_{1j}x_j$	
$C_i : \dots$	<i>column j-th</i>
$a_{mj}x_j$	
$m = n$	<i>square system</i>
$b_1 = \dots = b_m = 0$	<i>homogeneous system</i> .

Una *soluzione* di \mathfrak{L} è una n -upla $(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n/\mathbb{C}^n$ tale che ponendo $x_1 = \alpha_1, \dots, x_n = \alpha_n$ tutte le equazioni di \mathfrak{L} sono soddisfatte. L'*insieme delle soluzioni* di \mathfrak{L} si denota con S . Usando una terminologia tratta dalla Fisica, le equazioni rappresentano i *vincoli* del sistema \mathfrak{L} , mentre le incognite rappresentano i *gradi di libertà* di \mathfrak{L} . Vedremo in seguito che un sistema lineare può soltanto avere 0 *soluzioni*, 1 *soluzione* oppure ∞ *soluzioni*.

Dato un sistema lineare *a caso*, è *probabile* che:

- i) se $m > n$ vi siano 0 soluzioni;
- ii) se $m = n$ vi sia 1 soluzione;
- iii) se $m < n$ vi siano ∞ soluzioni.

Esempio. i) $m = 2, n = 1$: il sistema $\begin{cases} x = 1 \\ x = 2 \end{cases}$ ha 0 soluzioni

ii) $m = n = 2$: il sistema $\begin{cases} x + y = 1 \\ x - y = 0 \end{cases}$ ha 1 soluzione: $(\frac{1}{2}, \frac{1}{2})$

iii) $m = 1, n = 2$: il sistema $x + y = 0$ ha ∞ soluzioni: $(a, -a), a \in \mathbb{R}$.

E' chiaro che un sistema omogeneo ha sempre almeno 1 soluzione, ovvero la *soluzione banale* $(0, \dots, 0)$.

Il nostro scopo è

- a) individuare semplici criteri che permettano di decidere se un sistema lineare ha soluzioni oppure no, ed in caso affermativo quante

b) individuare semplici algoritmi per il calcolo delle soluzioni.

Operazioni elementari

Studiamo ora alcune semplici operazioni, dette *operazioni elementari*, che consentono di *trasformare* un sistema lineare, lasciandone *inalterate* le soluzioni. Tali operazioni agiscono sulle *righe* di \mathfrak{L} ; è chiaro che analoghe operazioni possono essere fatte sulle *colonne* di \mathfrak{L} , ma in questo caso le soluzioni non rimangono necessariamente inalterate.

i) operazione di *scambio* : $R_i \leftrightarrow R_j$.

E' chiaro che lo scambio di righe non altera le soluzioni di \mathfrak{L} .

ii) operazione di *moltiplicazione (per scalare)* : $R_i \rightarrow \lambda R_i$ con $\lambda \neq 0$, $\lambda \in \mathbb{R}/\mathbb{C}$.

Anche in questo caso è chiaro che la moltiplicazione di una riga per uno scalare $\lambda \neq 0$ non altera le soluzioni di \mathfrak{L} .

iii) operazione di *combinazione lineare* : $R_i \rightarrow R_i + \lambda R_j$ con $i \neq j$, $\lambda \in \mathbb{R}/\mathbb{C}$.

Tale operazione non altera le soluzioni di \mathfrak{L} . Infatti se $(\alpha_1, \dots, \alpha_n)$ è soluzione, allora sia R_i che R_j sono soddisfatte, e quindi anche $R_i + \lambda R_j$ è soddisfatta. Viceversa, se $(\beta_1, \dots, \beta_n)$ è soluzione del sistema \mathfrak{L}' trasformato mediante un'operazione di combinazione lineare, allora sia $R_i + \lambda R_j$ che R_j sono soddisfatte. Ma allora anche λR_j è soddisfatta, e quindi per sottrazione R_i è soddisfatta a sua volta. Segue che \mathfrak{L} e \mathfrak{L}' hanno le stesse soluzioni.

Osserviamo che le operazioni elementari sono *invertibili*, ovvero: se \mathfrak{L} è stato trasformato in \mathfrak{L}' mediante applicazione di un'operazione elementare, si può trasformare \mathfrak{L}' in \mathfrak{L} mediante applicazione di una opportuna operazione elementare. Le operazioni *inverse* delle tre operazioni elementari viste sopra sono

i) $R_j \leftrightarrow R_i$

ii) $R_i \rightarrow \frac{1}{\lambda} R_i$ ($\lambda \neq 0$)

iii) $R_i \rightarrow R_i - \lambda R_j$ ($i \neq j$),

come è facile verificare.

Definizione. Due sistemi lineari sono *equivalenti* se possono essere trasformati l'uno nell'altro mediante ripetute applicazioni di operazioni elementari.

Dalla discussione precedente abbiamo allora:

Proposizione 2.1. *Sistemi lineari equivalenti hanno le stesse soluzioni.*

Algoritmo di riduzione gaussiana

Passiamo all'*algoritmo di riduzione gaussiana*, che descrive un metodo standard di trasformazione di un sistema lineare in uno equivalente, ma decisamente più semplice per il calcolo delle soluzioni. Tale algoritmo si ispira alla ben nota tecnica di *sostituzione* delle incognite per la risoluzione dei sistemi lineari. Per semplicità scriviamo i sistemi lineari indicando solo i coefficienti e la colonna dei termini noti.

Esempio. $\begin{cases} 3x_1 + x_2 = -1 \\ -x_1 = 2 \end{cases}$ si scrive $\left(\begin{array}{cc|c} 3 & 1 & -1 \\ -1 & 0 & 2 \end{array} \right)$

Passo 1. *Individuare una riga, sia essa R_i , in cui il coefficiente della x_1 sia non nullo.*

Operare $R_i \leftrightarrow R_1$. Se tutti i coefficienti della x_1 sono nulli, passare alla x_2 e così via fino

a trovare la prima incognita con almeno un coefficiente non nullo. Portare tale coefficiente alla prima riga mediante opportuno scambio di righe.

Tale coefficiente non nullo si chiama *pivot* della incognita corrispondente e tale incognita prende il nome di *incognita pivotale*. Quindi se la prima colonna non è tutta nulla si tratterà del pivot della x_1 , altrimenti sarà il pivot della x_2 , o della x_3 ...

Esempio. Partendo dal sistema sulla sinistra otteniamo:

$$\left(\begin{array}{ccccc|c} 0 & 2 & 2 & 1 & -5 & 5 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 3 & 3 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 & 1 \end{array} \right) \quad \begin{array}{l} R_4 \leftrightarrow R_1 \\ \longrightarrow \end{array} \quad \left(\begin{array}{ccccc|c} 0 & 1 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 3 & 3 & -1 & 0 & 0 \\ 0 & 2 & 2 & 1 & -5 & 5 \end{array} \right)$$

La x_1 non ha pivot, ed il pivot della x_2 è in grassetto.

Osserviamo come il passo 1 contenga una possibilità di *scelta* nel caso vi siano più coefficienti non nulli per una incognita. Conviene, per comodità di calcolo nei passi successivi, scegliere come pivot il coefficiente *più semplice*, come è stato fatto nell'esempio. In tale esempio le altre possibilità di scelta nel passo 1 sono: $R_1 \leftrightarrow R_1$ (ovvero lasciare inalterato il sistema) e $R_3 \leftrightarrow R_1$.

Passo 2. Usare il pivot della x_1 (o della x_2 , ...) per annullare, mediante opportune combinazioni lineari di righe, tutti i coefficienti della x_1 (o della x_2 , ...) che sono al di sotto del pivot.

Esempio. Continuando l'esempio precedente otteniamo:

$$\left(\begin{array}{ccccc|c} 0 & 1 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 3 & 3 & -1 & 0 & 0 \\ 0 & 2 & 2 & 1 & -5 & 5 \end{array} \right) \quad \begin{array}{l} R_3 \rightarrow R_3 - 3R_1 \\ R_4 \rightarrow R_4 - 2R_1 \\ \longrightarrow \end{array} \quad \left(\begin{array}{ccccc|c} 0 & 1 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & -1 & 3 & -3 \\ 0 & 0 & 0 & 1 & -3 & 3 \end{array} \right)$$

Osserviamo come il passo 2 equivalga a ricavare la x_2 da R_1 e sostituirla in R_3 e R_4 .

Passo iterativo. Trascurare R_1 e applicare i passi 1 e 2 al sistema composto da R_2, \dots, R_m . Ripetere la procedura trascurando R_1 e R_2 ed applicando i passi 1 e 2 al sistema composto da R_3, \dots, R_m , e così via fino a quando non vi sono più righe o incognite cui applicare i passi 1 e 2.

Esempio. Proseguendo l'esempio precedente otteniamo:

$$\left(\begin{array}{ccccc|c} 0 & 1 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & -1 & 3 & -3 \\ 0 & 0 & 0 & 1 & -3 & 3 \end{array} \right) \quad \begin{array}{l} R_3 \rightarrow R_3 + R_2 \\ R_4 \rightarrow R_4 - R_2 \\ \longrightarrow \end{array} \quad \left(\begin{array}{ccccc|c} 0 & 1 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 4 & -2 \\ 0 & 0 & 0 & 0 & -4 & 2 \end{array} \right)$$

$$\begin{array}{l} R_4 \rightarrow R_4 + R_3 \\ \longrightarrow \end{array} \quad \left(\begin{array}{ccccc|c} 0 & 1 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 4 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

L'algoritmo termina in quanto non vi sono più incognite cui applicare il passo 1. Abbiamo quindi tre incognite pivotali, x_2 , x_4 e x_5 , con i rispettivi pivot 1, 1 e 4.

Al termine dell'algoritmo si perviene ad un sistema lineare *ridotto*, ovvero un sistema lineare in cui il primo coefficiente non nullo di ogni riga è più a sinistra del primo coefficiente non nullo della riga successiva. Ovviamente, i primi coefficienti non nulli di ogni riga sono i pivot. Osserviamo che mediante l'algoritmo di riduzione gaussiana abbiamo dimostrato il seguente

Teorema 2.1. *Ogni sistema lineare è equivalente ad un sistema lineare ridotto.*

Osserviamo inoltre che il sistema ridotto cui si perviene mediante l'algoritmo di riduzione gaussiana *non è unico*, in quanto il passo 1 permette di operare delle scelte.

Calcolo delle soluzioni e riduzione totale

E' chiaro che un sistema lineare ridotto è più semplice da risolvere rispetto ad un sistema generale; infatti si può partire dall'ultima riga e procedere a ritroso con il ben noto metodo di sostituzione delle incognite. Vediamo comunque la procedura generale di calcolo delle soluzioni di un sistema ridotto. Osserviamo per prima cosa che possono esserci delle *righe non-significative*, ovvero del tipo $0 = 0$, vedi R_4 del sistema ridotto dell'esempio. Tali righe possono essere *trascurate*, poichè il sistema formato dalle sole righe significative ha ovviamente le stesse soluzioni di quello di partenza.

A questo punto è molto semplice decidere se il sistema ha soluzioni. Scriviamo:

$$p = \text{numero delle incognite pivotali}$$

$$q = \text{numero delle righe significative};$$

ovviamente $q \geq p$.

Caso 1: $q > p$. L'ultima riga significativa sarà quindi del tipo $0 = b_q$, con $b_q \neq 0$, impossibile. Il sistema *non ha soluzioni*.

Caso 2: $q = p$. Il sistema *ha soluzioni*. Possiamo infatti trascurare le righe non-significative e portare a secondo membro le incognite non-pivotali. Supponiamo, per semplicità di notazione, che x_1, \dots, x_p siano le incognite pivotali e x_{p+1}, \dots, x_n quelle non-pivotali. A questo punto assegnamo alle incognite non-pivotali *valori arbitrari*, siano essi $x_{p+1} = \alpha_{p+1}, \dots, x_n = \alpha_n$. Otteniamo un sistema quadrato $p \times p$ ridotto con p incognite pivotali, del tipo

$$\left\{ \begin{array}{rcll} a_{11}x_1 + & \dots & \dots & = b_1 \\ & a_{22}x_2 + & \dots & = b_2 \\ & \dots & \dots & \dots \\ & a_{pp}x_p & = b_p \end{array} \right. \quad a_{i,i} \neq 0, \quad i = 1, \dots, p.$$

Tale sistema ha ovviamente 1 soluzione, come si vede risolvendo a partire dall'ultima riga.

Sia $(x_1, \dots, x_p) = (\alpha_1, \dots, \alpha_p)$ tale soluzione. Segue che il sistema di partenza ha la soluzione $(x_1, \dots, x_n) = (\alpha_1, \dots, \alpha_n)$.

Se $n > p$ abbiamo la possibilità di assegnare infiniti valori arbitrari alle $n - p$ incognite non-pivotali e, per ogni tale assegnazione, otteniamo 1 soluzione per le rimanenti incognite.

Otteniamo quindi ∞ soluzioni. Si usa sottolineare il fatto che ci sono $n - p$ possibilità di assegnazioni arbitrarie scrivendo che il sistema ha ∞^{n-p} soluzioni. Se $n = p$ abbiamo esattamente 1 soluzione, ed usiamo la convenzione $\infty^0 = 1$. Abbiamo quindi il seguente

Criterio. *i) $q > p \Rightarrow 0$ soluzioni ii) $q = p \Rightarrow \infty^{n-p}$ soluzioni.*

Nel caso *ii)* le soluzioni si ottengono nel modo descritto, ovvero dipendono da $n - p$ parametri cui si possono assegnare valori arbitrari.

Esempio. Nell'esempio precedente abbiamo $q = p = 3$ e $n = 5$. Le incognite non-pivotali sono x_1 e x_3 . Abbiamo quindi ∞^2 soluzioni, e precisamente $(a, \frac{1}{2} - b, b, \frac{3}{2}, -\frac{1}{2})$ al variare di $a, b \in \mathbb{R}$.

Per completezza descriviamo l'*algoritmo di riduzione gaussiana totale*, a partire da un sistema ridotto. Tale algoritmo porta il sistema in una forma in cui le soluzioni sono immediate. I passi sono i seguenti.

Passo 1. *Rendere uguale a 1 ogni pivot, mediante l'operazione di prodotto per scalare.*

Esempio. Mediante l'operazione $R_3 \rightarrow \frac{1}{4}R_3$ il nostro sistema diviene

$$\left(\begin{array}{ccccc|c} 0 & \mathbf{1} & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & \mathbf{1} & 1 & 1 \\ 0 & 0 & 0 & 0 & \mathbf{1} & -1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

Passo 2. *Partendo dall'ultima riga contenente un'incognita pivotale, annullare tutti i coefficienti al di sopra dei pivot mediante opportune operazioni di combinazione lineare di righe.*

Esempio. Nel nostro caso abbiamo da annullare solo i coefficienti al di sopra del pivot della x_5 .

$$\begin{array}{l} R_2 \rightarrow R_2 - R_3 \\ R_1 \rightarrow R_1 + R_3 \\ \hline \end{array} \quad \left(\begin{array}{ccccc|c} 0 & \mathbf{1} & 1 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 3/2 \\ 0 & 0 & 0 & 0 & \mathbf{1} & -1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

Osserviamo che il sistema così ottenuto è di immediata risoluzione: basta infatti portare a secondo membro le due incognite non-pivotali ed assegnare loro valori arbitrari. I corrispondenti valori delle incognite pivotali sono immediati.

L'algoritmo di riduzione gaussiana totale trasforma un sistema ridotto in uno equivalente *totalmente ridotto*, ovvero in un sistema lineare ridotto in cui i pivot sono tutti uguali a 1 e ogni altro coefficiente nella colonna di un pivot è uguale a 0. Osserviamo che è possibile dimostrare che il sistema totalmente ridotto è *unico*. Abbiamo quindi dimostrato il

Teorema 2.2. *Ogni sistema lineare è equivalente ad un sistema lineare totalmente ridotto.*

Concludiamo il paragrafo osservando che il metodo sopra esposto risponde alle richieste *a)* e *b)* fatte all'inizio del paragrafo. Tale metodo non è sempre il migliore per la risoluzione di un sistema lineare; vedremo in seguito altre tecniche che rispondono alle richieste *a)* e

b). Vedremo inoltre che tali tecniche sono spesso vantaggiose dal punto di vista teorico, e talvolta anche dal punto di vista pratico.

2.2. Matrici

Matrici

Una *matrice* è una tabella rettangolare di numeri (talvolta di simboli) formata da m righe e n colonne:

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} = (a_{ij}), \quad i = 1, \dots, m \quad \text{e} \quad j = 1, \dots, n, \quad a_{ij} \in \mathbb{R}/\mathbb{C}.$$

L'insieme delle matrici con m righe e n colonne, dette matrici $m \times n$, si denota con $M_{m \times n}(\mathbb{R}/\mathbb{C})$. Useremo le seguenti notazioni e definizioni:

a_{ij}	coefficienti o entrate
R_i	riga i -esima
C_j	colonna j -esima
$(a_1 \dots a_n)$	matrice riga
$\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}$	matrice colonna
$m = n$	matrice quadrata di ordine n
$\begin{pmatrix} a_{11} & \dots & \dots & \dots \\ \dots & a_{22} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & a_{nn} \end{pmatrix}$	diagonale della matrice $n \times n$
$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$	matrice identica, denotata con I
$\begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_n \end{pmatrix}$	matrice diagonale, denotata con Δ
$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}$	matrice triangolare superiore
$\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$	matrice triangolare inferiore.

Data una matrice A $m \times n$ e dati $m' \leq m$ e $n' \leq n$, ogni matrice ottenuta dall'intersezione di m' righe e n' colonne di A è una *sottomatrice* di A .

Introduciamo le *operazioni* tra matrici. Date $A, B \in M_{m \times n}(\mathbb{R}/\mathbb{C})$, con $A = (a_{ij})$ e $B = (b_{ij})$, e $\lambda \in \mathbb{R}/\mathbb{C}$ definiamo

$$A + B = (c_{ij}) \text{ con } c_{ij} = a_{ij} + b_{ij}$$

somma

$$\lambda A = (c_{ij}) \text{ con } c_{ij} = \lambda a_{ij}$$

prodotto per scalare.

Queste due operazioni godono delle proprietà usuali delle operazioni tra numeri: sono *commutative, associative* ed esistono

la matrice nulla $0 = \begin{pmatrix} 0 & \dots & 0 \\ 0 & \dots & 0 \\ 0 & \dots & 0 \end{pmatrix}$ e la matrice opposta $-A = (-a_{ij})$.

Più interessante è il *prodotto righe per colonne* di due matrici. Siano A matrice $m \times k$ e B matrice $k \times n$ (osservare che *il numero di colonne di A è uguale al numero di righe di B*).

Data una riga R_i di A e una colonna C_j di B definiamo il *prodotto di R_i per C_j* come

$$R_i \cdot C_j = (a_{i1} \dots a_{ik}) \begin{pmatrix} b_{1j} \\ \dots \\ b_{kj} \end{pmatrix} = \sum_{l=1}^k a_{il} b_{lj}.$$

Definiamo allora il prodotto righe per colonne come

$$AB = (c_{ij}) \text{ con } c_{ij} = R_i \cdot C_j, \quad i = 1, \dots, m \text{ e } j = 1, \dots, n.$$

Segue che AB è una matrice $m \times n$.

E' chiaro che, in generale, se si può fare il prodotto AB non è detto si possa fare il prodotto BA . Il caso più generale in cui si può fare sia AB che BA è quello di A matrice $m \times n$ e B matrice $n \times m$; in tal caso abbiamo che AB è $m \times m$ e BA è $n \times n$. In particolare, se A e B sono entrambe $n \times n$, allora è certamente possibile fare AB e BA . Osserviamo che il prodotto righe per colonne *non è commutativo*, ovvero, in generale, $AB \neq BA$; è addirittura possibile avere $AB = 0$ con $A, B \neq 0$.

Esempio. Siano $A = \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 1 & -2 \\ -1 & 2 \end{pmatrix}$. Si verifica che $AB = \begin{pmatrix} -2 & 4 \\ -1 & 2 \end{pmatrix}$ e $BA = 0$.

Osserviamo inoltre che

$$AI = IA = A \quad \text{e} \quad A0 = 0A = 0 \quad \text{per ogni } A$$

non appena le dimensioni di A , I e 0 permettano di fare i prodotti,

$$\Delta_1 \Delta_2 = \Delta_2 \Delta_1 = \begin{pmatrix} a_1 b_1 & & 0 \\ & \dots & \\ 0 & & a_n b_n \end{pmatrix} \text{ se } \Delta_1 = \begin{pmatrix} a_1 & & 0 \\ & \dots & \\ 0 & & a_n \end{pmatrix} \text{ e } \Delta_2 = \begin{pmatrix} b_1 & & 0 \\ & \dots & \\ 0 & & b_n \end{pmatrix}$$

e il prodotto di matrici triangolari (inferiori o superiori) è ancora una matrice triangolare (inferiore o superiore).

Una matrice A $n \times n$ è *invertibile* se esiste B $n \times n$ tale che $AB = BA = I$. In tal caso B è la matrice *inversa* di A e viene denotata con A^{-1} .

Proposizione 2.2. *Siano A e B matrici $n \times n$ invertibili. Allora AB è invertibile e $(AB)^{-1} = B^{-1}A^{-1}$.*

Dimostrazione. $(AB)(B^{-1}A^{-1}) = ABB^{-1}A^{-1} = AA^{-1} = I$, ed analogamente per $(B^{-1}A^{-1})(AB)$. \square

Osserviamo che una matrice invertibile non può avere righe o colonne *nulle*, come si verifica facilmente. Data una matrice $A = (a_{ij})$ $m \times n$ definiamo la matrice *trasposta* $A^T = (t_{ij})$ $n \times m$ con $(t_{ij}) = (a_{ji})$. In pratica A^T si ottiene da A scambiando le righe con le colonne. Non è difficile dimostrare che date le matrici A $m \times k$ e B $k \times n$ si ha

$$(AB)^T = B^T A^T; \quad (2.1)$$

controlliamo soltanto che le dimensioni di tali matrici siano coerenti:

$$AB \text{ } m \times n \Rightarrow (AB)^T \text{ } n \times m; \quad B^T \text{ } n \times k, \quad A^T \text{ } k \times m \Rightarrow B^T A^T \text{ } n \times m.$$

Una matrice quadrata A è *simmetrica* se $A = A^T$. Osserviamo che A è simmetrica se e solo se i suoi coefficienti a_{ij} sono *simmetrici rispetto alla diagonale*, ovvero $a_{ij} = a_{ji}$ per ogni i, j .

Definiamo infine tre tipi di matrici quadrate la cui forma, come vedremo in seguito, consente talvolta semplificazioni nei calcoli:

- i) matrice *diagonale a blocchi*: è una matrice nulla al di fuori di opportune sottomatrici quadrate centrate sulla diagonale
- ii) matrice *triangolare superiore a blocchi*: è nulla al di sotto di opportune sottomatrici quadrate centrate sulla diagonale
- iii) matrice *triangolare inferiore a blocchi*: è nulla al di sopra di opportune sottomatrici quadrate centrate sulla diagonale.

Tali sottomatrici quadrate, dette *blocchi*, vengono denotate con B_1, \dots, B_k .

Forma matriciale dei sistemi lineari

Per mezzo delle matrici è possibile scrivere i sistemi lineari in modo compatto e utile per il seguito. Dato un sistema lineare \mathfrak{L} poniamo

$$\begin{aligned} A &= \begin{pmatrix} a_{11} & \dots & a_{1n} \\ a_{m1} & \dots & a_{mn} \end{pmatrix} && \text{matrice dei coefficienti (o incompleta)} \\ b &= \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} && \text{colonna dei termini noti} \\ x &= \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} && \text{colonna delle incognite} \\ A|b &= \begin{pmatrix} a_{11} & \dots & a_{1n} & b_1 \\ \dots & \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} & b_m \end{pmatrix} && \text{matrice completa.} \end{aligned}$$

Il sistema lineare \mathfrak{L} si può allora scrivere nella forma

$$Ax = b \quad (2.2)$$

dove Ax è il prodotto (righe per colonne) di A per x . Non è difficile verificare che la (2.2) è equivalente al sistema lineare \mathfrak{L} ; nuovamente, controlliamo soltanto che le dimensioni delle matrici siano coerenti:

$$(m \times n)(n \times 1) = m \times 1.$$

La (2.2) è detta *forma matriciale* del sistema lineare \mathfrak{L} ; è una forma comoda e, come vedremo, *operativa*. Osserviamo infine come la (2.2) sia *formalmente* simile ad un'equazione lineare $ax = b$. E' ben noto che se $a \neq 0$ allora $x = a^{-1}b$; vedremo in seguito che, essenzialmente, la stessa regola di risoluzione vale anche per i sistemi lineari.

Matrici elementari e riduzione di matrici

Vediamo come le operazioni elementari descritte nel paragrafo precedente si possano effettuare mediante prodotto di opportune matrici. A tal scopo introduciamo le matrici *elementari*; data $A \in M_{m \times n}(\mathbb{R}/\mathbb{C})$, le matrici elementari sono:

- i) matrice di *scambio* E_{ij} : è la matrice che si ottiene dalla matrice identica I mediante $R_i \leftrightarrow R_j$. E' facile verificare che la matrice di scambio agisce nel modo seguente:
se E_{ij} è $m \times m$ allora $E_{ij}A = \text{matrice ottenuta da } A \text{ mediante } R_i \leftrightarrow R_j$
se E_{ij} è $n \times n$ allora $AE_{ij} = \text{matrice ottenuta da } A \text{ mediante } C_i \leftrightarrow C_j$;
- ii) matrice di *moltiplicazione (per scalare)* $E_i(\lambda)$, $\lambda \neq 0$ e $\lambda \in \mathbb{R}/\mathbb{C}$: è la matrice che si ottiene da I mediante $R_i \rightarrow \lambda R_i$ ed agisce nel modo seguente:
se $E_i(\lambda)$ è $m \times m$ allora $E_i(\lambda)A = \text{matrice ottenuta da } A \text{ mediante } R_i \rightarrow \lambda R_i$
se $E_i(\lambda)$ è $n \times n$ allora $AE_i(\lambda) = \text{matrice ottenuta da } A \text{ mediante } C_i \rightarrow \lambda C_i$;
- iii) matrice di *combinazione lineare* $E_{ij}(\lambda)$, $i \neq j$ e $\lambda \in \mathbb{R}/\mathbb{C}$: è la matrice ottenuta da I mediante $R_i \rightarrow R_i + \lambda R_j$, ed agisce nel modo seguente:
se $E_{ij}(\lambda)$ è $m \times m$ allora $E_{ij}(\lambda)A = \text{matrice ottenuta da } A \text{ mediante } R_i \rightarrow R_i + \lambda R_j$
se $E_{ij}(\lambda)$ è $n \times n$ allora $AE_{ij}(\lambda) = \text{matrice ottenuta da } A \text{ mediante } C_j \rightarrow C_j + \lambda C_i$ (notare lo *scambio di indici*).

Abbiamo quindi che la moltiplicazione a sinistra per una matrice elementare agisce sulle righe, mentre la moltiplicazione a destra agisce sulle colonne. Inoltre l'azione di una matrice elementare è analoga alla corrispondente operazione elementare. Non sorprende quindi che le matrici elementari siano *invertibili*; le matrici inverse sono:

- i) $E_{ij}^{-1} = E_{ij}$
- ii) $E_i(\lambda)^{-1} = E_i(\frac{1}{\lambda})$
- iii) $E_{ij}(\lambda)^{-1} = E_{ij}(-\lambda)$.

A questo punto è chiaro che la riduzione gaussiana dei sistemi lineari descritta nel paragrafo precedente ha un analogo nel caso delle matrici. Denoteremo la generica matrice elementare con la lettera E .

Definizione. i) Due matrici $A, B \in M_{m \times n}(\mathbb{R}/\mathbb{C})$ sono *equivalenti (per righe)* se esistono matrici elementari E_1, \dots, E_k tali che

$$E_k \cdots E_1 A = B.$$

ii) Una matrice A è *ridotta (per righe)* se il primo coefficiente non nullo di ogni riga è su una colonna più a sinistra del primo coefficiente non nullo della riga successiva. Tali coefficienti non nulli sono i *pivot* di A . Una matrice è *totalmente ridotta (per righe)* se è ridotta (per righe), se il primo coefficiente non nullo di ogni riga è uguale a 1 e se le colonne contenenti tali coefficienti hanno tutti gli altri coefficienti uguali a 0.

Analogamente al caso dei sistemi lineari possiamo dimostrare il

Teorema 2.3. *Ogni matrice è equivalente (per righe) ad una matrice ridotta (per righe) ed anche ad una matrice totalmente ridotta (per righe).*

Definizioni e risultati del tutto simili valgono anche nel caso delle *colonne*. Osserviamo inoltre che, come per i sistemi lineari, la riduzione di una matrice non è unica.

Le matrici elementari ed il processo di riduzione gaussiana consentono di ottenere una caratterizzazione delle matrici invertibili che sarà utile nel seguito, ad esempio per il calcolo della matrice inversa.

Teorema 2.4. *Sia A una matrice quadrata di ordine n . Sono equivalenti*

- i)* A è invertibile
- ii)* riducendo totalmente A (per righe) si ottiene I
- iii)* A è prodotto di matrici elementari.

Dimostrazione. *i)* \Rightarrow *ii)*. Osserviamo che riducendo A otteniamo una matrice che non può avere righe nulle. Infatti, se B è una matrice ridotta equivalente ad A , abbiamo $E_k \cdots E_1 A = B$; segue che B è invertibile in quanto prodotto di matrici invertibili e quindi non può avere righe nulle. Abbiamo quindi n pivot, che necessariamente stanno tutti sulla diagonale. E' chiaro quindi che riducendo totalmente A otteniamo la matrice identica I .

ii) \Rightarrow *iii)*. Riducendo A totalmente otteniamo $E_k \cdots E_1 A = I$, quindi $A = E_1^{-1} \cdots E_k^{-1}$; il risultato segue osservando che l'inversa di una matrice elementare è ancora una matrice elementare.

iii) \Rightarrow *i)*. Questo è ovvio grazie alla Proposizione 2.2, in quanto le matrici elementari sono invertibili. \square

2.3. Determinante e caratteristica

Determinante

Data una matrice quadrata A $n \times n$ denotiamo con A_{ij} la sottomatrice $(n-1) \times (n-1)$ di A ottenuta sopprimendo la riga i -esima R_i e la colonna j -esima C_j . Diamo la definizione di *determinante*:

$n = 1$: se $A = (a)$ allora $\det A = a$

$n = 2$: se $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ allora $\det A = ad - bc$

$n \geq 3$: $\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13} + \dots + (-1)^{n+1} a_{1n} \det A_{1n}$.

Tale definizione di determinante è *ricorsiva* in quanto il determinante di una matrice $n \times n$ si esprime per mezzo del determinante di matrici $(n - 1) \times (n - 1)$, e così via.

Si verifica facilmente che

$$\det I = 1$$

$$\det E_{ij} = -1 \quad (i \neq j)$$

$$\det E_i(\lambda) = \lambda$$

$$\det E_{ij}(\lambda) = 1 \quad (i \neq j)$$

$$\det \Delta = a_{11} \cdots a_{nn} \quad (\Delta \text{ matrice diagonale}).$$

E' chiaro che se la prima riga di A è nulla, allora $\det A = 0$. E' anche chiaro che se A ha una riga nulla allora $\det A = 0$. Questo è ovvio se $n = 1, 2$, mentre se $n \geq 3$ basta osservare che sviluppando $\det A$ per mezzo della definizione ricorsiva si perviene necessariamente ad un'espressione per $\det A$ in termini di determinanti di matrici aventi la prima riga nulla. Vale il seguente risultato, di immediata verifica nel caso di matrici diagonali.

Teorema 2.5. *Siano A, B matrici $n \times n$. Allora $\det(AB) = \det A \det B$.*

Abbiamo la seguente importante *caratterizzazione* delle matrici invertibili.

Teorema 2.6. *Una matrice A $n \times n$ è invertibile se e solo se $\det A \neq 0$. Inoltre, se A è invertibile allora $\det(A^{-1}) = (\det A)^{-1}$.*

Dimostrazione. Sia A invertibile; per il Teorema 2.4 abbiamo che $A = E_k \cdots E_1$ con E_i matrici elementari. Dal Teorema 2.5 deduciamo quindi che $\det A = \det E_k \cdots \det E_1 \neq 0$, in quanto le matrici elementari hanno determinante non nullo. Viceversa sia $\det A \neq 0$; la riduzione totale di A fornisce $E_k \cdots E_1 A = R$ e quindi per il Teorema 2.5 abbiamo $\det R \neq 0$. Ma l'unica matrice totalmente ridotta con determinante non nullo è chiaramente la matrice identica; quindi $R = I$ e $E_k \cdots E_1 A = I$, ovvero A è invertibile. Infine, se A è invertibile dal Teorema 2.5 abbiamo $\det A \det(A^{-1}) = \det I = 1$, e il teorema è dimostrato. \square

Estendendo la nozione di combinazione lineare di due righe, diciamo che una *combinazione lineare di righe* è un'espressione del tipo

$$\sum_{i=1}^k \lambda_i R_i , \quad \lambda_i \in \mathbb{R}/\mathbb{C}.$$

Un'analogia definizione vale per le colonne. Dal Teorema 2.5 e dalle proprietà delle matrici elementari otteniamo le proprietà seguenti:

- (i) scambiando due righe il determinante cambia segno
- (ii) moltiplicando una riga per λ il determinante viene moltiplicato per λ
- (iii) sommando ad una riga R_i una combinazione lineare di righe diverse da R_i il determinante non cambia
- (iv) se una riga è combinazione lineare di altre righe il determinante è nullo.

Osserviamo che le (i)–(iii) sono ovvie, e la (iv) si giustifica facilmente notando che in tal caso moltiplicando per opportune matrici di combinazione lineare si perviene ad una matri-

ce con una riga nulla.

Il teorema seguente fornisce utili formule per il calcolo del determinante.

Teorema di Laplace. *Sia A una matrice $n \times n$. Allora*

i) per ogni $i = 1, \dots, n$

$$\det A = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det A_{ij}$$

ii) per ogni $j = 1, \dots, n$

$$\det A = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det A_{ij}.$$

La *i*) prende il nome di *sviluppo secondo la riga i -esima* del determinante, mentre la *ii)* è lo *sviluppo secondo la colonna j -esima*. Tali sviluppi consentono di scegliere la riga o colonna più favorevole per il calcolo del determinante. Un fattore importante per tale scelta è il numero di coefficienti nulli in una riga o colonna; ad esempio, si verifica facilmente che se A ha una colonna nulla allora $\det A = 0$.

Grazie al teorema di Laplace abbiamo le seguenti ulteriori proprietà del determinante:

(v) se A è triangolare (inferiore o superiore) allora $\det A = a_{11} \cdots a_{nn}$

(vi) se A è diagonale a blocchi o triangolare a blocchi (inferiore o superiore) allora $\det A = \det B_1 \cdots \det B_k$.

Osserviamo inoltre che

$$E_{ij}^T = E_{ij}, \quad E_i(\lambda)^T = E_i(\lambda), \quad E_{ij}(\lambda)^T = E_{ji}(\lambda);$$

segue allora che se E è una matrice elementare

$$\det E^T = \det E. \tag{2.3}$$

In generale abbiamo

Teorema 2.7. *Sia A una matrice quadrata; allora $\det A^T = \det A$.*

Dimostrazione. Supponiamo che $\det A \neq 0$. Allora per il Teorema 2.6 abbiamo che A è invertibile, e quindi $A = E_1 \cdots E_k$ per il Teorema 2.4. Quindi $A^T = E_k^T \cdots E_1^T$ per la (2.1) ed ancora $\det A^T = \det E_k^T \cdots \det E_1^T = \det E_k \cdots \det E_1 = \det A$ per il Teorema 2.5 e la (2.3). Quindi il teorema è dimostrato se $\det A \neq 0$. Se $\det A = 0$ allora A non è invertibile. Ma allora anche A^T non è invertibile; infatti se A^T fosse invertibile avremmo $A^T B = I$ per un'opportuna matrice B e quindi, prendendo la trasposta di entrambi i lati, $B^T A = I$ per la (2.1), da cui seguirebbe che A è invertibile. Abbiamo quindi $\det A^T = 0$, ed il teorema è dimostrato. \square

Caratteristica

Data una matrice A $m \times n$, un *minore* di ordine k di A è una sottomatrice quadrata $k \times k$ di A ; è chiaro che $k \leq \min(m, n)$.

Definizione. Una matrice A ha *caratteristica* (o *rango*) uguale a k se

- i) esiste un minore M di A di ordine k con $\det M \neq 0$;
- ii) ogni minore di A di ordine $\geq k+1$ ha determinante nullo.

La caratteristica di A viene denotata con $\rho(A)$, e ovviamente $0 \leq \rho(A) \leq \min(m, n)$. Osserviamo che dal Teorema 2.7 segue che $\rho(A) = \rho(A^T)$.

Per il calcolo di $\rho(A)$ è utile introdurre il concetto di *orlato* di un minore: se M è un minore di ordine k , allora un orlato di M è un qualunque minore M' di ordine $k+1$ ottenuto aggiungendo una riga ed una colonna ad M . Vale allora il seguente

Teorema di Kronecker. *Se una matrice A ha un minore M di ordine k con $\det M \neq 0$ e se $\det M' = 0$ per ogni orlato M' di M , allora $\rho(A) = k$.*

Il teorema di Kronecker semplifica il calcolo della caratteristica: si può infatti iniziare con un minore di ordine 2 a determinante non nullo (se esiste!) e progressivamente orlare tale minore fino a raggiungere la situazione descritta nel teorema di Kronecker. Alternativamente, si può iniziare dai minori di ordine massimo possibile, sperando di trovarne uno con determinante non nullo.

Siano A $m \times n$, B $n \times n$ e B' $m \times m$, con $\det B, \det B' \neq 0$; allora

$$\rho(AB) = \rho(B'A) = \rho(A). \quad (2.4)$$

Infatti è facile verificare la (2.4) se B e B' sono matrici elementari, e la (2.4) segue nel caso generale dal Teorema 2.4.

Dalla (2.4) deduciamo che se *riducendo* la matrice A otteniamo la matrice B , allora

$$\rho(A) = \rho(B);$$

se necessario, possiamo quindi limitarci a matrici ridotte per il calcolo della caratteristica.

Il risultato seguente fornisce un'importante proprietà della caratteristica di una matrice ridotta.

Teorema 2.8. *Sia A una matrice ridotta. Allora $\rho(A)$ è uguale al numero di righe non nulle di A .*

Dimostrazione. E' chiaro che il numero di righe non nulle di A è uguale al numero di pivot in A , sia esso p . D'altra parte, il minore M di ordine p ottenuto intersecando le righe e le colonne su cui stanno i pivot è triangolare superiore con elementi non nulli sulla diagonale (i pivot), e quindi ha determinante non nullo. Inoltre, orlando tale minore si introduce necessariamente una riga nulla, quindi ogni orlato di M ha determinante nullo. Dal teorema di Kronecker segue allora che $\rho(A) = p$. \square

2.4. Sistemi lineari: teorema di Rouchè-Capelli

Teorema di Rouchè-Capelli

Dato un sistema lineare $Ax = b$, ricordiamo che A è una matrice $m \times n$, le incognite sono n , la matrice completa viene denotata con $A|b$, i pivot del sistema ridotto mediante riduzione

gaussiana sono p e le righe significative di tale sistema ridotto sono q .

Un’alternativa al metodo di riduzione gaussiana per la determinazione del numero di soluzioni è fornita dal

Teorema di Rouchè-Capelli. *Un sistema lineare $Ax = b$ ha soluzioni se e solo se $\rho(A) = \rho(A|b)$. In tal caso le soluzioni sono ∞^{n-k} , dove $k = \rho(A) = \rho(A|b)$.*

Dimostrazione. Osserviamo che, dopo aver ridotto il sistema lineare, la nuova matrice dei coefficienti A' è ridotta ed ha p righe non nulle. Se $p = q$ anche la nuova matrice completa $A'|b'$ è ridotta ed ha lo stesso numero di righe non nulle di A' . Se invece $p < q$, riducendo $A|b$ si ottengono chiaramente $p+1$ righe non nulle per la matrice ridotta $A'|b'$. Il teorema segue allora dal Criterio del paragrafo 2.1 e dal Teorema 2.8. \square

Prima di passare ad un metodo di calcolo delle soluzioni basato sul teorema di Rouchè-Capelli, osserviamo che c’è una relazione tra caratteristica e combinazione lineare di righe o colonne di una matrice. Abbiamo infatti il

Teorema 2.9. *Sia A una matrice $m \times n$ con $\rho(A) = k$ e sia M un minore di ordine k con $\det M \neq 0$. Allora*

- i) *ogni riga di A al di fuori di M si può scrivere in modo unico come combinazione lineare delle righe di A all’interno di M*
- ii) *ogni colonna di A al di fuori di M si può scrivere in modo unico come combinazione lineare delle colonne di A all’interno di M .*

Dimostrazione. Consideriamo solo il caso delle colonne, quello delle righe essendo analogo. Per semplicità supponiamo che M sia composto dall’intersezione delle prime k righe e k colonne di A . Dobbiamo allora dimostrare che ogni colonna C_j con $j > k$ si può scrivere in modo unico come

$$C_j = \sum_{i=1}^k x_i C_i , \quad x_i \in \mathbb{R}/\mathbb{C}. \quad (2.5)$$

Ovviamente (2.5) è un sistema $m \times k$ in cui matrice dei coefficienti e matrice completa hanno la stessa caratteristica, uguale a k . Per il teorema di Rouchè-Capelli abbiamo allora $\infty^{k-k} = \infty^0 = 1$ soluzione, ed il teorema è dimostrato. \square

Osserviamo che con il linguaggio degli spazi vettoriali, vedi capitolo 3, il Teorema 2.9 si può formulare dicendo che le righe (colonne) all’interno di M sono *linearmente indipendenti* e le righe (colonne) al di fuori di M sono *linearmente dipendenti* dalle precedenti.

Calcolo delle soluzioni

Consideriamo per primo il caso di un sistema lineare *quadrato* $Ax = b$ con $\det A \neq 0$. Dal teorema di Rouchè-Capelli otteniamo che tale sistema ha 1 soluzione; inoltre tale soluzione si può calcolare moltiplicando a sinistra per A^{-1} entrambi i lati di $Ax = b$:

$$Ax = b \Leftrightarrow Ix = A^{-1}b \Leftrightarrow x = A^{-1}b. \quad (2.6)$$

Vediamo ora come il calcolo delle soluzioni di un *qualunque* sistema lineare si possa riportare al calcolo della soluzione di un opportuno sistema lineare quadrato a determinante non nullo; un tale sistema è detto *sistema di Cramer*. Il metodo di seguito riportato è un'alternativa al metodo basato sulla riduzione gaussiana, visto in precedenza.

Sia $Ax = b$ un sistema lineare con $\rho(A) = \rho(A|b) = k$ e sia M un minore di ordine k di A con $\det M \neq 0$. Procediamo allora nel modo seguente:

a) trascurare le righe al di fuori di M

infatti, per il Teorema 2.9 tali righe sono combinazione lineare di quelle in M e possono quindi essere ridotte a righe non-significative per mezzo di opportune operazioni di combinazione lineare;

b) portare al lato destro le colonne al di fuori di M , ed assegnare alle incognite in esse contenute valori arbitrari

osserviamo che, essenzialmente, il passo b) corrisponde alla procedura seguita nel Caso 2 del paragrafo 2.1;

c) per ogni tale assegnazione calcolare la soluzione del sistema di Cramer così ottenuto in questo modo otteniamo 1 soluzione per ogni assegnazione di valori alle $n - k$ incognite portate al lato destro; otteniamo quindi ∞^{n-k} soluzioni in totale. Inoltre, il Teorema 2.8 implica che $k = p$ nel caso di un sistema che ammetta almeno una soluzione. Abbiamo quindi ritrovato, con un metodo diverso, il risultato fornito dalla riduzione gaussiana ed espresso mediante il Criterio del paragrafo 2.1.

Calcolo della matrice inversa

Abbiamo visto che, per quanto riguarda il calcolo delle soluzioni, ogni sistema lineare viene riportato ad un opportuno sistema di Cramer, la cui soluzione è fornita dalla (2.6). Proponiamo pertanto due *tecniche di calcolo* della matrice inversa.

1° *metodo: riduzione totale*. Dal Teorema 2.4 abbiamo che riducendo totalmente A ottieniamo I ; in termini matriciali abbiamo $E_k \cdots E_1 A = I$ e quindi

$$A^{-1} = E_k \cdots E_1 I.$$

In altre parole, applicando ad I le operazioni che portano alla riduzione totale di A ottieniamo A^{-1} . Da un punto di vista *pratico*, il calcolo di A^{-1} si fa nel modo seguente. Si inizia affiancando le matrici A ed I :

$$(A|I);$$

si eseguono su A le operazioni di riduzione totale e contemporaneamente si esegue su I ogni operazione fatta su A . Alla fine della riduzione totale si ottiene

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

ovvero la matrice inversa.

2° metodo: *metodo di Laplace*. Usando il teorema di Laplace si può dimostrare che

$$A^{-1} = \frac{1}{\det A} (A^*)^T \quad (2.7)$$

dove A^* è la cosiddetta *matrice dei complementi algebrici*:

$$A^* = (a_{ij}^*) = ((-1)^{i+j} \det A_{ij}), \quad i, j = 1, \dots, n.$$

La (2.7) fornisce quindi un'espressione per A^{-1} ottenibile mediante il calcolo di n^2 determinanti $(n-1) \times (n-1)$ e del determinante di A .

Concludiamo il paragrafo fornendo esplicitamente la *soluzione* di un sistema di Cramer ottenuta mediante le (2.6) e (2.7). Denotando con A_j la matrice ottenuta da A sostituendo la j -esima colonna C_j con la colonna dei termini noti b , la soluzione di $Ax = b$ è data da

$$x = \begin{pmatrix} x_1 \\ \dots \\ x_n \end{pmatrix} = \frac{1}{\det A} \begin{pmatrix} \det A_1 \\ \dots \\ \det A_n \end{pmatrix}.$$

2.5. Fattorizzazione LU

Fattorizzazione LU

In questo paragrafo descriviamo un metodo di risoluzione per i sistemi di Cramer particolarmente vantaggioso in alcune situazioni concrete. Premettiamo la definizione di *fattorizzazione LU*: una matrice A $n \times n$ ammette fattorizzazione LU se esistono una matrice L $n \times n$ triangolare inferiore con tutti 1 sulla diagonale e una matrice U $n \times n$ triangolare superiore tali che

$$A = LU.$$

Abbiamo il

Teorema 2.10. *Sia A una matrice $n \times n$ riducibile senza scambi di righe. Allora A ammette fattorizzazione LU.*

Dimostrazione. Applicando la riduzione gaussiana ad A otteniamo

$$E_k \cdots E_1 A = U \quad (2.8)$$

con U triangolare superiore. Poichè non abbiamo fatto uso di matrici di scambio, le matrici E_1, \dots, E_k sono tutte del tipo $E_{ij}(\lambda)$ con $i > j$, e quindi triangolari inferiori. Inoltre, il prodotto di matrici triangolari inferiori è ancora triangolare inferiore, quindi la (2.8) diviene

$$L_1 A = U \quad (2.9)$$

con L_1 triangolare inferiore. Osserviamo inoltre che E_1, \dots, E_k hanno tutti 1 sulla diagonale, quindi anche L_1 ha tutti 1 sulla diagonale. Segue che L_1 è invertibile, e non è difficile verificare che $L = L_1^{-1}$ è a sua volta triangolare inferiore con tutti 1 sulla diagonale. Abbiamo allora che la (2.9) diviene

$$A = LU,$$

e il teorema è dimostrato. \square

Osserviamo che non è difficile dimostrare che le matrici L ed U nel Teorema 2.10 sono univocamente determinate. L'enunciato del Teorema 2.10 contiene l'ipotesi di riducibilità di A senza scambi di righe. Osserviamo che senza tale ipotesi il Teorema 2.10 è falso; infatti, senza ipotesi su A si può soltanto dimostrare che esiste un *prodotto di matrici di scambio* E tale che EA ammette fattorizzazione LU . In pratica, prima si operano opportuni scambi di righe su A in modo da poter poi ridurre EA senza scambi di righe. Osserviamo infine che l'ipotesi di riducibilità di A senza scambi di righe può essere rimpiazzata da altre ipotesi, che non approfondiamo.

La costruzione esplicita delle matrici L ed U si ottiene con un algoritmo molto simile a quello per il calcolo della matrice inversa, presentato nel paragrafo 2.4 come primo metodo.

Utilizzo della fattorizzazione LU

Vediamo infine l'utilità della fattorizzazione LU . Supponiamo di dover risolvere un *grande numero* di sistemi di Cramer del tipo

$$Ax = b_j, \quad j = 1, \dots, J,$$

ovvero con la matrice A fissata ma con b_j che assume un grande numero di valori diversi. Il metodo visto nel paragrafo 2.4 richiederebbe il calcolo di $A^{-1}b_j$ per ogni valore di b_j , e questo può risultare oneroso.

Utilizzando la fattorizzazione $A = LU$ si ha invece

$$LUx = b_j, \quad j = 1, \dots, J;$$

ponendo

$$Ux = y \tag{2.10}$$

possiamo prima calcolare y mediante

$$Ly = b_j \tag{2.11}$$

e poi calcolare x mediante la (2.10). Poiché sia L che U sono matrici triangolari, i sistemi (2.10) e (2.11) sono di *veloce risoluzione* e, se J è molto grande, questa tecnica risulta essere più economica rispetto alla precedente.

Osserviamo infine che il caso in cui A non ammette fattorizzazione LU non presenta problemi. In tal caso infatti $A' = EA$ ammette una tale fattorizzazione, dove E è prodotto di matrici di scambio; inoltre, il sistema $Ax = b_j$ è equivalente al sistema $A'x = b'_j$, dove la colonna b'_j è ottenuta da b_j operando i dovuti scambi di righe, e si può quindi procedere come sopra.

Capitolo 3. SPAZI VETTORIALI E TRASFORMAZIONI LINEARI

3.1. Spazi vettoriali

Spazi vettoriali

Negli insiemi \mathbb{R}^n e \mathbb{C}^n si possono definire le operazioni di *somma* e *prodotto per scalare* in modo del tutto simile a quanto fatto per le matrici: basta infatti pensare una n -upla di \mathbb{R}^n o \mathbb{C}^n come una matrice riga oppure come una matrice colonna. In generale, abbiamo la nozione di *spazio vettoriale*.

Definizione. Un insieme V in cui sono definite le operazioni di *somma* $u + v$, $u, v \in V$, e *prodotto per scalare* λv , $v \in V$ e $\lambda \in \mathbb{R}/\mathbb{C}$, si dice *spazio vettoriale su \mathbb{R}/\mathbb{C}* se tali operazioni soddisfano le proprietà seguenti:

- i) (*associatività*) $(u + v) + w = u + (v + w)$, $\lambda(\mu v) = (\lambda\mu)v$ per $u, v, w \in V$ e $\lambda, \mu \in \mathbb{R}/\mathbb{C}$
- ii) (*commutatività*) $u + v = v + u$ per $u, v \in V$
- iii) (*distributività*) $\lambda(u + v) = \lambda u + \lambda v$, $(\lambda + \mu)v = \lambda v + \mu v$ per $u, v \in V$ e $\lambda, \mu \in \mathbb{R}/\mathbb{C}$
- iv) (*esistenza dello zero*) esiste $0 \in V$ tale che $v + 0 = v$ per ogni $v \in V$
- v) (*esistenza dell'opposto*) per ogni $v \in V$ esiste $-v \in V$ tale che $v + (-v) = 0$
- vi) $0v = 0$ e $1v = v$ per ogni $v \in V$.

In altre parole, uno spazio vettoriale è un insieme in cui sono definite le operazioni di somma e prodotto per scalare, e tali operazioni soddisfano le usuali proprietà di somma e prodotto tra numeri. Gli elementi di uno spazio vettoriale si chiamano *vettori*.

Vediamo alcuni esempi particolarmente importanti di spazi vettoriali.

a) \mathbb{R}^n : come già osservato definiamo

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n) \quad \text{e} \quad \lambda(x_1, \dots, x_n) = (\lambda x_1, \dots, \lambda x_n) \quad (3.1)$$

con $\lambda \in \mathbb{R}$; con tali operazioni \mathbb{R}^n diviene uno spazio vettoriale su \mathbb{R} .

b) \mathbb{C}^n : possiamo definire le stesse operazioni di (3.1), ma questa volta possiamo avere $\lambda \in \mathbb{R}/\mathbb{C}$; nel primo caso \mathbb{C}^n sarà uno spazio vettoriale su \mathbb{R} , nel secondo caso su \mathbb{C} .

c) $\mathbb{R}[z]$, $\mathbb{C}[z]$: con le usuali operazioni i polinomi divengono uno spazio vettoriale, su \mathbb{R} nel caso di $\mathbb{R}[z]$ e su \mathbb{R}/\mathbb{C} nel caso di $\mathbb{C}[z]$. Si verifica che anche $P_k(\mathbb{R})$ e $P_k(\mathbb{C})$, rispettivamente i polinomi di grado $\leq k$ a coefficienti in \mathbb{R} e \mathbb{C} , sono spazi vettoriali su \mathbb{R} e su \mathbb{R}/\mathbb{C} , rispettivamente.

d) $M_{m \times n}(\mathbb{R}/\mathbb{C})$: con le operazioni di somma e prodotto per scalare introdotte nel Capitolo 2, tali insiemi di matrici sono spazi vettoriali su \mathbb{R} nel primo caso e su \mathbb{R}/\mathbb{C} nel secondo.

Dipendenza lineare

Dato uno spazio vettoriale V su \mathbb{R}/\mathbb{C} , $v_1, \dots, v_n \in V$ e $\lambda_1, \dots, \lambda_n \in \mathbb{R}/\mathbb{C}$, un'espressione del tipo $\lambda_1 v_1 + \dots + \lambda_n v_n$ è una *combinazione lineare* dei vettori v_1, \dots, v_n con *coefficienti* $\lambda_1, \dots, \lambda_n$. L'insieme di tutte le combinazioni lineari dei vettori v_1, \dots, v_n si denota con $L(v_1, \dots, v_n)$ e si chiama *spazio generato* da v_1, \dots, v_n , ovvero

$$L(v_1, \dots, v_n) = \{\lambda_1 v_1 + \dots + \lambda_n v_n, \lambda_i \in \mathbb{R}/\mathbb{C}, j = 1, \dots, n\}.$$

Si verifica che $L(v_1, \dots, v_n)$ è uno spazio vettoriale su \mathbb{R}/\mathbb{C} . Diciamo che i vettori v_1, \dots, v_n sono *linearmente indipendenti* se

$$\lambda_1 v_1 + \dots + \lambda_n v_n = 0 \quad \Rightarrow \quad \lambda_1 = \dots = \lambda_n = 0,$$

ovvero l'unica combinazione lineare nulla di v_1, \dots, v_n è quella *banale*, con coefficienti tutti nulli. Nel caso opposto i vettori sono *linearmente dipendenti*, ovvero esiste una combinazione lineare nulla di v_1, \dots, v_n con almeno un coefficiente non nullo.

Esempio. Si verifica facilmente che $(0, 1)$ e $(1, 1)$ sono linearmente indipendenti in \mathbb{R}^2 , mentre $(1, 1)$ e $(-2, -2)$ sono linearmente dipendenti.

Il risultato seguente mette in relazione i concetti di combinazione e dipendenza lineare.

Proposizione 3.1. *I vettori $v_1, \dots, v_n \in V$ sono linearmente dipendenti se e solo se almeno uno di essi è combinazione lineare dei rimanenti.*

Dimostrazione. Se v_1, \dots, v_n sono linearmente dipendenti esiste una combinazione lineare $\lambda_1 v_1 + \dots + \lambda_n v_n = 0$ con almeno un coefficiente non nullo, sia esso λ_1 ; segue che

$$v_1 = -\frac{\lambda_2}{\lambda_1} v_2 - \dots - \frac{\lambda_n}{\lambda_1} v_n,$$

ovvero v_1 è combinazione lineare di v_2, \dots, v_n . Viceversa se, ad esempio, v_1 è combinazione lineare di v_2, \dots, v_n abbiamo $v_1 = \lambda_2 v_2 + \dots + \lambda_n v_n$, da cui

$$v_1 - \lambda_2 v_2 - \dots - \lambda_n v_n = 0,$$

ovvero v_1, \dots, v_n sono linearmente dipendenti. \square

E' chiaro che se un vettore $v_1 \in V$ è combinazione lineare dei vettori $v_2, \dots, v_n \in V$, allora $L(v_1, \dots, v_n) = L(v_2, \dots, v_n)$. Questo vale più in generale, nel senso che è sempre possibile "sfoltire" in modo opportuno un insieme di vettori lasciando inalterato lo spazio generato. Il *metodo degli scarti successivi* fornisce una procedura per tale operazione; dati $v_1, \dots, v_n \in V$ procediamo nel modo seguente:

- a) scartiamo tutti i vettori nulli tra v_1, \dots, v_n
- b) denotiamo con w_1 il primo tra i vettori rimasti
- c) scartiamo i vettori che sono combinazione lineare di w_1 e denotiamo con w_2 il primo tra i vettori rimasti
- d) scartiamo i vettori che sono combinazione lineare di w_1 e w_2 e denotiamo con w_3 il primo tra i vettori rimasti, e così via.

Al termine di tale procedura otteniamo un sottoinsieme $\{w_1, \dots, w_k\}$ di $\{v_1, \dots, v_n\}$ che, grazie alla Proposizione 3.1, chiaramente gode delle proprietà seguenti:

- i) i vettori w_1, \dots, w_k sono *linearmente indipendenti*
- ii) $L(w_1, \dots, w_k) = L(v_1, \dots, v_n)$.

Esempio. Applicando il metodo degli scarti successivi ai vettori $(0, 0), (1, 0), (3, 0), (0, 0), (1, 1)$ e $(2, 1)$ si ottengono i vettori $(1, 0)$ e $(1, 1)$, come si verifica facilmente.

Basi e dimensione

Le proprietà appena viste suggeriscono le seguenti definizioni:

- i) un insieme di vettori $\{v_1, \dots, v_n\}$ si dice *insieme libero* se v_1, \dots, v_n sono linearmente indipendenti
- ii) i vettori $v_1, \dots, v_n \in V$ formano un *sistema* (o *insieme*) di generatori di V se $V = L(v_1, \dots, v_n)$
- iii) uno spazio vettoriale V si dice *finitamente generato* se esistono $v_1, \dots, v_n \in V$ tali che $V = L(v_1, \dots, v_n)$
- vi) un insieme libero di generatori di V si dice *base* di V .

Una base B di V formata dai vettori v_1, \dots, v_n si denota con $B = \langle v_1, \dots, v_n \rangle$. Osserviamo che il metodo degli scarti successivi fornisce un *algoritmo* per costruire una base di uno spazio vettoriale del tipo $L(v_1, \dots, v_n)$: infatti, i vettori w_1, \dots, w_k che rimangono alla fine della procedura di scarto formano chiaramente una base di $L(v_1, \dots, v_n)$. Da tale osservazione deduciamo immediatamente il

Teorema 3.1. *Ogni spazio vettoriale finitamente generato ha una base.*

D'ora in poi considereremo soltanto *spazi vettoriali finitamente generati*, a meno di esplicita avvertenza. Osserviamo comunque che molti dei risultati che otterremo valgono anche nel caso di spazi vettoriali non finitamente generati. Osserviamo inoltre che tra gli esempi visti in precedenza soltanto $\mathbb{R}[z]$ e $\mathbb{C}[z]$ non sono finitamente generati.

Un esempio importante di base degli spazi vettoriali \mathbb{R}^n e \mathbb{C}^n è la *base canonica* $K = \langle e_1, \dots, e_n \rangle$, ovvero una base particolarmente semplice e utile, definita da

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

Osserviamo che K è base di \mathbb{C}^n come spazio vettoriale su \mathbb{C} , ma non su \mathbb{R} . La base canonica di \mathbb{C}^n come spazio vettoriale su \mathbb{R} è $\tilde{K} = \langle e_1, \dots, e_n, f_1, \dots, f_n \rangle$, dove

$$f_1 = (i, 0, \dots, 0), \quad f_2 = (0, i, 0, \dots, 0), \dots, f_n = (0, \dots, 0, i);$$

c'è quindi effettivamente una differenza tra \mathbb{C}^n come spazio vettoriale su \mathbb{R} e su \mathbb{C} .

E' facile vedere che uno spazio vettoriale può avere più di una base; ad esempio $\langle e_1, e_2 \rangle$ e $\langle (1, 1), (-1, 1) \rangle$ sono due basi di \mathbb{R}^2 . Vale però il seguente importante

Teorema 3.2. *Tutte le basi di uno spazio vettoriale V sono formate dallo stesso numero di vettori.*

Il Teorema 3.2 porta al concetto di *dimensione* di uno spazio vettoriale: la dimensione di V , denotata con $\dim V$, è il numero di elementi di una base di V .

Esempio. $\dim \mathbb{R}^n = n$, $\dim \mathbb{C}^n = n$ come spazio vettoriale su \mathbb{C} e $\dim \mathbb{C}^n = 2n$ come spazio vettoriale su \mathbb{R} ; abbiamo già determinato la base canonica di tali spazi. $\dim M_{m \times n}(\mathbb{R}) = mn$, $\dim M_{m \times n}(\mathbb{C}) = mn$ come spazio vettoriale su \mathbb{C} e $\dim M_{m \times n}(\mathbb{C}) = 2mn$ come spazio vettoriale su \mathbb{R} ; la base canonica di tali spazi, nei primi due casi, è formata dalle mn matrici aventi un coefficiente uguale a 1 e tutti gli altri uguali a 0, mentre nel terzo caso è formata dalle $2mn$ matrici aventi un coefficiente uguale a 1 o i e tutti gli altri uguali a zero. $\dim P_k(\mathbb{R}) = k + 1$, $\dim P_k(\mathbb{C}) = k + 1$ come spazio vettoriale su \mathbb{C} e $\dim P_k(\mathbb{C}) = 2(k + 1)$ come spazio vettoriale su \mathbb{R} ; la base canonica, nei primi due casi, è formata dai $k + 1$ polinomi $1, x, \dots, x^k$, mentre nel terzo caso è formata dai $2(k + 1)$ polinomi $1, i, x, ix, \dots, x^k, ix^k$.

Osserviamo come nel caso di \mathbb{R}^n il concetto di dimensione appena introdotto coincida con il concetto geometrico intuitivo di dimensione, perlomeno quando $n = 1, 2, 3$. Osserviamo inoltre che lo spazio vettoriale *banale* $V = \{0\}$, ovvero lo spazio vettoriale formato dal solo vettore nullo, ha $\dim V = 0$.

Abbiamo infine le proprietà seguenti

Proposizione 3.2. *Sia V uno spazio vettoriale con $\dim V = n$. Allora*

- i) *da ogni sistema di generatori si può estrarre una base*
- ii) *ogni insieme libero può essere completato a una base*
- iii) *n è il massimo numero di vettori linearmente indipendenti in V*
- iv) *n è il minimo numero di generatori di V*
- v) *n generatori di V formano una base*
- vi) *n vettori linearmente indipendenti di V formano una base.*

Dimostrazione. i). Segue immediatamente dal metodo degli scarti successivi.

ii). Sia $\{v_1, \dots, v_k\}$ un insieme libero e $\langle w_1, \dots, w_n \rangle$ una base di V ; applicando il metodo degli scarti successivi all'insieme $\{v_1, \dots, v_k, w_1, \dots, w_n\}$ si ottiene una base di V che, per costruzione, contiene i vettori v_1, \dots, v_k e quindi completa v_1, \dots, v_k ad una base di V .

iii). Se esistessero $\geq n + 1$ vettori linearmente indipendenti potremmo completare tale insieme ad una base di V , ottenendo così una base con $\geq n + 1$ elementi, assurdo.

iv). Se esistessero $\leq n - 1$ generatori di V potremmo estrarre una base da tali vettori, assurdo.

v). Se tali generatori non formassero una base potremmo da essi estrarre una base con $\leq n - 1$ elementi, assurdo.

vi). Se tali vettori non formassero una base potremmo completarli ad una base con $\geq n + 1$ elementi, assurdo. \square

Coordinate

L'importanza della nozione di base di uno spazio vettoriale è dovuta al seguente

Teorema 3.3. *Sia V uno spazio vettoriale e $B = \langle v_1, \dots, v_n \rangle$ una sua base. Allora ogni*

vettore $v \in V$ si scrive in modo unico come

$$v = a_1 v_1 + \dots + a_n v_n. \quad (3.2)$$

Dimostrazione. Poichè v_1, \dots, v_n formano un sistema di generatori, un'espressione di v del tipo (3.2) esiste certamente. Supponiamo per assurdo che

$$v = a_1 v_1 + \dots + a_n v_n = b_1 v_1 + \dots + b_n v_n$$

con $(a_1, \dots, a_n) \neq (b_1, \dots, b_n)$. Sottraendo le due espressioni otteniamo allora

$$(a_1 - b_1)v_1 + \dots + (a_n - b_n)v_n = 0,$$

i cui coefficienti $a_i - b_i, i = 1, \dots, n$, non sono tutti nulli, assurdo. \square

Osserviamo che non è difficile dimostrare che vale anche il viceversa del Teorema 3.3, ovvero: se ogni vettore $v \in V$ si scrive in modo unico come combinazione lineare dei vettori $v_1, \dots, v_n \in V$, allora $\langle v_1, \dots, v_n \rangle$ è una base di V . Da questa osservazione e da quella che segue la (2.5) deduciamo quindi che la caratteristica $\rho(A)$ di una matrice A coincide con il massimo numero di righe, o di colonne, linearmente indipendenti di A .

La n -upla (a_1, \dots, a_n) in (3.2) prende il nome di n -upla delle *coordinate* di v rispetto alla base B . Conviene scrivere tali coordinate nella forma di una *matrice colonna*, detta anche *vettore colonna*, ovvero nella forma

$$v_B = \begin{pmatrix} a_1 \\ \dots \\ a_n \end{pmatrix}.$$

Il vettore colonna v_B è detto *colonna delle coordinate* di v rispetto alla base B . Osserviamo che l'*ordine* dei vettori v_1, \dots, v_n nella base B è importante, in quanto cambiando tale ordine cambia la colonna delle coordinate; in altre parole, due basi formate dagli stessi vettori presi in ordine diverso sono considerate due basi distinte.

La notazione delle n -uple sotto forma di vettori colonna sarà utile in seguito. Pertanto denoteremo spesso i vettori di \mathbb{R}^n e di \mathbb{C}^n sotto forma di vettori colonna.

Sia V uno spazio vettoriale su \mathbb{R}/\mathbb{C} con $\dim V = n$ e sia B una base. La (3.2) definisce un'applicazione *bigettiva* da V a $\mathbb{R}^n/\mathbb{C}^n$ tramite

$$v \rightarrow v_B. \quad (3.3)$$

Osserviamo che tale applicazione è *coerente con le operazioni* in V e in $\mathbb{R}^n/\mathbb{C}^n$, nel senso che

$$(u + v)_B = u_B + v_B \quad \text{e} \quad (\lambda v)_B = \lambda v_B.$$

Possiamo quindi pensare lo spazio vettoriale V come *identificato*, tramite la (3.3), a $\mathbb{R}^n/\mathbb{C}^n$; vedremo nel seguito che la (3.3) è un *isomorfismo* tra V e $\mathbb{R}^n/\mathbb{C}^n$. Osserviamo inoltre che se $B = \langle v_1, \dots, v_n \rangle$ allora $v_{1_B} = e_1, \dots, v_{n_B} = e_n$, e che i vettori u_1, \dots, u_m sono linearmente indipendenti in V se e solo se i vettori colonne u_{1_B}, \dots, u_{m_B} sono linearmente indipendenti in $\mathbb{R}^n/\mathbb{C}^n$. Abbiamo quindi che gli spazi vettoriali \mathbb{R}^n e \mathbb{C}^n sono i *prototipi* di spazi vettoriali, rispettivamente su \mathbb{R} e su \mathbb{C} . In altre parole, uno spazio vettoriale V su \mathbb{R}/\mathbb{C} con $\dim V = n$ può essere pensato, rispettivamente, come \mathbb{R}^n o \mathbb{C}^n .

Vediamo un primo esempio dell'utilità delle coordinate. Sia, come in precedenza, V uno spazio vettoriale su \mathbb{R}/\mathbb{C} , B una sua base e $\dim V = n$. Dati $u_1, \dots, u_m \in V$ consideriamo la matrice $n \times m$

$$C = (u_{1_B} \dots u_{m_B}),$$

le cui colonne sono formate dalle colonne delle coordinate di u_1, \dots, u_m . Abbiamo il

Teorema 3.4. $\dim L(u_1, \dots, u_m) = \rho(C)$.

Dimostrazione. Grazie all'identificazione tra V e $\mathbb{R}^n/\mathbb{C}^n$ è sufficiente dimostrare che

$$\dim L(u_{1_B}, \dots, u_{m_B}) = \rho(C).$$

Sia M un minore di ordine $\rho(C)$ della matrice C con $\det M \neq 0$. Dalla *ii)* del Teorema 2.9 abbiamo che ogni colonna di C al di fuori di M si scrive in modo unico come combinazione lineare delle colonne che entrano in M ; per quanto osservato subito dopo la dimostrazione del Teorema 3.3 abbiamo allora che tali colonne formano una base di $L(u_{1_B}, \dots, u_{m_B})$, e il teorema è dimostrato. \square

Il Teorema 3.4 fornisce quindi un utile strumento per il calcolo della dimensione degli spazi vettoriali del tipo $L(u_1, \dots, u_m)$. Inoltre, esso fornisce un'alternativa al metodo degli scarti successivi per individuare una base per tali spazi vettoriali; infatti, i vettori le cui colonne delle coordinate (rispetto ad una base qualunque) entrano in un minore di C di ordine $\rho(C)$ a determinante non nullo costituiscono una base per lo spazio vettoriale in questione.

Sistemi omogenei

Esempi importanti di spazi vettoriali sono forniti dalle soluzioni dei sistemi omogenei; abbiamo infatti il

Teorema 3.5. *Sia $A \in M_{m \times n}(\mathbb{R}/\mathbb{C})$. Allora le soluzioni del sistema omogeno $Ax = 0$ formano uno spazio vettoriale su \mathbb{R}/\mathbb{C} di dimensione $n - \rho(A)$.*

Dimostrazione. E' chiaro che l'insieme S delle soluzioni del sistema $Ax = 0$ forma uno spazio vettoriale; infatti se α e β sono soluzioni, anche $\alpha + \beta$ e $\lambda\alpha$, $\lambda \in \mathbb{R}/\mathbb{C}$, sono soluzioni. Il metodo di risoluzione dei sistemi lineari mediante l'algoritmo di riduzione gaussiana (o mediante il teorema di Rouchè-Capelli) consiste nel portare a lato destro le incognite non-pivotali (o quelle non contenute in un minore M di ordine $\rho(A)$ con $\det M \neq 0$), assegnare loro valori arbitrari ed infine calcolare il valore delle rimanenti incognite in funzione di

tali assegnazioni. Ponendo $d = n - \rho(A)$ e supponendo, per semplicità, che le incognite non-pivotali siano x_1, \dots, x_d , abbiamo quindi che un sistema di generatori di S è dato dai d vettori

$$(1, 0, \dots, 0, \alpha_{d+1}^{(1)}, \dots, \alpha_n^{(1)}), (0, 1, 0, \dots, 0, \alpha_{d+1}^{(2)}, \dots, \alpha_n^{(2)}), \dots, (0, \dots, 0, 1, \alpha_{d+1}^{(d)}, \dots, \alpha_n^{(d)}),$$

dove $\alpha_j^{(i)}, j = d + 1, \dots, n$ e $i = 1, \dots, d$ sono i valori delle incognite pivotali calcolati in base a tali assegnazioni. Ma tali vettori sono linearmente indipendenti, ed il teorema è dimostrato. \square

Le soluzioni di un sistema non-omogeneo *non* formano uno spazio vettoriale; infatti, ad esempio, la n -upla $(0, \dots, 0)$ *non* è soluzione di un tale sistema lineare. Vedremo in seguito che le soluzioni di un sistema non-omogeneo sono comunque *collegate* ad uno spazio vettoriale. Osserviamo infine che il Teorema 3.5 fornisce un supporto teorico alla notazione $\infty^{n-\rho(A)}$ usata per indicare il numero di soluzioni: infatti, nel caso di un sistema omogeneo il numero $n - \rho(A)$ coincide con la dimensione dello spazio vettoriale delle soluzioni.

Sottospazi

In precedenza abbiamo visto esempi di spazi vettoriali contenuti in uno spazio vettoriale più grande. Diamo allora la definizione di *sottospazio*: dato uno spazio vettoriale V su \mathbb{R}/\mathbb{C} , un sottinsieme $W \subset V$ è un sottospazio di V se $\lambda u + \mu v \in W$ per ogni $u, v \in W$ e $\lambda, \mu \in \mathbb{R}/\mathbb{C}$. In altre parole, un sottospazio è un sottinsieme *chiuso* rispetto alle operazioni in V ; osserviamo che $0 \in W$ e che W è a sua volta uno spazio vettoriale su \mathbb{R}/\mathbb{C} .

Esempi di sottospazi di uno spazio vettoriale V sono gli spazi $L(v_1, \dots, v_n)$ generati da n vettori $v_1, \dots, v_n \in V$; in particolare le soluzioni di un sistema lineare $Ax = 0$, con $A \in M_{m \times n}(\mathbb{R}/\mathbb{C})$, formano un sottospazio di $\mathbb{R}^n/\mathbb{C}^n$ di dimensione $n - \rho(A)$.

Non è difficile verificare che vale la

Proposizione 3.3. *Sia W sottospazio di uno spazio vettoriale V . Allora*

- i) $\dim W \leq \dim V$
- ii) se $\dim W = \dim V$ allora $W = V$.

Introduciamo ora due *operazioni* tra sottospazi. Dati W_1 e W_2 sottospazi dello spazio vettoriale V definiamo l'*intersezione* di W_1 e W_2 nel modo usuale, ovvero come $W_1 \cap W_2$, e la *somma* di W_1 e W_2 come

$$W_1 + W_2 = \{w_1 + w_2 \text{ con } w_i \in W_i, i = 1, 2\}.$$

E' facile verificare che $W_1 \cap W_2$ e $W_1 + W_2$ sono *sottospazi* di V . Ad esempio, è chiaro che se $W_1 = L(u_1, \dots, u_m)$ e $W_2 = L(v_1, \dots, v_n)$ allora $W_1 + W_2 = L(u_1, \dots, u_m, v_1, \dots, v_n)$. In generale, le dimensioni di $W_1 \cap W_2$ e $W_1 + W_2$ sono legate dal

Teorema di Grassmann. *Siano W_1 e W_2 sottospazi di uno spazio vettoriale V . Allora*

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2).$$

Più in generale si definisce

$$W_1 + \dots + W_r = \{w_1 + \dots + w_r \text{ con } w_i \in W_i, i = 1, \dots, r\},$$

dove W_1, \dots, W_r sono sottospazi di uno spazio vettoriale V , e si verifica facilmente che $W_1 + \dots + W_r$ è a sua volta un sottospazio di V . Talvolta useremo la notazione

$$W_1 + \dots + W_r = \sum_{i=1}^r W_i.$$

Particolare importanza ha il concetto di *somma diretta* dei sottospazi W_1, \dots, W_r : la somma $W_1 + \dots + W_r$ è diretta se ogni vettore $v \in W_1 + \dots + W_r$ si scrive in modo *unico* come $v = w_1 + \dots + w_r$ con $w_i \in W_i, i = 1, \dots, r$; la somma diretta si denota con $W_1 \oplus \dots \oplus W_r$.

Esempio. Se $V = \langle v_1, \dots, v_n \rangle$ è una base di V allora $V = L(v_1) \oplus \dots \oplus L(v_n)$.

Se $r = 2$ è chiaro che la somma $W_1 + W_2$ è diretta se e solo se $W_1 \cap W_2 = \{0\}$; infatti, se esiste un vettore $v \in W_1 \cap W_2, v \neq 0$, possiamo scrivere $v = v + 0 = 0 + v$ e quindi la somma $W_1 + W_2$ non è diretta, e viceversa. In generale vale il

Teorema 3.6. Siano W_1, \dots, W_r sottospazi di uno spazio vettoriale V e B_1, \dots, B_r basi di W_1, \dots, W_r , rispettivamente. Allora sono equivalenti:

i) $W = W_1 \oplus \dots \oplus W_r$

ii) $W = W_1 + \dots + W_r$ e per ogni $j = 1, \dots, r$ si ha $W_j \cap (\sum_{\substack{i=1 \\ i \neq j}}^r W_i) = \{0\}$

iii) $\bigcup_{i=1}^r B_i$ è una base di W

iv) $W = W_1 + \dots + W_r$ e $\dim W = \sum_{i=1}^r \dim W_i$.

Osserviamo che la somma di W_1, \dots, W_r è diretta se e solo se per ogni scelta di vettori $w_1 \in W_1, \dots, w_r \in W_r$ l'insieme $\{w_1, \dots, w_r\}$ è libero. Osserviamo infine che le iii) e iv) del Teorema 3.6 chiariscono il concetto di somma diretta; in altre parole, $W = W_1 \oplus \dots \oplus W_r$ se W_1, \dots, W_r sono gli "addendi minimi" per ottenere il sottospazio somma W .

3.2. Trasformazioni lineari

Trasformazioni lineari e matrici associate

Siano V e W spazi vettoriali su \mathbb{R}/\mathbb{C} ; un'applicazione $\varphi : V \rightarrow W$ si dice *trasformazione lineare* se

$$\varphi(\lambda u + \mu v) = \lambda\varphi(u) + \mu\varphi(v)$$

per ogni $u, v \in V$ e $\lambda, \mu \in \mathbb{R}/\mathbb{C}$. In altre parole, una trasformazione lineare è un'applicazione da V in W che "rispetta" le operazioni in V e W . Osserviamo che $\varphi(0) = 0$ per ogni trasformazione lineare φ .

Esempio. $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ data da $\varphi((x, y)) = (x + y, x - y)$ è una trasformazione lineare, mentre $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ data da $\varphi((x, y)) = (x + y + 1, x - y)$ non lo è.

Osserviamo che vi sono *due modi* per fornire una trasformazione lineare $\varphi : V \rightarrow W$: il primo, quello ovvio, consiste nel fornire $\varphi(v)$ per ogni $v \in V$, come nell'esempio appena visto. Il secondo modo fa uso delle proprietà delle trasformazioni lineari; data una base $B = \langle v_1, \dots, v_n \rangle$ di V è sufficiente fornire $\varphi(v_1), \dots, \varphi(v_n)$: infatti, ogni $v \in V$ si scrive come $v = a_1v_1 + \dots + a_nv_n$ e quindi per linearità

$$\varphi(v) = a_1\varphi(v_1) + \dots + a_n\varphi(v_n). \quad (3.4)$$

Esempio. Scegliendo $B = K$ nell'esempio precedente si ha $\varphi(e_1) = (1, 1)$ e $\varphi(e_2) = (1, -1)$; i due modi di fornire φ sono del tutto equivalenti.

La (3.4) suggerisce la possibilità di associare delle matrici ad una trasformazione lineare $\varphi : V \rightarrow W$. Se $\dim V = n$, $\dim W = m$ e $E = \langle v_1, \dots, v_n \rangle$, $F = \langle w_1, \dots, w_m \rangle$ sono basi di V e W rispettivamente, abbiamo infatti

$$\begin{aligned} \varphi(v_1) &= a_{11}w_1 + \dots + a_{m1}w_m \\ &\dots \\ \varphi(v_n) &= a_{1n}w_1 + \dots + a_{mn}w_m \end{aligned}$$

per opportuni coefficienti $a_{ij} \in \mathbb{R}/\mathbb{C}$. Alla trasformazione lineare φ associamo allora la matrice

$$M_\varphi^{E,F} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \in M_{m \times n}(\mathbb{R}/\mathbb{C})$$

detta *matrice associata a φ mediante le basi E, F* . In altre parole, $M_\varphi^{E,F}$ ha come colonna j -esima la colonna delle coordinate di $\varphi(v_j)$ rispetto alla base F .

Esempio. Sia φ come nell'esempio precedente; allora $M_\varphi^{K,K} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

L'interesse della matrice associata $M_\varphi^{E,F}$ deriva dal fatto che essa rispecchia *a livello di coordinate* la trasformazione lineare φ , nel seguente senso: dato un vettore $v \in V$, non è difficile verificare che tra le coordinate di v rispetto alla base E e quelle di $\varphi(v)$ rispetto alla base F sussiste la relazione

$$\varphi(v)_F = M_\varphi^{E,F}v_E. \quad (3.5)$$

In altre parole, fissate le basi E e F , la matrice $M_\varphi^{E,F}$ opera, mediante moltiplicazione, sulle *coordinate* dei vettori nello stesso modo in cui la trasformazione lineare φ opera sui *vettori*. Osserviamo che nel caso speciale $V = \mathbb{R}^n$, $W = \mathbb{R}^m$ e E, F basi canoniche, la matrice $M_\varphi^{K,K}$ e la trasformazione lineare φ operano nello *stesso modo*, ovvero

$$\varphi(v) = M_\varphi^{K,K}v,$$

in quanto la colonna delle coordinate di un vettore coincide col vettore stesso in questo caso.

La procedura di associare matrici a trasformazioni lineari può essere *invertita*. Dati due spazi vettoriali V e W su \mathbb{R}/\mathbb{C} con $\dim V = n$ e $\dim W = m$, una base $E = \langle v_1, \dots, v_n \rangle$ di V e una base $F = \langle w_1, \dots, w_m \rangle$ di W ed infine una matrice $M = (a_{ij}) \in M_{m \times n}(\mathbb{R}/\mathbb{C})$, definiamo

$$\begin{aligned}\varphi(v_1) &= a_{11}w_1 + \dots + a_{m1}w_m \\ \vdots \\ \varphi(v_n) &= a_{1n}w_1 + \dots + a_{mn}w_m.\end{aligned}$$

Estendendo φ per linearità si ottiene una trasformazione lineare $\varphi : V \rightarrow W$ che soddisfa

$$M = M_{\varphi}^{E,F}.$$

La trasformazione lineare $\phi : V \rightarrow W$ prende il nome di *trasformazione lineare associata a M mediante le basi E, F* . Segue quindi che, fissati V, W, E e F come sopra, c'è una corrispondenza tra le trasformazioni lineari $\varphi : V \rightarrow W$ e le matrici $M \in M_{m \times n}(\mathbb{R}/\mathbb{C})$: ad ogni trasformazione lineare φ corrisponde, nel modo sopra descritto, una matrice M e viceversa.

Trasformazione composta e inversa

Consideriamo ora tre spazi vettoriali V, W e U su \mathbb{R}/\mathbb{C} con basi rispettive E, F e G e due trasformazioni lineari $\varphi : V \rightarrow W$ e $\psi : W \rightarrow U$. In questa situazione è possibile considerare l'applicazione composta $\psi \circ \varphi : V \rightarrow U$; vale il seguente

Teorema 3.7. Siano φ e ψ le trasformazioni lineari sopra descritte. Allora $\psi \circ \varphi : V \rightarrow U$ è una trasformazione lineare ed inoltre

$$M_{\psi \circ \varphi}^{E,G} = M_\psi^{F,G} M_\varphi^{E,F}.$$

Si verifica facilmente che le dimensioni di tali matrici sono coerenti. Abbiamo quindi che la corrispondenza tra matrici e trasformazioni lineari è *operativa*, nel senso che alla *trasformazione composta* $\psi \circ \varphi$ corrisponde il prodotto delle matrici associate alle trasformazioni lineari ψ e φ .

Una trasformazione lineare bigettiva $\varphi : V \rightarrow W$ si dice *isomorfismo*. Abbiamo già visto nel paragrafo precedente un esempio importante di isomorfismo: se V è uno spazio vettoriale su \mathbb{R}/\mathbb{C} con $\dim V = n$ e $B = \langle v_1, \dots, v_n \rangle$ è una sua base, allora l'applicazione (3.3) è un isomorfismo. Vale il seguente

Teorema 3.8. Sia $\varphi : V \rightarrow W$ un isomorfismo, $\dim V = n$, $\dim W = m$ e E, F basi di V e W rispettivamente. Allora $\varphi^{-1} : W \rightarrow V$ è un isomorfismo, $m = n$, $M_\varphi^{E,F}$ è invertibile e $M_{\varphi^{-1}}^{F,E} = (M_\varphi^{E,F})^{-1}$. Viceversa, se $\varphi : V \rightarrow W$ è una trasformazione lineare con $m = n$ e $M_\varphi^{E,F}$ è invertibile, allora φ è un isomorfismo.

L'isomorfismo φ^{-1} si chiama *trasformazione inversa* dell'isomorfismo φ ; il Teorema 3.8 fornisce un ulteriore esempio dell'operatività della corrispondenza tra trasformazioni lineari e matrici.

Matrice di passaggio

Consideriamo un importante caso speciale della situazione sopra descritta; precisamente,

date due basi $B = \langle v_1, \dots, v_n \rangle$ e $B' = \langle v'_1, \dots, v'_n \rangle$ di uno spazio vettoriale V con $\dim V = n$, consideriamo

$$id_V : (V, B) \rightarrow (V, B'),$$

dove con (V, B) intendiamo lo spazio vettoriale V munito della base B . La matrice associata sarà allora

$$P = M_{id_V}^{B, B'} = \begin{pmatrix} p_{11} & \dots & p_{1n} \\ \dots & \dots & \dots \\ p_{n1} & \dots & p_{nn} \end{pmatrix},$$

detta *matrice di passaggio* da B' a B e denotata talvolta con $P_{B' \rightarrow B}$ per evidenziare le basi in questione. Chiaramente, i coefficienti p_{ij} sono determinati da

$$\begin{aligned} v_1 &= p_{11}v'_1 + \dots + p_{n1}v'_n \\ &\dots \\ v_n &= p_{1n}v'_1 + \dots + p_{nn}v'_n \end{aligned}$$

ovvero la *colonna j -esima della matrice di passaggio P* è costituita dalla *colonna delle coordinate del vettore v_j rispetto alla base B'* . Inoltre, considerando le basi $B = \langle v_1, \dots, v_n \rangle$ e $B' = \langle v'_1, \dots, v'_n \rangle$ come matrici riga $1 \times n$, abbiamo

$$B = B'P,$$

che giustifica il nome dato alla matrice P .

Dal Teorema 3.8 abbiamo che P è invertibile e che P^{-1} è la *matrice di passaggio da B a B'* , ovvero

$$P_{B \rightarrow B'} = (P_{B' \rightarrow B})^{-1}. \quad (3.6)$$

Inoltre, dalla (3.5) abbiamo che per ogni $v \in V$

$$v_{B'} = Pv_B, \quad (3.7)$$

ovvero la matrice di passaggio da B' a B *agisce sulle coordinate rispetto a B e le trasforma nelle coordinate rispetto a B'* . Ovviamente abbiamo anche la relazione inversa

$$v_B = P^{-1}v_{B'},$$

in accordo con la (3.6).

La matrice di passaggio consente di stabilire una relazione tra le matrici associate ad una stessa trasformazione lineare φ mediante basi diverse. Siano infatti E, E' basi di V e F, F' basi di W ; data la trasformazione lineare $\varphi : V \rightarrow W$ consideriamo anche $id_V : V \rightarrow V$, $id_W : W \rightarrow W$ e la trasformazione composta

$$(V, E') \rightarrow (V, E) \rightarrow (W, F) \rightarrow (W, F').$$

Dal Teorema 3.7 otteniamo allora

$$M_{\varphi}^{E',F'} = P_{F' \rightarrow F} M_{\varphi}^{E,F} P_{E \rightarrow E'}. \quad (3.8)$$

In particolare, poichè le matrici di passaggio sono invertibili, dalla (2.4) abbiamo

$$\rho(M_{\varphi}^{E',F'}) = \rho(M_{\varphi}^{E,F}), \quad (3.9)$$

ovvero tutte le matrici associate ad una stessa trasformazione lineare hanno la stessa caratteristica.

Consideriamo infine il caso speciale di una trasformazione lineare $\varphi : V \rightarrow V$; date due basi B e B' di V , la (3.8) diviene

$$M_{\varphi}^{B,B} = P_{B \rightarrow B'} M_{\varphi}^{B',B'} P_{B' \rightarrow B}$$

ovvero, denotando nuovamente con $P = P_{B' \rightarrow B}$ la matrice di passaggio da B' a B , abbiamo

$$M_{\varphi}^{B,B} = P^{-1} M_{\varphi}^{B',B'} P. \quad (3.10)$$

Le considerazioni fin qui fatte sulle matrici di passaggio saranno particolarmente utili nel Capitolo 5.

Nucleo e immagine

Data una trasformazione lineare $\varphi : V \rightarrow W$ definiamo

$$\text{Ker } \varphi = \{v \in V : \varphi(v) = 0\} \subset V \quad \text{e} \quad \text{Im } \varphi = \text{immagine di } \varphi \subset W,$$

rispettivamente *nucleo* e *immagine* di φ ; si verifica che $\text{Ker } \varphi$ e $\text{Im } \varphi$ sono sottospazi, rispettivamente, di V e W . Tali sottospazi forniscono importanti informazioni sulla trasformazione lineare φ .

Proposizione 3.4. *Sia $\varphi : V \rightarrow W$ una trasformazione lineare. Allora*

- i) φ è iniettiva se e solo se $\dim \text{Ker } \varphi = 0$
- ii) φ è surgettiva se e solo se $\dim \text{Im } \varphi = \dim W$.

Dimostrazione. i). Poichè $\varphi(0) = 0$, se φ è iniettiva allora $\text{Ker } \varphi = \{0\}$ e quindi $\dim \text{Ker } \varphi = 0$. Viceversa, se $\dim \text{Ker } \varphi = 0$ allora $\text{Ker } \varphi = \{0\}$ e quindi se $\varphi(u) = \varphi(v)$ abbiamo $\varphi(u - v) = 0$, da cui $u = v$ ovvero φ è iniettiva.

ii). Ovvia. \square

In particolare, φ è un isomorfismo se e solo se $\dim \text{Ker } \varphi = 0$ e $\dim \text{Im } \varphi = \dim W$.

Le dimensioni di $\text{Ker } \varphi$ e $\text{Im } \varphi$ sono calcolabili per mezzo delle matrici associate a φ ; abbiamo infatti

Teorema 3.9. *Sia $\varphi : V \rightarrow W$ una trasformazione lineare con $\dim V = n$ e siano $E = \langle v_1, \dots, v_n \rangle$ una base di V e M una matrice associata a φ mediante due basi qualunque di V e W . Allora*

$$\text{Im } \varphi = L(\varphi(v_1), \dots, \varphi(v_n)) \quad \text{e} \quad \dim \text{Im } \varphi = \rho(M).$$

Dimostrazione. Abbiamo visto in precedenza che ogni vettore $\varphi(v)$ si può scrivere come combinazione lineare dei vettori $\varphi(v_1), \dots, \varphi(v_n)$, quindi $\text{Im } \varphi = L(\varphi(v_1), \dots, \varphi(v_n))$. Sia ora F una base di W ; le colonne di $M_\varphi^{E,F}$ sono costituite dalle colonne delle coordinate di $\varphi(v_1), \dots, \varphi(v_n)$ rispetto alla base F e quindi dal Teorema 3.4 segue che $\dim \text{Im } \varphi = \rho(M_\varphi^{E,F})$. Il teorema segue allora dalla (3.9). \square

Teorema 3.10. *Sia $\varphi : V \rightarrow W$ una trasformazione lineare con $\dim V = n$ e sia M una matrice associata a φ mediante due basi qualunque di V e W . Allora*

$$\dim \text{Ker } \varphi = n - \rho(M).$$

Dimostrazione. Siano E ed F basi di V e W , rispettivamente. Dalla (3.5) abbiamo che

$$\text{Ker } \varphi = \{v \in V : M_\varphi^{E,F} v_E = 0\}.$$

Ma $M_\varphi^{E,F} v_E = 0$ è un sistema lineare omogeneo e quindi per il Teorema 3.5 il sottospazio di $\mathbb{R}^n / \mathbb{C}^n$ delle soluzioni ha dimensione $n - \rho(M_\varphi^{E,F})$. Segue che $\dim \text{Ker } \varphi = n - \rho(M_\varphi^{E,F})$, e il teorema segue dalla (3.9). \square

Dai Teoremi 3.9 e 3.10 otteniamo immediatamente il

Teorema della dimensione. *Sia $\varphi : V \rightarrow W$ una trasformazione lineare. Allora*

$$\dim \text{Ker } \varphi + \dim \text{Im } \varphi = \dim V.$$

Abbiamo quindi, in particolare, che φ è un isomorfismo se e solo se $\dim \text{Ker } \varphi = 0$ e $\dim V = \dim W$.

Osserviamo infine le seguenti proprietà di una trasformazione lineare $\varphi : V \rightarrow W$:

- i) se $\varphi(v_1), \dots, \varphi(v_k)$ sono linearmente indipendenti, allora v_1, \dots, v_k sono linearmente indipendenti
- ii) se φ è iniettiva e v_1, \dots, v_k sono linearmente indipendenti, allora $\varphi(v_1), \dots, \varphi(v_k)$ sono linearmente indipendenti
- iii) se φ è un isomorfismo, allora $E = \langle v_1, \dots, v_n \rangle$ è una base di V se e solo se $F = \langle \varphi(v_1), \dots, \varphi(v_n) \rangle$ è una base di W .

3.3. Sistemi lineari e trasformazioni lineari

Equazioni lineari

Data una trasformazione lineare $\varphi : V \rightarrow W$ ed un vettore $b \in W$ consideriamo l'*equazione lineare*

$$\varphi(x) = b; \tag{3.11}$$

Una *soluzione* dell'equazione (3.11) è un vettore $x \in V$ che soddisfa la (3.11). Osserviamo per prima cosa che, ovviamente,

(i) l'equazione (3.11) ha soluzione se e solo se $b \in \text{Im } \varphi$.

Dati un vettore $v \in V$ e un insieme $U \subset V$ definiamo l'insieme *traslato* $U + v = \{u + v : u \in U\}$. Supponiamo ora che $b \in \text{Im } \varphi$ e che $x_0 \in V$ sia una soluzione dell'equazione (3.11); abbiamo allora

(ii) l'insieme delle soluzioni dell'equazione (3.11) è $\text{Ker } \varphi + x_0$.

Infatti, se $v \in \text{Ker } \varphi$ abbiamo $\varphi(v + x_0) = \varphi(v) + \varphi(x_0) = \varphi(x_0) = b$, ovvero $v + x_0$ è soluzione della (3.11). Viceversa, se $x \in V$ soddisfa la (3.11) poniamo $v = x - x_0$; abbiamo allora $\varphi(v) = \varphi(x) - \varphi(x_0) = b - b = 0$, ovvero $v \in \text{Ker } \varphi$.

Le (i) e (ii) forniscono un metodo di risoluzione dell'equazione lineare (3.11): se $b \in \text{Im } \varphi$ e x_0 è una soluzione qualunque della (3.11), tutte le soluzioni della (3.11) si ottengono traslando di x_0 il nucleo di φ .

Sistemi lineari

Il metodo di risoluzione dell'equazione (3.11) fornisce a sua volta un metodo alternativo per la risoluzione dei sistemi lineari.

Dato un sistema lineare $Ax = b$ con $A \in M_{m \times n}(\mathbb{R}/\mathbb{C})$ consideriamo la trasformazione lineare $\varphi : \mathbb{R}^n/\mathbb{C}^n \rightarrow \mathbb{R}^m/\mathbb{C}^m$ associata alla matrice A mediante le basi canoniche, ovvero

$$\varphi(x) = Ax$$

per la (3.5); il sistema lineare $Ax = b$ è quindi equivalente all'equazione lineare (3.11). Osservando che $\text{Ker } \varphi$ coincide con il sottospazio di \mathbb{R}^n delle soluzioni del *sistema omogeneo associato* $Ax = 0$, il metodo sopra descritto stabilisce che le soluzioni del sistema lineare $Ax = b$, se esistono, si possono ottenere calcolando una soluzione x_0 e poi traslando di x_0 le soluzioni del sistema omogeneo associato. In altre parole, denotando con S e S_0 rispettivamente le soluzioni dei sistemi lineari $Ax = b$ e $Ax = 0$, se $S \neq \emptyset$ allora

$$S = S_0 + x_0. \tag{3.12}$$

Osserviamo infine che la condizione di risolubilità dell'equazione (3.11) espressa dalla (i) coincide, in questo caso, con la condizione di risolubilità espressa dal teorema di Rouchè-Capelli. Infatti, dal Teorema 3.9 abbiamo che $\text{Im } \varphi = L(Ae_1, \dots, Ae_n)$; ma Ae_1, \dots, Ae_n non sono altro che le colonne C_1, \dots, C_n della matrice A e quindi

$$b \in \text{Im } \varphi \Leftrightarrow b \in L(C_1, \dots, C_n) \Leftrightarrow \rho(A) = \rho(A|b)$$

per quanto visto nel paragrafo precedente.

Capitolo 4. GEOMETRIA ANALITICA, parte I

4.1. Vettori geometrici

Vettori geometrici

Lo spazio vettoriale \mathbb{R}^3 può essere identificato con lo spazio euclideo tridimensionale dotato degli assi cartesiani x, y, z : un vettore $v = (a, b, c) \in \mathbb{R}^3$ si identifica con il punto P di coordinate (a, b, c) e viceversa. In tale identificazione gli spazi vettoriali \mathbb{R}^2 e \mathbb{R} coincidono rispettivamente con il piano formato dagli assi x, y e la retta individuata dall'asse x ; pertanto nel seguito considereremo soltanto il caso di \mathbb{R}^3 , gli spazi vettoriali \mathbb{R}^2 e \mathbb{R} essendo identificati con tali sottospazi di \mathbb{R}^3 .

Un'altra visualizzazione dei vettori di \mathbb{R}^3 , molto utile nelle applicazioni, è la seguente: un vettore $v = (a, b, c) \in \mathbb{R}^3$ viene identificato con la *freccia* uscente dall'origine O degli assi cartesiani ed avente la punta coincidente con il punto P di coordinate (a, b, c) . Tali frecce prendono il nome di *vettori geometrici*, che vengono denotati con \mathbf{v} per distinguerli dai vettori $v \in \mathbb{R}^3$ definiti nel capitolo precedente.

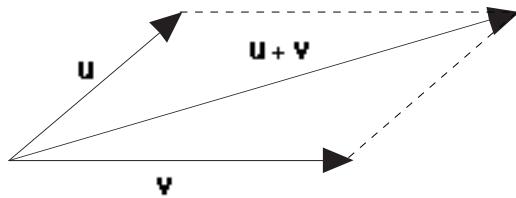
I vettori geometrici sono determinati da tre entità:

- i) *direzione*, ovvero la direzione della retta su cui giace il segmento OP
- ii) *verso*, ovvero il verso indicato dalla freccia
- iii) *modulo*, ovvero la lunghezza del segmento OP .

Le coordinate (a, b, c) del punto P prendono il nome di *coordinate* (o *componenti*) di \mathbf{v} , e scriveremo $\mathbf{v} = (a, b, c)$; abbiamo quindi che il modulo di \mathbf{v} , denotato con $|\mathbf{v}|$, vale $|\mathbf{v}| = \sqrt{a^2 + b^2 + c^2}$. Osserviamo che i vettori di \mathbb{R}^2 e \mathbb{R} sono rispettivamente quelli del tipo $(a, b, 0)$ e $(a, 0, 0)$.

Dato un punto P di coordinate (a, b, c) denotiamo con \mathbf{OP} il vettore geometrico sopra descritto, le cui coordinate sono ovviamente (a, b, c) . L'origine O è detta *punto di applicazione dei vettori* \mathbf{v} ; è chiaro che dato un qualunque punto P_0 di coordinate (a_0, b_0, c_0) possiamo definire i *vettori geometrici applicati in* P_0 come le frecce uscenti da P_0 ed aventi la punta coincidente con un punto P di coordinate (a, b, c) . Tali frecce si denotano con $\mathbf{P}_0\mathbf{P}$; è chiaro che $\mathbf{P}_0\mathbf{P}$ ha lo stesso verso e modulo del vettore geometrico \mathbf{v} di coordinate $(a - a_0, b - b_0, c - c_0)$ e giace su una retta parallela alla direzione di \mathbf{v} . Segue che i vettori geometrici $\mathbf{P}_0\mathbf{P}$ e \mathbf{v} differiscono soltanto per il punto di applicazione; nel seguito identificheremo spesso i vettori che differiscono solo per il punto di applicazione, e pertanto diremo che $\mathbf{P}_0\mathbf{P}$ ha coordinate $(a - a_0, b - b_0, c - c_0)$.

Le *operazioni* tra vettori di \mathbb{R}^3 hanno la seguente interpretazione geometrica. Dati $\mathbf{u} = (a, b, c)$ e $\mathbf{v} = (a', b', c')$, il vettore geometrico *somma* è $\mathbf{u} + \mathbf{v} = (a + a', b + b', c + c')$ ed è rappresentato, secondo la *regola del parallelogramma*, dalla diagonale del parallelogramma individuato da \mathbf{u} e \mathbf{v} , come da figura.



Osserviamo che vale la *diseguaglianza triangolare* $|\mathbf{u} + \mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|$, che si verifica facilmente notando che la lunghezza di un lato di un triangolo non supera la somma delle lunghezze degli altri due lati. Inoltre, dato $\lambda \in \mathbb{R}$ il *prodotto per scalare* $\lambda\mathbf{v} = (\lambda a, \lambda b, \lambda c)$ ha la direzione di \mathbf{v} , verso uguale o opposto a seconda del segno di λ e modulo $|\lambda||\mathbf{v}|$.

I *versori* degli assi cartesiani sono rispettivamente i vettori $\mathbf{i} = (1, 0, 0)$, $\mathbf{j} = (0, 1, 0)$ e $\mathbf{k} = (0, 0, 1)$; è chiaro che ogni vettore geometrico $\mathbf{v} = (a, b, c)$ si scrive come $\mathbf{v} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$.

Osserviamo infine che l'interpretazione geometrica delle operazioni di somma e prodotto per scalare fornisce un'interpretazione geometrica del concetto di dipendenza lineare. Abbiamo infatti che

$$\begin{aligned} \text{due vettori sono linearmente dipendenti} &\Leftrightarrow \text{hanno la stessa direzione} \\ \text{tre vettori sono linearmente dipendenti} &\Leftrightarrow \text{sono complanari}. \end{aligned}$$

Prodotto scalare e proiezioni

Dati due vettori \mathbf{u} e \mathbf{v} definiamo il *prodotto scalare*

$$\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}||\mathbf{v}| \cos \theta$$

dove θ è l'angolo formato dai vettori \mathbf{u} e \mathbf{v} . Tale angolo viene definito come l'angolo orientato in senso antiorario che porta il vettore \mathbf{u} a sovrapporsi al vettore \mathbf{v} , nel piano individuato dai vettori \mathbf{u} e \mathbf{v} . Abbiamo quindi che il prodotto scalare di due vettori è un *numero reale*; il prodotto scalare soddisfa le seguenti proprietà

- i) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
- ii) $(\mathbf{u} \cdot \mathbf{u})^{\frac{1}{2}} = |\mathbf{u}|$
- iii) $\mathbf{u} \cdot \mathbf{v} = 0 \Leftrightarrow \mathbf{u}$ e \mathbf{v} sono *ortogonali*
- iv) se $\mathbf{u} = (a, b, c)$ e $\mathbf{v} = (a', b', c')$ allora $\mathbf{u} \cdot \mathbf{v} = aa' + bb' + cc'$.

L'interesse del prodotto scalare deriva principalmente dalla proprietà iii): infatti tale proprietà caratterizza l'*ortogonalità*, che è un importante concetto geometrico. Se i vettori \mathbf{u} e \mathbf{v} sono ortogonali scriveremo $\mathbf{u} \perp \mathbf{v}$.

Il prodotto scalare consente di determinare in modo semplice la *proiezione (ortogonale)* $\mathbf{P}\mathbf{u}$ di un vettore \mathbf{u} su un vettore \mathbf{v} (o meglio sulla *direzione* di \mathbf{v}). Dalla trigonometria è infatti facile vedere che

$$\mathbf{P}\mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{v}|^2} \mathbf{v}.$$

Abbiamo quindi che

- i) $\mathbf{P}\mathbf{u}$ ha la direzione di \mathbf{v}
- ii) $|\mathbf{P}\mathbf{u}| = |\mathbf{u}||\cos \theta|$
- iii) $\mathbf{u} - \mathbf{P}\mathbf{u}$ è ortogonale a $\mathbf{P}\mathbf{u}$.

In modo analogo possiamo definire

la *proiezione (ortogonale)* $\mathbf{P}\mathbf{u}$ di \mathbf{u} su un piano π : se \mathbf{v}_1 e \mathbf{v}_2 sono due vettori *ortogonali* che individuano π , la proiezione è data da

$$\mathbf{P}\mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{v}_1}{|\mathbf{v}_1|^2} \mathbf{v}_1 + \frac{\mathbf{u} \cdot \mathbf{v}_2}{|\mathbf{v}_2|^2} \mathbf{v}_2.$$

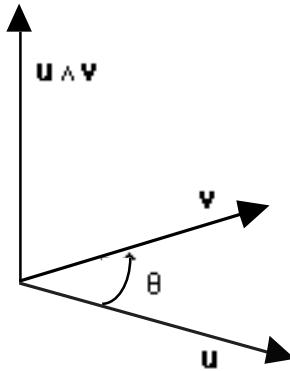
Denotando nuovamente con θ l'angolo formato da \mathbf{u} e $\mathbf{P}\mathbf{u}$ abbiamo allora che

- i) $\mathbf{P}\mathbf{u}$ giace su π
- ii) $|\mathbf{P}\mathbf{u}| = |\mathbf{u}||\cos \theta|$
- iii) $\mathbf{u} - \mathbf{P}\mathbf{u}$ è ortogonale a $\mathbf{P}\mathbf{u}$.

Prodotto vettoriale e prodotto misto

Dati due vettori \mathbf{u} e \mathbf{v} indichiamo con π il piano da essi individuato e definiamo il *prodotto vettoriale* $\mathbf{u} \wedge \mathbf{v}$ nel modo seguente

- i) $\mathbf{u} \wedge \mathbf{v}$ ha direzione ortogonale a π
- ii) $\mathbf{u} \wedge \mathbf{v}$ ha verso individuato dalla testa di un osservatore con il piede destro su \mathbf{u} , quello sinistro su \mathbf{v} e la schiena rivolta verso l'origine
- iii) $|\mathbf{u} \wedge \mathbf{v}| = |\mathbf{u}||\mathbf{v}||\sin \theta|$, θ essendo l'angolo formato da \mathbf{u} e \mathbf{v} .



Il prodotto vettoriale gode delle seguenti proprietà

- i) $\mathbf{u} \wedge \mathbf{v} = -\mathbf{v} \wedge \mathbf{u}$
- ii) $\mathbf{u} \wedge \mathbf{v} = 0 \Leftrightarrow \mathbf{u}$ e \mathbf{v} sono *parallel*i
- iii) $|\mathbf{u} \wedge \mathbf{v}|$ è uguale all'area del parallelogramma individuato da \mathbf{u} e \mathbf{v} .
- iv) se $\mathbf{u} = (a, b, c)$ e $\mathbf{v} = (a', b', c')$ allora

$$\mathbf{u} \wedge \mathbf{v} = \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a & b & c \\ a' & b' & c' \end{pmatrix}$$

v) $\mathbf{u} \wedge \mathbf{v}$ è ortogonale al piano individuato da \mathbf{u} e \mathbf{v} .

Abbiamo quindi che prodotto scalare e prodotto vettoriale consentono di caratterizzare le importanti nozioni geometriche di ortogonalità e parallelismo; inoltre, il prodotto vettoriale consente di costruire un vettore ortogonale a due vettori dati.

Introduciamo infine il *prodotto misto* di tre vettori \mathbf{u} , \mathbf{v} e \mathbf{w} come $(\mathbf{u} \wedge \mathbf{v}) \cdot \mathbf{w}$; quindi il prodotto misto è un numero reale e si verifica che

$$(\mathbf{u} \wedge \mathbf{v}) \cdot \mathbf{w} = \det \begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix},$$

dove $\mathbf{u} = (a, b, c)$, $\mathbf{v} = (a', b', c')$ e $\mathbf{w} = (a'', b'', c'')$. Si verifica inoltre che

- i) $|(\mathbf{u} \wedge \mathbf{v}) \cdot \mathbf{w}|$ è uguale al volume del parallelepipedo individuato dai vettori \mathbf{u} , \mathbf{v} e \mathbf{w}
- ii) $(\mathbf{u} \wedge \mathbf{v}) \cdot \mathbf{w} = 0 \Leftrightarrow \mathbf{u}$, \mathbf{v} e \mathbf{w} sono *complanari*.

In definitiva, l'annullarsi del prodotto scalare, vettoriale e misto ha il seguente significato geometrico

- i) $\mathbf{u} \cdot \mathbf{v} = 0 \Leftrightarrow \mathbf{u}$ e \mathbf{v} sono ortogonali
- ii) $\mathbf{u} \wedge \mathbf{v} = 0 \Leftrightarrow \mathbf{u}$ e \mathbf{v} sono paralleli
- iii) $(\mathbf{u} \wedge \mathbf{v}) \cdot \mathbf{w} = 0 \Leftrightarrow \mathbf{u}$, \mathbf{v} e \mathbf{w} sono complanari.

4.2. Geometria lineare nel piano

Coordinate

In analogia con i numeri complessi introduciamo due tipi di coordinate in \mathbb{R}^2 , ovvero le coordinate *cartesiane* e *polari*. Un punto P del piano è individuato in coordinate cartesiane mediante l'*ascissa* x e l'*ordinata* y , $P = (x, y)$, e in coordinate polari mediante il *modulo* ρ e l'*argomento* θ , $P = (\rho, \theta)$; in analogia con le (1.2) e (1.3) abbiamo le seguenti *formule di passaggio* tra le due coordinate:

$$\begin{cases} x = \rho \cos \theta \\ y = \rho \sin \theta \end{cases} \quad \begin{cases} \rho = \sqrt{a^2 + b^2} \\ \cos \theta = \frac{a}{\rho} \\ \sin \theta = \frac{b}{\rho}. \end{cases}$$

Consideriamo ora un *sistema di riferimento cartesiano Oxy*, dove O denota l'*origine* $(0, 0)$. Tale sistema può essere trasformato in un altro sistema di riferimento cartesiano $O'XY$ mediante due trasformazioni base: la *traslazione* e la *rotazione*. La traslazione è descritta semplicemente mediante le coordinate in Oxy della nuova origine $O' = (a, b)$; conseguentemente, le *formule di passaggio* tra i due sistemi di riferimento sono date da

$$\begin{cases} x = X + a \\ y = Y + b \end{cases} \quad \begin{cases} X = x - a \\ Y = y - b. \end{cases} \tag{4.1}$$

La rotazione è descritta mediante l'angolo θ , orientato in senso antiorario, che l'asse delle ascisse X forma con l'asse delle ascisse x ; dalla trigonometria si deduce che le *formule di passaggio* in questo caso sono

$$\begin{cases} x = X \cos \theta - Y \sin \theta \\ y = X \sin \theta + Y \cos \theta \end{cases} \quad \begin{cases} X = x \cos \theta + y \sin \theta \\ Y = -x \sin \theta + y \cos \theta. \end{cases} \quad (4.2)$$

Traslazione e rotazione possono essere combinate in modo da ottenere la più generale trasformazione del sistema Oxy , ovvero la *rototraslazione*. In questo caso le *formule di passaggio* si ottengono combinando le (4.1) e (4.2):

$$\begin{cases} x = a + X \cos \theta - Y \sin \theta \\ y = b + X \sin \theta + Y \cos \theta \end{cases} \quad \begin{cases} X = (x - a) \cos \theta + (y - b) \sin \theta \\ Y = -(x - a) \sin \theta + (y - b) \cos \theta. \end{cases}$$

Formule analoghe possono essere ottenute anche nel caso delle coordinate polari.

Osserviamo infine che la *distanza* tra due punti $P_1 = (x_1, y_1)$ e $P_2 = (x_2, y_2)$ vale

$$d(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2},$$

mentre il loro *punto medio* M ha coordinate

$$M = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2} \right).$$

La retta

Una retta r nel piano è individuata da un punto $P_0 = (x_0, y_0) \in r$ e da un *vettore normale* $\mathbf{v} = (a, b)$ ortogonale a r ; abbiamo infatti che

$$P \in r \Leftrightarrow \mathbf{P}_0 \mathbf{P} \text{ è ortogonale a } \mathbf{v}$$

e quindi dalle proprietà del prodotto scalare otteniamo che le coordinate del generico punto $P = (x, y) \in r$ soddisfano l'equazione

$$a(x - x_0) + b(y - y_0) = 0.$$

Segue che, in generale, l'*equazione cartesiana* della retta nel piano è

$$ax + by + c = 0. \quad (4.3)$$

I coefficienti a e b hanno significato geometrico; precisamente, il vettore $\mathbf{v} = (a, b)$ è ortogonale alla retta.

Le rette parallele all'asse y hanno equazione $x = x_0$; l'equazione di tutte le altre rette può essere scritta nella forma

$$y = mx + q,$$

dove i coefficienti m e q hanno il seguente significato geometrico: m è il *coefficiente angolare*, ovvero $m = \tan \theta$ dove θ è l'angolo (orientato in senso antiorario) che la retta forma con l'asse x , mentre q è la *quota* a cui la retta interseca l'asse y . In particolare, le rette parallele all'asse x hanno equazione $y = y_0$.

Un altro modo di individuare una retta r è mediante un punto $P_0 = (x_0, y_0) \in r$ e un *vettore direzionale* $\mathbf{v} = (l, m)$ parallelo a r ; abbiamo infatti che

$$P \in r \Leftrightarrow \mathbf{P}_0 P \text{ è parallelo a } \mathbf{v}$$

e quindi da quanto visto nel paragrafo precedente otteniamo che l'*equazione parametrica* della retta nel piano è

$$\begin{cases} x = x_0 + lt \\ y = y_0 + mt \end{cases} \quad t \in \mathbb{R}. \quad (4.4)$$

Anche in questo caso i coefficienti hanno un ovvio significato geometrico.

Osserviamo che l'equazione cartesiana e l'equazione parametrica di una retta *non sono uniche*; ad esempio, due equazioni $ax + by + c = 0$ e $a'x + b'y + c' = 0$ rappresentano la stessa retta se e solo se $(a', b', c') = \lambda(a, b, c)$ per qualche $\lambda \in \mathbb{R}$, e considerazioni analoghe valgono nel caso dell'equazione parametrica.

Osserviamo infine che è molto semplice passare dall'equazione cartesiana a quella parametrica, e viceversa: nel primo caso basta porre $x = t$ (oppure $y = t$) e calcolare il corrispondente valore di y (oppure di x) dalla (4.3), mentre nel secondo è sufficiente eliminare la t dalla (4.4).

Problemi geometrici

Consideriamo alcuni *problemi geometrici* relativi alla retta.

(a) *retta per due punti*. La retta passante per due punti *distinti* $P_0 = (x_0, y_0)$ e $P_1 = (x_1, y_1)$ ha vettore direzionale $\mathbf{v} = (x_1 - x_0, y_1 - y_0)$; la sua equazione parametrica è quindi

$$\begin{cases} x = x_0 + t(x_1 - x_0) \\ y = y_0 + t(y_1 - y_0). \end{cases}$$

Se $(x_1 - x_0)(y_1 - y_0) \neq 0$ l'equazione cartesiana è

$$\frac{x - x_0}{x_1 - x_0} = \frac{y - y_0}{y_1 - y_0},$$

mentre se $x_1 = x_0$ (rispettivamente $y_1 = y_0$) l'equazione cartesiana è

$$x = x_0 \quad (\text{rispettivamente } y = y_0).$$

(b) *rette parallele e ortogonali*. Due rette r e r' sono parallele (rispettivamente ortogonali) se e solo se i loro vettori normali o direzionali sono paralleli (rispettivamente ortogonali). Con le notazioni introdotte sopra abbiamo quindi

r è parallela a $r' \Leftrightarrow (a, b) = \lambda(a', b')$, oppure $(l, m) = \lambda(l', m')$, oppure $m = m'$
 r è ortogonale a $r' \Leftrightarrow aa' + bb' = 0$, oppure $ll' + mm' = 0$, oppure $m = -\frac{1}{m'}$.

(c) *intersezione di due rette.* L'intersezione di due rette r e r' si può ottenere mettendo a sistema le equazioni cartesiane di r e r' . Osserviamo come le tre possibilità previste in modo algebrico dal teorema di Rouchè-Capelli in questo caso, ovvero 0 soluzioni, 1 soluzione o ∞^1 soluzioni, coincidano dal punto di vista geometrico rispettivamente con il caso di due rette distinte aventi vettori normali paralleli (e quindi parallele), di due rette incidenti o di due rette coincidenti.

(d) *angolo tra due rette.* Per angolo tra due rette r e r' intendiamo quello $\leq \frac{\pi}{2}$ tra i due angoli determinati da r e r' . Denotando con θ tale angolo e con $\mathbf{v} = (a, b, 0)$ e $\mathbf{v}' = (a', b', 0)$ i vettori normali (o direzionali) di r e r' , dalle formule per il prodotto scalare e vettoriale viste nel paragrafo precedente otteniamo

$$\cos \theta = \frac{|\mathbf{v} \cdot \mathbf{v}'|}{|\mathbf{v}| |\mathbf{v}'|}, \quad \sin \theta = \frac{|\mathbf{v} \wedge \mathbf{v}'|}{|\mathbf{v}| |\mathbf{v}'|} \quad \text{e} \quad \operatorname{tg} \theta = \frac{|\mathbf{v} \wedge \mathbf{v}'|}{|\mathbf{v} \cdot \mathbf{v}'|}.$$

(e) *fascio di rette.* Per *fascio di rette* con *centro* in $P_0 = (x_0, y_0)$ si intende l'insieme di tutte le rette passanti per P_0 . Vi sono due modi per determinare l'equazione di tutte le rette del fascio. Note le equazioni di due rette r e r' del fascio, siano esse $ax + by + c = 0$ e $a'x + b'y + c' = 0$, le equazioni di tutte le rette del fascio si ottengono da

$$\lambda(ax + by + c) + \mu(a'x + b'y + c') = 0$$

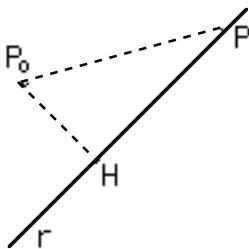
al variare di $\lambda, \mu \in \mathbb{R}$. Osserviamo che se r e r' sono *parallele* allora tale equazione fornisce tutte le rette parallele a r e r' . Osserviamo inoltre che, per quanto visto in precedenza, tali equazioni rappresentano la stessa retta se il rapporto $\frac{\lambda}{\mu}$ è costante.

Alternativamente, le equazioni delle rette del fascio si ottengono da

$$y - y_0 = m(x - x_0)$$

al variare di $m \in \mathbb{R}$; ovviamente a tali rette va aggiunta la retta di equazione $x = x_0$.

(f) *distanza punto-retta.* Dati un punto $P_0 = (x_0, y_0)$ e una retta r di equazione $ax+by+c=0$, la distanza $d(P_0, r)$ di P_0 da r è la lunghezza del segmento che congiunge P_0 con H , vedi figura.



Abbiamo quindi che $d(P_0, r)$ è il modulo della proiezione di $\mathbf{P}_0\mathbf{P}$, dove $P = (\alpha, \beta)$ è un punto qualunque su r , sul vettore normale $\mathbf{v} = (a, b)$ di r , ovvero

$$d(P_0, r) = \frac{|\mathbf{v} \cdot \mathbf{P}_0\mathbf{P}|}{|\mathbf{v}|}. \quad (4.5)$$

Per calcolare $d(P_0, r)$ osserviamo che

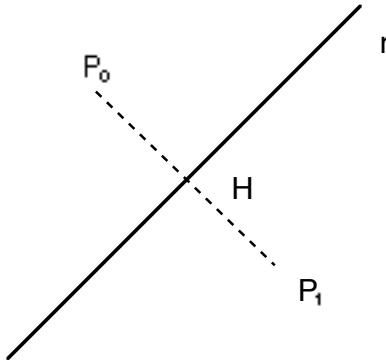
$$\mathbf{v} \cdot \mathbf{P}_0\mathbf{P} = a(x_0 - \alpha) + b(y_0 - \beta) = ax_0 + by_0 - (a\alpha + b\beta); \quad (4.6)$$

ma $P \in r$, quindi $a\alpha + b\beta = -c$ e dalle (4.5) e (4.6) otteniamo

$$d(P_0, r) = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}.$$

Osserviamo che allo stesso risultato si perviene costruendo la retta s passante per P_0 ed ortogonale ad r , e calcolando poi l'intersezione tra r e s .

(g) *punto simmetrico rispetto a una retta*. Dati un punto $P_0 = (x_0, y_0)$ e una retta r , per calcolare il punto $P_1 = (x_1, y_1)$ simmetrico di P_0 rispetto a r basta osservare che H è il punto medio di P_0 e P_1 , vedi figura.



Segue che $H = \left(\frac{x_0+x_1}{2}, \frac{y_0+y_1}{2}\right)$ e quindi per ottenere le coordinate di P_1 basta calcolare quelle di H , come intersezione tra r e la retta s passante per P_0 ed ortogonale a r .

(h) *asse di un segmento*. L'asse di un segmento P_1P_2 è la retta ortogonale a P_1P_2 passante per il suo punto medio. L'equazione dell'asse si calcola facilmente in quanto sono noti un punto di passaggio, ovvero il punto medio M di P_1P_2 , e un vettore ortogonale, ovvero il vettore $\mathbf{P}_1\mathbf{P}_2$.

4.3. Coniche in forma canonica

La circonferenza

La *circonferenza* è il *luogo geometrico* dei punti che hanno una data distanza r da un punto fissato C ; r e C sono rispettivamente il *raggio* e il *centro* della circonferenza. Se $C = (x_0, y_0)$, dal teorema di Pitagora abbiamo che l'equazione della circonferenza è

$$(x - x_0)^2 + (y - y_0)^2 = r^2. \quad (4.7)$$

Sviluppando i quadrati vediamo che la (4.7) è del tipo

$$x^2 + y^2 + ax + by + c = 0, \quad (4.8)$$

e viceversa la (4.8) è l'equazione della circonferenza con

$$\text{centro } C = \left(-\frac{a}{2}, -\frac{b}{2}\right) \text{ e raggio } r = \sqrt{\frac{a^2 + b^2}{4} - c};$$

la (4.8) rappresenta quindi una circonferenza a *punti reali* se e solo se $a^2 + b^2 > 4c$. Le (4.7) e (4.8) sono due forme dell'*equazione cartesiana* della circonferenza. Osserviamo che moltiplicando per $\lambda \in \mathbb{R}$, $\lambda \neq 0$, i coefficienti della (4.8) si ottiene un'altra equazione cartesiana della stessa circonferenza.

Dalla trigonometria abbiamo che l'*equazione parametrica* della circonferenza di centro $C = (x_0, y_0)$ e raggio r è

$$\begin{cases} x = x_0 + r \cos t \\ y = y_0 + r \sin t \end{cases} \quad t \in \mathbb{R}.$$

Consideriamo ora alcuni *problemi geometrici* riguardanti la circonferenza.

(a) *circonferenza per tre punti*. E' ben noto che esiste una sola circonferenza passante per tre punti *non allineati*. Dati tre punti $P_i = (x_i, y_i)$, $i = 1, 2, 3$, non allineati vi sono due semplici metodi per calcolare l'equazione della circonferenza passante per tali punti:

i) sostituire le coordinate dei punti di passaggio nella (4.8); in questo modo si ottiene un sistema di Cramer 3×3 nelle incognite a, b e c

ii) calcolare le equazioni degli assi dei segmenti P_1P_2 e P_2P_3 ; l'intersezione di tali assi è il centro C della circonferenza cercata, ed il raggio non è altro che la distanza tra C e uno qualunque dei punti P_i .

(b) *intersezione retta-circonferenza*. Vi sono tre posizioni di una retta r rispetto ad una circonferenza \mathcal{C} : *i)* r *interseca* \mathcal{C} in due punti distinti; *ii)* r è *tangente* a \mathcal{C} , ovvero interseca \mathcal{C} in due punti coincidenti; *iii)* r è *esterna* a \mathcal{C} , ovvero interseca \mathcal{C} in due punti complessi. Per calcolare i punti di intersezione basta mettere a sistema le equazioni cartesiane di r e di \mathcal{C} , ottenendo un semplice sistema *non lineare* di due equazioni in due incognite.

(c) *rette tangenti passanti per un punto*. Data una circonferenza \mathcal{C} e un punto P_0 *esterno* a \mathcal{C} vi sono due rette passanti per P_0 e tangenti a \mathcal{C} . Per calcolare l'equazione di tali rette basta imporre che la generica retta del fascio di centro P_0 sia tangente a \mathcal{C} , ovvero che il sistema formato dalle equazioni cartesiane di \mathcal{C} e della generica retta del fascio abbia una sola soluzione; in questo modo si ottengono i coefficienti delle due rette tangenti.

Se P_0 è *interno* a \mathcal{C} non vi sono ovviamente tangenti a \mathcal{C} passanti per P_0 , mentre se $P_0 \in \mathcal{C}$ c'è una sola tangente, la cui equazione può essere calcolata come nel caso precedente.

Osserviamo che è semplice verificare se un punto $P_0 = (a, b)$ è interno, esterno o appartenente ad una circonferenza \mathcal{C} di equazione $(x - x_0)^2 + (y - y_0)^2 = r^2$: nel primo

caso si ha $(a - x_0)^2 + (b - y_0)^2 < r^2$, nel secondo $(a - x_0)^2 + (b - y_0)^2 > r^2$ e nel terzo $(a - x_0)^2 + (b - y_0)^2 = r^2$.

(d) *circonferenze tangenti ad una retta in un punto*. Vi sono infinite circonference tangenti ad una retta r in un punto $P_0 \in r$: tali circonference hanno centro C sulla retta s passante per P_0 ed ortogonale ad r , e raggio uguale alla distanza tra C e P_0 .

Coniche

Dalla (4.8) vediamo che la circonferenza è il luogo geometrico dei punti le cui coordinate soddisfano un'equazione di secondo grado. In generale, una *conica* è il luogo geometrico dei punti le cui coordinate soddisfano un'equazione del tipo

$$\Gamma(x, y) = ax^2 + by^2 + cxy + dx + ey + f = 0. \quad (4.9)$$

Come già osservato nel caso della circonferenza, un'equazione del tipo (4.9) può non avere soluzioni reali, come ad esempio $x^2 + 2y^2 + 1 = 0$, ed in tal caso darà luogo ad una *conica immaginaria*. Un altro esempio di “patologia” ottenibile da un'equazione del tipo (4.9) è dato da $3x^2 + y^2 = 0$, la cui unica soluzione reale è $(0, 0)$; pertanto tale equazione rappresenta una conica con un solo punto reale. Osserviamo inoltre che un'equazione del tipo

$$(ax + by + c)(a'x + b'y + c') = 0$$

è ancora della forma (4.9); in questo caso il luogo geometrico che essa rappresenta è l'unione delle due rette di equazione $ax + by + c = 0$ e $a'x + b'y + c' = 0$.

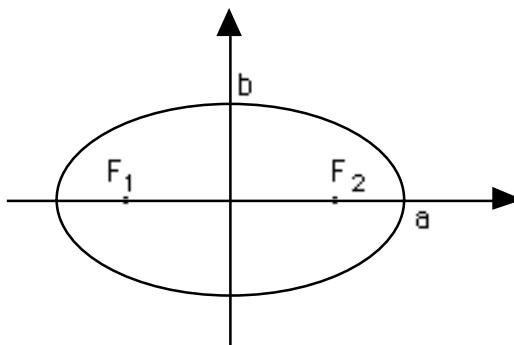
Nel Capitolo 7 studieremo le coniche in modo sistematico; per il momento ci limitiamo alla descrizione delle principali proprietà delle coniche fondamentali, ovvero l'*ellisse*, la *parabola* e l'*iperbole*, nella loro forma più semplice. Tali coniche sono dette coniche *non-degeneri* in *forma canonica*.

Ellisse

L'*ellisse* è il luogo geometrico dei punti P tali che la somma delle distanze di P da due punti fissi F_1 e F_2 è costante. I punti F_1 e F_2 sono detti *fuochi* dell'ellisse. Nel caso in cui

$$d(P, F_1) + d(P, F_2) = 2a$$

e i fuochi sono $F_1 = (-\sqrt{a^2 - b^2}, 0)$ e $F_2 = (\sqrt{a^2 - b^2}, 0)$ con $a > b > 0$ si ottiene l'ellisse della figura.



Non è difficile verificare che l'equazione di tale ellisse è

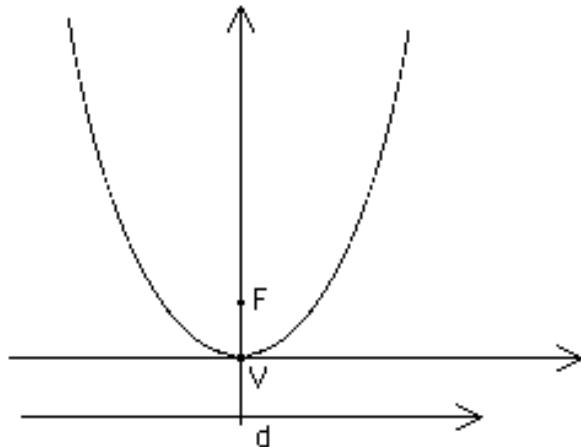
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \quad (4.10)$$

Le quattro intersezioni con gli assi cartesiani, ovvero i punti $(\pm a, 0)$ e $(0, \pm b)$, sono i *vertici* dell'ellisse mentre l'origine è il *centro* dell'ellisse.

Se invece $0 < a < b$, la (4.10) rappresenta un'ellisse i cui fuochi sono sull'asse y ed i ruoli di a e b sono scambiati. Osserviamo che la circonferenza corrisponde al caso speciale in cui $a = b$, ovvero i due fuochi coincidono; questo si verifica facilmente sia geometricamente che ricorrendo all'equazione (4.10). Ricordiamo infine che la (4.10) prende il nome di *forma canonica* dell'equazione dell'ellisse.

Parabola

La *parabola* è il luogo geometrico dei punti P equidistanti da un punto fisso F e da una retta d . Il punto F è detto *fuoco* e la retta d *direttrice* della parabola. Nel caso in cui $F = (0, \frac{1}{4a})$ e d ha equazione $y = -\frac{1}{4a}$ con $a > 0$ si ottiene la parabola della figura.



Si verifica facilmente che l'equazione di tale parabola è

$$y = ax^2. \quad (4.11)$$

Il punto $V = (0, 0)$ è il *vertice* della parabola e l'asse y è l'*asse* della parabola.

Se $a < 0$ si ottiene una parabola con la concavità rivolta verso il basso, simmetrica della precedente rispetto all'asse x . La (4.11) è la forma canonica dell'equazione della parabola.

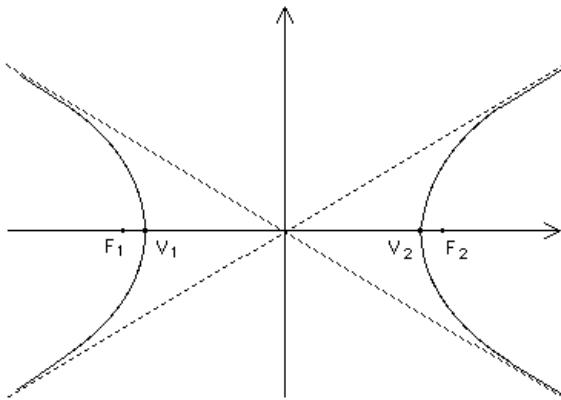
Iperbole

L'*iperbole* è il luogo geometrico dei punti P tali che il valore assoluto della differenza delle distanze di P da due punti fissi F_1 e F_2 è costante; i punti F_1 e F_2 sono i *fuochi* dell'iperbole.

Nel caso in cui

$$|d(P, F_1) - d(P, F_2)| = 2a$$

con $a > 0$ e i fuochi sono $F_1 = (-\sqrt{a^2 + b^2}, 0)$ e $F_2 = (\sqrt{a^2 + b^2}, 0)$ con $b > 0$ si ottiene l'iperbole della figura,



ed ancora una volta si verifica facilmente che l'equazione di tale iperbole è

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \quad (4.12)$$

I punti $V_1 = (-a, 0)$ e $V_2 = (a, 0)$ sono i *vertici* dell'iperbole; le due rette tratteggiate nella figura sono gli *asintoti* dell'iperbole ed hanno equazione $\frac{x}{a} \pm \frac{y}{b} = 0$, e l'origine è il *centro* dell'iperbole.

Contrariamente al caso di ellisse e parabola, l'iperbole è formata da due *rami* disgiunti. Osserviamo che l'iperbole i cui rami occupano i “quadranti” superiore e inferiore formati dagli asintoti ha equazione

$$-\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1; \quad (4.13)$$

le (4.12) e (4.13) sono la forma canonica dell'equazione dell'iperbole.

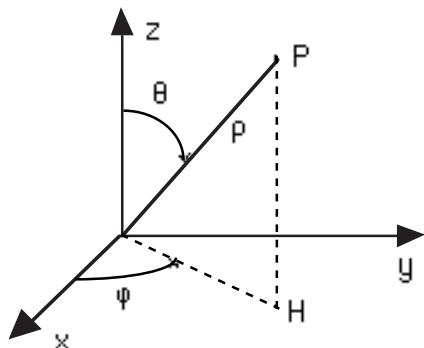
Osserviamo infine che *problemi geometrici* analoghi a quelli considerati nel caso della circonferenza possono essere posti anche nel caso di ellisse, parabola e iperbole; la loro risoluzione si ottiene con ragionamenti analoghi.

4.4. Geometria lineare nello spazio

Coordinate

Le coordinate *cartesiane* nello spazio sono del tutto analoghe a quelle nel piano: ogni punto P è individuato dalle proiezioni

x, y, z sugli assi cartesiani; scriviamo $P = (x, y, z)$. Le coordinate *sferiche* fanno uso dei due angoli φ e θ e del modulo ρ , vedi figura; in questo caso scriviamo $P = (\rho, \varphi, \theta)$. Infine, le coordinate *cilindriche* fanno uso dell'angolo θ , del modulo ρ della proie-



zione di P sul piano individuato dagli assi x, y e dall'altezza z , vedi figura; in questo caso scriviamo $P = (\rho, \theta, z)$. Non è difficile determinare le *formule di passaggio* tra le coordinate cartesiane, sferiche e cilindriche.

Un *sistema di riferimento cartesiano* $Oxyz$, dove $O = (0, 0, 0)$ denota l'origine degli assi, può essere trasformato in un altro sistema di riferimento cartesiano $O'XYZ$ mediante due trasformazioni base: la *traslazione* e la *rotazione*. Per descrivere la traslazione è sufficiente conoscere le coordinate in $Oxyz$ della nuova origine $O' = (a, b, c)$, e le *formule di passaggio* tra i due sistemi di riferimento sono date da

$$\begin{cases} x = X + a \\ y = Y + b \\ z = Z + c \end{cases} \quad \begin{cases} X = x - a \\ Y = y - b \\ Z = z - c. \end{cases} \quad (4.14)$$

Per descrivere la rotazione conviene introdurre i *versori* $\mathbf{i}, \mathbf{j}, \mathbf{k}$ e $\mathbf{I}, \mathbf{J}, \mathbf{K}$ dei due sistemi di riferimento e, dato un punto P , osservare che il vettore \mathbf{OP} si scrive nei due sistemi come

$$\mathbf{OP} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} = X\mathbf{I} + Y\mathbf{J} + Z\mathbf{K}. \quad (4.15)$$

Poichè le coordinate di P sono determinate dalle proiezioni del vettore \mathbf{OP} sui versori, dalla (4.15) abbiamo

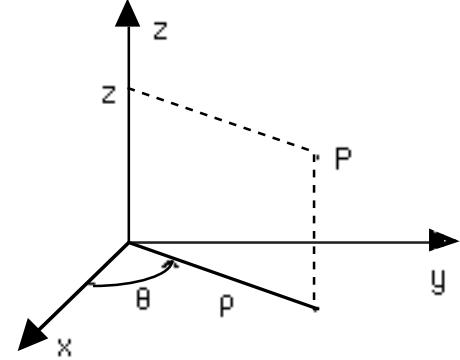
$$\begin{cases} x = \mathbf{OP} \cdot \mathbf{i} = X\mathbf{I} \cdot \mathbf{i} + Y\mathbf{J} \cdot \mathbf{i} + Z\mathbf{K} \cdot \mathbf{i} = p_{11}X + p_{21}Y + p_{31}Z \\ y = \mathbf{OP} \cdot \mathbf{j} = X\mathbf{I} \cdot \mathbf{j} + Y\mathbf{J} \cdot \mathbf{j} + Z\mathbf{K} \cdot \mathbf{j} = p_{12}X + p_{22}Y + p_{32}Z \\ z = \mathbf{OP} \cdot \mathbf{k} = X\mathbf{I} \cdot \mathbf{k} + Y\mathbf{J} \cdot \mathbf{k} + Z\mathbf{K} \cdot \mathbf{k} = p_{13}X + p_{23}Y + p_{33}Z. \end{cases} \quad (4.16)$$

La (4.16) può essere scritta in modo sintetico considerando la *matrice di rotazione* 3×3

$$P = (p_{ij}) = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}.$$

Osserviamo infatti che $\langle \mathbf{i}, \mathbf{j}, \mathbf{k} \rangle$ e $\langle \mathbf{I}, \mathbf{J}, \mathbf{K} \rangle$ sono due basi di \mathbb{R}^3 (la prima è in realtà la base canonica K) e, per quanto visto nel paragrafo 3.2, P è la *matrice di passaggio* da $\langle \mathbf{I}, \mathbf{J}, \mathbf{K} \rangle$ a $\langle \mathbf{i}, \mathbf{j}, \mathbf{k} \rangle$; abbiamo quindi

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = P \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (4.17)$$



Osserviamo inoltre che le due basi in questione sono formate da vettori *ortogonali* fra loro e di modulo uguale a 1; nel Capitolo 6 vedremo che ciò implica che la matrice P è *ortogonale*, ovvero $P^{-1} = P^T$. Abbiamo quindi

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = P^T \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}, \quad (4.18)$$

e le (4.17) e (4.18) forniscono le formule di passaggio nel caso della rotazione.

Si verifica facilmente che se A è una matrice ortogonale, allora $\det A = \pm 1$; nel caso della matrice P abbiamo

$$\det P = 1$$

in quanto le basi $\langle \mathbf{i}, \mathbf{j}, \mathbf{k} \rangle$ e $\langle \mathbf{I}, \mathbf{J}, \mathbf{K} \rangle$ sono *destrorse*, ovvero $\mathbf{k} = \mathbf{i} \wedge \mathbf{j}$ e $\mathbf{K} = \mathbf{I} \wedge \mathbf{J}$.

Osserviamo che i coefficienti p_{ij} della matrice P hanno significato geometrico, esprimibile in termini degli angoli formati dai vettori $\mathbf{i}, \mathbf{j}, \mathbf{k}$ e $\mathbf{I}, \mathbf{J}, \mathbf{K}$.

Osserviamo inoltre che la rotazione di un sistema di riferimento cartesiano nel piano può essere trattata in modo del tutto analogo, mediante una *matrice di rotazione* 2×2 .

Analogamente al caso di \mathbb{R}^2 , traslazione e rotazione possono essere combinate in modo da ottenere la *rototraslazione*, che rappresenta la più generale trasformazione del sistema di riferimento *Oxyz*; le formule di passaggio, che non riportiamo, si ottengono combinando le (4.14), (4.17) e (4.18).

Osserviamo infine che la *distanza* tra due punti $P_1 = (x_1, y_1, z_1)$ e $P_2 = (x_2, y_2, z_2)$ vale

$$d(P_1 P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2},$$

e il loro *punto medio* M ha coordinate

$$M = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}, \frac{z_1 + z_2}{2} \right).$$

Il piano

Un piano π è individuato da un punto $P_0 = (x_0, y_0, z_0) \in \pi$ e da un *vettore normale* $\mathbf{u} = (a, b, c)$ ortogonale a π ; abbiamo infatti che

$$P \in \pi \Leftrightarrow \mathbf{P}_0 \mathbf{P} \text{ è ortogonale a } \mathbf{u},$$

ovvero le coordinate del generico punto $P = (x, y, z) \in \pi$ soddisfano l'equazione

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0.$$

Segue che, in generale, l'*equazione cartesiana* del piano è

$$ax + by + cz + d = 0. \quad (4.19)$$

I coefficienti a, b, c hanno significato geometrico; precisamente, il vettore $\mathbf{u} = (a, b, c)$ è ortogonale al piano.

Un altro modo per individuare un piano π è mediante tre punti *non allineati* $P_i = (x_i, y_i, z_i) \in \pi$, $i = 0, 1, 2$; infatti

$$P \in \pi \Leftrightarrow \mathbf{P}_0\mathbf{P}, \mathbf{P}_0\mathbf{P}_1 \text{ e } \mathbf{P}_0\mathbf{P}_2 \text{ sono complanari,}$$

ovvero per quanto visto nel paragrafo 4.1

$$P \in \pi \Leftrightarrow \mathbf{P}_0\mathbf{P} \text{ è combinazione lineare di } \mathbf{P}_0\mathbf{P}_1 \text{ e } \mathbf{P}_0\mathbf{P}_2.$$

Segue che le coordinate del generico punto $P = (x, y, z) \in \pi$ soddisfano

$$\begin{cases} x - x_0 = s(x_1 - x_0) + t(x_2 - x_0) \\ y - y_0 = s(y_1 - y_0) + t(y_2 - y_0) \\ z - z_0 = s(z_1 - z_0) + t(z_2 - z_0) \end{cases} \quad s, t \in \mathbb{R} \quad (4.20)$$

e quindi, in generale, l'*equazione parametrica* del piano è

$$\begin{cases} x = x_0 + sl + tl' \\ y = y_0 + sm + tm' \\ z = z_0 + sn + tn' \end{cases} \quad s, t \in \mathbb{R}. \quad (4.21)$$

I coefficienti (x_0, y_0, z_0) e (l, m, n) , (l', m', n') hanno il significato geometrico, rispettivamente, di coordinate di un punto P_0 sul piano e di due vettori linearmente indipendenti sul piano.

Analogamente al caso della retta in \mathbb{R}^2 , l'*equazione cartesiana* e l'*equazione parametrica* di un piano *non sono uniche*. Inoltre, è semplice passare dall'*equazione cartesiana* a quella *parametrica*, e viceversa. Ad esempio, nel primo caso (se $c \neq 0$) basta porre $x = s$, $y = t$ e calcolare il corrispondente valore di z dalla (4.19), mentre nel secondo è sufficiente eliminare la s e la t dalla (4.21).

La retta

L'*equazione cartesiana* di una retta r nello spazio si ottiene mettendo a sistema le equazioni cartesiane di due piani π_1 e π_2 tali che $r = \pi_1 \cap \pi_2$; segue che l'*equazione cartesiana* della retta nello spazio è del tipo

$$\begin{cases} ax + by + cz + d = 0 \\ a'x + b'y + c'z + d' = 0. \end{cases} \quad (4.22)$$

Poichè la retta r è ortogonale ad entrambi i vettori normali $\mathbf{u} = (a, b, c)$ e $\mathbf{u}' = (a', b', c')$ dei due piani π e π' , un *vettore direzionale* di r , ovvero un vettore parallelo a r , è dato da $\mathbf{v} = \mathbf{u} \wedge \mathbf{u}'$.

L'*equazione parametrica* della retta nello spazio si ottiene in modo analogo, ed ha proprietà analoghe, a quella della retta nel piano: data una retta r , un punto $P_0 = (x_0, y_0, z_0) \in r$ e un vettore direzionale $\mathbf{v} = (l, m, n)$ di r , l'*equazione parametrica* è

$$\begin{cases} x = x_0 + lt \\ y = y_0 + mt \\ z = z_0 + nt \end{cases} \quad t \in \mathbb{R}. \quad (4.23)$$

Osserviamo come l'equazione parametrica del piano, che è un luogo geometrico bidimensionale, sia determinata da *due parametri*, mentre per la retta, che è monodimensionale, ne basti *uno* soltanto.

Il passaggio da equazione cartesiana a equazione parametrica, e viceversa, si opera con le modalità usuali. Ad esempio, nel primo caso, se $\det \begin{pmatrix} b & c \\ b' & c' \end{pmatrix} \neq 0$, si pone $x = t$ e poi si calcolano i valori di y e z dalla (4.22). Nel secondo caso si elimina la t dalla (4.23); ad esempio, se $lmn \neq 0$ si ottiene la (4.22) nella forma

$$\frac{x - x_0}{l} = \frac{y - y_0}{m} = \frac{z - z_0}{n},$$

mentre se una o due delle coordinate l, m, n sono nulle si pone uguale a zero il numeratore corrispondente.

Problemi geometrici

Consideriamo alcuni *problemi geometrici* relativi a rette e piani nello spazio.

(a) *piano per tre punti*. Siano $P_i = (x_i, y_i, z_i)$, $i = 0, 1, 2$, tre punti *non allineati*. E' chiaro che $P = (x, y, z)$ appartiene al piano π individuato da P_0, P_1 e P_2 se e solo se $\mathbf{P}_0\mathbf{P}$, $\mathbf{P}_0\mathbf{P}_1$ e $\mathbf{P}_0\mathbf{P}_2$ sono complanari; per quanto visto nel paragrafo 4.1 abbiamo allora che l'equazione cartesiana di π è

$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \end{vmatrix} = 0.$$

L'equazione parametrica di π è fornita dalla (4.20).

(b) *rette e piani paralleli e ortogonali*. Siano π, π' due piani con vettori normali $\mathbf{u} = (a, b, c)$, $\mathbf{u}' = (a', b', c')$ rispettivamente, e r, r' due rette con vettori direzionali $\mathbf{v} = (l, m, n)$ e $\mathbf{v}' = (l', m', n')$ rispettivamente. E' chiaro che

$$\begin{aligned} \pi \text{ è parallelo a } \pi' &\Leftrightarrow \mathbf{u} \text{ è parallelo a } \mathbf{u}' \Leftrightarrow (a, b, c) = \lambda(a', b', c') \\ \pi \text{ è ortogonale a } \pi' &\Leftrightarrow \mathbf{u} \text{ è ortogonale a } \mathbf{u}' \Leftrightarrow aa' + bb' + cc' = 0 \\ r \text{ è parallela a } r' &\Leftrightarrow \mathbf{v} \text{ è parallelo a } \mathbf{v}' \Leftrightarrow (l, m, n) = \lambda(l', m', n') \\ r \text{ è ortogonale a } r' &\Leftrightarrow \mathbf{v} \text{ è ortogonale a } \mathbf{v}' \Leftrightarrow ll' + mm' + nn' = 0 \\ r \text{ è parallela a } \pi &\Leftrightarrow \mathbf{v} \text{ è ortogonale a } \mathbf{u} \Leftrightarrow al + bm + cn = 0 \\ r \text{ è ortogonale a } \pi &\Leftrightarrow \mathbf{v} \text{ è parallelo a } \mathbf{u} \Leftrightarrow (l, m, n) = \lambda(a, b, c). \end{aligned}$$

(c) *retta per due punti*. Siano $P_i = (x_i, y_i, z_i)$, $i = 0, 1$, due punti *distinti*. In analogia con il caso della retta nel piano, l'equazione parametrica della retta r passante per P_0 e P_1 è

$$\begin{cases} x = x_0 + t(x_1 - x_0) \\ y = y_0 + t(y_1 - y_0) \\ z = z_0 + t(z_1 - z_0) \end{cases} \quad t \in \mathbb{R},$$

mentre l'equazione cartesiana di r è

$$\frac{x - x_0}{x_1 - x_0} = \frac{y - y_0}{y_1 - y_0} = \frac{z - z_0}{z_1 - z_0}, \quad (4.24)$$

con l'usuale convenzione che se uno o due dei denominatori della (4.24) si annullano, allora si pongono uguali a zero i numeratori corrispondenti.

(d) *intersezione tra piani e rette*. L'intersezione tra piani e rette si può ottenere mettendo a sistema le equazioni cartesiane di tali piani e rette. In questo caso si hanno un massimo di ∞^2 soluzioni (intersezione tra due piani coincidenti) e un minimo di 0 soluzioni (intersezione di rette o piani paralleli e distinti e intersezione di rette *sghembe*, cioè di rette non parallele e non incidenti).

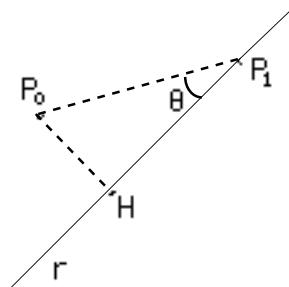
(e) *distanza punto-piano e punto-retta*. Dati un punto $P_0 = (x_0, y_0, z_0)$ e un piano π di equazione $ax + by + cz + d = 0$, la distanza $d(P_0, \pi)$ di P_0 da π si definisce e si calcola in completa analogia con il punto (f) del paragrafo 4.2, ottenendo

$$d(P_0, \pi) = \frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}}.$$

La distanza $d(P_0, r)$ di P_0 dalla retta r passante per $P_1 = (x_1, y_1, z_1)$ ed avente vettore direzionale $\mathbf{v} = (l, m, n)$ è chiaramente uguale alla lunghezza del segmento P_0H , vedi figura.

Dalla trigonometria abbiamo allora che

$$d(P_0, r) = |\mathbf{P}_0 \mathbf{P}_1| \sin \theta = \frac{|\mathbf{P}_0 \mathbf{P}_1 \wedge \mathbf{v}|}{|\mathbf{v}|}.$$



Osserviamo che $d(P_0, \pi)$ ($d(P_0, r)$) può essere calcolata anche costruendo la retta (il piano) passante per P_0 e ortogonale a π (a r), in analogia con il punto (f) del paragrafo 4.2.

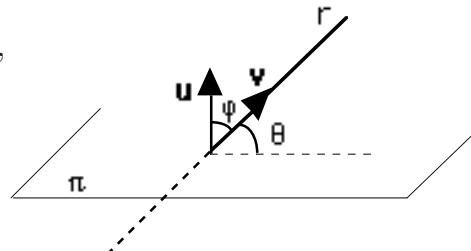
(f) *punto simmetrico rispetto a un piano, a una retta o a un punto*. Il problema è analogo al punto (g) del paragrafo 4.2. Dato un punto $P_0 = (x_0, y_0, z_0)$, il punto simmetrico $P_1 = (x_1, y_1, z_1)$ di P_0 rispetto al piano π (alla retta r o al punto M) si calcola imponendo che la proiezione H di P_0 su π (su r o M stesso) sia il punto medio di P_0 e P_1 . Nei primi due casi, le coordinate di H si calcolano facilmente come intersezione tra π (tra r) e la retta (il piano) passante per P_0 e ortogonale a π (a r).

(g) *angolo tra rette e piani*. L'angolo θ tra due rette (due piani) incidenti si definisce come il minore tra i due angoli formati dalle rette (dai piani). Per il calcolo di tale angolo θ si procede in analogia con il punto (d) del paragrafo 4.2, ottenendo

$$\cos \theta = \frac{|\mathbf{v} \cdot \mathbf{v}'|}{|\mathbf{v}| |\mathbf{v}'|}, \quad \sin \theta = \frac{|\mathbf{v} \wedge \mathbf{v}'|}{|\mathbf{v}| |\mathbf{v}'|} \quad \text{e} \quad \operatorname{tg} \theta = \frac{|\mathbf{v} \wedge \mathbf{v}'|}{|\mathbf{v} \cdot \mathbf{v}'|},$$

dove \mathbf{v} e \mathbf{v}' sono i vettori direzionali delle due rette (i vettori normali dei due piani).

L'angolo θ tra una retta r e un piano π incidenti è definito come in figura. Per calcolarlo



osserviamo che $\theta = \frac{\pi}{2} - \varphi$, dove φ è l'angolo tra il vettore normale \mathbf{u} di π e il vettore direzionale \mathbf{v} di r , e quindi

$$\cos \theta = \sin \varphi = \frac{|\mathbf{u} \wedge \mathbf{v}|}{|\mathbf{u}| |\mathbf{v}|}, \quad \sin \theta = \cos \varphi = \frac{|\mathbf{u} \cdot \mathbf{v}|}{|\mathbf{u}| |\mathbf{v}|} \quad \text{e} \quad \operatorname{tg} \theta = \frac{|\mathbf{u} \cdot \mathbf{v}|}{|\mathbf{u} \wedge \mathbf{v}|}.$$

(h) *fascio di piani*. Dati due piani *incidenti* π e π' , rispettivamente di equazioni $ax + by + cz + d = 0$ e $a'x + b'y + c'z + d' = 0$, il *fascio di piani* individuato da π e π' è formato da tutti i piani passanti per la retta intersezione di π e π' . Alternativamente, data una retta r (che possiamo pensare come intersezione di due piani π e π') il fascio di piani *incernierato* su r è formato da tutti i piani passanti per r . In entrambi i casi le equazioni di tutti i piani del fascio si ottengono da

$$\lambda(ax + by + cz + d) + \mu(a'x + b'y + c'z + d') = 0 \quad (4.25)$$

al variare di $\lambda, \mu \in \mathbb{R}$. Osserviamo che se π e π' sono *paralleli*, allora la (4.25) fornisce l'equazione di tutti i piani paralleli a π e π' . Osserviamo inoltre che la (4.25) rappresenta lo stesso piano se il rapporto $\frac{\lambda}{\mu}$ è costante.

(i) *proiezione di una retta su un piano*. Data una retta r e un piano π , la *proiezione (ortogonale)* r' di r su π è la retta intersezione tra π ed il piano passante per r ortogonale a π . Tale definizione consente il calcolo dell'equazione cartesiana di r' , usando il fascio di piani incernierato su r .

(l) *retta incidente ed ortogonale a due rette*. Se le due rette r e r' non sono sghembe, il problema della costruzione di una retta s incidente ed ortogonale a r e r' è di semplice risoluzione. Se r e r' sono *sghembe*, siano \mathbf{v} e \mathbf{v}' rispettivamente i vettori direzionali di r e r' . Chiaramente, la retta s avrà la direzione di $\mathbf{u} = \mathbf{v} \wedge \mathbf{v}'$. *Intuitivamente*, s si può ottenere nel modo seguente: si parte da una retta s' con vettore direzionale \mathbf{u} ed “appoggiata” a r , e poi si fa “scorrere” s' su r finché non incontra r' ; questa è la retta s cercata. L'equazione cartesiana di s si può calcolare nel modo seguente: siano

$\pi = \text{piano passante per } r \text{ con vettore normale ortogonale a } \mathbf{u}$

$\pi' = \text{piano passante per } r' \text{ con vettore normale ortogonale a } \mathbf{u}$;

allora s si ottiene come intersezione di π e π' . Ovviamente, le equazioni di π e π' si calcolano facendo uso dei fasci di piani incernierati su r e r' .

(m) *distanza di due rette*. Nuovamente se r e r' non sono sghembe la loro distanza $d(r, r')$ si calcola facilmente. Se r e r' sono *sghembe*, e \mathbf{v} e \mathbf{v}' sono i rispettivi vettori direzionali, la distanza $d(r, r')$ tra r e r' si definisce come la lunghezza del segmento i cui estremi sono le intersezioni della retta s , incidente e ortogonale a r e r' , con r e r' . Siano $P \in r$ e $P' \in r'$ due punti qualunque e $\mathbf{u} = \mathbf{v} \wedge \mathbf{v}'$; allora $d(r, r')$ non è altro che il modulo della proiezione di $\mathbf{P}\mathbf{P}'$ su \mathbf{u} , ovvero

$$d(r, r') = \frac{|\mathbf{P}\mathbf{P}' \cdot \mathbf{u}|}{|\mathbf{u}|}.$$

Se r e r' sono fornite mediante le loro equazioni parametriche, allora basta imporre che il vettore \mathbf{PP}' sia parallelo a \mathbf{u} , dove $P = P(t)$ e $P' = P'(t')$ denotano il generico punto su r e r' rispettivamente. Osserviamo che i valori di t e t' così trovati individuano i punti di intersezione della retta s con r e r' ; in questo modo si ottiene quindi anche un metodo alternativo per il calcolo dell'equazione di s , come retta passante per tali punti.

4.5. Quadriche in forma canonica

La sfera

La *sfera* è il luogo geometrico dei punti che hanno una data distanza r da un punto fissato C ; r e C sono rispettivamente il *raggio* e il *centro* della sfera. L'equazione della sfera è quindi

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2, \quad (4.26)$$

dove $C = (x_0, y_0, z_0)$. Sviluppando i quadrati vediamo che la (4.26) è del tipo

$$x^2 + y^2 + z^2 + ax + by + cz + d = 0, \quad (4.27)$$

e viceversa la (4.27) è l'equazione della sfera con

$$\text{centro } C = \left(-\frac{a}{2}, -\frac{b}{2}, -\frac{c}{2}\right) \text{ e raggio } r = \sqrt{\frac{a^2 + b^2 + c^2}{4} - d};$$

la (4.27) rappresenta quindi una sfera a *punti reali* se e solo se $a^2 + b^2 + c^2 > 4d$. Le (4.26) e (4.27) sono due forme dell'*equazione cartesiana* della sfera; osserviamo che, ancora una volta, moltiplicando per $\lambda \in \mathbb{R}$, $\lambda \neq 0$, i coefficienti della (4.27) si ottiene un'altra equazione della stessa sfera.

L'*equazione parametrica* della sfera di centro $C = (x_0, y_0, z_0)$ e raggio r è

$$\begin{cases} x = x_0 + r \cos \varphi \cos \theta \\ y = y_0 + r \sin \varphi \cos \theta \\ z = z_0 + r \sin \theta \end{cases} \quad \varphi, \theta \in \mathbb{R},$$

come si verifica usando le coordinate sferiche.

Consideriamo ora alcuni *problemi geometrici* riguardanti la sfera.

(a) *sfera per quattro punti*. E' ben noto che esiste una sola sfera passante per quattro punti *non complanari*. Il modo più semplice per calcolare l'equazione cartesiana della sfera passante per i quattro punti non complanari $P_i = (x_i, y_i, z_i)$, $i = 1, 2, 3, 4$, consiste nel sostituire le coordinate di tali punti nella (4.27) e risolvere il sistema di Cramer 4×4 nelle incognite a, b, c, d così ottenuto.

(b) *intersezione tra retta o piano e sfera*. Per calcolare i punti di intersezione tra una retta s (un piano π) e una sfera S basta mettere a sistema le equazioni cartesiane di s (di π) e di S . Se S ha centro C e raggio r , è chiaro che si hanno le possibilità seguenti: i) s (π)

interseca \mathcal{S} in due punti (lungo una circonferenza) se e solo se $d(C, s) < r$ ($d(C, \pi) < r$); *ii)* s (π) è *tangente* a \mathcal{S} , ovvero interseca \mathcal{S} in un solo punto, se e solo se $d(C, s) = r$ ($d(C, \pi) = r$); *iii)* s è *esterna* (π è *esterno*) a \mathcal{S} se e solo se $d(C, s) > r$ ($d(C, \pi) > r$).

(c) *rette e piani tangenti ad una sfera.* Sia \mathcal{S} la sfera di centro $C = (a, b, c)$ e raggio r e $P_0 = (x_0, y_0, z_0)$ un punto nello spazio. E' chiaro che

i) se $d(P_0, C) < r$ non esistono rette o piani passanti per P_0 e tangenti a \mathcal{S}

ii) se $d(P_0, C) = r$, ovvero $P_0 \in \mathcal{S}$, esistono un piano π e infinite rette passanti per P_0 e tangenti a \mathcal{S} , e tali rette stanno tutte sul piano π ; inoltre, il piano π e le rette tangenti sono caratterizzate dal passaggio per P_0 e dal fatto di essere ortogonali al vettore $\mathbf{P}_0\mathbf{C}$

iii) se $d(P_0, C) > r$ esistono infinite rette e infiniti piani passanti per P_0 e tangenti a \mathcal{S} ; tali rette s e tali piani π sono caratterizzati dal passaggio per P_0 e dal fatto che $d(C, s) = r$ e $d(C, \pi) = r$, oppure dal passaggio per P_0 e dal fatto che la loro intersezione con \mathcal{S} è ridotta ad un solo punto.

Osserviamo che l'equazione del piano π del punto *ii)* è

$$(x_0 - a)(x - x_0) + (y_0 - b)(y - y_0) + (z_0 - c)(z - z_0) = 0$$

o equivalentemente

$$\frac{\partial f}{\partial x}(x_0, y_0, z_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0, z_0)(y - y_0) + \frac{\partial f}{\partial z}(x_0, y_0, z_0)(z - z_0) = 0,$$

dove $f(x, y, z) = 0$ è l'equazione cartesiana della sfera \mathcal{S} .

Osserviamo infine che, data una retta s , non esistono piani passanti per s e tangenti a \mathcal{S} se s *interseca* \mathcal{S} in due punti (ovvero $d(C, s) < r$), c'è un solo piano passante per s e tangente a \mathcal{S} se s è *tangente* a \mathcal{S} (ovvero $d(C, s) = r$), mentre esistono due piani passanti per s tangenti a \mathcal{S} se s è *esterna* a \mathcal{S} (ovvero $d(C, s) > r$). In quest'ultimo caso tali due piani appartengono al fascio di piani incernierati su s . Il calcolo dell'equazione del o dei piani tangenti a \mathcal{S} e passanti per s si opera in base a principi geometrici del tutto simili a quelli dei punti *ii)* e *iii)* visti sopra.

(d) *circonferenze nello spazio.* Una circonferenza \mathcal{C} nello spazio è il luogo geometrico dei punti che hanno data distanza r da un punto fissato C e che appartengono a un piano π passante per C . Segue che la circonferenza \mathcal{C} può sempre essere rappresentata come intersezione del piano π con un'opportuna sfera, ad esempio la sfera di centro C e raggio r . Una retta s è tangente alla circonferenza \mathcal{C} se è contenuta nel piano π e interseca \mathcal{C} in un solo punto. I problemi geometrici relativi alla circonferenza nello spazio si possono risolvere in base alle proprietà geometriche appena descritte.

Quadratiche

Abbiamo visto che la sfera è il luogo geometrico dei punti le cui coordinate soddisfano una certa equazione di secondo grado. In generale, una *quadrifica* è il luogo dei punti le cui

coordinate soddisfano un'equazione del tipo

$$\Gamma(x, y, z) = ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + l = 0. \quad (4.28)$$

Come nel caso delle coniche, un'equazione del tipo (4.28) può non avere soluzioni reali, ad esempio $x^2 + y^2 + z^2 + 1 = 0$, ed in tal caso darà luogo ad una *quadrica immaginaria*, oppure può avere una sola soluzione reale, ad esempio $x^2 + 3y^2 + 2z^2 = 0$. Inoltre, un'equazione del tipo

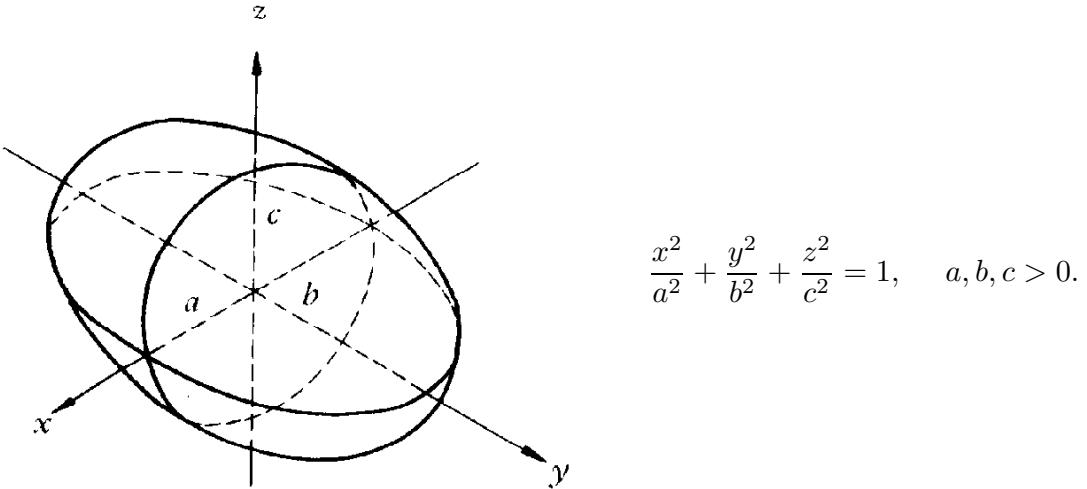
$$(ax + by + cz + d)(a'x + b'y + c'z + d') = 0$$

è ancora del tipo (4.28), e rappresenta l'unione dei due piani di equazione $ax + by + cz + d = 0$ e $a'x + b'y + c'z + d' = 0$.

Le quadriche verranno studiate in modo sistematico nel Capitolo 7; in questo paragrafo ci limitiamo alla descrizione delle principali proprietà delle quadriche fondamentali nella loro forma più semplice, ovvero delle quadriche *non degeneri* in *forma canonica*.

Ellisoidi

L'*ellissoide* in forma canonica si presenta ed ha equazione come in figura:



I tre segmenti in figura, di lunghezza rispettivamente a, b e c sono i *semiassi* dell'ellissoide. Non è difficile vedere che i *piani coordinati* (ovvero $x = 0, y = 0$ e $z = 0$) e gli *assi coordinati* (ovvero l'asse x, y e z) sono rispettivamente *piani di simmetria* e *assi di simmetria* dell'ellissoide; inoltre, l'origine è il *centro di simmetria*. E' chiaro che intersecando l'ellissoide con piani paralleli ai piani coordinati, ovvero del tipo $x = x_0, y = y_0$ e $z = z_0$, si ottengono delle *ellissi*, a punti reali se $|x_0| < a, |y_0| < b$ e $|z_0| < c$.

Se due dei semiassi sono uguali l'ellissoide diviene una *superficie di rotazione* attorno all'asse corrispondente al terzo semiasse, mentre se $a = b = c$ l'ellissoide si riduce ad una *sfera*.

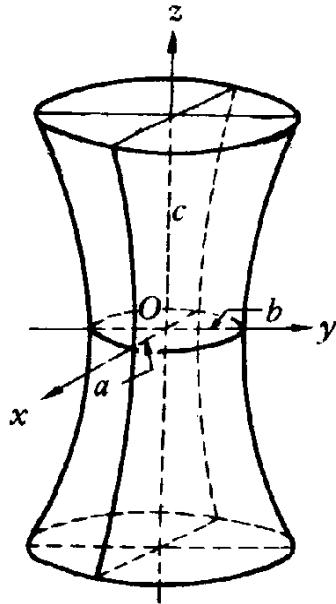
L'*ellissoide immaginario* in forma canonica ha equazione

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = -1, \quad a, b, c > 0$$

ed evidentemente non ha alcun punto reale.

Iperboloide a una falda

L'iperboloide a una falda in forma canonica si presenta ed ha ha equazione come in figura:

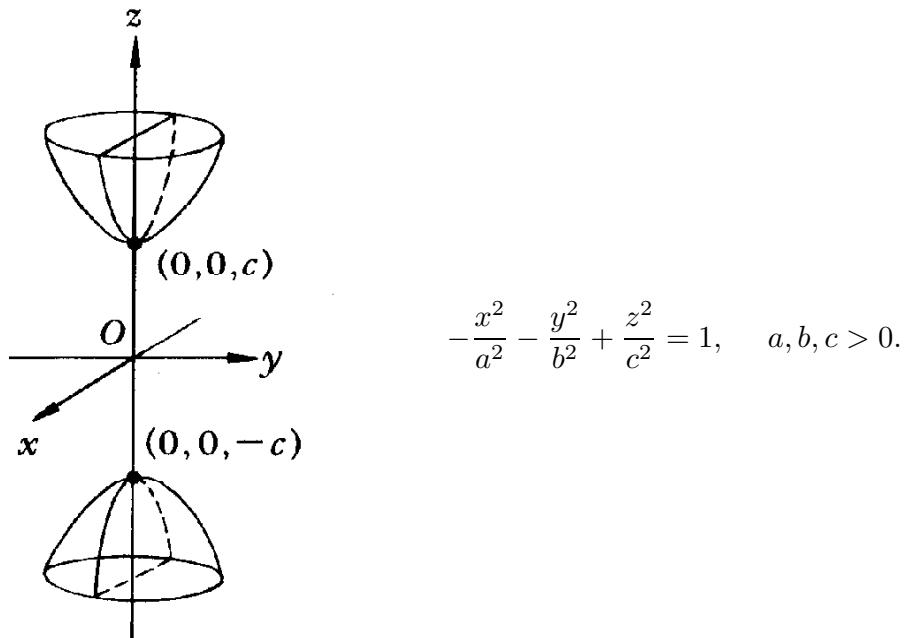


$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1, \quad a, b, c > 0.$$

Non è difficile vedere che i piani coordinati e l'origine O sono *elementi di simmetria* per l'iperboloide a una falda. Le intersezioni con i piani $x = x_0$ e $y = y_0$ sono *iperboli*, mentre quelle con i piani $z = z_0$ sono *ellissi*. Infine, se $a = b$ l'iperboloide a una falda diviene una *superficie di rotazione* attorno all'asse z .

Iperboloide a due falde

L'iperboloide a due falde in forma canonica si presenta ed ha equazione come in figura:

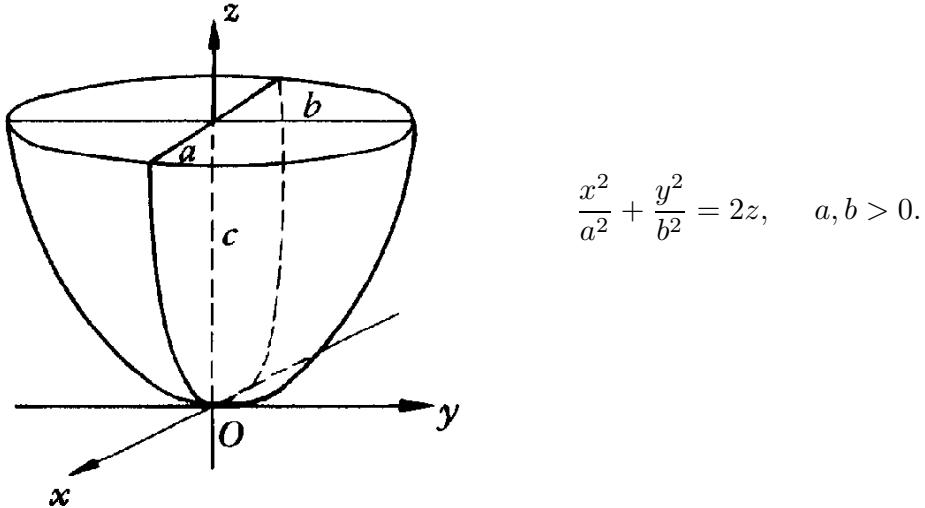


L'iperboloide a due falde ha gli stessi *elementi di simmetria* di quello a una falda. Le intersezioni con i piani $x = x_0$ e $y = y_0$ sono *iperboli*, mentre quelle con i piani $z = z_0$

sono *ellissi*, a punti reali se $|z_0| > c$. Infine, se $a = b$ l'iperboloido a due falde diviene una *superficie di rotazione* attorno all'asse z .

Paraboloide ellittico

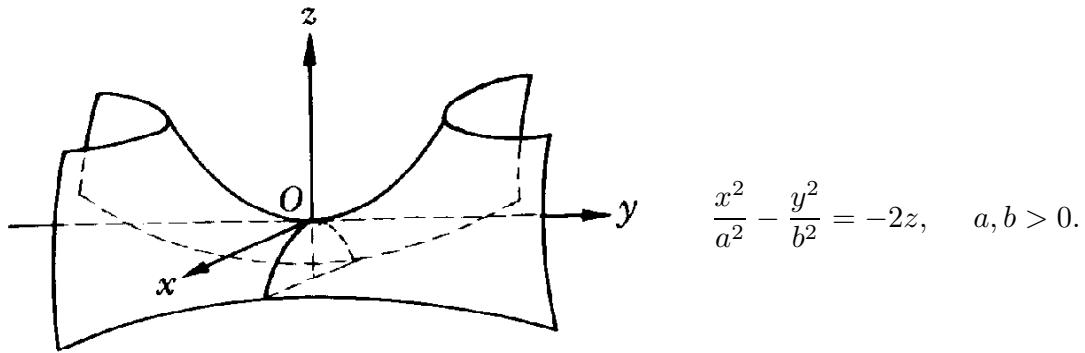
Il *paraboloide ellittico* in forma canonica si presenta ed ha equazione come in figura:



Si verifica che i piani $x = 0$ e $y = 0$ sono *piani di simmetria*, l'asse z è *asse di simmetria* e non c'è alcun centro di simmetria. Le intersezioni con i piani $x = x_0$ e $y = y_0$ sono *parbole*, mentre quelle con i piani $z = z_0$ sono *ellissi*, a punti reali se $z_0 > 0$. Infine, il paraboloide ellittico diviene una *superficie di rotazione* attorno all'asse z se $a = b$.

Paraboloide iperbolico

Il *paraboloide iperbolico* in forma canonica si presenta ed ha equazione come in figura:



Il paraboloide iperbolico ha gli stessi *elementi di simmetria* di quello ellittico. Le intersezioni con i piani $x = x_0$ e $y = y_0$ sono *parbole*, le prime rivolte verso l'alto e le seconde verso il basso, mentre le intersezioni con i piani $z = z_0$ sono *iperboli*; osserviamo che l'iperbole corrispondente a $z_0 = 0$ degenera nella coppia di rette $bx \pm ay = z = 0$. Infine, il paraboloide iperbolico non è mai una superficie di rotazione.

Osserviamo infine che *problemi geometrici* analoghi a quelli considerati nel caso della sfera possono essere posti e risolti anche nel caso delle altre quadriche sopra elencate; la loro risoluzione si ottiene con ragionamenti analoghi.

Capitolo 5. DIAGONALIZZAZIONE

5.1. Trasformazioni lineari e matrici diagonalizzabili

Definizioni

Sia V uno spazio vettoriale su \mathbb{R}/\mathbb{C} con $\dim V = n$ e $\varphi : V \rightarrow V$ una trasformazione lineare. L'utilità di avere una base B di V rispetto alla quale la matrice associata $M_\varphi^B = M_\varphi^{B,B}$ sia *diagonale* è evidente, ed in effetti talvolta si può scegliere opportunamente una base B tale che M_φ^B sia diagonale.

Esempio. Sia $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ definita da $\varphi(x, y) = (x + y, x - y)$. Consideriamo la base canonica K e la base $B = \langle (1, \sqrt{2} - 1), (1, -\sqrt{2} - 1) \rangle$. Si verifica facilmente che

$$M_\varphi^K = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{e} \quad M_\varphi^B = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & -\sqrt{2} \end{pmatrix}.$$

Una trasformazione lineare si dice *diagonalizzabile* se esiste una base B di V tale che $M_\varphi^B = \Delta$ sia diagonale. Se B' è un'altra base di V , dalla (3.10) abbiamo che

$$M_\varphi^{B'} = P^{-1} M_\varphi^B P, \quad (5.1)$$

dove P è la matrice di passaggio da B a B' .

In base alla (5.1) diciamo che due matrici quadrate A e B dello stesso ordine sono *simili* se esiste una matrice *invertibile* P tale che

$$A = P^{-1} B P \quad (\text{o equivalentemente } B = P A P^{-1}).$$

Diciamo inoltre che A è *diagonalizzabile* se A è simile a una matrice diagonale Δ , ovvero se

$$A = P^{-1} \Delta P \quad \text{oppure} \quad \Delta = P^{-1} A P \quad \text{con } \Delta \text{ diagonale e } P \text{ invertibile.}$$

Dalla (5.1) vediamo che tutte le matrici associate ad una stessa trasformazione lineare φ sono tra loro simili e, se φ è diagonalizzabile, tali matrici sono tutte diagonalizzabili. Viceversa, supponiamo che la matrice M_φ^B sia diagonalizzabile, ovvero

$$\Delta = P^{-1} M_\varphi^B P \quad \text{con } \Delta \text{ diagonale e } P \text{ invertibile.} \quad (5.2)$$

Per mezzo della base $B = \langle v_1, \dots, v_n \rangle$ e della matrice P possiamo costruire una nuova base B' di V definita da

$$B' = BP, \quad (5.3)$$

dove il significato di BP è descritto nel paragrafo 3.2. La matrice P è quindi la matrice di passaggio da B a B' e pertanto la (5.1) fornisce

$$M_\varphi^{B'} = P^{-1} M_\varphi^B P = P^{-1} (P \Delta P^{-1}) P = \Delta,$$

ovvero φ è diagonalizzabile. Abbiamo quindi dimostrato la

Proposizione 5.1. *Una trasformazione lineare $\varphi : V \rightarrow V$ è diagonalizzabile se e solo se la matrice M_φ^B associata a φ mediante una qualunque base B di V è diagonalizzabile.*

Osserviamo che le (5.2) e (5.3) forniscono un metodo per trovare una base di V rispetto alla quale la matrice associata ad una trasformazione lineare diagonalizzabile φ è diagonale.

Autovalori e autovettori

Dalla definizione di trasformazione lineare diagonalizzabile otteniamo facilmente che $\varphi : V \rightarrow V$ è diagonalizzabile se e solo se esistono una base $B = \langle v_1, \dots, v_n \rangle$ di V e $\lambda_1, \dots, \lambda_n \in \mathbb{R}/\mathbb{C}$ tali che

$$\varphi(v_1) = \lambda_1 v_1, \dots, \varphi(v_n) = \lambda_n v_n. \quad (5.4)$$

In tal caso è chiaro che $\lambda_1, \dots, \lambda_n$ formano la diagonale della matrice M_φ^B .

In base alla (5.4) diciamo che $\lambda \in \mathbb{R}/\mathbb{C}$ è un *autovalore* di $\varphi : V \rightarrow V$ se esiste $v \in V$, $v \neq 0$, tale che

$$\varphi(v) = \lambda v;$$

in tal caso v è un *autovettore* di φ associato all'autovalore λ . Osserviamo che, mentre un autovettore è per definizione sempre *non nullo*, un autovalore può essere nullo; è inoltre chiaro che

$$\lambda = 0 \text{ è autovalore di } \varphi \Leftrightarrow \dim \text{Ker} \varphi \geq 1.$$

In particolare, $\lambda = 0$ è autovalore di φ se e solo se φ non è iniettiva.

Possiamo quindi riformulare la (5.4) mediante la

Proposizione 5.2. *Una trasformazione lineare $\varphi : V \rightarrow V$ è diagonalizzabile se e solo se V ha una base costituita da autovettori di φ .*

Per mezzo del concetto di autovalore e autovettore nel seguito di questo capitolo daremo una soluzione ai seguenti problemi

(a) trovare un criterio per verificare se una trasformazione lineare $\varphi : V \rightarrow V$ è diagonalizzabile

(b) nel caso φ sia diagonalizzabile, costruire una base di V costituita da autovettori di φ .

Una problematica del tutto analoga si può porre nel caso delle matrici. Data una matrice $A \in M_{n \times n}(\mathbb{R}/\mathbb{C})$, il problema della diagonalizzabilità di A si può ricondurre a quello della diagonalizzabilità della trasformazione lineare $\varphi : \mathbb{R}^n/\mathbb{C}^n \rightarrow \mathbb{R}^n/\mathbb{C}^n$ associata ad A mediante le basi canoniche. In particolare, diremo quindi che $\lambda \in \mathbb{R}/\mathbb{C}$ è *autovalore* di A se esiste $x \in \mathbb{R}^n/\mathbb{C}^n$, $x \neq 0$, tale che

$$Ax = \lambda x,$$

ed in tal caso x è un *autovettore* di A associato a λ .

Osserviamo che i problemi (a) e (b) sopra esposti sono *equivalenti*, nel caso delle matrici, ai seguenti

- (a') trovare un criterio per verificare se una matrice quadrata A è diagonalizzabile
- (b') nel caso A sia diagonalizzabile, costruire una sua *diagonalizzazione*, ovvero una matrice diagonale Δ e una matrice invertibile P tali che $\Delta = P^{-1}AP$.

In definitiva, la diagonalizzabilità di una trasformazione lineare o di una matrice sono due aspetti dello stesso problema. Nel seguito formuleremo i risultati in termini di trasformazioni lineari, ed illustreremo il significato degli stessi nel caso delle matrici.

Infine, osserviamo esplicitamente che la teoria sviluppata in questo capitolo si applica a trasformazioni lineari φ di uno spazio vettoriale V in *se stesso*, ovvero $\varphi : V \rightarrow V$, ed alle matrici M_φ^B associate a φ mediante la *stessa base* B su dominio e codominio.

Polinomio caratteristico

Consideriamo per primo il problema della ricerca degli autovalori. Data $A \in M_{n \times n}(\mathbb{R}/\mathbb{C})$ consideriamo il *polinomio caratteristico* di A

$$P_A(x) = \det(A - xI),$$

dove I è la matrice identica di ordine n ; è chiaro che $\deg P_A(x) = n$. La prima proprietà di $P_A(x)$ è data dalla

Proposizione 5.3. *Se A e B sono simili allora $P_A(x) = P_B(x)$.*

Dimostrazione. Sia $A = P^{-1}BP$; chiaramente abbiamo che $I = P^{-1}IP$, quindi

$$A - xI = P^{-1}BP - xP^{-1}IP = P^{-1}(B - xI)P$$

da cui

$$\det(A - xI) = \det P^{-1} \det(B - xI) \det P = \det(B - xI)$$

e la Proposizione 5.3 è dimostrata. \square

La (5.1) e la Proposizione 5.3 consentono quindi di definire il *polinomio caratteristico* $P_\varphi(x)$ di una trasformazione lineare $\varphi : V \rightarrow V$, ponendo $P_\varphi(x) = P_M(x)$ dove M è la matrice associata a φ mediante una base qualunque di V . L'importanza del polinomio caratteristico dipende dal fatto che esso consente il calcolo degli autovalori. Abbiamo infatti

Teorema 5.1. *Sia V uno spazio vettoriale su \mathbb{R}/\mathbb{C} e $\varphi : V \rightarrow V$ una trasformazione lineare. Allora gli autovalori di φ sono tutte e sole le radici di $P_\varphi(x)$ in \mathbb{R}/\mathbb{C} .*

Dimostrazione. Sia B una base di V e $M = M_\varphi^B$. Allora $\lambda \in \mathbb{R}/\mathbb{C}$ è autovalore di φ se e solo se esiste $v \neq 0$ tale che $\varphi(v) - \lambda v = 0$. Ma $\varphi(v) - \lambda v = 0$ è equivalente a $Mv_B - \lambda v_B = 0$, ovvero $(M - \lambda I)v_B = 0$. Poiché $v \neq 0$ abbiamo che $v_B \neq 0$, quindi il sistema quadrato $(M - \lambda I)x = 0$ ha una soluzione non banale, ovvero $\det(M - \lambda I) = 0$. Abbiamo quindi che $(M - \lambda I)v_B = 0$ se e solo se $P_\varphi(\lambda) = \det(M - \lambda I) = 0$, ed il teorema è dimostrato. \square

Osserviamo che se V è uno spazio vettoriale su \mathbb{C} *tutte* le radici di $P_\varphi(x)$ sono autovalori di φ . Osserviamo inoltre che il Teorema 5.1, nel caso delle matrici, afferma che gli autovalori di $A \in M_{n \times n}(\mathbb{R}/\mathbb{C})$ sono tutte e sole le radici in \mathbb{R}/\mathbb{C} di $P_A(x)$.

Autospazi

Dato un autovalore λ della trasformazione lineare $\varphi : V \rightarrow V$ definiamo l'*autospazio* V_λ di φ associato a λ come

$$V_\lambda = \{v \in V : \varphi(v) = \lambda v\},$$

ovvero l'insieme degli autovettori associati a λ più il vettore nullo. Si verifica che V_λ è *sottospazio* di V ; inoltre V_λ si caratterizza come

$$V_\lambda = \text{Ker}(\varphi - \lambda \text{id}),$$

dove $\text{id} : V \rightarrow V$ è la trasformazione lineare identica.

Nel caso di una matrice $A \in M_{n \times n}(\mathbb{R}/\mathbb{C})$, l'autospazio V_λ associato all'autovalore λ è ovviamente

$$V_\lambda = \{x \in \mathbb{R}^n/\mathbb{C}^n : Ax = \lambda x\} = \{x \in \mathbb{R}^n/\mathbb{C}^n : (A - \lambda I)x = 0\} \quad (5.5)$$

ovvero il sottospazio di $\mathbb{R}^n/\mathbb{C}^n$ formato dalle soluzioni del sistema lineare $(A - \lambda I)x = 0$.

Osserviamo che la costruzione degli autospazi V_λ è molto semplice. Nel caso di autospazi di matrici ciò è immediato, poiché tali autospazi sono soluzioni di opportuni sistemi lineari omogenei; inoltre, dalla (5.5) e dal Teorema 3.10 abbiamo che

$$\dim V_\lambda = n - \rho(A - \lambda I). \quad (5.6)$$

Segue che per determinare una *base* di V_λ è sufficiente costruire $n - \rho(A - \lambda I)$ soluzioni linearmente indipendenti del sistema lineare $(A - \lambda I)x = 0$.

Nel caso di autospazi di trasformazioni lineari basta considerare la matrice M associata a φ rispetto a una qualunque base B di V e risolvere il sistema lineare $(M - \lambda I)x = 0$, dove I è la matrice identica di ordine $\dim V$. Tali soluzioni forniscono le *coordinate* rispetto alla base B dei vettori di V_λ , dalle quali si può immediatamente risalire ai vettori stessi. Anche in questo caso abbiamo

$$\dim V_\lambda = \dim V - \rho(M - \lambda I), \quad (5.7)$$

e quindi una *base* di V_λ si ottiene costruendo $\dim V - \rho(M - \lambda I)$ soluzioni linearmente indipendenti del sistema lineare $(M - \lambda I)x = 0$ e poi usando tali soluzioni come coordinate rispetto a B per ottenere la base di V_λ cercata.

Passiamo ora alla principale proprietà degli autospazi V_λ .

Teorema 5.2. *Siano $\lambda_1, \dots, \lambda_r$ gli autovalori distinti di una trasformazione lineare φ . Allora la somma degli autospazi $V_{\lambda_1}, \dots, V_{\lambda_r}$ è diretta.*

Dimostrazione. Siano B_1, \dots, B_r basi degli autospazi $V_{\lambda_1}, \dots, V_{\lambda_r}$ rispettivamente. Per il Teorema 3.6 è sufficiente dimostrare che dati comunque $v_1 \in B_1, \dots, v_r \in B_r$, i vettori v_1, \dots, v_r sono linearmente indipendenti. Procediamo per induzione su r . Se $r = 1$ il risultato è ovvio; supponiamo quindi che il risultato valga per $r - 1$ vettori, e sia

$$a_1v_1 + \dots + a_rv_r = 0. \quad (5.8)$$

Abbiamo quindi

$$0 = \varphi(a_1v_1 + \dots + a_rv_r) = a_1\varphi(v_1) + \dots + a_r\varphi(v_r) = a_1\lambda_1v_1 + \dots + a_r\lambda_rv_r. \quad (5.9)$$

Dalle (5.8) e (5.9) per sottrazione otteniamo

$$0 = a_1\lambda_1v_1 + \dots + a_r\lambda_rv_r - \lambda_r(a_1v_1 + \dots + a_rv_r) = a_1(\lambda_1 - \lambda_r)v_1 + \dots + a_{r-1}(\lambda_{r-1} - \lambda_r)v_{r-1}.$$

Per ipotesi gli autovalori $\lambda_1, \dots, \lambda_r$ sono distinti, e per l'ipotesi induttiva i vettori v_1, \dots, v_{r-1} sono linearmente indipendenti; segue che

$$a_1 = \dots = a_{r-1} = 0$$

e quindi la (5.8) diviene

$$a_rv_r = 0.$$

Ma $v_r \neq 0$ in quanto autovettore, da cui anche $a_r = 0$ e il teorema è dimostrato. \square

Dal Teorema 5.2 possiamo dedurre una prima condizione *sufficiente*, ma non necessaria, per la diagonalizzabilità di una trasformazione lineare.

Corollario 5.1. *Sia $\dim V = n$ e sia $\varphi : V \rightarrow V$ una trasformazione lineare con n autovalori distinti. Allora φ è diagonalizzabile.*

Dimostrazione. Dal Teorema 5.2 abbiamo che la somma di $V_{\lambda_1}, \dots, V_{\lambda_n}$ è diretta, dove $\lambda_1, \dots, \lambda_n$ sono gli n autovalori distinti. Ma $\dim V_{\lambda_j} \geq 1$ per ogni $j = 1, \dots, n$, quindi da una parte $\dim(V_{\lambda_1} \oplus \dots \oplus V_{\lambda_n}) \geq n$, mentre $\dim(V_{\lambda_1} \oplus \dots \oplus V_{\lambda_n}) \leq n$ in quanto $\dim V = n$. Segue che

$$\dim(V_{\lambda_1} \oplus \dots \oplus V_{\lambda_n}) = n,$$

da cui

$$V_{\lambda_1} \oplus \dots \oplus V_{\lambda_n} = V \quad (5.10)$$

per la Proposizione 3.3. Dalla (5.10) abbiamo che V possiede una base di autovettori di φ , e quindi il risultato segue dalla Proposizione 5.2. \square

Nel caso di una matrice $A \in M_{n \times n}(\mathbb{R}/\mathbb{C})$ il Teorema 5.2 afferma che la somma degli autospazi di A è diretta, e il Corollario 5.1 stabilisce che A è certamente diagonalizzabile se ha n autovalori distinti.

5.2. Diagonalizzazione

Criterio di diagonalizzabilità

A questo punto abbiamo gli strumenti per la risoluzione dei problemi (a) e (b), e quindi anche (a') e (b'), sopra esposti. Infatti, grazie al Teorema 5.1 possiamo calcolare gli autovalori di una trasformazione lineare $\varphi : V \rightarrow V$ (di una matrice quadrata A), e la Proposizione 5.2 e il Teorema 5.2 mostrano che φ (che A) è diagonalizzabile se e solo se

$$V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_r} = V \quad (= \mathbb{R}^n / \mathbb{C}^n), \quad (5.11)$$

dove $\lambda_1, \dots, \lambda_r$ sono gli autovalori *distinti* di φ (di A).

Dalla (5.11) e dal Teorema 3.6 otteniamo immediatamente il

Criterio di diagonalizzabilità. *Siano $\dim V = n$ e $\varphi : V \rightarrow V$ una trasformazione lineare (oppure sia A una matrice quadrata di ordine n), e siano $\lambda_1, \dots, \lambda_r$ gli autovalori distinti di φ (oppure di A). Allora φ (oppure A) è diagonalizzabile se e solo se*

$$\dim V_{\lambda_1} + \cdots + \dim V_{\lambda_r} = n. \quad (5.12)$$

Osserviamo che le dimensioni degli autospazi V_{λ_j} , $j = 1, \dots, r$, sono facilmente calcolabili mediante le formule (5.6) e (5.7); ad esempio, tali formule mostrano che la (5.12) è equivalente a

$$\sum_{j=1}^r \rho_j = n(r - 1),$$

dove $\rho_j = \rho(M - \lambda_j I)$ nel caso di una trasformazione lineare φ con matrice associata M rispetto ad una base qualunque, oppure $\rho_j = \rho(A - \lambda_j I)$ nel caso di una matrice A .

Il criterio di diagonalizzabilità risolve quindi i problemi (a) e (a'). Enunciamo ora, senza riportarne la giustificazione (peraltro semplice), una *variante* del criterio di diagonalizzabilità. Denotando con μ_j la *molteplicità* dell'autovalore λ_j come radice del polinomio caratteristico $P_\varphi(x)$ (oppure $P_A(x)$), vale il seguente

Teorema 5.3. $1 \leq \dim V_{\lambda_j} \leq \mu_j$ per ogni $j = 1, \dots, r$.

Poichè la somma delle molteplicità di tutte le radici distinte di un polinomio è uguale al grado del polinomio stesso, è chiaro che se V è uno spazio vettoriale su \mathbb{R} (oppure $A \in M_{n \times n}(\mathbb{R})$) e $P_\varphi(x)$ (oppure $P_A(x)$) ha almeno una radice *non reale* allora

$$\mu_1 + \cdots + \mu_r < n$$

e quindi per il Teorema 5.3

$$\dim V_{\lambda_1} + \cdots + \dim V_{\lambda_r} \leq \mu_1 + \cdots + \mu_r < n,$$

ovvero φ (oppure A) *non è diagonalizzabile* poichè la (5.12) non è verificata. Questa osservazione consente talvolta di concludere che φ (oppure A) non è diagonalizzabile *senza* dover calcolare le dimensioni degli autospazi.

Esempio. Sia $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Allora $P_A(x) = x^2 + 1$, le cui radici sono $\pm i$. Per il Corollario 5.1 abbiamo quindi che A è diagonalizzabile come matrice *complessa*, mentre A non è diagonalizzabile come matrice *reale*. In altre parole, se consideriamo $A \in M_{2 \times 2}(\mathbb{C})$ allora A è diagonalizzabile ed è simile, ad esempio, alla matrice $\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$, mentre se consideriamo $A \in M_{2 \times 2}(\mathbb{R})$ allora A non è diagonalizzabile.

Ragionamenti del tutto analoghi mostrano l'importanza del fatto che lo spazio vettoriale V sia su \mathbb{R} o su \mathbb{C} nel contesto della diagonalizzabilità di una trasformazione lineare $\varphi : V \rightarrow V$.

Osserviamo infine che il Teorema 5.3 porta alla seguente *variante del criterio di diagonalizzabilità*: sia V uno spazio vettoriale su \mathbb{R}/\mathbb{C} e sia $\varphi : V \rightarrow V$ una trasformazione lineare (oppure $A \in M_{n \times n}(\mathbb{R}/\mathbb{C})$) e siano $\lambda_1, \dots, \lambda_r$ gli autovalori *distinti* di φ (oppure di A). Allora φ (oppure A) è diagonalizzabile se e solo se

- i) tutte le radici di $P_\varphi(x)$ (oppure di $P_A(x)$) sono in \mathbb{R}/\mathbb{C}
- ii) $\dim V_{\lambda_j} = \mu_j$ per ogni $j = 1, \dots, r$.

Metodo di diagonalizzazione

Affrontiamo ora i problemi (b) e (b'). Posto $\dim V = n$, vediamo per prima cosa un algoritmo per la costruzione esplicita di una base di autovettori di una trasformazione lineare $\varphi : V \rightarrow V$ diagonalizzabile, anche se questo è già implicito in quanto visto nel paragrafo precedente. L'algoritmo è il seguente:

- i) mediante la matrice M associata a φ rispetto ad una qualunque base B di V calcolare $P_\varphi(x)$ e quindi gli autovalori *distinti* $\lambda_1, \dots, \lambda_r$ di φ
- ii) calcolare $\dim V_{\lambda_j}$, $j = 1, \dots, r$, mediante la (5.7)
- iii) per ogni $j = 1, \dots, r$ calcolare una base del sottospazio formato dalle soluzioni del sistema omogeneo $(M - \lambda_j I)x = 0$, ovvero calcolare $\dim V_j$ soluzioni linearmente indipendenti di tale sistema lineare
- iv) per ogni $j = 1, \dots, r$ usare le n -uple così trovate come coordinate rispetto alla base B per costruire la base di autovettori $\langle w_1, \dots, w_n \rangle$ cercata.

E' ovvio che la matrice associta a φ rispetto alla base di autovettori $\langle w_1, \dots, w_n \rangle$ è diagonale. Osserviamo che è importante ricordare l'*ordine* con cui sono stati ottenuti i vettori w_1, \dots, w_n . Precisamente, la matrice diagonale associata a φ mediante la base $\langle w_1, \dots, w_n \rangle$ avrà al posto i -esimo sulla diagonale l'autovalore corrispondente all'autovettore w_i ; in altre parole, al posto i -esimo su tale diagonale si troverà l'autovalore λ per mezzo del quale l'autovettore w_i è stato calcolato al punto iv).

Quanto sopra mostra che non vi è un'unica matrice diagonale associata a φ ; più precisamente, data una base di autovettori $\langle w_1, \dots, w_n \rangle$ di φ si possono costruire altre basi di autovettori *permutando* i vettori w_1, \dots, w_n , e le relative matrici diagonali associate si otterranno permutando conseguentemente gli autovalori posti sulla diagonale.

La diagonalizzazione esplicita di una matrice $A \in M_{n \times n}(\mathbb{R}/\mathbb{C})$ si ottiene in modo del tutto analogo a quanto visto sopra. L'algoritmo è il seguente:

i) calcolare $P_A(x)$ e quindi gli autovalori *distinti* $\lambda_1, \dots, \lambda_r$ di A

ii) calcolare $\rho_j = n - \rho(A - \lambda_j I)$ per ogni $j = 1, \dots, r$

iii) per ogni $j = 1, \dots, r$ calcolare una base delle soluzioni del sistema lineare omogeneo $(A - \lambda_j I)x = 0$, ovvero ρ_j soluzioni linearmente indipendenti di tale sistema lineare

iv) ordinare consecutivamente le n -uple trovate al punto precedente, partendo da quelle corrispondenti all'autovalore λ_1 e così via. Tali n -uple formano una base B di $\mathbb{R}^n/\mathbb{C}^n$. La matrice diagonale Δ cercata ha gli autovalori $\lambda_1, \dots, \lambda_r$ sulla diagonale, ciascuno ripetuto ρ_j volte, partendo dall'autovalore λ_1 e così via. La matrice P tale che $\Delta = P^{-1}AP$ è quindi la *matrice di passaggio* dalla base canonica K alla base B e pertanto, per quanto visto nel paragrafo 3.2, le colonne di P sono le n -uple di cui sopra, nell'ordine considerato.

In altre parole, se $\mathbf{x}_1, \dots, \mathbf{x}_n$ sono i vettori colonna calcolati al punto *iii)* presi nell'ordine descritto al punto *iv*), abbiamo

$$P = (\mathbf{x}_1, \dots, \mathbf{x}_n).$$

Osserviamo ancora una volta che permutando l'ordine degli autovalori $\lambda_1, \dots, \lambda_r$ occorre permutare conseguentemente la loro posizione sulla diagonale di Δ ed anche le colonne di P . Segue quindi che la diagonalizzazione di una matrice A *non è unica*, ma è unica a meno di permutazioni dei suoi autovalori distinti $\lambda_1, \dots, \lambda_r$.

Tornando all'esempio sopra esposto abbiamo quindi che $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ è simile sia a $\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$ che a $\begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}$, e le matrici P relative alle due matrici diagonali in questione hanno le colonne tra loro scambiate.

Esempio. La matrice $\begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ non è diagonalizzabile su \mathbb{R} e su \mathbb{C} . Diagonalizzare su \mathbb{R} la matrice $\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$.

Capitolo 6. PRODOTTO SCALARE E FORME QUADRATICHE

6.1. Prodotto scalare

Prodotto scalare

Nel Capitolo 4 abbiamo visto che il prodotto scalare di due vettori geometrici $\mathbf{u} = (a, b, c)$ e $\mathbf{v} = (a', b', c')$ è $\mathbf{u} \cdot \mathbf{v} = aa' + bb' + cc'$. La nozione di prodotto scalare si può estendere agli spazi vettoriali \mathbb{R}^n e \mathbb{C}^n . Dati due vettori $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, pensati come vettori colonna $\mathbf{x} = (x_1, \dots, x_n)^T$ e $\mathbf{y} = (y_1, \dots, y_n)^T$, il loro *prodotto scalare (euclideo)* è

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i = \mathbf{x}^T \mathbf{y} \quad (\text{prodotto di matrici}).$$

Si verifica facilmente che valgono le proprietà seguenti

- i) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ (*simmetria*)
- ii) $\langle a\mathbf{x} + b\mathbf{y}, \mathbf{z} \rangle = a \langle \mathbf{x}, \mathbf{z} \rangle + b \langle \mathbf{y}, \mathbf{z} \rangle$ (*linearità*)
- iii) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ per ogni $\mathbf{x} \in \mathbb{R}^n$ e $\langle \mathbf{x}, \mathbf{x} \rangle = 0 \Leftrightarrow \mathbf{x} = 0$ (*positività*).

Nel caso di \mathbb{C}^n (come spazio vettoriale su \mathbb{C}) si pone una definizione simile a quella di \mathbb{R}^n , ed in effetti le due definizioni coincidono per vettori di \mathbb{C}^n a coordinate reali; dati $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ definiamo

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i \overline{y_i} = \mathbf{x}^T \overline{\mathbf{y}}.$$

Osserviamo che la i) diviene in questo caso

$$i') \langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle} \quad (\text{antisimmetria})$$

mentre le altre proprietà rimangono invariate, la iii) grazie alla presenza del *coniugato*.

Grazie alla iii) definiamo la *norma (indotta dal prodotto scalare)*

$$\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{\frac{1}{2}}$$

di un vettore $\mathbf{x} \in \mathbb{R}^n / \mathbb{C}^n$. Dalla iii) abbiamo che $\|\mathbf{x}\| \geq 0$ e, alla luce del caso $n = 3$, la norma di $\mathbf{x} - \mathbf{y}$ ha il significato di *distanza* tra \mathbf{x} e \mathbf{y} :

$$\|\mathbf{x} - \mathbf{y}\| = \text{distanza tra } \mathbf{x} \text{ e } \mathbf{y}.$$

Proprietà

Prodotto scalare e norma soddisfano tre importanti proprietà.

I. Diseguaglianza di Cauchy-Schwarz

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\| \quad \text{per ogni } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Dimostrazione. Se $\mathbf{x} = 0$ o $\mathbf{y} = 0$ la diseguaglianza è ovvia. Supponiamo $\mathbf{y} \neq 0$ e consideriamo $\|\mathbf{x} + \lambda \mathbf{y}\|^2$ con $\lambda \in \mathbb{R}$; abbiamo

$$0 \leq \|\mathbf{x} + \lambda \mathbf{y}\|^2 = \langle \mathbf{x} + \lambda \mathbf{y}, \mathbf{x} + \lambda \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle + 2\lambda \langle \mathbf{x}, \mathbf{y} \rangle + \lambda^2 \langle \mathbf{y}, \mathbf{y} \rangle$$

$$= \|\mathbf{x}\|^2 + 2\lambda <\mathbf{x}, \mathbf{y}> + \lambda^2 \|\mathbf{y}\|^2.$$

Abbiamo quindi che il polinomio di secondo grado

$$\lambda^2 \|\mathbf{y}\|^2 + 2\lambda <\mathbf{x}, \mathbf{y}> + \|\mathbf{x}\|^2$$

è sempre ≥ 0 ; segue che il suo discriminante è ≤ 0 , ovvero

$$<\mathbf{x}, \mathbf{y}>^2 - \|\mathbf{x}\|^2 \|\mathbf{y}\|^2 \leq 0$$

e la diseguaglianza segue. \square

II. Diseguaglianza triangolare

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \quad \text{per ogni } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Dimostrazione. Dalla diseguaglianza di Cauchy-Schwarz abbiamo

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &= <\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}>^2 = <\mathbf{x}, \mathbf{x}>^2 + 2 <\mathbf{x}, \mathbf{y}> + <\mathbf{y}, \mathbf{y}>^2 \\ &\leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 \end{aligned}$$

e la diseguaglianza segue. \square

La diseguaglianza triangolare prende il nome dalla ben nota diseguaglianza tra lati di un triangolo, nel caso $n = 3$.

III. Formula di aggiunzione

$$< A\mathbf{x}, \mathbf{y}> = <\mathbf{x}, A^T \mathbf{y}> \quad \text{per } \mathbf{x} \in \mathbb{C}^n, \mathbf{y} \in \mathbb{C}^m \text{ e } A \in M_{m \times n}(\mathbb{R}).$$

Dimostrazione. Abbiamo

$$< A\mathbf{x}, \mathbf{y}> = (A\mathbf{x})^T \bar{\mathbf{y}} = \mathbf{x}^T (A^T \bar{\mathbf{y}}) = <\mathbf{x}, A^T \mathbf{y}>.$$

\square

Norma

Abbiamo visto che il prodotto scalare *induce* una norma; è utile introdurre un concetto di norma *indipendentemente* dal prodotto scalare. Definiamo pertanto una *norma* su $\mathbb{R}^n / \mathbb{C}^n$ mediante le seguenti proprietà:

- i) $\|\mathbf{x}\| \geq 0$ e $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = 0$
- ii) $\|\lambda \mathbf{x}\| = |\lambda| \|\mathbf{x}\|$
- iii) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

Analogamente a quanto accade per la norma indotta dal prodotto scalare, la quantità $\|\mathbf{x} - \mathbf{y}\|$ ha il significato di *distanza* tra \mathbf{x} e \mathbf{y} . Concludiamo il paragrafo con alcuni esempi notevoli di norme su $\mathbb{R}^n / \mathbb{C}^n$:

- i) $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$, dove $\mathbf{x} = (x_1, \dots, x_n)^T$
- ii) $\|\mathbf{x}\|_2$ = norma indotta dal prodotto scalare
- iii) $\|\mathbf{x}\|_\infty = \max_{i=1, \dots, n} |x_i|$, dove $\mathbf{x} = (x_1, \dots, x_n)^T$.

Si verifica facilmente che si tratta effettivamente di norme in tutti e tre i casi.

6.2. Ortogonalità

Ortogonalità

Così come accade in \mathbb{R}^3 , il prodotto scalare consente di introdurre il concetto di ortogonalità tra vettori di $\mathbb{R}^n/\mathbb{C}^n$ mediante

$$\mathbf{x} \perp \mathbf{y} \Leftrightarrow \langle \mathbf{x}, \mathbf{y} \rangle = 0.$$

Osserviamo anzitutto che l'ortogonalità è una forma forte di indipendenza lineare.

Proposizione 6.1. *Se $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^n/\mathbb{C}^n$ sono vettori non nulli a due a due ortogonali allora $\mathbf{x}_1, \dots, \mathbf{x}_k$ sono linearmente indipendenti.*

Dimostrazione. Supponiamo che $\lambda_1 \mathbf{x}_1 + \dots + \lambda_k \mathbf{x}_k = 0$; allora per ogni $i = 1, \dots, k$

$$0 = \langle \lambda_1 \mathbf{x}_1 + \dots + \lambda_k \mathbf{x}_k, \mathbf{x}_i \rangle = \lambda_1 \langle \mathbf{x}_1, \mathbf{x}_i \rangle + \dots + \lambda_k \langle \mathbf{x}_k, \mathbf{x}_i \rangle = \lambda_i \|\mathbf{x}_i\|_2^2.$$

Segue che $\lambda_i = 0$ per $i = 1, \dots, k$ e la Proposizione 6.1 è dimostrata. \square

Un esempio importante di vettori ortogonali è dato dalla base canonica K di $\mathbb{R}^n/\mathbb{C}^n$; infatti, si verifica facilmente che

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Abbiamo quindi che i vettori della base canonica $\langle \mathbf{e}_1, \dots, \mathbf{e}_n \rangle$ sono a due a due ortogonali e hanno norma = 1. In generale, diciamo che una base $B = \langle \mathbf{x}_1, \dots, \mathbf{x}_n \rangle$ di $\mathbb{R}^n/\mathbb{C}^n$ è una *base ortonormale* se $\mathbf{x}_i \perp \mathbf{x}_j$ per $i \neq j$ e $\|\mathbf{x}_i\|_2 = 1$ per $i = 1, \dots, n$. Abbiamo quindi che la base canonica K è ortonormale.

Osserviamo che il *calcolo delle coordinate* rispetto ad una *base ortonormale* è molto semplice. Sia infatti $B = \langle \mathbf{x}_1, \dots, \mathbf{x}_n \rangle$ una base ortonormale di $\mathbb{R}^n/\mathbb{C}^n$; abbiamo allora

$$\begin{aligned} \langle \mathbf{x}, \mathbf{x}_i \rangle &= \langle a_1 \mathbf{x}_1 + \dots + a_n \mathbf{x}_n, \mathbf{x}_i \rangle \\ &= a_1 \langle \mathbf{x}_1, \mathbf{x}_i \rangle + \dots + a_n \langle \mathbf{x}_n, \mathbf{x}_i \rangle = a_i \langle \mathbf{x}_i, \mathbf{x}_i \rangle = a_i. \end{aligned}$$

In altre parole

$$a_i = \langle \mathbf{x}, \mathbf{x}_i \rangle, \quad i = 1, \dots, n.$$

Algoritmo di ortonormalizzazione di Gram-Schmidt

Data l'utilità delle basi ortonormali ci si può chiedere come fare per costruirle. Il seguente algoritmo, detto di *ortonormalizzazione di Gram-Schmidt*, consente di trasformare una

base qualunque di $\mathbb{R}^n/\mathbb{C}^n$ in una base ortonormale. Sia $B = \langle \mathbf{x}_1, \dots, \mathbf{x}_n \rangle$ una base di $\mathbb{R}^n/\mathbb{C}^n$; l'algoritmo procede nel modo seguente: iniziamo ponendo

$$\mathbf{y}_1 = \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|_2}, \quad \text{quindi} \quad \|\mathbf{y}_1\|_2 = 1.$$

Poniamo poi

$$\mathbf{y}'_2 = \mathbf{x}_2 - \langle \mathbf{x}_2, \mathbf{y}_1 \rangle \mathbf{y}_1 \quad \text{e} \quad \mathbf{y}_2 = \frac{\mathbf{y}'_2}{\|\mathbf{y}'_2\|_2};$$

si verifica facilmente che $\mathbf{y}_2 \perp \mathbf{y}_1$ e $\|\mathbf{y}_2\|_2 = 1$. Proseguiamo ponendo

$$\mathbf{y}'_3 = \mathbf{x}_3 - \langle \mathbf{x}_3, \mathbf{y}_1 \rangle \mathbf{y}_1 - \langle \mathbf{x}_3, \mathbf{y}_2 \rangle \mathbf{y}_2 \quad \text{e} \quad \mathbf{y}_3 = \frac{\mathbf{y}'_3}{\|\mathbf{y}'_3\|_2}$$

e nuovamente si verifica facilmente che $\mathbf{y}_3 \perp \mathbf{y}_2$, $\mathbf{y}_3 \perp \mathbf{y}_1$ e $\|\mathbf{y}_3\|_2 = 1$.

Procedendo iterativamente in questo modo otteniamo una *base ortonormale*

$$E = \langle \mathbf{y}_1, \dots, \mathbf{y}_n \rangle$$

di $\mathbb{R}^n/\mathbb{C}^n$, come si voleva.

Esempio. Ortonormalizzare la base $\langle (1, 2), (-1, 1) \rangle$ di \mathbb{R}^2 ; applicando l'algoritmo si trova $\mathbf{y}_1 = (\frac{\sqrt{5}}{5}, \frac{2\sqrt{5}}{5})$, $\mathbf{y}_2 = (-\frac{2\sqrt{5}}{5}, \frac{\sqrt{5}}{5})$.

Ortogonalità tra sottospazi

Il concetto di ortogonalità tra vettori si può estendere ai sottospazi di $\mathbb{R}^n/\mathbb{C}^n$. Siano W_1 e W_2 sottospazi di $\mathbb{R}^n/\mathbb{C}^n$; diciamo che W_1 e W_2 sono *ortogonali*, e scriviamo $W_1 \perp W_2$, se

$$\mathbf{x} \perp \mathbf{y} \quad \text{per ogni } \mathbf{x} \in W_1 \text{ e } \mathbf{y} \in W_2.$$

Inoltre, dato un sottospazio W di $\mathbb{R}^n/\mathbb{C}^n$ definiamo l'*ortogonale* W^\perp di W come

$$W^\perp = \{ \mathbf{y} \in \mathbb{R}^n/\mathbb{C}^n : \langle \mathbf{y}, \mathbf{x} \rangle = 0 \quad \text{per ogni } \mathbf{x} \in W \};$$

si verifica che W^\perp è un *sottospazio* di $\mathbb{R}^n/\mathbb{C}^n$.

L'importanza dell'ortogonale risiede nelle due proprietà seguenti, la cui dimostrazione viene omessa:

- i) $(W^\perp)^\perp = W$
- ii) $\mathbb{R}^n/\mathbb{C}^n = W \oplus W^\perp$.

In particolare, la seconda proprietà sarà utile in seguito.

6.3. Diagonalizzazione delle matrici simmetriche reali

Matrici ortogonali e simmetriche

Abbiamo già introdotto le matrici ortogonali nel Capitolo 4: $A \in M_{n \times n}(\mathbb{R})$ è *ortogonale* se $A^{-1} = A^T$. Osserviamo che se A è ortogonale allora $\det A = \pm 1$ in quanto

$$1 = \det I = \det(AA^{-1}) = \det(AA^T) = \det A \det A^T = (\det A)^2.$$

Inoltre, si verifica facilmente che

$$A, B \in M_{n \times n}(\mathbb{R}) \text{ ortogonali} \Rightarrow AB \text{ ortogonale.}$$

Le matrici ortogonali possono essere caratterizzate come segue.

Proposizione 6.2. *$A \in M_{n \times n}(\mathbb{R})$ è ortogonale se e solo se le sue colonne (o le sue righe) formano una base ortonormale di \mathbb{R}^n .*

Dimostrazione. Analizziamo soltanto il caso delle colonne. Siano $\mathbf{x}_1, \dots, \mathbf{x}_n$ le colonne di A e sia $A^T A = (a_{ij})$. Poichè la riga i -esima di A^T non è altro che la colonna i -esima di A , abbiamo

$$a_{ij} = \mathbf{x}_i^T \mathbf{x}_j = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$

e quindi $A^T A = I$ se e solo se $\mathbf{x}_1, \dots, \mathbf{x}_n$ sono n vettori di norma 1 a due a due ortogonali, ovvero le colonne di A formano una base ortonormale di \mathbb{R}^n . \square

Dalla Proposizione 6.2 segue facilmente che le matrici ortogonali possono essere caratterizzate anche come le *matrici di passaggio* tra basi ortonormali di \mathbb{R}^n .

Una matrice $A \in M_{n \times n}(\mathbb{R})$ è *simmetrica* se $A = A^T$, ovvero se i suoi coefficienti a_{ij} sono *simmetrici* rispetto alla diagonale, $a_{ij} = a_{ji}$. È facile costruire una matrice simmetrica a partire da una matrice qualunque; infatti, data $A \in M_{m \times n}(\mathbb{R})$ le matrici $A^T A$ e AA^T sono simmetriche in quanto

$$(A^T A)^T = A^T (A^T)^T = A^T A$$

ed analogamente per AA^T . Osserviamo anche che se A è una matrice simmetrica di ordine n e B è una *qualsiasi* matrice quadrata di ordine n , allora $B^T AB$ è ancora *simmetrica*; infatti

$$(B^T AB)^T = B^T A(B^T)^T = B^T AB. \quad (6.1)$$

Diagonalizzazione delle matrici simmetriche reali

Dal criterio di diagonalizzabilità abbiamo che condizione *necessaria* affinchè una matrice quadrata reale sia diagonalizzabile è che il suo polinomio caratteristico abbia radici tutte reali. Nel caso delle matrici simmetriche abbiamo la

Proposizione 6.3. *Sia $A \in M_{n \times n}(\mathbb{R})$ una matrice simmetrica. Allora tutte le radici di $P_A(x)$ sono reali.*

Dimostrazione. Consideriamo A come matrice *complessa*; le radici di $P_A(x)$ sono allora tutte quante autovalori di A . Sia λ un tale autovalore e $\mathbf{x} \in \mathbb{C}^n$ un suo autovettore; dalla formula di aggiunzione abbiamo allora

$$\lambda \|\mathbf{x}\|_2^2 = \lambda \langle \mathbf{x}, \mathbf{x} \rangle = \langle \lambda \mathbf{x}, \mathbf{x} \rangle = \langle A\mathbf{x}, \mathbf{x} \rangle = \langle \mathbf{x}, A^T \mathbf{x} \rangle$$

$$= \langle \mathbf{x}, A\mathbf{x} \rangle = \langle \mathbf{x}, \lambda \mathbf{x} \rangle = \bar{\lambda} \|\mathbf{x}\|_2^2.$$

Poichè $\|\mathbf{x}\|_2^2 \neq 0$, abbiamo $\lambda = \bar{\lambda}$ e il risultato segue. \square

A questo punto possiamo dimostrare che le matrici simmetriche reali sono *sempre* diagonalizzabili; in realtà, tali matrici sono diagonalizzabili in modo speciale, ovvero mediante matrici *ortogonalni*. Poniamo quindi la definizione seguente: due matrici $A, B \in M_{n \times n}(\mathbb{R})$ sono *ortogonalmente simili* se esiste una matrice *ortogonale* P tale che

$$A = P^{-1}BP = P^TBP.$$

In tal caso, se A è simmetrica lo è anche B e viceversa, per quanto visto sopra.

Teorema 6.1. *Ogni matrice $A \in M_{n \times n}(\mathbb{R})$ simmetrica è ortogonalmente simile ad una matrice diagonale reale.*

Dimostrazione. Procediamo per induzione su n . Se $n = 1$ non c'è nulla da dimostrare. Supponiamo il teorema vero per matrici di ordine $n - 1$ e sia $\lambda \in \mathbb{R}$ un autovalore della matrice $A \in M_{n \times n}(\mathbb{R})$ simmetrica. Consideriamo l'autospazio V_λ ; data una base ortonormale di V_λ , sia B una base *ortonormale* di \mathbb{R}^n che la completa; ciò è sempre possibile grazie all'algoritmo di ortonormalizzazione di Gram-Schmidt. Sia $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ la trasformazione lineare associata ad A mediante la base canonica K , $\varphi(\mathbf{x}) = A\mathbf{x}$, e scriviamo

$$M = M_\varphi^B.$$

Poichè A ed M sono associate a φ tramite le basi ortonormali K e B , abbiamo

$$A = P^{-1}MP \tag{6.2}$$

dove P è la matrice di passaggio da B a K e quindi, per la Proposizione 6.2, P è *ortogonale*. Inoltre, per la (6.1) la matrice M è *simmetrica*.

Osserviamo che, poichè B completa la base di V_λ , la prima colonna di M è $(\lambda, 0, \dots, 0)^T$; quindi, poichè M è simmetrica, la prima riga di M è $(\lambda, 0, \dots, 0)$. Segue che

$$M = \begin{pmatrix} \lambda & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & M' & \\ 0 & & & \end{pmatrix} \tag{6.3}$$

dove M' è una matrice reale *simmetrica* di ordine $n - 1$.

Per l'ipotesi induttiva, M' è quindi ortogonalmente simile ad una matrice diagonale reale, ovvero

$$M' = R^{-1}\Delta'R \tag{6.4}$$

con R ortogonale e Δ' diagonale, entrambe di ordine $n - 1$. Non è difficile verificare che dalle (6.3) e (6.4) si deduce che

$$M = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & R^{-1} & \\ 0 & & & \end{pmatrix} \begin{pmatrix} \lambda & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \Delta' & \\ 0 & & & \end{pmatrix} \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & R & \\ 0 & & & \end{pmatrix} = Q^{-1}\Delta Q \tag{6.5}$$

con Q ortogonale e Δ diagonale, entrambe di ordine n .

Dalle (6.2) e (6.5) abbiamo allora

$$A = P^{-1}(Q^{-1}\Delta Q)P = (QP)^{-1}\Delta(QP)$$

con Δ diagonale e QP ortogonale poichè il prodotto di matrici ortogonali è a sua volta ortogonale, e il teorema è dimostrato. \square

Metodo di diagonalizzazione

Il Teorema 6.1 assicura che ogni matrice simmetrica reale è diagonalizzabile mediante matrici ortogonali. Per stabilire un *algoritmo di diagonalizzazione* per le matrici simmetriche reali abbiamo bisogno del seguente

Teorema 6.2. *Gli autospazi di una matrice simmetrica reale sono a due a due ortogonali.*

Dimostrazione. Siano A una matrice simmetrica reale, λ_1 e λ_2 due autovalori *distinti* di A e $\mathbf{x}_1 \in V_{\lambda_1}$, $\mathbf{x}_2 \in V_{\lambda_2}$, $\mathbf{x}_1, \mathbf{x}_2 \neq 0$. Dalla formula di aggiunzione abbiamo

$$\begin{aligned} \lambda_1 < \mathbf{x}_1, \mathbf{x}_2 > &= < \lambda_1 \mathbf{x}_1, \mathbf{x}_2 > = < A\mathbf{x}_1, \mathbf{x}_2 > = < \mathbf{x}_1, A\mathbf{x}_2 > \\ &= < \mathbf{x}_1, \lambda_2 \mathbf{x}_2 > = \lambda_2 < \mathbf{x}_1, \mathbf{x}_2 >, \end{aligned}$$

da cui

$$(\lambda_1 - \lambda_2) < \mathbf{x}_1, \mathbf{x}_2 > = 0;$$

poichè $\lambda_1 \neq \lambda_2$ abbiamo $< \mathbf{x}_1, \mathbf{x}_2 > = 0$, e il teorema è dimostrato. \square

A questo punto abbiamo gli strumenti per enunciare l'*algoritmo di diagonalizzazione*, mediante matrici ortogonali, di una matrice simmetrica $A \in M_{n \times n}(\mathbb{R})$. In realtà, tale algoritmo è quasi identico all'algoritmo generale di diagonalizzazione. Si procede nel modo seguente:

- i) calcolare gli autovalori *distinti* $\lambda_1, \dots, \lambda_r$ di A
- ii) calcolare una base di ogni V_{λ_i}
- iii) ortonormalizzare tali basi; per il Teorema 6.2 l'unione di tali basi ortonormali forma un base ortonormale B di \mathbb{R}^n
- iv) a questo punto proseguire come nel capitolo precedente, ottenendo le matrici P e Δ .

In questo caso P è ortogonale, in quanto matrice di passaggio tra le due basi ortonormali B e K . In altre parole, per diagonalizzare una matrice simmetrica reale mediante matrici ortogonali è sufficiente ortonormalizzare le basi degli autospazi; per il Teorema 6.2 tali autospazi sono tra loro ortogonali, quindi l'unione delle singole basi ortonormali forma una base ortonormale di \mathbb{R}^n .

6.4. Forme quadratiche

Forme quadratiche

Una *forma quadratica* in n variabili è un polinomio *omogeneo* $F(x_1, \dots, x_n)$ di grado 2 a coefficienti reali nelle n variabili x_1, \dots, x_n , ovvero

$$F(x_1, \dots, x_n) = \sum_{i,j=1}^n a_{ij}x_i x_j, \quad a_{ij} \in \mathbb{R}, \quad x_i \in \mathbb{R}.$$

Poichè $x_i x_j = x_j x_i$, riarrangiando i coefficienti a_{ij} possiamo sempre supporre che

$$a_{ij} = a_{ji},$$

cosa che d'ora in poi assumeremo *sempre*. Osserviamo inoltre che $F(0, \dots, 0) = 0$.

Data una forma quadratica $F(x_1, \dots, x_n)$ consideriamo la matrice

$$Q = (a_{ij});$$

è chiaro che Q è *simmetrica* e si verifica facilmente che la forma quadratica $F(x_1, \dots, x_n)$ può essere espressa mediante la matrice Q nel modo seguente:

$$F(x_1, \dots, x_n) = (x_1, \dots, x_n) Q (x_1, \dots, x_n)^T. \quad (6.6)$$

La scrittura (6.6) consente di interpretare la matrice Q come la *matrice associata* alla forma quadratica $F(x_1, \dots, x_n)$ rispetto alla base canonica K di \mathbb{R}^n . Denotando come al solito con \mathbf{x} il vettore colonna $(x_1, \dots, x_n)^T$, la (6.6) si riscrive come

$$F(x_1, \dots, x_n) = \mathbf{x}^T Q \mathbf{x}. \quad (6.7)$$

Sia ora $B = \langle v_1, \dots, v_n \rangle$ una base qualunque di \mathbb{R}^n , dove i vettori v_1, \dots, v_n sono ovviamente vettori colonna di \mathbb{R}^n ; per quanto visto nel Paragrafo 3.2, denotando con $\mathbf{y} = (y_1, \dots, y_n)^T$ le *coordinate* di \mathbf{x} rispetto alla base B abbiamo

$$\mathbf{x} = P \mathbf{x}_B = P \mathbf{y} \quad (6.8)$$

dove P è la matrice di passaggio da K a B e quindi

$$P = (v_1, \dots, v_n). \quad (6.9)$$

Dalle (6.7) e (6.8) otteniamo allora

$$F(x_1, \dots, x_n) = \mathbf{x}^T Q \mathbf{x} = (P \mathbf{y})^T Q (P \mathbf{y}) = \mathbf{y}^T (P^T Q P) \mathbf{y},$$

e quindi $P^T Q P$ è la *matrice associata* a $F(x_1, \dots, x_n)$ rispetto alla base B .

In altre parole, se la forma quadratica $F(x_1, \dots, x_n)$ ha matrice Q associata rispetto a K e se B è un'altra base di \mathbb{R}^n allora

$$F(x_1, \dots, x_n) = \mathbf{x}_B^T R \mathbf{x}_B \quad (6.10)$$

dove R è la matrice associata rispetto alla base B , legata a Q mediante la relazione

$$R = P^T Q P, \quad (6.11)$$

P essendo la matrice di passaggio da K a B data dalla (6.9); osserviamo inoltre che R è ancora *simmetrica*.

La (6.11) suggerisce la definizione seguente: due matrici $A, B \in M_{n \times n}(\mathbb{R})$ sono *congruenti* se esiste una matrice *invertibile* P tale che

$$A = P^T B P;$$

la (6.11) afferma quindi che le matrici associate alla stessa forma quadratica mediante basi diverse sono tra loro congruenti.

Osserviamo come la nozione di congruenza e similitudine tra matrici siano molto somiglianti. In effetti, abbiamo

$$A \text{ e } B \text{ ortogonalmente simili} \Rightarrow A \text{ e } B \text{ simili e congruenti},$$

ma in generale non vi sono altre relazioni tra similitudine e congruenza. Vedremo però che le matrici congruenti conservano *qualche* proprietà delle matrici simili.

Carattere di definizione

L'oggetto principale del nostro studio sulle forme quadratiche è il *carattere di definizione* di $F(x_1, \dots, x_n)$; diremo che $F(x_1, \dots, x_n)$ è

- *definita positiva* se $F(x_1, \dots, x_n) > 0$ per ogni $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$
- *semidefinita positiva* se $F(x_1, \dots, x_n) \geq 0$ per ogni $\mathbf{x} \in \mathbb{R}^n$ e $F(x_1, \dots, x_n) = 0$ per qualche $\mathbf{x} \neq 0$
- *definita negativa* se $F(x_1, \dots, x_n) < 0$ per ogni $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$
- *semidefinita negativa* se $F(x_1, \dots, x_n) \leq 0$ per ogni $\mathbf{x} \in \mathbb{R}^n$ e $F(x_1, \dots, x_n) = 0$ per qualche $\mathbf{x} \neq 0$
- *indefinita* se esistono \mathbf{x} e \mathbf{y} tali che $F(x_1, \dots, x_n) > 0$ e $F(y_1, \dots, y_n) < 0$.

Vediamo ora come il carattere di definizione di $F(x_1, \dots, x_n)$ sia deducibile dal *segno* degli autovalori della matrice Q associata a $F(x_1, \dots, x_n)$ rispetto alla base canonica. Per il Teorema 6.1 la matrice Q è ortogonalmente simile ad una matrice diagonale, ovvero

$$\Delta = P^{-1} Q P = P^T Q P \quad \text{con } P \text{ ortogonale e } \Delta \text{ diagonale};$$

consideriamo allora la base B definita da

$$B = K P$$

e osserviamo che B è *ortonormale* e P è la matrice di passaggio da K a B . Dalle (6.10) e (6.11) abbiamo allora che, posto $\mathbf{x}_B = (y_1, \dots, y_n)^T = \mathbf{y}$,

$$F(x_1, \dots, x_n) = \mathbf{x}^T Q \mathbf{x} = \mathbf{x}_B^T \Delta \mathbf{x}_B = \mathbf{y}^T \Delta \mathbf{y} = \lambda_1 y_1^2 + \dots + \lambda_n y_n^2 \quad (6.12)$$

dove $\lambda_1, \dots, \lambda_n$ sono gli *autovalori* (non necessariamente distinti) di Q .

Un ragionamento del tutto analogo si può fare a partire dall'espressione di $F(x_1, \dots, x_n)$ rispetto ad una base B qualunque di \mathbb{R}^n . La (6.12) prende il nome di *diagonalizzazione* della forma quadratica $F(x_1, \dots, x_n)$.

A questo punto è immediato osservare che la (6.12) fornisce il *criterio* seguente per la determinazione del carattere di definizione: $F(x_1, \dots, x_n)$ è

- definita positiva $\Leftrightarrow \lambda_i > 0$ per ogni $i = 1, \dots, n$
- semidefinita positiva $\Leftrightarrow \lambda_i \geq 0$ per ogni $i = 1, \dots, n$ ed esiste un i con $\lambda_i = 0$
- definita negativa $\Leftrightarrow \lambda_i < 0$ per ogni $i = 1, \dots, n$
- semidefinita negativa $\Leftrightarrow \lambda_i \leq 0$ per ogni $i = 1, \dots, n$ ed esiste un i con $\lambda_i = 0$
- indefinita \Leftrightarrow esistono $\lambda_i > 0$ e $\lambda_j < 0$.

Segue quindi che, noti gli autovalori della matrice Q (o di una qualunque matrice associata a $F(x_1, \dots, x_n)$), è noto il carattere di definizione della forma quadratica $F(x_1, \dots, x_n)$; in realtà è sufficiente conoscere il *segno* di tali autovalori, che nella pratica è un problema decisamente più semplice da risolvere.

Segno degli autovalori di una matrice simmetrica

Osserviamo che il carattere di definizione dipende dalla forma quadratica $F(x_1, \dots, x_n)$, non dalla matrice simmetrica usata per esprimere $F(x_1, \dots, x_n)$ rispetto ad una certa base. D'altra parte, abbiamo appena visto che il carattere di definizione dipende dal segno degli autovalori di una qualunque matrice associata alla forma quadratica. Queste osservazioni portano alla conclusione che ci deve essere uno *stretto legame* tra il segno degli autovalori delle matrici associate alla stessa forma quadratica. Notiamo che se B è una base ortonormale, allora le matrici associate a $F(x_1, \dots, x_n)$ mediante la base canonica e la base B sono congruenti ma anche ortogonalmente simili, quindi hanno in realtà gli *stessi* autovalori, non soltanto autovalori dello stesso segno. In generale, matrici simmetriche congruenti non hanno necessariamente gli stessi autovalori, ma hanno autovalori dello *stesso segno*.

Per precisare questo concetto consideriamo la *segnatura* (p, q, z) di una matrice simmetrica $A \in M_{n \times n}(\mathbb{R})$, ovvero

p = numero di autovalori positivi di A contati con molteplicità

q = numero di autovalori negativi di A contati con molteplicità

z = numero di autovalori nulli di A contati con molteplicità;

osserviamo per inciso che $\rho(A) = p + q$. Si dimostra il seguente

Teorema di inerzia di Sylvester. *Due matrici simmetriche reali sono congruenti se e solo se hanno la stessa segnatura.*

In particolare, matrici simmetriche reali congruenti hanno autovalori dello stesso segno; questo porta a ridurre il problema del carattere di definizione di una forma quadratica a quello della determinazione del segno degli autovalori di una qualunque matrice *congruente* a Q .

Calcolo del segno degli autovalori

Abbiamo quindi il problema del calcolo del segno degli autovalori di Q , o di una qualunque matrice congruente a Q . Il primo metodo di calcolo del segno è quello ovvio, ovvero calcolare gli autovalori di Q .

Talvolta questo non è agevole nella pratica, quindi dobbiamo trovare un metodo alternativo. Una buona alternativa è offerta dal risultato seguente, di cui omettiamo la dimostrazione.

Regola di Cartesio. *Sia $P(x) = a_n x^n + \cdots + a_1 x + a_0$ un polinomio a coefficienti reali di grado n con tutte le radici reali. Allora il numero di radici positive di $P(x)$, contate con la loro molteplicità, è uguale al numero di variazioni di segno nella successione dei coefficienti non nulli di $P(x)$.*

Poichè il numero di radici *nulle* di $P(x)$ è facilmente calcolabile (è infatti uguale al minimo grado a cui compare la x), la regola di Cartesio applicata al polinomio caratteristico $P_Q(x)$ della matrice Q fornisce immediatamente il segno degli autovalori di Q e quindi la *segnatura* della forma quadratica.

Esempio. Sia $F(x, y, z)$ la forma quadratica associata mediante la base canonica K alla matrice $Q = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}$. Il polinomio caratteristico è $P_Q(x) = -x^3 + 2x^2$, quindi la segnatura è $(1,0,2)$ e conseguentemente $F(x, y, z)$ è semidefinita positiva.

Capitolo 7. GEOMETRIA ANALITICA, parte II

7.1. Coniche e quadriche

Coniche

Ad una conica Γ di equazione

$$\Gamma(x, y) = a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0$$

si associano le due matrici *simmetriche*

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \quad \text{matrice } incompleta$$

$$B = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix} \quad \text{matrice } completa.$$

È chiaro che A è la matrice associata alla forma quadratica $a_{11}x^2 + 2a_{12}xy + a_{22}y^2$, detta *parte quadratica* dell'equazione della conica Γ . E' inoltre chiaro che il polinomio $\Gamma(x, y)$ può essere scritto in *forma matriciale* mediante la matrice B nel modo seguente

$$\Gamma(x, y) = (x, y, 1)B \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}.$$

Diciamo che una conica è *degenera* se

$$\det B = 0;$$

si può dimostrare che

$$\det B = 0 \Leftrightarrow \Gamma(x, y) = (ax + by + c)(a'x + b'y + c')$$

con $a, b, c \in \mathbb{R}$ e $a', b', c' \in \mathbb{R}$ oppure $a, b, c \in \mathbb{C}$ e $a' = \bar{a}$, $b' = \bar{b}$ e $c' = \bar{c}$. Segue che le coniche *degeneri* possono essere

- a) una coppia di rette reali
- b) una coppia di rette complesse coniugate.

La *classificazione* e il *riconoscimento* delle coniche si può effettuare per mezzo delle due matrici A e B ; ad esempio, abbiamo appena visto che il valore di $\det B$ consente di distinguere le coniche degeneri e non-degeneri.

Per quel che riguarda le coniche *non-degeneri*, si può dimostrare che le possibilità sono

- a) ellisse
- b) parabola
- c) iperbole;

tali tre casi possono essere individuati mediante la matrice A , nel modo seguente

- a) ellisse $\Leftrightarrow \det A > 0$
- b) parabola $\Leftrightarrow \det A = 0$
- c) iperbole $\Leftrightarrow \det A < 0$.

Questo è ovvio se la conica è in *forma canonica*, in quanto in tal caso la matrice A è diagonale e

- a) ellisse $\Leftrightarrow A$ ha i due autovalori dello stesso segno
- b) parabola $\Leftrightarrow A$ un autovalore nullo
- c) iperbole $\Leftrightarrow A$ ha un autovalore positivo e uno negativo,

e quanto asserito segue calcolando $\det A$ come prodotto dei due autovalori.

Osserviamo per inciso che se A è una qualunque matrice di ordine n *diagonalizzabile* e $\lambda_1, \dots, \lambda_n$ sono i suoi autovalori allora

$$\det A = \lambda_1 \cdots \lambda_n;$$

infatti si ha $A = P^{-1}\Delta P$, da cui $\det A = \det(P^{-1}) \det \Delta \det P = \det \Delta = \lambda_1 \cdots \lambda_n$.

In generale, la classificazione e relativo riconoscimento delle coniche illustrata in precedenza si può giustificare, in modo *costruttivo*, effettuando esplicitamente la *riduzione a forma canonica* di una generica conica non-degenera.

Dal punto di vista *geometrico*, tale riduzione si esegue in due fasi:

- i) si opera una *rotazione* in modo da rendere gli assi della conica paralleli agli assi coordinati
- ii) si opera una *traslazione* in modo da far coincidere gli assi della conica ruotata con gli assi coordinati.

Dal punto di vista *algebrico*, le due operazioni sopra descritte si effettuano nel modo seguente:

- i) si *diagonalizza* la parte quadratica di $\Gamma(x, y)$ mediante una matrice ortogonale P con $\det P = 1$
- ii) si opera il *cambio di coordinate* (rotazione) di cui la matrice P è la matrice di rotazione; conseguentemente, il coefficiente a_{12} della nuova equazione è nullo
- iii) si *completano i quadrati* in tale equazione e si opera il cambio di coordinate (traslazione) risultante.

Dopo tali operazioni si ottiene una nuova equazione della conica, e tale equazione coincide con la forma canonica della conica stessa, che quindi è stata ridotta a forma canonica. Osserviamo che se il termine misto xy non compare nell'equazione originale, allora il procedimento si semplifica, riducendosi semplicemente al completamento dei quadrati.

Esempio. $\Gamma(x, y) = 2x^2 + 4xy + 5y^2 + 2x$. Allora

$$A = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix} \quad B = \begin{pmatrix} 2 & 2 & 1 \\ 2 & 5 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$\det B \neq 0$, quindi conica non-degenera

$\det A > 0$, quindi ellisse

gli autovalori di A sono $\lambda_1 = 1$ e $\lambda_2 = 6$

base ortonormale di V_{λ_1} è $\left\langle -\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} \right\rangle$

base ortonormale di V_{λ_2} è $\left\langle \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \right\rangle$

matrice $\tilde{P} = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 & 1 \\ 1 & 2 \end{pmatrix}$; poiché $\det \tilde{P} = -1$ si cambia segno:

$$P = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \quad \text{e} \quad (x, y)^T = P(\bar{X}, \bar{Y})^T,$$

ovvero

$$\begin{cases} x &= \frac{1}{\sqrt{5}}(2\bar{X} + \bar{Y}) \\ y &= \frac{1}{\sqrt{5}}(-\bar{X} + 2\bar{Y}). \end{cases}$$

La nuova equazione è quindi

$$\bar{X}^2 + 6\bar{Y}^2 + \frac{2}{\sqrt{5}}(2\bar{X} + \bar{Y}) = 0;$$

completamento dei quadrati:

$$\begin{cases} X &= \bar{X} + \frac{2}{\sqrt{5}} \\ Y &= \bar{Y} + \frac{1}{6\sqrt{5}} \end{cases}$$

da cui l'equazione in forma canonica è

$$X^2 + 6Y^2 = \frac{5}{6}.$$

Osserviamo senza entrare nei dettagli che alla base della classificazione delle coniche non-degeneri c'è il concetto di *segnatura* della matrice A ; nel caso delle coniche la segnatura di A si esprime semplicemente per mezzo del *segno* di $\det A$, ma già nel caso delle quadriche la segnatura di A gioca un ruolo primario.

Quadratiche

Una teoria analoga a quella appena illustrata vale per le quadriche. Data una quadrica $\Gamma(x, y, z) = 0$ consideriamo le matrici *simmetriche*

$$A = (a_{ij})_{i,j=1,2,3} \quad B = (a_{ij})_{i,j=1,2,3,4}$$

formate con i coefficienti, come nel caso delle coniche; A è quindi la matrice associata alla *parte quadratica* di $\Gamma(x, y, z)$, mentre

$$\Gamma(x, y, z) = (x, y, z, 1)B \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}.$$

Una quadrica è *degenera* se

$$\det B = 0,$$

e si può dimostrare che le quadriche *degeneri* sono

- a) coppie di piani (eventualmente complessi coniugati)
- b) coni quadrici
- c) cilindri quadrici;

definizioni ed esempi si vedranno nelle Esercitazioni.

La *classificazione* delle quadriche *non-degeneri* è quella presentata, in forma canonica, nel Paragrafo 4.5, ed il *riconoscimento* di una generica quadrica non-degenere si può fare usando opportune proprietà delle matrici A e B . Come già osservato in precedenza, il tipo di quadrica dipende essenzialmente dalla *segnatura* della parte quadratica di $\Gamma(x, y, z)$. Inoltre, la *riduzione a forma canonica* delle quadriche si opera in modo del tutto simile a quanto visto nel caso delle coniche. Esempi ed approfondimenti verranno svolti nelle Esercitazioni.

7.2. Cenni su linee e superfici nello spazio

Linee e superfici

In questo paragrafo diamo alcuni brevi cenni sulle equazioni di *linee e superfici* nello spazio.

Usualmente, l'equazione *parametrica*

$$\begin{cases} x = x(t) \\ y = y(t) \\ z = z(t) \end{cases} \quad \text{con } x(t), y(t), z(t) \text{ funzioni continue da } \mathbb{R} \text{ in } \mathbb{R} \quad (7.1)$$

rappresenta una *linea* \mathcal{L} in \mathbb{R}^3 , mentre l'equazione *parametrica*

$$\begin{cases} x = x(t, u) \\ y = y(t, u) \\ z = z(t, u) \end{cases} \quad \text{con } x(t, u), y(t, u), z(t, u) \text{ funzioni continue da } \mathbb{R}^2 \text{ in } \mathbb{R}$$

rappresenta una *superficie* \mathcal{S} in \mathbb{R}^3 ; si vedano, ad esempio, le equazioni parametriche di rette e piani nello spazio.

L'equazione *cartesiana* di una superficie è del tipo

$$f(x, y, z) = 0 \quad \text{con } f : \mathbb{R}^3 \rightarrow \mathbb{R} \text{ continua,} \quad (7.2)$$

mentre quella di una linea è del tipo

$$\begin{cases} f(x, y, z) = 0 \\ g(x, y, z) = 0 \end{cases} \quad \text{con } f, g : \mathbb{R}^3 \rightarrow \mathbb{R} \text{ continue.}$$

Diciamo che una linea \mathcal{L} , di equazione parametrica (7.1), *giace* sulla superficie \mathcal{S} , di equazione cartesiana (7.2), se

$$f(x(t), y(t), z(t)) = 0 \quad \text{per ogni } t \in \mathbb{R}. \quad (7.3)$$

In particolare, la (7.3) fornisce una condizione affinchè una linea sia *piana*: la linea \mathcal{L} giace sul piano π di equazione $ax + by + cz + d = 0$ se

$$ax(t) + by(t) + cz(t) + d = 0 \quad \text{per ogni } t \in \mathbb{R}.$$

Cilindri

Un *cilindro* nello spazio è una superficie esprimibile come luogo di rette (dette *generatrici*) tutte parallele a un dato vettore $\mathbf{v} = (l, m, n)$. Una *direttrice* del cilindro è una linea \mathcal{L} che interseca ogni generatrice del cilindro; se \mathcal{L} è data dall'equazione (7.1) allora l'equazione *parametrica* del cilindro è

$$\begin{cases} x &= x(t) + lu \\ y &= y(t) + mu \\ z &= z(t) + nu \end{cases} \quad \text{con } t, u \in \mathbb{R}.$$

Coni

Un *cono* nello spazio è una superficie esprimibile come luogo di rette (dette *generatrici*) tutte passanti per un punto $V = (x_0, y_0, z_0)$ detto *vertice*. Anche nel caso del cono si può parlare di *direttrice* \mathcal{L} , e l'equazione *parametrica* del cono di vertice V e direttrice \mathcal{L} data dalla (7.1) è

$$\begin{cases} x &= x_0 + (x(t) - x_0)u \\ y &= y_0 + (y(t) - y_0)u \\ z &= z_0 + (z(t) - z_0)u \end{cases} \quad \text{con } t, u \in \mathbb{R}.$$

Superfici di rotazione

Una *superficie di rotazione* si ottiene facendo ruotare una linea \mathcal{L} attorno a una retta r . Fissato un punto $C \in r$, la superficie di rotazione si può rappresentare come intersezione tra

- la sfera di centro C e raggio CP
- il piano ortogonale a r passante per P ,

dove P è un punto generico su \mathcal{L} . Tale rappresentazione fornisce l'equazione della superficie di rotazione.

Capitolo 8. ELEMENTI DI CALCOLO NUMERICO

8.1. Norme matriciali

Norme matriciali

Definiamo una *norma matriciale*, ovvero una norma sulle matrici $M_{n \times n}(\mathbb{R})$, mediante le proprietà seguenti:

- i) $\|A\| \geq 0$ e $\|A\| = 0 \Leftrightarrow A = 0$
- ii) $\|\lambda A\| = |\lambda| \|A\|$
- iii) $\|A + B\| \leq \|A\| + \|B\|$
- iv) $\|AB\| \leq \|A\| \|B\|$.

Osserviamo come tali proprietà siano del tutto analoghe a quelle che definiscono una *norma vettoriale*, ed in effetti ogni norma vettoriale su \mathbb{R}^n induce una norma matriciale su $M_{n \times n}(\mathbb{R})$ nel modo seguente

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}; \quad (8.1)$$

si verifica che la (8.1) definisce una norma su $M_{n \times n}(\mathbb{R})$, e tale norma si dice *norma indotta* dalla norma su \mathbb{R}^n .

E' semplice definire l'analogo matriciale delle norme $\|\cdot\|_1$ e $\|\cdot\|_\infty$ introdotte nel Paragrafo 6.1; data $A = (a_{ij}) \in M_{n \times n}(\mathbb{R})$ definiamo

$$\|A\|_1 = \max_{j=1, \dots, n} \sum_{i=1}^n |a_{ij}|$$

$$\|A\|_\infty = \max_{i=1, \dots, n} \sum_{j=1}^n |a_{ij}|,$$

ovvero $\|A\|_1$ e $\|A\|_\infty$ sono rispettivamente il massimo delle $\|\cdot\|_1$ delle colonne e delle righe di A . Si può verificare che $\|\cdot\|_1$ e $\|\cdot\|_\infty$ sono *indotte* dalle $\|\cdot\|_1$ e $\|\cdot\|_\infty$ su \mathbb{R}^n .

La definizione della norma $\|\cdot\|_2$ delle matrici $M_{n \times n}(\mathbb{R})$ è più laboriosa, ed il calcolo più complicato rispetto alle $\|\cdot\|_1$ e $\|\cdot\|_\infty$. Data $A \in M_{n \times n}(\mathbb{R})$, ricordiamo che $A^T A$ è simmetrica; inoltre la forma quadratica

$$F(x_1, \dots, x_n) = x^T (A^T A) x$$

è *definita* o *semidefinita positiva*. Infatti

$$x^T (A^T A) x = (Ax)^T (Ax)$$

e denotando con

$$y = Ax = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

abbiamo quindi

$$(Ax)^T(Ax) = y_1^2 + \cdots + y_n^2.$$

Per quanto visto nel Paragrafo 6.4 la matrice $A^T A$ ha allora autovalori $\alpha_1, \dots, \alpha_n \geq 0$; tali autovalori prendono il nome di *valori singolari* della matrice A . A questo punto possiamo definire la norma $\| \cdot \|_2$ di A come

$$\|A\|_2 = \max_{j=1,\dots,n} \sqrt{\alpha_j};$$

si può verificare che $\| \cdot \|_2$ è una norma matriciale, ed inoltre $\| \cdot \|_2$ è *indotta* dalla $\| \cdot \|_2$ in \mathbb{R}^n . Osserviamo che se A è *simmetrica* allora $A^T A = A^2$, quindi i suoi valori singolari non sono altro che i *quadrati* dei suoi *autovalori* $\lambda_1, \dots, \lambda_n$; quindi nel caso di una matrice A simmetrica abbiamo

$$\|A\|_2 = \max_{j=1,\dots,n} |\lambda_j|.$$

A titolo di esempio introduciamo una norma, la *norma di Schur*, che non è indotta da alcuna norma su \mathbb{R}^n ; tale norma è definita come

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

L'interesse della norma di Schur deriva dal fatto che, mentre la $\| \cdot \|_2$ è di difficile calcolo, la $\| \cdot \|_S$ si calcola facilmente, ed inoltre se $A \in M_{n \times n}(\mathbb{R})$ abbiamo

$$\|A\|_2 \leq \|A\|_S \leq \sqrt{n} \|A\|_2;$$

tale proprietà consente di utilizzare la norma di Schur al posto della $\| \cdot \|_2$ in alcune applicazioni.

Osserviamo infine che

$$\|I\|_1 = \|I\|_2 = \|I\|_\infty = 1$$

mentre

$$\|I\|_S = \sqrt{n}.$$

Vediamo ora alcune proprietà delle norme:

i) se A è ortogonale, allora $\|A\|_2 = 1$; infatti $A^T A = I$ e quindi i valori singolari di A sono tutti uguali a 1

ii) se $\| \cdot \|$ è una norma indotta, allora $\|I\| = 1$; infatti $\|Ix\| = \|x\|$ per ogni $x \in \mathbb{R}^n$

iii) se $\| \cdot \|$ è una norma indotta, allora $\|Ax\| \leq \|A\| \|x\|$; infatti $\|A\|$ è l'estremo superiore di $\frac{\|Ax\|}{\|x\|}$

iv) se $\| \cdot \|$ è una norma qualunque, allora $\|I\| \geq 1$; infatti $\|A\| = \|IA\| \leq \|I\| \|A\|$, da cui $\|I\| \geq 1$

v) se $\|\cdot\|$ è una norma qualunque e A è invertibile, allora $\|A\|\|A^{-1}\| \geq 1$; infatti $\|A\|\|A^{-1}\| \geq \|AA^{-1}\| = \|I\| \geq 1$.

Condizionamento

Definiamo infine il *numero di condizionamento* di una matrice A *invertibile* rispetto a una data norma matriciale $\|\cdot\|$ come

$$\text{cond}(A) = \|A\|\|A^{-1}\|.$$

Osserviamo che se A è simmetrica e invertibile, allora gli autovalori di A^{-1} sono $\lambda_1^{-1}, \dots, \lambda_n^{-1}$, dove $\lambda_1, \dots, \lambda_n$ sono gli autovalori di A (abbiamo già osservato che gli autovalori di A^2 sono $\lambda_1^2, \dots, \lambda_n^2$ e questo fatto vale in generale: gli autovalori di A^k , $k \in \mathbb{N}$, sono $\lambda_1^k, \dots, \lambda_n^k$); abbiamo quindi che se A è simmetrica e invertibile allora

$$\|A\|_2 = \max_j |\lambda_j| \quad \text{e} \quad \|A^{-1}\|_2 = \max_j \left| \frac{1}{\lambda_j} \right|.$$

Ordinando gli autovalori di A nel modo seguente

$$|\lambda_1| \leq |\lambda_2| \leq \dots \leq |\lambda_n|$$

abbiamo quindi che

$$\text{cond}_2(A) = \left| \frac{\lambda_n}{\lambda_1} \right|$$

per ogni matrice A simmetrica e invertibile.

8.2. Sistemi lineari: stabilità

Perturbamento dei sistemi lineari

Consideriamo ora il problema della stabilità delle soluzioni di un sistema lineare

$$Ax = b \quad A \in M_{n \times n}(\mathbb{R}) \quad \text{invertibile}, \tag{8.2}$$

ovvero il problema del comportamento della soluzione x quando la matrice A o il vettore b vengano *perturbati*, dove perturbare una matrice o un vettore significa alterarne i coefficienti di quantità piccole rispetto ai coefficienti stessi. Tale problema è di grande importanza nella pratica in quanto i dati sono spesso dati approssimati (ovvero i coefficienti sono noti a meno di opportune approssimazioni) ed è importante sapere come tali approssimazioni si trasmettono sulla soluzione. In altre parole, il problema è: di quanto varia la soluzione quando i dati del sistema sono alterati da errori o approssimazioni opportunamente piccoli?

Perturbamento del termine noto

Supponiamo di *perturbare il termine noto* b , ovvero consideriamo il sistema (8.2) ed il sistema in cui il termine noto b è rimpiazzato da $b + \delta_b$; δ_b rappresenta la (piccola) perturbazione di b e di conseguenza la soluzione di tale sistema non sarà più x ma $x + \delta_x$. In altre parole, abbiamo

$$A(x + \delta_x) = b + \delta_b, \quad (8.3)$$

ed il problema è controllare in qualche modo la quantità δ_x . Dalle (8.2) e (8.3) abbiamo che δ_x soddisfa

$$A\delta_x = \delta_b,$$

ovvero

$$\delta_x = A^{-1}\delta_b.$$

Data una norma vettoriale $\|\cdot\|$ su \mathbb{R}^n e la norma matriciale *indotta* $\|\cdot\|$, abbiamo allora

$$\|b\| = \|Ax\| \leq \|A\| \|x\|$$

$$\|\delta_x\| = \|A^{-1}\delta_b\| \leq \|A^{-1}\| \|\delta_b\|;$$

segue che

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

ovvero

$$\frac{\|\delta_x\|}{\|x\|} \leq \text{cond}(A) \frac{\|\delta_b\|}{\|b\|}. \quad (8.4)$$

La (8.4) fornisce allora una *maggiorazione* per l'*errore relativo* (ovvero $\frac{\|\delta_x\|}{\|x\|}$) che si introduce nella soluzione del sistema rimpiazzando b con $b + \delta_b$, ovvero perturbando il termine noto b . La (8.4) mostra che tale errore relativo è tanto più piccolo quanto più piccolo è il condizionamento della matrice A . In altre parole, un sistema lineare con matrice A *ben condizionata*, ovvero con $\text{cond}(A)$ *piccolo*, è *meno sensibile* agli errori introdotti nel termine noto b , ovvero un tale sistema è *stabile* rispetto agli errori.

Perturbamento della matrice

Consideriamo ora l'analogo problema in cui *viene perturbata la matrice* A , ovvero si rimpiazza A con $A + \delta_A$; conseguentemente otterremo una nuova soluzione $x + \delta_x$ soddisfacente

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

ovvero

$$Ax + A\delta_x + \delta_A(x + \delta_x) = b. \quad (8.5)$$

Dalle (8.2) e (8.5) otteniamo allora

$$A\delta_x = -\delta_A(x + \delta_x)$$

ovvero

$$\delta_x = -A^{-1}\delta_A(x + \delta_x);$$

passando alle norme come nel caso precedente abbiamo

$$\|\delta_x\| \leq \|A^{-1}\| \|\delta_A\| \|x + \delta_x\| = \|A^{-1}\| \|A\| \frac{\|\delta_A\|}{\|A\|} \|x + \delta_x\|$$

da cui

$$\frac{\|\delta_x\|}{\|x + \delta_x\|} \leq \text{cond}(A) \frac{\|\delta_A\|}{\|A\|}. \quad (8.6)$$

Anche in questo caso si vede che il sistema è stabile se $\text{cond}(A)$ è piccolo; questo è un fatto generale, nel senso che un sistema è poco sensibile agli errori immessi nei dati quando il condizionamento della matrice A è piccolo.