Chapter 6

Prodotti scalari, proiezioni ortogonali, isometrie lineari

6.1 Prodotti scalari, angoli e norme

Nel Capitolo 1, formula (1.19), abbiamo introdotto un prodotto tra terne o coppie di numeri reali mediante le formule

$$x \cdot y = x_1 y_1 + x_2 y_2 + x_3 y_3, \quad x \cdot y = x_1 y_1 + x_2 y_2$$

chiamandolo prodotto scalare in quanto associa a due vettori di $\mathbb{R}^3$ o $\mathbb{R}^2$ un numero reale (ovvero uno scalare). L'importanza di tale operazione consiste nel fatto che mediante di essa possiamo calcolare lunghezze e angoli tra vettori geometrici applicati, mediante le formule

$$\sqrt{x \cdot x}, \quad \cos\alpha = \frac{x \cdot y}{\sqrt{x \cdot x}\sqrt{y \cdot y}} \tag{6.1}$$

dove la prima ci dà la lunghezza di un vettore $\vec{OP}$ le cui coordinate (rispetto a una base ortonormale) sono date da x, la seconda ci dà l'angolo α convesso (cioè compreso tra 0 e π) tra due vettori non nulli $\vec{OP}$ e $\vec{OQ}$ di coordinate x e y (sempre rispetto a una base ortonormale).
Vista l'importanza di tale operazione, ne abbiamo visto (pagina 20 e seguente) le proprietà più importanti, ovvero

(i) $x \cdot (y + y') = x \cdot y + x \cdot y'$

(ii) $(x + x') \cdot y = x \cdot y + x' \cdot y$

(iii) $x \cdot (cy) = (cx) \cdot y = c(x \cdot y)$

(iv) $x \cdot y = y \cdot x$

Ora, in molti contesti della matematica e delle sue applicazioni si incontrano operazioni tra vettori di uno spazio vettoriale che si comportano in modo analogo al prodotto $x \cdot y$ di $\mathbb{R}^3$ e $\mathbb{R}^2$: per fare un esempio riguardante l'analisi e più precisamente lo spazio vettoriale delle funzioni (pagina 28), si ha che date due funzioni f e g da un intervallo $[a, b] \subset \mathbb{R}$ in $\mathbb{R}$, l'integrale $\int_a^b fgdx$ può essere pensato come un prodotto $f \cdot g$ tra f e g che ci dà come risultato un numero reale e che, è facile vederlo, ha le stesse proprietà $(i) - (iv)$ viste sopra per il prodotto scalare di $\mathbb{R}^3$ e $\mathbb{R}^2$ (ad esempio, per la (iii), dato un numero reale c si ha $f \cdot (cg) = \int_a^b fcgdx = c \int_a^b fgdx = c(f \cdot g)$, dove abbiamo potuto portare lo scalare c fuori dall'integrale perché è costante rispetto alla variabile di integrazione x).
Quando si ha a disposizione un'operazione di questo tipo, si possono introdurre nello spazio vettoriale, tramite formule analoghe alle (6.1), delle nozioni di lunghezza e angolo e quindi tutte le nozioni collegate (ortogonalità, proiezioni ortogonali etc.) che si rivelano molto utili nelle più svariate applicazioni.

Vogliamo allora fare qualcosa di simile a quello che abbiamo fatto quando siamo passati dai vettori geometrici alla nozione astratta di vettore e spazio vettoriale: introdurre una nozione astratta di prodotto scalare, che avrà come caso particolare il prodotto scalare di $\mathbb{R}^3$ e $\mathbb{R}^2$ visto nel Capitolo 1, ma che comprenderà numerosi altri esempi importanti, che conviene quindi studiare insieme in un'unica teoria generale.
Diamo allora la seguente

Definizione 6.1. Sia V un $\mathbb{K}$-spazio vettoriale. Una funzione $f : V \times V \to \mathbb{K}$ (ovvero che associa a due vettori uno scalare) che soddisfi le quattro proprietà

(1) $f(v_1 + v_2, w) = f(v_1, w) + f(v_2, w)$ per ogni $v_1, v_2, w \in V$

(2) $f(v, w_1 + w_2) = f(v, w_1) + f(v, w_2)$ per ogni $v, w_1, w_2 \in V$

(3) $f(v, cw) = f(cv, w) = cf(v, w)$ per ogni $v, w \in V$ e ogni $c \in \mathbb{K}$

si dice *forma bilineare su V*.
Se, inoltre, per ogni $v, w \in V$ si ha

(4) $f(v, w) = f(w, v)$

allora la forma bilineare si dice *simmetrica*.
Infine, se $\mathbb{K} = \mathbb{R}$, una forma bilineare simmetrica si dice *definita positiva* se

(5) $f(v, v) \geq 0$ e $f(v, v) = 0$ solo se v è il vettore nullo $\bar{0}$

Una forma bilineare simmetrica e definita positiva su un $\mathbb{R}$-spazio vettoriale V si dice *prodotto scalare su* V.
Useremo anche la notazione $v \cdot w$ alternativa a $f(v, w)$.

Osservazione 6.2. (1) La richiesta che $\mathbb{K} = \mathbb{R}$ per la definizione di forma definita positiva si spiega con il fatto che in un generico campo la nozione di "maggiore o uguale a zero" può non avere senso (ad esempio, se $\mathbb{K}$ è il campo $\mathbb{C}$ dei numeri complessi, su $\mathbb{K}$ non ha senso dire che un numero complesso è maggiore o uguale a un altro).

(2) Il nome di forma bilineare assegnato a una funzione che soddisfa (1)-(3) della Definizione 6.1 è dovuto al fatto che la funzione è "doppiamente lineare", in quanto la (1) e la (3) ci dicono che ha esattamente le proprietà che deve avere una applicazione lineare se facciamo variare solo il primo argomento della coppia (v, w) e non guardiamo il secondo (le somme si spezzano e gli scalari escono fuori), e analogamente la (2) e la (3) ci dicono che f si comporta come un'applicazione lineare se facciamo variare solo il secondo argomento della coppia (v, w) e non guardiamo il primo.

(3) Dalla proprietà (3) della definizione di forma bilineare discende che $f(v, w) = 0$ se almeno uno dei due vettori v o w è uguale al vettore nullo $\bar{0}$. Infatti, dal momento che il vettore nullo può essere scritto come $\bar{0} = 0u$, per un qualunque vettore u di V, si può scrivere $f(\bar{0}, w) = f(0u, w) = 0f(u, w) = 0$, e analogamente $f(v, \bar{0}) = f(v, 0u) = 0f(v, u) = 0$

Le proprietà (1)-(4) della Definizione 6.1 corrispondono esattamente alle (i)-(iv) richiamate sopra per il prodotto scalare di $\mathbb{R}^2$ e $\mathbb{R}^3$ e ne rappresentano quindi la generalizzazione astratta che cercavamo; la proprietà (5) è invece implicita nel prodotto scalare di $\mathbb{R}^2$ e $\mathbb{R}^3$ in quanto per esso si ha chiaramente che il prodotto $x \cdot x = x_1^2 + x_2^2 + x_3^2$ tra una terna e se stessa, essendo somma di quadrati, è maggiore o uguale di zero e nulla se e solo se $x_1 = x_2 = x_3 = 0$, ovvero se e solo se x è il vettore nullo di $\mathbb{R}^3$. Nella definizione astratta di prodotto scalare tale proprietà deve essere esplicitamente richiesta perché vogliamo definire la lunghezza di un vettore v in uno spazio vettoriale dotato di un prodotto scalare tramite una formula analoga alla prima delle (6.1), ovvero $\sqrt{v \cdot v}$, e chiaramente non potremmo mettere $v \cdot v$ sotto radice se non sapessimo che esso è maggiore o uguale di zero.

Uno spazio vettoriale reale dotato di un prodotto scalare si dice *spazio vettoriale euclideo*. Vediamo ora come definire in uno spazio vettoriale euclideo qualunque le nozioni di lunghezza e angolo tra vettori:

Definizione 6.3. In uno spazio vettoriale V euclideo, definiamo la *lunghezza* (o *norma*) di un vettore v come

$$\|v\| = \sqrt{v \cdot v} \tag{6.2}$$

e *l'angolo convesso* tra due vettori non nulli v e w come quell'unico $\theta \in \mathbb{R}$, compreso tra 0 e π (inclusi) tale che

$$\cos\theta = \frac{v \cdot w}{\|v\|\|w\|} \tag{6.3}$$

Le formule (6.2) e (6.3) sono proprio le analoghe delle (6.1) usate per calcolare lunghezze e angoli di vettori geometrici.

Ora, mentre la formula (6.2) ha chiaramente senso perché grazie alla proprietà (5) della Definizione 6.1 la quantità $v \cdot v$ è maggiore o uguale di zero e quindi può stare sotto radice, la (6.3) richiede maggiore attenzione: più precisamente, affinché essa abbia senso è necessario che chiunque siano $v, w \in V$ non nulli, la quantità $\frac{v \cdot w}{\|v\|\|w\|}$ sia compresa tra -1 e $+1$, altrimenti essa non potrebbe essere uguale a un coseno.
Questo è garantito dalla seguente, importante

Proposizione 6.4. Sia V uno spazio vettoriale euclideo e siano $v, w \in V$. Allora, si ha

$$|v \cdot w| \leq \|v\|\|w\| \tag{6.4}$$

La (6.4) è nota come *disuguaglianza di Cauchy-Schwarz*. Prima di dimostrarla, vediamo che essa ci garantisce che la definizione di angolo data tramite la (6.3) è ben posta: infatti, per definizione di valore assoluto, la (6.4) equivale a dire

$$-\|v\|\|w\| \leq v \cdot w \leq \|v\|\|w\|$$

Nel caso in cui v e w siano non nulli, possiamo dividere per $\|v\|\|w\|$ questa doppia disuguaglianza ottenendo

$$-1 \leq \frac{v \cdot w}{\|v\|\|w\|} \leq 1$$

che ci garantisce quindi che $\frac{v \cdot w}{\|v\|\|w\|}$ è uguale a un coseno.
Dimostriamo ora la proposizione.

NO

Proof. Osserviamo prima che la disuguaglianza vale sicuramente se il vettore w è uguale al vettore nullo, in quanto in tal caso entrambi i membri sono uguali a zero (e la disuguaglianza $0 \leq 0$ è ovviamente verificata): infatti, il primo membro diventa $|v \cdot \bar{0}|$ che è uguale a 0 per la (3) dell'Osservazione 6.2, e il secondo membro diventa $\|v\|\|\bar{0}\| = 0$ in quanto $\|\bar{0}\| = \sqrt{\bar{0} \cdot \bar{0}} = 0$ (sempre per la (3) dell'Osservazione 6.2).

Ci resta quindi da far vedere che la (6.4) vale anche per $w \neq \bar{0}$. A questo scopo, applichiamo al vettore[1] $u = v - \frac{v \cdot w}{w \cdot w} w$ la proprietà $u \cdot u \geq 0$ (che vale in quanto il prodotto scalare è definito positivo), ovvero

$$\left(v - \frac{v \cdot w}{w \cdot w} w\right) \cdot \left(v - \frac{v \cdot w}{w \cdot w} w\right) \geq 0 \tag{6.5}$$

Mostreremo ora come, sviluppando questa disuguaglianza, si ricavi esattamente la disuguaglianza di Cauchy-Schwarz.

Per facilitarci i calcoli, denotiamo $b = \frac{v \cdot w}{w \cdot w}$, e quindi $u = v - bw$. Quindi, la (6.5) con questa notazione semplificata si scrive

$$(v - bw) \cdot (v - bw) \geq 0 \tag{6.6}$$

Per le proprietà delle forme bilineari (e in particolare la (1) della Definizione 6.1) il primo membro si spezza come

$$(v - bw) \cdot (v - bw) = v \cdot (v - bw) + (-bw) \cdot (v - bw) =$$

(applicando a ognuno dei due addendi la (2) della Definizione 6.1)

$$= v \cdot v + v \cdot (-bw) + (-bw) \cdot v + (-bw) \cdot (-bw) =$$

(portando fuori lo scalare $-b$ da tutti gli addendi, come prevede la (3) della Definizione 6.1)

$$= v \cdot v - bv \cdot w - bw \cdot v + b^2 w \cdot w =$$

(essendo il prodotto scalare una forma bilineare *simmetrica*, i due addendi centrali sono uguali)

$$= v \cdot v - 2bv \cdot w + b^2 w \cdot w.$$

Ricordandosi che avevamo denotato $b = \frac{v \cdot w}{w \cdot w}$, quest'ultima espressione diventa

[1] Possiamo mettere $w \cdot w$ a denominatore senza il rischio che sia uguale a zero in quanto il prodotto scalare è definito positivo, e quindi $w \cdot w$ si annulla solo se $w = \bar{0}$, mentre ora stiamo supponendo $w \neq \bar{0}$.

NO

$$= v \cdot v - 2\frac{v \cdot w}{w \cdot w} v \cdot w + \left(\frac{v \cdot w}{w \cdot w}\right)^2 w \cdot w =$$

$$= v \cdot v - 2\frac{(v \cdot w)^2}{w \cdot w} + \frac{(v \cdot w)^2}{(w \cdot w)^2} w \cdot w = v \cdot v - 2\frac{(v \cdot w)^2}{w \cdot w} + \frac{(v \cdot w)^2}{w \cdot w} =$$

$$v \cdot v - \frac{(v \cdot w)^2}{w \cdot w}$$

Quindi la (6.6) equivale a

$$v \cdot v - \frac{(v \cdot w)^2}{w \cdot w} \geq 0.$$

Moltiplicando entrambi i membri di questa disuguaglianza per $w \cdot w$ (possiamo farlo senza modificare il segno della disuguaglianza in quanto $w \cdot w > 0$ perché il prodotto scalare è definito positivo) si ottiene

$$(v \cdot v)(w \cdot w) - (v \cdot w)^2 \geq 0$$

ovvero

$$(v \cdot w)^2 \leq (v \cdot v)(w \cdot w).$$

Estraendo la radice a entrambi i membri di questa disuguaglianza si ha, essendo $\sqrt{(v \cdot w)^2} = |v \cdot w|$, $\sqrt{v \cdot v} = \|v\|$ e $\sqrt{w \cdot w} = \|w\|$, proprio

$$|v \cdot w| \leq \|v\| \|w\|.$$

La dimostrazione è conclusa. □

Osservazione 6.5. La richiesta che θ sia compreso tra 0 e π nella (6.3) serve a garantire che la $\cos\theta = \frac{v \cdot w}{\|v\|\|w\|}$ determini un unico θ (le equazioni del tipo $\cos\theta = a$ hanno in generale infinite soluzioni).

Oltre a giustificare la definizione di angolo tra vettori, la disuguaglianza di Cauchy-Schwarz si usa per dimostrare la seguente importante proprietà della norma, detta *disuguaglianza triangolare*.

$$\|v + w\| \leq \|v\| + \|w\| \tag{6.7}$$

Infatti, per definizione di norma, si ha

$$\|v + w\|^2 = (v + w) \cdot (v + w) =$$

(per la (1) e la (2) della Definizione 6.1)

$$= v \cdot (v + w) + w \cdot (v + w) = v \cdot v + v \cdot w + w \cdot v + w \cdot w =$$

(per la simmetria del prodotto scalare)

$$= v \cdot v + 2v \cdot w + w \cdot w = \|v\|^2 + 2v \cdot w + \|w\|^2. \tag{6.8}$$

Ma la disuguaglianza di Cauchy-Schwarz ci dice che $-\|v\|\|w\| \leq v \cdot w \leq \|v\|\|w\|$: usando la disuguaglianza di destra, dalla (6.8) si trova

$$\|v + w\|^2 = \|v\|^2 + 2v \cdot w + \|w\|^2 \leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2.$$

Il termine a secondo membro di questa disuguaglianza è il quadrato del binomio $\|v\| + \|w\|$, quindi si conclude che

$$\|v + w\|^2 \leq (\|v\| + \|w\|)^2$$

e la disuguaglianza richiesta si trova estraendo la radice di primo e secondo membro.

La disuguaglianza triangolare è una delle proprietà più importanti della norma. Altre due proprietà fondamentali sono le seguenti:

(1) $\|v\| \geq 0$ e $\|v\| = 0$ solamente se v è il vettore nullo.

Infatti, $\|v\| = \sqrt{v \cdot v}$ implica che $\|v\| \geq 0$ in quanto la radice quadrata è sempre un numero reale non negativo, e inoltre $\|v\| = \sqrt{v \cdot v} = 0$ implica $v \cdot v = 0$, che a sua volta vale solo per il vettore nullo in quanto il prodotto scalare è definito positivo.

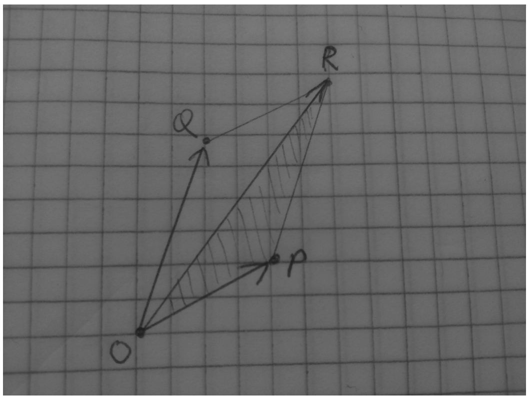
(2) per ogni vettore $v \in V$ e ogni scalare $c \in \mathbb{R}$, si ha $\|cv\| = |c|\|v\|$.

Infatti, dalla definizione di norma e dalla proprietà (3) della Definizione 6.1 si ha

$$\|cv\| = \sqrt{(cv) \cdot (cv)} = \sqrt{c^2(v \cdot v)} = \sqrt{c^2}\sqrt{v \cdot v} = |c|\|v\|.$$

Osservazione 6.6. Le proprietà (1) e (2), assieme alla disuguaglianza triangolare, vengono spesso assunte come definizione di norma nei casi in cui si è interessati a introdurre una nozione di lunghezza su uno spazio vettoriale ma non si ha bisogno di calcolare angoli (che invece richiedono un prodotto scalare).

Che la (6.7) sia una proprietà che ci si aspetta da una lunghezza si può giustificare notando che nel caso particolare dello spazio vettoriale dei vettori geometrici applicati in un punto O essa corrisponde esattamente alla disuguaglianza che afferma che il lato di un triangolo è minore della somma degli altri due (chiamata appunto in geometria euclidea disuguaglianza triangolare). Infatti, consideriamo due vettori $\vec{OP}$ e $\vec{OQ}$, e la loro somma $\vec{OR} = \vec{OP} + \vec{OQ}$, come nel disegno seguente:



La disuguaglianza triangolare della geometria euclidea, applicata al triangolo ORP, ci dice che la lunghezza $\|\vec{OR}\| = \|\vec{OP} + \vec{OQ}\|$ del lato OR è minore o uguale della somma della lunghezza $\|\vec{OP}\|$ del lato OP più la lunghezza del lato RP. Ma la lunghezza di quest'ultimo, essendo il quadrilatero $OPRQ$ un parallelogramma, è uguale alla lunghezza $\|\vec{OQ}\|$ del vettore $\vec{OQ}$. Quindi si ha proprio $\|\vec{OP} + \vec{OQ}\| \leq \|\vec{OP}\| + \|\vec{OQ}\|$, ovvero la (6.7) nel caso $v = \vec{OP}$ e $w = \vec{OQ}$.

6.2 Prodotti scalari in coordinate

Ora, analogamente a quanto già abbiamo fatto quando abbiamo studiato le applicazioni lineari (Paragrafo 4.2), mostriamo come si traduce un prodotto scalare e più in generale una forma bilineare in coordinate rispetto a una base.
Sia V un $\mathbb{K}$-spazio vettoriale e sia f una forma bilineare su V. Fissata una base $B = \{v_1, v_2, \ldots, v_n\}$, ogni vettore si identifica con la n-upla delle sue coordinate rispetto a B: dati allora due vettori v e w, denotiamo le loro coordinate rispetto a B rispettivamente $x = (x_1, x_2, \ldots, x_n)$ e $y = (y_1, y_2, \ldots, y_n)$. Questo, per definizione di coordinate, significa che $v = x_1v_1 + x_2v_2 + \cdots + x_nv_n$ e $w = y_1v_1 + y_2v_2 + \cdots + y_nv_n$. Vogliamo ora calcolare il valore $f(v, w)$ che otteniamo quando applichiamo la forma f a v e w in funzione delle coordinate x e y. Si ha

$$f(v, w) = f(x_1v_1 + x_2v_2 + \cdots + x_nv_n,\ y_1v_1 + y_2v_2 + \cdots + y_nv_n) \qquad (6.9)$$

Per le proprietà (1) e (2) della Definizione 6.1 di forma bilineare, che sostanzialmente ci dicono che la f si comporta come un prodotto che ha la proprietà distributiva, la (6.9) si spezza come somma

$$f(x_1v_1, y_1v_1) + f(x_1v_1, y_2v_2) + \cdots$$

di addendi del tipo $f(x_iv_i, y_jv_j)$, con i e j che vanno da 1 a n, ovvero, usando per comodità la notazione di sommatoria,

$$f(v, w) = \sum_{i,j=1}^{n} f(x_iv_i, y_jv_j) \tag{6.10}$$

Per la proprietà (3) della Definizione 6.1, per ogni addendo $f(x_iv_i, x_jv_j)$ lo scalare x_i che moltiplica v_i può essere portato fuori, e analogamente per lo scalare y_j che moltiplica v_j, e quindi

$$f(v, w) = \sum_{i,j=1}^{n} x_iy_jf(v_i, v_j) \tag{6.11}$$

La (6.11) è l'espressione che cercavamo, che ci permette di calcolare $f(v, w)$ una volta note le coordinate x di v e le coordinate y di w, a patto di conoscere i valori $f(v_i, v_j)$ che la f assume sui vettori della base: tali valori formano le entrate a_{ij} di una matrice A, definita proprio come $a_{ij} = f(v_i, v_j)$, che chiameremo *matrice associata alla forma bilineare f rispetto alla base* $B = \{v_1, \ldots, v_n\}$.

Abbiamo quindi concluso che ogni forma bilineare, in coordinate, si traduce sempre in un'espressione del tipo

$$f(v, w) = \sum_{i,j=1}^{n} a_{ij}x_iy_j \tag{6.12}$$

Osservazione 6.7. Nel caso in cui $V = \mathbb{K}^n$ e la base B è quella canonica formata dai vettori $v_1 = (1, 0, \ldots, 0), v_2 = (0, 1, \ldots, 0), \ldots, v_n = (0, 0, \ldots, 1)$, per ogni vettore $v = (x_1, x_2, \ldots, x_n)$ le sue coordinate rispetto a B coincidono con le sue componenti $x_1, x_2, \ldots, x_n$ in quanto $(x_1, x_2, \ldots, x_n) = x_1(1, 0, \ldots, 0) + x_2(0, 1, \ldots, 0) + \cdots + x_n(0, 0, \ldots, 1)$ (si veda pagina 44), e quindi nella (6.12) non vi è più distinzione tra i vettori v, w e le loro coordinate x, y, ovvero possiamo scrivere che su $\mathbb{K}^n$ una qualunque forma bilineare è $f(x, y) = \sum_{i,j=1}^{n} a_{ij}x_iy_j$, per qualche matrice A di entrate a_{ij}.
In particolare, la forma bilineare su $\mathbb{R}^n$

$$f(x, y) = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

in cui a_{ij} è 0 se $i \neq j$ e 1 se $i = j$ (e la matrice A è quindi la matrice identica) si vede facilmente soddisfare le proprietà $f(x, y) = f(y, x)$ e $f(x, x) = x_1^2 + x_2^2 + \cdots + x_n^2 \geq 0$ (con $f(x, x) = 0$ solo per il vettore nullo $(0, 0, \ldots, 0)$) ed è quindi un prodotto scalare, che chiameremo *prodotto scalare standard di* $\mathbb{R}^n$ (per $n = 2$ e $n = 3$ si tratta esattamente del prodotto scalare che ci è servito per calcolare lunghezze e angoli tra vettori geometrici del piano e dello spazio tridimensionale rispetto a una base ortonormale).

Ora, usando la (6.12), ci proponiamo di trovare dei criteri che, in termini della matrice a_{ij}, ci dicano quando la forma f è simmetrica e quando è definita positiva (e quindi quando è un prodotto scalare).
Per quello che riguarda la simmetria, in base alla Definizione 6.1, si ha che f è simmetrica se e solo se si ha $f(v, w) = f(w, v)$ per ogni $v, w \in V$, ovvero, tenendo conto della traduzione in coordinate data dalla (6.12), se e solo se

$$\sum_{i,j=1}^{n} a_{ij} x_i y_j = \sum_{i,j=1}^{n} a_{ij} y_i x_j$$

per ogni x e ogni y.
Allo scopo di confrontare correttamente i due membri di questa uguaglianza, vogliamo riscrivere il secondo membro in modo che in esso compaia $x_i y_j$ (come nella sommatoria a primo membro) anziché $y_i x_j$: in effetti, questo può essere fatto semplicemente scambiando gli indici[2] i e j, ottenendo quindi

$$\sum_{i,j=1}^{n} a_{ij} x_i y_j = \sum_{i,j=1}^{n} a_{ji} x_i y_j$$

Affinchè i due membri di questa uguaglianza siano uguali per ogni scelta di x e y, deve succedere che il coefficiente davanti a ogni addendo $x_i y_j$ sia lo stesso, da una parte e dall'altra, ovvero $a_{ij} = a_{ji}$ per tutti gli indici i e j, ovvero la matrice associata alla forma deve soddisfare la condizione della seguente

Definizione 6.8. Una matrice $A \in M_n(\mathbb{K})$ di entrate a_{ij} si dice *simmetrica* se si ha $a_{ij} = a_{ji}$ per ogni $i, j = 1, \ldots, n$.

Alla luce di questa definizione, abbiamo quindi che *una forma bilineare f è simmetrica se e solo se la sua matrice associata rispetto a una base è una matrice simmetrica.*
Ad esempio, la seguente forma bilineare su $\mathbb{R}^2$

[2]Infatti, in una sommatoria siamo liberi di chiamare gli indici sui quali sommiamo in qualunque modo, e quindi anche di chiamare j quello che avevamo chiamato i e viceversa chiamare i quello che avevamo denotato j.

$$f(x,y) = x_1y_1 + 2x_1y_2 - x_2y_1 + 3x_2y_2$$

è data dalla matrice

$$\det\begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}$$

(l'entrata a_{ij} è semplicemente il coefficiente di x_iy_j) e quindi non è simmetrica, in quanto $a_{12} = 2 \neq -1 = a_{21}$. In effetti, si vede ad esempio che

$$f((1,0),(0,1)) = 1 \cdot 0 + 2\ 1 \cdot 1 - 0 \cdot 0 + 3\ 0 \cdot 1 = 2$$
$$f((0,1),(1,0)) = 0 \cdot 1 + 2\ 0 \cdot 0 - 1 \cdot 1 + 3\ 1 \cdot 0 = -1$$

e quindi $f((1,0),(0,1)) \neq f((0,1),(1,0))$.

Invece, la forma

$$f(x,y) = x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2$$

è simmetrica in quanto la sua matrice

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

lo è.

Ora, daremo dei criteri per determinare quando una forma bilineare simmetrica è anche definita positiva (quindi quando è un prodotto scalare). Dal momento che ogni forma simmetrica su uno spazio vettoriale reale in coordinate si rappresenta come una forma su $\mathbb{R}^n$ del tipo $f(x,y) = \sum_{i,j=1}^n a_{ij}x_iy_j$, sarà sufficiente avere un criterio per queste ultime. Il primo è dato dal seguente

Teorema 6.9. La forma bilineare $f(x,y) = \sum_{i,j=1}^n a_{ij}x_iy_j$ su $\mathbb{R}^n$ determinata dalla matrice simmetrica $A = (a_{ij})$ è definita positiva se e solo se gli autovalori della matrice A sono tutti positivi.

Daremo un'idea della dimostrazione del teorema a pagina 280. Vediamo intanto subito un esempio.

Esempio 6.10. Consideriamo la forma bilineare simmetrica su $\mathbb{R}^2$ data da $f(x,y) = 3x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2$, i cui coefficienti determinano la matrice $A = \begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}$. Gli autovalori di A sono le soluzioni dell'equazione caratteristica

$$\det\begin{pmatrix} 3-\lambda & -1 \\ -1 & 2-\lambda \end{pmatrix} = (3-\lambda)(2-\lambda) - 1 = \lambda^2 - 5\lambda + 5 = 0$$

ovvero, usando la formula risolutiva per le equazioni di secondo grado, $\lambda_1 = \frac{5+\sqrt{5}}{2}$ e $\lambda_2 = \frac{5-\sqrt{5}}{2}$. Essendo $\lambda_1 > 0$ e $\lambda_2 > 0$, in base al Teorema 6.9 la forma è definita positiva, e quindi è un prodotto scalare su $\mathbb{R}^2$.

Consideriamo invece la forma bilineare simmetrica su $\mathbb{R}^2$ data da $f(x, y) = x_1y_1 + 2x_1y_2 + 2x_2y_1 + x_2y_2$, i cui coefficienti determinano la matrice $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$. Gli autovalori di A sono le soluzioni dell'equazione caratteristica

$$\det \begin{pmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{pmatrix} = (1-\lambda)^2 - 4 = \lambda^2 - 2\lambda - 3 = 0$$

ovvero, usando la formula risolutiva per le equazioni di secondo grado, $\lambda_1 = -1$ e $\lambda_2 = 3$. Dal momento che gli autovalori non sono entrambi positivi, in base al Teorema 6.9 la forma non è definita positiva.
Questo significa che non è vero che $f(x, x) = x_1^2 + 4x_1x_2 + x_2^2$ è maggiore di zero per ogni vettore non nullo e uguale a zero solo per il vettore nullo $(0, 0)$.
A conferma di ciò, prendiamo ad esempio $x = (1, -1)$: allora

$$f(x, x) = 1^2 + 4 \cdot 1 \cdot (-1) + (-1)^2 = -2 < 0.$$

Un secondo criterio usato per capire se una forma bilineare simmetrica è definita positiva si basa sul concetto di *sottomatrice principale*.
Ricordiamo che una sottomatrice di una matrice data A è una matrice che si ottiene eliminando da A delle righe e delle colonne e mantenendo le rimanenti.
Ad esempio, se nella matrice

$$A = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 3 & -1 \\ -2 & -1 & 5 \end{pmatrix} \tag{6.13}$$

scegliamo la prima e la terza riga e la seconda e la terza colonna (cancellando quindi la seconda riga e la prima colonna) otteniamo la sottomatrice quadrata di ordine due $\begin{pmatrix} 0 & -2 \\ -1 & 5 \end{pmatrix}$.
Tra tutte le sottomatrici quadrate di una matrice A, si dicono *principali* quelle che si ottengono scegliendo un insieme di righe e prendendo poi le colonne corrispondenti agli stessi indici: ad esempio, scegliendo sempre nella matrice (6.13) la prima e la terza riga e la prima e la terza colonna si ottiene la sottomatrice quadrata principale $\begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix}$; la sottomatrice trovata sopra, sceligendo la prima e la terza riga ma la seconda e la terza colonna, invece, non è principale.

Tra tutte le sottomatrici principali di una matrice quadrata di ordine n, si chiamano *sottomatrici principali di nord-ovest* [3] quelle che si ottengono scegliendo le prime k righe e le prime k colonne, con k un numero tra 1 e n.

Ad esempio, le sottomatrici principali di nord-ovest della matrice (6.13) sono

$$(1), \quad \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 0 & -2 \\ 0 & 3 & -1 \\ -2 & -1 & 5 \end{pmatrix} \tag{6.14}$$

la prima ottenuta scegliendo la prima riga e la prima colonna, la seconda le prime due righe e le prime due colonne, la terza scegliendo le tre righe e le tre colonne (si osservi quindi che tra le sottomatrici principali di una matrice quadrata A vi è sempre A stessa).
Siamo ora pronti a enunciare il seguente criterio

Teorema 6.11. La forma bilineare $f(x,y) = \sum_{i,j=1}^{n} a_{ij}x_iy_j$ su $\mathbb{R}^n$ determinata dalla matrice simmetrica $A = (a_{ij})$ è definita positiva se e solo se tutte le sottomatrici principali di nord-ovest di A hanno determinante > 0

Ad esempio, la prima matrice dell'Esempio 6.10, ovvero $A = \begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}$, verifica il criterio del Teorema 6.11, in quanto i determinanti delle sue matrici principali di nord-ovest sono $\det(3) = 3 > 0$ e $\det \begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix} = 5 > 0$, mentre la seconda, ovvero $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, non verifica il criterio in quanto $\det(1) = 1 > 0$ e $\det \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} = -3 < 0$.

Se una forma bilineare simmetrica $f : V \times V \to \mathbb{R}$ su uno spazio vettoriale reale V non è definita positiva, essa può essere:

(1) *semidefinita positiva* se si ha $f(v,v) \geq 0$ per ogni $v \in V$, ed esiste almeno un vettore $v \neq \bar{0}$ per cui $f(v,v) = 0$

(2) *definita negativa* se si ha $f(v,v) \leq 0$ per ogni $v \in V$, e $f(v,v) = 0$ solo se $v = \bar{0}$

(3) *semidefinita negativa* se si ha $f(v,v) \leq 0$ per ogni $v \in V$, ed esiste almeno un vettore $v \neq \bar{0}$ per cui $f(v,v) = 0$

[3] Il motivo di tale nome è che si tratta delle sottomatrici che si ottengono, partendo dall'entrata in alto a sinistra della matrice (quindi, in un'ipotetica bussola, a nord-ovest) e aggiungendo volta per volta una riga e una colonna.

(4) *indefinita* se esistono almeno un $v \in V$ per cui $f(v, v) > 0$ e un $w \in V$ per cui $f(w, w) < 0$

Analisi Matematica 2

Capire se una forma è definita positiva o, in caso contrario, di quale dei tipi (1)-(4) essa sia, è importante ad esempio in analisi matematica, dove si usano queste nozioni per determinare massimi, minimi e altri tipi di punti critici di una funzione di più variabili: più precisamente, ma senza entrare in dettagli, ad ogni punto critico di una funzione $F : \mathbb{R}^n \to \mathbb{R}$ viene associata una opportuna matrice simmetrica di ordine n, e il tipo di forma f determinato da questa matrice ci dice se tale punto critico è un massimo, un minimo, o di altro tipo (ad esempio, sarà un minimo se la matrice è definita positiva). Un altro campo in cui le forme non definite positive svolgono un ruolo fondamentale è quello della teoria della relatività ristretta, dove ad esempio la forma bilineare simmetrica $f(x, y) = x_1y_1 + x_2y_2 + x_3y_3 - x_4y_4$ su $\mathbb{R}^4$, che è indefinita, serve a formalizzare matematicamente la struttura dello spazio-tempo.

Ci chiediamo a questo punto se esistono criteri, analoghi dei teoremi 6.9 e 6.11 visti sopra per le forme definite positive, per determinare se una forma simmetrica è di uno dei tipi (1)-(4) detti. La risposta è affermativa e più precisamente abbiamo i seguenti risultati, che non dimostriamo.

Teorema 6.12. La forma bilineare $f(x, y) = \sum_{i,j=1}^{n} a_{ij}x_iy_j$ su $\mathbb{R}^n$ determinata dalla matrice simmetrica $A = (a_{ij})$ è

(1) semidefinita positiva se e solo se tutti gli autovalori di A sono ≥ 0 e esiste almeno un autovalore nullo

(2) definita negativa se e solo se tutti gli autovalori di A sono < 0

(3) semidefinita negativa se e solo se tutti gli autovalori di A sono ≤ 0 e esiste almeno un autovalore nullo

(4) indefinita se e solo se A ammette almeno un autovalore > 0 e un autovalore < 0.

Teorema 6.13. La forma bilineare $f(x, y) = \sum_{i,j=1}^{n} a_{ij}x_iy_j$ su $\mathbb{R}^n$ determinata dalla matrice simmetrica $A = (a_{ij})$ è

(1) semidefinita positiva se e solo se tutte le sottomatrici principali di A hanno determinante ≥ 0 (e ne esiste una con determinante nullo)

(2) definita negativa se e solo se tutte le sue sottomatrici principali di nord-ovest di ordine dispari hanno determinante < 0 e quelle principali di nord-ovest di ordine pari hanno determinante > 0

(3) semidefinita negativa se e solo se tutte le sue sottomatrici principali di ordine dispari hanno determinante ≤ 0 e quelle principali di ordine pari hanno determinante ≥ 0 (e ne esiste una con determinante nullo)

(4) indefinita se e solo se non si verifica nessuna delle precedenti.

Vediamo un esempio per illustrare il Teorema 6.13: consideriamo la forma bilineare su $\mathbb{R}^2$ data da $f(x,y) = x_1y_1+2x_1y_2+2x_2y_1+4x_2y_2$, i cui coefficienti determinano la matrice $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$: i determinanti delle sue matrici principali sono $\det(1) = 1 > 0$, $\det(4) = 4 > 0$ e $\det \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} = 0$, quindi siamo nel caso (1) e la forma è semidefinita positiva. Per verificare che effettivamente essa soddisfa la condizione espressa nella definizione, notiamo che

$$f(x,x) = x_1^2 + 4x_1x_2 + 4x_2^2 = (x_1 + 2x_2)^2$$

e quindi $f(x,x) \geq 0$ per ogni $x = (x_1, x_2)$ ma esistono vettori non nulli per cui $f(x,x) = 0$: infatti, essendo $f(x,x) = (x_1 + 2x_2)^2$, basta scegliere un vettore $x = (x_1, x_2)$ per cui $x_1 + 2x_2 = 0$, ad esempio $x = (-2, 1)$.

Osservazione 6.14. Si noti che nel caso di una forma bilineare simmetrica f su $\mathbb{R}^2$ il Teorema 6.13 risulta particolarmente semplice da applicare, in quanto in tal caso la matrice associata è una matrice $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ di ordine 2 e abbiamo solamente tre sottomatrici principali: (a_{11}), (a_{22}) e A stessa (quelle di nordovest sono quindi solo (a_{11}) e A stessa).
Abbiamo quindi

(1) f è definita positiva se e solo se $a_{11} > 0$ e $\det(A) > 0$

(2) f è definita negativa se e solo se $a_{11} < 0$ e $\det(A) > 0$

(3) f è semidefinita positiva se e solo se $a_{11} \geq 0$, $a_{22} \geq 0$ e $\det(A) \geq 0$ (e almeno una di queste quantità è zero)

(4) f è semidefinita negativa se e solo se $a_{11} \leq 0$, $a_{22} \leq 0$ e $\det(A) \geq 0$ (e almeno una di queste quantità è zero)

(5) f è indefinita in tutti gli altri casi

Osservazione 6.15. Quando poniamo $y = x$ in una forma bilineare simmetrica $f(x,y) = \sum_{i,j=1}^{n} a_{ij}x_iy_j$ otteniamo una somma di monomi di secondo

grado[4] in $x_1, \dots, x_n$: una tale espressione si dice anche *forma quadratica di* $\mathbb{R}^n$. Ad esempio, se $f(x,y) = x_1y_1 - 2x_1y_2 - 2x_2y_1 + 3x_2y_2$, ponendo $y = x$ si ottiene la forma quadratica $q(x) = f(x,x) = x_1^2 - 4x_1x_2 + 3x_2^2$ (dal momento che la forma quadratica dipende solo da una n-upla x e non da due n-uple x e y come la forma bilineare di partenza si può scrivere $q(x)$ invece che $f(x,x)$). Data una forma quadratica $q(x)$, si dice che essa è definita positiva, semidefinita positiva, definita negativa, semidefinita negativa, indefinita se $q(x)$ soddisfa le stesse condizioni che, nelle analoghe definizioni per una forma bilineare simmetrica f, deve soddisfare $f(x,x)$ (ad esempio, q si dice definita positiva se $q(x) \geq 0$ e $q(x) = 0$ solo se x è il vettore nullo, e analogamente per le altre).
Per determinare il tipo di una forma quadratica $q(x)$ basta risalire alla forma bilineare simmetrica $f(x,y)$ da cui proviene (ovvero tale che $q(x) = f(x,x)$) e applicare alla matrice di f i criteri visti nei teoremi 6.9, 6.11, 6.12 e 6.13. Per fare ciò, basta guardare i coefficienti dei monomi che compongono la forma $q(x)$, ma bisogna prestare attenzione al fatto seguente: un monomio di $q(x)$ del tipo cx_ix_j con $i \neq j$ è risultato della somma dei *due* monomi $a_{ij}x_iy_j$ e $a_{ji}x_jy_i$ di f quando si pone $y = x$, ovvero $cx_ix_j = a_{ij}x_ix_j + a_{ji}x_jx_i = 2a_{ij}x_ix_j$ (dove abbiamo usato il fatto che f è simmetrica e quindi $a_{ij} = a_{ji}$). Quindi l'entrata a_{ij} della matrice della forma da cui proviene q è uguale a $\frac{c}{2}$, dove c è il coefficiente di x_ix_j in $q(x)$ (e sarà anche $a_{ji} = \frac{c}{2}$, visto che A è simmetrica). Ad esempio, la forma quadratica $q(x) = x_1^2 - 4x_1x_2 + 3x_2^2$, come abbiamo detto, viene dalla forma bilineare $f(x,y) = x_1y_1 - 2x_1y_2 - 2x_2y_1 + 3x_2y_2$ che ha come matrice $\begin{pmatrix} 1 & -2 \\ -2 & 3 \end{pmatrix}$, e come si vede le entrate a_{12} e a_{21} sono uguali al coefficiente -4 di x_1x_2 in $q(x)$ diviso per due. Per quello che riguarda invece le entrate della matrice che stanno sulla diagonale, come si vede in questo stesso esempio esse coincidono con i coefficienti dei quadrati x_i^2 che compaiono in $q(x)$: questo perché gli addendi cx_i^2 in $q(x)$ si ottengono quando si pone $y = x$ nell'unico monomio $a_{ii}x_iy_i$ della forma bilineare, e quindi $a_{ii} = c$.

Per fare un ulteriore esempio, consideriamo la forma quadratica su $\mathbb{R}^3$ data da $q(x) = x_1^2 - 7x_3^2 - 5x_1x_2 + 6x_2x_3$: allora si ha che $q(x) = f(x,x)$, dove f è la forma bilineare $f(x,y) = \sum_{i,j=1}^{n} a_{ij}x_iy_j$ con matrice A data da $\begin{pmatrix} 1 & -\frac{5}{2} & 0 \\ -\frac{5}{2} & 0 & 3 \\ 0 & 3 & -7 \end{pmatrix}$, e per capire il tipo di q basta applicare i criteri visti a tale matrice.

Infine, segnaliamo che, data una matrice simmetrica A con entrate a_{ij}, si dice che A è definita positiva, semidefinita positiva, definita negativa,

[4] Ovvero un *polinomio omogeneo*, tale cioè che tutti i monomi hanno lo stesso grado

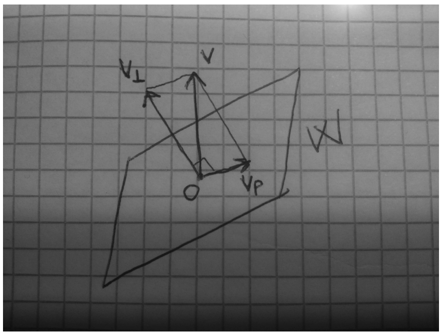
semidefinita negativa, indefinita se lo è la forma $f(x, y) = \sum_{i,j=1}^{n} a_{ij} x_i y_j$.

6.3 Proiezioni ortogonali

Dato uno spazio vettoriale euclideo V (cioè dotato di un prodotto scalare) e un suo sottospazio W, si ha che ogni vettore $v \in V$ può essere decomposto come somma

$$v = v_P + v_\perp \tag{6.15}$$

di un vettore v_P che appartiene a W e di un vettore $v_\perp$ ortogonale a ogni vettore di W (si dice anche che $v_\perp$ è *ortogonale a* W), come nel seguente disegno che rappresenta la decomposizione di un vettore v dello spazio tridimensionale V_O^3 dei vettori applicati su un suo sottospazio W di dimensione 2 (ovvero un piano che passa per O):



La componente v_P della decomposizione si dice la *proiezione ortogonale di* v *su* W, mentre $v_\perp$ è la sua *componente ortogonale*.
Vogliamo ora una formula per calcolare la proiezione v_P dati v e W.

Iniziamo con il caso più semplice, quello in cui il sottospazio W ha dimensione 1, ovvero è generato da un solo vettore w.
Dal momento che per definizione la proiezione v_P appartiene a W, sarà sicuramente $v_P = cw$ per qualche $c \in \mathbb{R}$, ovvero la decomposizione (6.15) si può riscrivere come

$$v = cw + v_\perp \tag{6.16}$$

Ma allora eseguendo il prodotto scalare tra v e w otteniamo

$$v \cdot w = (cw + v_\perp) \cdot w = cw \cdot w + v_\perp \cdot w = cw \cdot w \tag{6.17}$$

dove nella seconda uguaglianza abbiamo sfruttato le proprietà delle forme bilineari e nella terza il fatto che $v_\perp$ è per definizione ortogonale a ogni vettore di W, e in particolare a w, ovvero $v_\perp \cdot w = 0$. Dividendo entrambi i membri della (6.17) per $w \cdot w$ e ricordando la definizione di norma otteniamo allora

$$c = \frac{v \cdot w}{w \cdot w} = \frac{v \cdot w}{\|w\|^2}$$

Ricordando quindi che $v_P = cw$, abbiamo finalmente la formula per la proiezione ortogonale di un vettore v su un sottospazio W di dimensione 1 generato da w:

$$v_P = \frac{v \cdot w}{\|w\|^2} w \tag{6.18}$$

e combinando con la (6.15) abbiamo anche l'espressione della componente ortogonale $v_\perp$:

$$v_\perp = v - \frac{v \cdot w}{\|w\|^2} w \tag{6.19}$$

Ora, vediamo come il problema si complica quando supponiamo in generale che W abbia dimensione $k > 1$: sia $w_1, w_2, \dots, w_k$ una base di W. Come prima, dal momento che per definizione la proiezione v_P appartiene a W, essa si potrà sicuramente scrivere come combinazione dei vettori della base di W, ovvero $v_P = c_1 w_1 + c_2 w_2 + \cdots + c_k w_k$, con $c_1, c_2, \dots, c_k \in \mathbb{R}$, ovvero la decomposizione (6.15) si può stavolta riscrivere come

$$v = c_1 w_1 + c_2 w_2 + \cdots + c_k w_k + v_\perp \tag{6.20}$$

Se eseguiamo allora il prodotto scalare tra v e ciascuno dei vettori di base $w_1, w_2, \dots, w_k$, otteniamo

$$\begin{gathered}
v \cdot w_1 = (c_1 w_1 + c_2 w_2 + \cdots + c_k w_k + v_\perp) \cdot w_1 = c_1 w_1 \cdot w_1 + c_2 w_2 \cdot w_1 + \cdots + c_k w_k \cdot w_1 + v_\perp \cdot w_1 \\
v \cdot w_2 = (c_1 w_1 + c_2 w_2 + \cdots + c_k w_k + v_\perp) \cdot w_2 = c_1 w_1 \cdot w_2 + c_2 w_2 \cdot w_2 + \cdots + c_k w_k \cdot w_2 + v_\perp \cdot w_2 \\
\vdots \\
v \cdot w_1 = (c_1 w_1 + c_2 w_2 + \cdots + c_k w_k + v_\perp) \cdot w_k = c_1 w_1 \cdot w_k + c_2 w_2 \cdot w_k + \cdots + c_k w_k \cdot w_k + v_\perp \cdot w_k
\end{gathered} \tag{6.21}$$

(dove come nel caso di dimensione 1 abbiamo sfruttato le proprietà delle forme bilineari). Tenendo conto del fatto che $v_\perp$ è per definizione ortogonale a ogni vettore di W, e in particolare a $w_1, w_2, \dots, w_k$, ovvero $v_\perp \cdot w_1 = 0, v_\perp \cdot w_2 = 0, \dots, v_\perp \cdot w_k = 0$, le (6.21) diventano il seguente sistema di equazioni

$$\begin{cases} c_1 w_1 \cdot w_1 + c_2 w_2 \cdot w_1 + \cdots + c_k w_k \cdot w_1 = v \cdot w_1 \\ c_1 w_1 \cdot w_2 + c_2 w_2 \cdot w_2 + \cdots + c_k w_k \cdot w_2 = v \cdot w_2 \\ \vdots \\ c_1 w_1 \cdot w_k + c_2 w_2 \cdot w_k + \cdots + c_k w_k \cdot w_k = v \cdot w_k \end{cases} \tag{6.22}$$

Si vede allora che mentre per la proiezione su un sottospazio di dimensione 1 avevamo bisogno di ricavare un solo coefficiente c, che abbiamo ottenuto facilmente, per la proiezione su un sottospazio di dimensione k dobbiamo a priori ricavare i k coefficienti $c_1, c_2, \dots, c_k$, e per farlo dobbiamo risolvere un sistema di equazioni.
In realtà, come vedremo, il problema può essere semplificato, ma prima di entrare nei dettagli vediamo un esempio per illustrare il sistema (6.22).

Esempio 6.16. Consideriamo $\mathbb{R}^3$ dotato del prodotto scalare standard $x \cdot y = x_1 y_1 + x_2 y_2 + x_3 y_3$ e il suo sottospazio bidimensionale W generato dai vettori $w_1 = (2, 1, 0)$, $w_2 = (0, -1, 1)$. Determiniamo la proiezione del vettore $v = (1, 1, 1)$ su W. Per $k = 2$, il sistema (6.22) diventa

$$\begin{cases} c_1 w_1 \cdot w_1 + c_2 w_2 \cdot w_1 = v \cdot w_1 \\ c_1 w_1 \cdot w_2 + c_2 w_2 \cdot w_2 = v \cdot w_2 \end{cases}$$

ovvero, calcolando i prodotti scalari

$$\begin{cases} 5c_1 - c_2 = 3 \\ -c_1 + 2c_2 = 0 \end{cases}$$

che risolto (ad esempio con il metodo di Cramer) ci dà $c_1 = \frac{2}{3}$, $c_2 = \frac{1}{3}$.
Quindi la proiezione è

$$v_P = c_1 w_1 + c_2 w_2 = \frac{2}{3}(2, 1, 0) + \frac{1}{3}(0, -1, 1) = (\frac{4}{3}, \frac{1}{3}, \frac{1}{3})$$

Ora, come abbiamo accennato, si può fare un'ipotesi aggiuntiva per semplificare il problema: se i vettori della base $w_1, w_2, \dots, w_k$ di W fossero tutti ortogonali tra loro, ovvero $w_i \cdot w_j = 0$ ogni volta che prendiamo due indici i e j tra 1 e k, con $i \neq j$, allora il sistema (6.22) diventerebbe semplicemente

$$\begin{cases} c_1 w_1 \cdot w_1 = v \cdot w_1 \\ c_2 w_2 \cdot w_2 = v \cdot w_2 \\ \vdots \\ c_k w_k \cdot w_k = v \cdot w_k \end{cases} \tag{6.23}$$

da cui possiamo trovare $c_1, c_2, \dots, c_k$ indipendentemente una dall'altra come nel caso di dimensione 1, ovvero

$$c_1 = \frac{v \cdot w_1}{w_1 \cdot w_1} = \frac{v \cdot w_1}{\|w_1\|^2}$$
$$c_2 = \frac{v \cdot w_2}{w_2 \cdot w_2} = \frac{v \cdot w_2}{\|w_2\|^2}$$
$$\vdots$$
$$c_k = \frac{v \cdot w_k}{w_k \cdot w_k} = \frac{v \cdot w_k}{\|w_k\|^2}$$

Ricordando che avevamo posto $v_P = c_1 w_1 + c_2 w_2 + \cdots + c_k w_k$ (e che $v_\perp = v - v_P$) possiamo riassumere tutto nel seguente

Teorema 6.17. Sia V uno spazio vettoriale euclideo e sia W un sottospazio vettoriale di dimensione k generato da una base $w_1, w_2, \ldots, w_k$ composta da vettori ortogonali tra loro. Allora, dato un vettore $v \in V$, la proiezione v_P di v su W e la componente ortogonale $v_\perp$ sono dati dalla formule seguenti:

$$v_P = \frac{v \cdot w_1}{\|w_1\|^2} w_1 + \frac{v \cdot w_2}{\|w_2\|^2} w_2 + \cdots + \frac{v \cdot w_k}{\|w_k\|^2} w_k \tag{6.24}$$

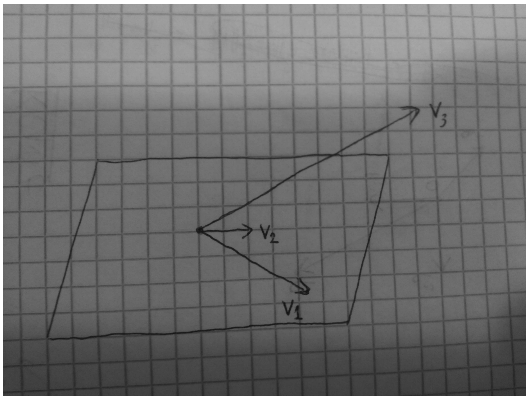
$$v_\perp = v - \frac{v \cdot w_1}{\|w_1\|^2} w_1 - \frac{v \cdot w_2}{\|w_2\|^2} w_2 - \cdots - \frac{v \cdot w_k}{\|w_k\|^2} w_k \tag{6.25}$$

Una base formata da vettori ortogonali tra loro si dice *base ortogonale*. Il teorema mostra uno dei tanti vantaggi di usare basi di questo tipo (ne vedremo un altro più avanti nella Proposizione 6.20) quando si è in uno spazio vettoriale euclideo.

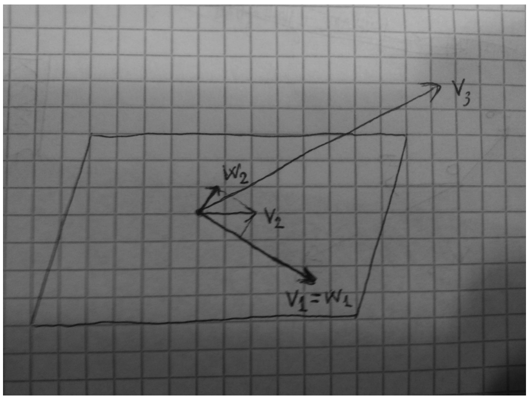
Ci chiediamo allora come si può, dato uno spazio vettoriale euclideo V (o un suo sottospazio), costruire una sua base ortogonale. La risposta è data dal cosiddetto *procedimento di ortogonalizzazione di Gram-Schmidt*.

Tale metodo consiste nel partire da una qualunque base $\{v_1, v_2, \ldots, v_n\}$ di V e costruire dei nuovi vettori $w_1, w_2, \ldots, w_n$ da sostituire al posto dei v_i e che costituiranno una base ortogonale.

Per illustrare il procedimento, supponiamo a titolo di esempio che V sia lo spazio V_O^3 dei vettori applicati e la base sia data da v_1, v_2, v_3 come nel disegno seguente



(dove per chiarezza abbiamo messo in evidenza il piano su cui si trovano v_1 e v_2). Dobbiamo sostituire i vettori v_1, v_2, v_3 con tre nuovi vettori w_1, w_2, w_3 che formino una base ortogonale di V_O^3.
Iniziamo ponendo $w_1 = v_1$ (ovvero, non modifichiamo il primo vettore v_1 della base data: questo ha senso in quanto l'obiettivo è trovare una base ortogonale, quindi l'importante è che i vettori seguenti siano ortogonali a w_1 e tra loro).
Per ottenere w_2, decomponiamo v_2 come nel disegno seguente



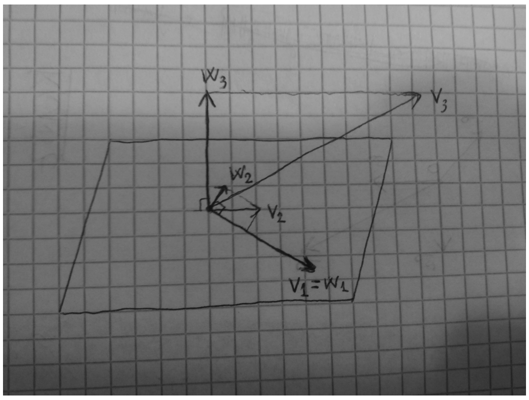
come somma della sua proiezione sul sottospazio di dimensione 1 generato da w_1 più la componente ortogonale, che è data dalla (6.19): scegliendo w_2 uguale proprio a tale componente, e ponendo cioè

$$w_2 = v_2 - \frac{v_2 \cdot w_1}{\|w_1\|^2} w_1 \tag{6.26}$$

abbiamo ottenuto due vettori w_1 e w_2 che giacciono sullo stesso piano di v_1 e v_2 e che sono tra loro ortogonali.

Ora non ci resta che sostituire il terzo vettore v_3 con un vettore w_3 che sia ortogonale sia a w_1 che a w_2. Ma questo può essere fatto, analogamente a quanto fatto per ottenere w_2, scegliendo w_3 uguale alla componente ortogonale di v_3 stavolta rispetto al sottospazio generato da w_1 e w_2 (rappresentato dal piano in figura): poichè w_1 e w_2 sono ortogonali grazie ai primi due passi del procedimento, essi sono una base ortogonale di tale sottospazio e siamo autorizzati a usare la formula (6.25) che in questo caso particolare diventa

$$w_3 = v_3 - \frac{v_3 \cdot w_1}{\|w_1\|^2} w_1 - \frac{v_3 \cdot w_2}{\|w_2\|^2} w_2 \tag{6.27}$$



Abbiamo quindi ottenuto la base ortogonale $\{w_1, w_2, w_3\}$ cercata.
Osservando la struttura delle formule (6.26) e (6.27) deduciamo che, se ci fosse un quarto vettore v_4, per ottenere al suo posto un vettore ortogonale ai primi tre w_1, w_2, w_3 già ottenuti basterebbe sostituirlo con la sua componente ortogonale rispetto al sottospazio generato da w_1, w_2, w_3, ovvero

$$w_4 = v_4 - \frac{v_4 \cdot w_1}{\|w_1\|^2} w_1 - \frac{v_4 \cdot w_2}{\|w_2\|^2} w_2 - \frac{v_4 \cdot w_3}{\|w_3\|^2} w_3$$

(dove per calcolare tale componente ortogonale abbiamo potuto usare la (6.25) in quanto i vettori w_1, w_2, w_3 erano una base ortogonale del sottospazio da loro generato) e così via: possiamo dire che in generale si ha allora che il procedimento consiste nel partire da $w_1 = v_1$ e sostituire via via $v_2, v_3, \ldots, v_n$ in base alla formula

$$w_k = v_k - \frac{v_k \cdot w_1}{\|w_1\|^2} w_1 - \frac{v_k \cdot w_2}{\|w_2\|^2} w_2 - \cdots - \frac{v_k \cdot w_{k-1}}{\|w_{k-1}\|^2} w_{k-1}$$

per $k = 2, \ldots, n$, ovvero, usato la notazione di sommatoria

$$w_k = v_k - \sum_{j=1}^{k-1} \frac{v_k \cdot w_j}{\|w_j\|^2} w_j$$

Osservazione 6.18. Una volta che si ha a disposizione una base $\{w_1, \dots, w_n\}$ di vettori tra loro ortogonali, ottenuta mediante il procedimento di ortogonalizzazione di Gram-Schmidt, se volessimo ottenere una *base ortonormale*, cioè nella quale i vettori oltre a essere ortogonali hanno anche tutti lunghezza 1, basterebbe "aggiustare" le lunghezze dei vettori $w_1, \dots, w_n$ dividendoli per la loro lunghezza, ovvero

$$\left\{\frac{w_1}{\|w_1\|}, \frac{w_2}{\|w_2\|}, \dots, \frac{w_n}{\|w_n\|}\right\}$$

sarebbe la base ortonormale cercata.
Infatti, osserviamo che per la proprietà (2) della norma vista a pagina 231, se divido un vettore v per la sua norma $\|v\|$ ho che il nuovo vettore così ottenuto ha norma

$$\left\|\frac{v}{\|v\|}\right\| = \left\|\frac{1}{\|v\|}v\right\| = \frac{1}{\|v\|}\|v\| = 1.$$

Un vettore di norma 1 si chiama anche *versore*; quando si divide un vettore v per la sua norma allo scopo di farlo diventare di lunghezza 1 si dice anche che si è *normalizzato* v.

Esempio 6.19. Supponiamo di voler calcolare una base ortonormale del sottospazio W di $\mathbb{R}^4$ (dotato del prodotto scalare standard $x \cdot y = x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4$) generato dai vettori

$$v_1 = (1, 0, 1, 0),\ v_2 = (0, 2, -1, 1),\ v_3 = (1, 1, 1, 1)$$

Applichiamo dapprima ai tre vettori il procedimento di ortogonalizzazione di Gram-Schmidt per ottenere una base ortogonale w_1, w_2, w_3. Si ha

$$w_1 = v_1 = (1, 0, 1, 0)$$

$$w_2 = v_2 - \frac{v_2 \cdot w_1}{\|w_1\|^2}w_1 = (0, 2, -1, 1) - \frac{0 \cdot 1 + 2 \cdot 0 + (-1) \cdot 1 + 1 \cdot 0}{1^2 + 0^2 + 1^2 + 0^2}(1, 0, 1, 0) =$$

$$= (0, 2, -1, 1) - \frac{-1}{2}(1, 0, 1, 0) = \left(\frac{1}{2}, 2, -\frac{1}{2}, 1\right)$$

e infine

$$w_3 = v_3 - \frac{v_3 \cdot w_1}{\|w_1\|^2}w_1 - \frac{v_3 \cdot w_2}{\|w_2\|^2}w_2 =$$

$$= (1,1,1,1) - \frac{1\cdot 1 + 1\cdot 0 + 1\cdot 1 + 1\cdot 0}{1^2+0^2+1^2+0^2}(1,0,1,0) - \frac{1\cdot\frac{1}{2} + 1\cdot 2 + 1\cdot(-\frac{1}{2}) + 1\cdot 1}{(\frac{1}{2})^2+2^2+(-\frac{1}{2})^2+1^2}\left(\frac{1}{2},2,-\frac{1}{2},1\right) =$$

$$= (1,1,1,1) - \frac{2}{2}(1,0,1,0) - \frac{3}{\frac{11}{2}}\left(\frac{1}{2},2,-\frac{1}{2},1\right) = \left(-\frac{3}{11},-\frac{1}{11},\frac{3}{11},\frac{5}{11}\right)$$

I vettori

$$w_1 = (1,0,1,0),\ w_2 = \left(\frac{1}{2},2,-\frac{1}{2},1\right),\ w_3 = \left(-\frac{3}{11},-\frac{1}{11},\frac{3}{11},\frac{5}{11}\right)$$

così ottenuti formano dunque una base ortogonale del sottospazio W.
Osserviamo che se due vettori v e v' sono ortogonali (ovvero $v\cdot v' = 0$) lo sono sicuramente anche due qualunque loro multipli cv e $c'v'$, in quanto per le proprietà del prodotto scalare si ha $(cv)\cdot(c'v') = cc'(v\cdot v) = cc'0 = 0$: quindi, se vogliamo ad esempio ottenere una base ortogonale più semplice per i calcoli, possiamo moltiplicare il vettore w_2 per 2 e il vettore w_3 per 11, ottenendo

$$w_1 = (1,0,1,0),\quad w_2 = (1,4,-1,2),\quad w_3 = (-3,-1,3,5).$$

Infine, per ottenere una base *ortonormale* di W, ovvero in cui oltre a essere ortogonali i vettori hanno norma 1, basta come abbiamo detto sopra dividere ognuno dei vettori ottenuti per la sua norma. Si ha

$$\|w_1\| = \sqrt{1^2+0^2+1^2+0^2} = \sqrt{2}$$
$$\|w_2\| = \sqrt{1^2+4^2+(-1)^2+2^2} = \sqrt{22}$$
$$\|w_3\| = \sqrt{(-3)^2+(-1)^2+3^2+5^2} = \sqrt{44}$$

Una base ortonormale del sottospazio W è quindi data da

$$w_1' = \frac{1}{\sqrt{2}}(1,0,1,0),\quad w_2' = \frac{1}{\sqrt{22}}(1,4,-1,2),\quad w_3' = \frac{1}{\sqrt{44}}(-3,-1,3,5).$$

Come abbiamo detto, usare basi ortogonali e in particolare ortonormali presenta vari vantaggi. Uno di questi consiste nel calcolo delle coordinate:

Proposizione 6.20. Sia V uno spazio vettoriale euclideo e sia $B = \{v_1,\dots,v_n\}$ una base ortonormale di V. Per ogni vettore $v \in V$, le coordinate $x_1,\dots,x_n$ di v rispetto a B sono date da $x_i = v\cdot v_i$.

Proof. Per definizione, $x_1, \dots, x_n$ sono le coordinate di v rispetto a $\{v_1, \dots, v_n\}$ se e solo se

$$v = x_1 v_1 + \dots + x_n v_n.$$

Quindi, per le proprietà di bilinearità del prodotto scalare si trova

$$v \cdot v_i = (x_1 v_1 + \dots + x_n v_n) \cdot v_i = x_1 v_1 \cdot v_i + \dots + x_n v_n \cdot v_i$$

Ma, essendo la base ortonormale, i prodotti scalari $v_1 \cdot v_i, \dots, v_n \cdot v_i$ che compaiono nella somma a secondo membro di questa uguaglianza sono tutti zero tranne $v_i \cdot v_i$ che, essendo uguale alla norma di v_i al quadrato, è uguale a 1. Quindi

$$v \cdot v_i = x_1 v_1 \cdot v_i + \dots + x_n v_n \cdot v_i = x_i v_i \cdot v_i = x_i$$

che è quello che volevamo dimostrare. $\square$

Matematica Applicata

Osservazione 6.21. Risultati analoghi alla Proposizione 6.20 sono validi anche negli spazi vettoriali infinito-dimensionali, per esempio gli spazi di funzioni, che come abbiamo detto all'inizio del capitolo possono essere dotati di prodotti scalari integrali del tipo $f \cdot g = \int_a^b fg dx$.
In tale contesto, avere a disposizione basi ortornormali si dimostra di grandissima utilità: ad esempio, le funzioni periodiche $\sin(nx)$ con n intero maggiore o uguale a 1 e $\cos(mx)$ con m intero maggiore o uguale a 0 sono ortogonali rispetto al prodotto scalare integrale sull'intervallo $[0, 2\pi]$ e quindi, normalizzate, danno una base ortonormale per lo spazio delle funzioni periodiche su tale intervallo, sotto alcune ipotesi (ad esempio, basta che f sia derivabile). Quindi, ogni tale f può essere scritta come loro combinazione lineare, e per trovare i coefficienti della combinazione (ovvero le coordinate) basta, analogamente a quanto affermato nella Proposizione 6.20, calcolare i prodotti scalari tra f e $\sin(nx)$ o $\cos(mx)$, ovvero integrali del tipo $\int f(x)\sin(nx)dx$ e $\int f(x)\cos(mx)dx$. L'importanza di questo esempio consiste nelle numerose applicazioni fisiche e ingegneristiche, visto che ad esempio le onde sonore sono rappresentate proprio da funzioni periodiche.

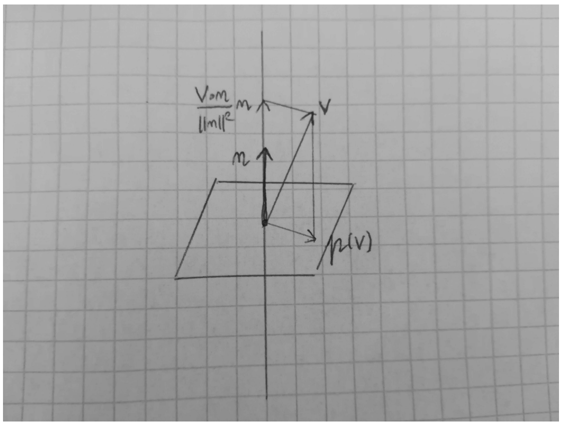
6.4 Proiezioni, riflessioni e rotazioni nello spazio ordinario

La formula generale (6.24) per la proiezione ortogonale di un vettore in uno spazio vettoriale euclideo V su un sottospazio W di dimensione maggiore di 1 prevede che si debba calcolare prima una base ortogonale del sottospazio

stesso. Tuttavia, nel caso particolare in cui $V = V_O^3$ è lo spazio tridimensionale dei vettori applicati in O e W è dato da un piano che passa per O, espresso in coordinate rispetto a una base ortonormale tramite un'equazione cartesiana, allora questo non è necessario: infatti, supponiamo che il piano sia rappresentato dall'equazione $Ax+By+Cz = 0$. Come sappiamo, la terna (A, B, C) dei coefficienti delle incognite rappresenta un vettore n normale, ovvero perpendicolare al piano.
Allora, dato un vettore v, la sua proiezione sulla direzione di n (cioè sul sottospazio unidimensionale generato da n) è data, in base alla formula (6.18), da $\frac{v \cdot n}{\|n\|^2} v$. Ma allora, visto che n è ortogonale al piano, la proiezione ortogonale sul piano è semplicemente la componente ortogonale a n, ovvero

$\frac{v \cdot n}{\|n\|^2} n$

$$p(v) = v - \frac{v \cdot n}{\|n\|^2} n \tag{6.28}$$



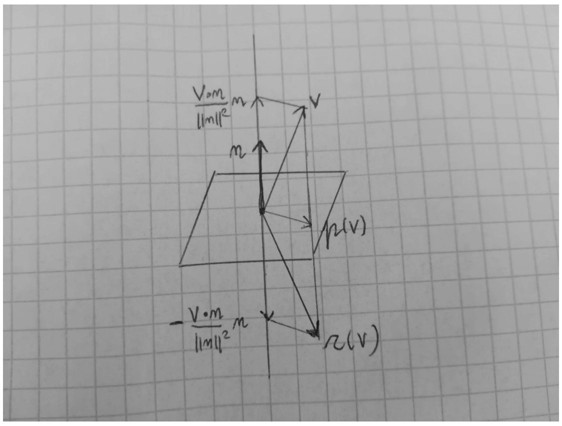
Esempio 6.22. Consideriamo ad esempio il piano per O rappresentato (in coordinate rispetto a una base ortonormale) dall'equazione cartesiana $x+y+z = 0$. Il vettore normale è dato da $n = (1, 1, 1)$, quindi, dato un qualunque vettore v, rappresentato in coordinate da (x, y, z), la (6.28) diventa

$$p(x, y, z) = (x, y, z) - \frac{x \cdot 1 + y \cdot 1 + z \cdot 1}{1^2 + 1^2 + 1^2}(1, 1, 1) = (x, y, z) - \frac{x+y+z}{3}(1, 1, 1) =$$
$$= (x, y, z) - \left(\frac{x+y+z}{3}, \frac{x+y+z}{3}, \frac{x+y+z}{3}\right) =$$
$$= \left(\frac{2}{3}x - \frac{1}{3}y - \frac{1}{3}z, -\frac{1}{3}x + \frac{2}{3}y - \frac{1}{3}z, -\frac{1}{3}x - \frac{1}{3}y + \frac{2}{3}z\right)$$

e quindi la matrice che rappresenta la proiezione (rispetto al sistema di riferimento fissato) è

$$\begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

Quanto appena visto ci permette di ricavare subito anche una formula per la riflessione rispetto a un piano dato di cui conosciamo la normale. Infatti, come si vede nel seguente disegno,



la riflessione $r(v)$ di un vettore v rispetto a un piano si ottiene semplicemente come somma $p(v)-v_\perp$ della proiezione ortogonale $p(v)$ sul piano con *l'opposto* $-v_\perp$ della componente normale: essendo, in base alla (6.28), la proiezione uguale a $p(v) = v - \frac{v \cdot n}{\|n\|^2}n$ mentre la componente normale è uguale alla proiezione di v in direzione n, ovvero $v_\perp = \frac{v \cdot n}{\|n\|^2}n$, si ha $r(v) = v - \frac{v \cdot n}{\|n\|^2}n - \frac{v \cdot n}{\|n\|^2}n$ ovvero

$$r(v) = v - 2\frac{v \cdot n}{\|n\|^2}n \tag{6.29}$$

Esempio 6.23. Consideriamo lo stesso piano per O dell'esempio precedente dato dall'equazione cartesiana $x + y + z = 0$, con $n = (1, 1, 1)$. Dato un qualunque vettore v, rappresentato in coordinate da (x, y, z), la (6.29) diventa

$$r(x, y, z) = (x, y, z) - 2\frac{x \cdot 1 + y \cdot 1 + z \cdot 1}{1^2 + 1^2 + 1^2}(1, 1, 1) = (x, y, z) - 2\frac{x + y + z}{3}(1, 1, 1) =$$

$$= (x, y, z) - \left(2\frac{x + y + z}{3}, 2\frac{x + y + z}{3}, 2\frac{x + y + z}{3}\right) =$$

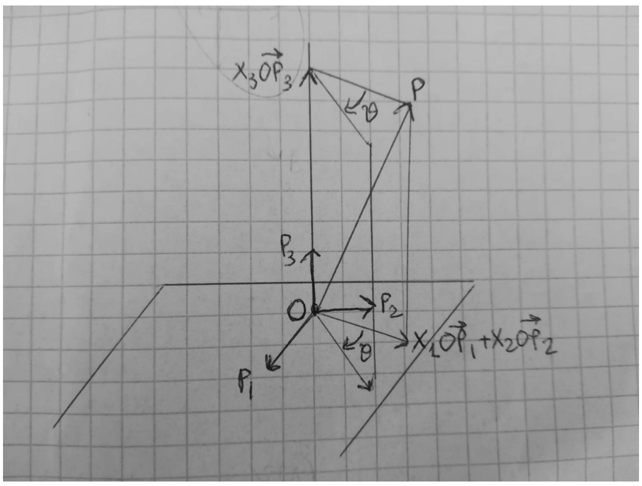
$$= \left(\frac{1}{3}x - \frac{2}{3}y - \frac{2}{3}z, -\frac{2}{3}x + \frac{1}{3}y - \frac{2}{3}z, -\frac{2}{3}x - \frac{2}{3}y + \frac{1}{3}z\right)$$

e quindi la matrice che rappresenta la riflessione (rispetto al sistema di riferimento fissato) è

$$\begin{pmatrix} \frac{1}{3} & -\frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ -\frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

Dopo le proiezioni ortogonali e le riflessioni, un altro importante esempio di applicazioni lineari dallo spazio tridimensionale V_O^3 in sè che abbiamo visto all'inizio del Capitolo 4 è stato quello delle rotazioni attorno a un asse (ovvero attorno a una retta passante per O). Impareremo ora a determinare la matrice che rappresenta la rotazione di un angolo dato attorno a un asse dato.

Il caso più semplice è quello delle rotazioni attorno a uno degli assi coordinati, ovvero attorno alle rette per O aventi come direzione uno dei tre vettori $\vec{OP_1}, \vec{OP_2}, \vec{OP_3}$ della base ortonormale scelta per fissare il sistema di riferimento, che chiameremo rispettivamente *asse x, asse y e asse z*.
Iniziamo dalla rotazione attorno all'asse z: come si vede nel seguente disegno



quando facciamo ruotare di un angolo θ attorno all'asse z un vettore $\vec{OP}$ di coordinate (x_1, x_2, x_3) (ovvero $\vec{OP} = x_1\vec{OP_1} + x_2\vec{OP_2} + x_3\vec{OP_3}$) la sua componente $x_3\vec{OP_3}$ rimane invariata, mentre la componente $x_1\vec{OP_1} + x_2\vec{OP_2}$ che si trova sul piano che contiene $\vec{OP_1}$ e $\vec{OP_2}$ ruota dello stesso angolo θ attorno a O.
Ma noi abbiamo già visto (Esempio 4.3 (1)) che ruotando $x_1\vec{OP_1} + x_2\vec{OP_2}$ in un piano di un angolo θ, le sue coordinate x_1, x_2 si trasformano secondo la (4.13), ovvero $(x_1, x_2) \mapsto (\cos\theta x_1 - \sin\theta x_2, \sin\theta x_1 + \cos\theta x_2)$, quindi la

componente $x_1\vec{OP_1}+x_2\vec{OP_2}$ si trasforma in $(\cos\theta x_1-\sin\theta x_2)\vec{OP_1}+(\sin\theta x_1+\cos\theta x_2)\vec{OP_2}$. Combinando questo col fatto che la componente $x_3\vec{OP_3}$ rimane invariata, concludiamo quindi che ruotando $\vec{OP}=x_1\vec{OP_1}+x_2\vec{OP_2}+x_3\vec{OP_3}$ di un angolo θ attorno all'asse z otteniamo il vettore

$$(\cos\theta x_1-\sin\theta x_2)\vec{OP_1}+(\sin\theta x_1+\cos\theta x_2)\vec{OP_2}+x_3\vec{OP_3}$$

ovvero in coordinate la rotazione attorno all'asse z si esprime come

$$(x_1,x_2,x_3)\mapsto(\cos\theta x_1-\sin\theta x_2,\sin\theta x_1+\cos\theta x_2,x_3) \tag{6.30}$$

o, usando la notazione del prodotto di matrici,

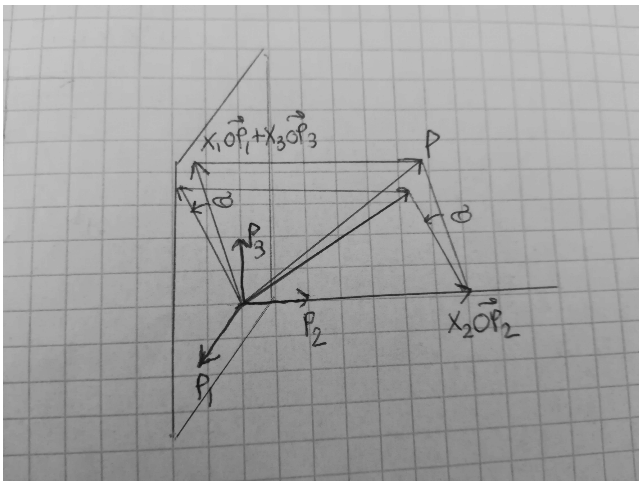
$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\mapsto\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

La matrice

$$R_\theta^z=\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{6.31}$$

è quindi la matrice che rappresenta la rotazione di un angolo θ attorno all'asse z (rispetto a una base ortonormale).

Con un ragionamento analogo si trovano le matrici che rappresentano le rotazioni attorno agli altri assi: come si vede nel seguente disegno



quando facciamo ruotare di un angolo θ attorno all'asse y un vettore $\vec{OP}$ di coordinate (x_1,x_2,x_3) (ovvero $\vec{OP}=x_1\vec{OP_1}+x_2\vec{OP_2}+x_3\vec{OP_3}$) stavolta

è la sua componente $x_2\vec{OP}_2$ a rimanere invariata, mentre è la componente $x_1\vec{OP}_1 + x_3\vec{OP}_3$ che si trova sul piano che contiene $\vec{OP}_1$ e $\vec{OP}_3$ a ruotare dello stesso angolo θ attorno a O in tale piano.
Analogamente a quanto visto sopra, le coordinate x_1, x_3 si trasformano secondo la (4.13), ovvero $(x_1, x_3) \mapsto (\cos\theta x_1 - \sin\theta x_3, \sin\theta x_1 + \cos\theta x_3)$, e combinando questo col fatto che la componente $x_2\vec{OP}_2$ rimane invariata, concludiamo quindi che ruotando $\vec{OP} = x_1\vec{OP}_1 + x_2\vec{OP}_2 + x_3\vec{OP}_3$ di un angolo θ attorno all'asse y otteniamo il vettore

$$(\cos\theta x_1 - \sin\theta x_3)\vec{OP}_1 + x_2\vec{OP}_2 + (\sin\theta x_1 + \cos\theta x_3)\vec{OP}_3$$

ovvero in coordinate la rotazione attorno all'asse y si esprime come

$$(x_1, x_2, x_3) \mapsto (\cos\theta x_1 - \sin\theta x_3, x_2, \sin\theta x_1 + \cos\theta x_3)$$

o, usando la notazione del prodotto di matrici,

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

La matrice che rappresenta la rotazione di un angolo θ attorno all'asse y (rispetto a una base ortonormale) è quindi

$$R^y_\theta = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \tag{6.32}$$

Infine, quando facciamo ruotare di un angolo θ attorno all'asse x un vettore $\vec{OP}$ di coordinate (x_1, x_2, x_3) (ovvero $\vec{OP} = x_1\vec{OP}_1 + x_2\vec{OP}_2 + x_3\vec{OP}_3$) sarà la sua componente $x_1\vec{OP}_1$ a rimanere invariata e la componente $x_2\vec{OP}_2 + x_3\vec{OP}_3$ a trasformarsi secondo la (4.13), quindi ruotando $\vec{OP}$ otteniamo

$$x_1\vec{OP}_1 + (\cos\theta x_2 - \sin\theta x_3)\vec{OP}_2 + (\sin\theta x_2 + \cos\theta x_3)\vec{OP}_3$$

ovvero in coordinate la rotazione attorno all'asse x si esprime come

$$(x_1, x_2, x_3) \mapsto (x_1, \cos\theta x_2 - \sin\theta x_3, \sin\theta x_2 + \cos\theta x_3)$$

o, usando la notazione del prodotto di matrici,

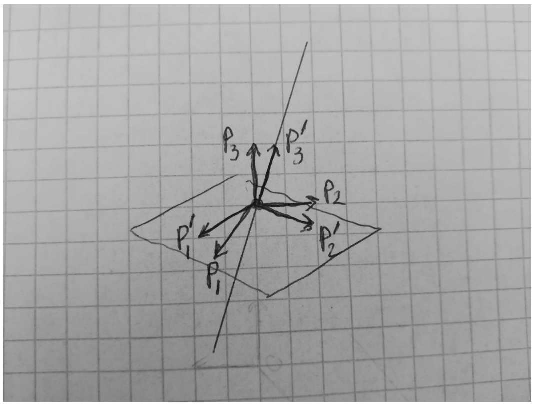
$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

La matrice che rappresenta la rotazione di un angolo θ attorno all'asse x (rispetto a una base ortonormale) è quindi

$$R_\theta^x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \tag{6.33}$$

Ci poniamo ora il problema di determinare la matrice che, rispetto a una base ortonormale fissata, rappresenta la rotazione f di un angolo θ attorno a un asse qualunque.

L'idea è la seguente: scelti un vettore $\vec{OP_3'}$ dell'asse di rotazione che abbia lunghezza 1 e due vettori $\vec{OP_1'}, \vec{OP_2'}$ sul piano perpendicolare all'asse che siano a loro volta perpendicolari tra loro e di lunghezza 1, come nel disegno seguente:



i tre vettori $\vec{OP_1'}, \vec{OP_2'}, \vec{OP_3'}$ così scelti formano una nuova base ortonormale B' (diversa dalla base ortonormale $B = \{\vec{OP_1}, \vec{OP_2}, \vec{OP_3}\}$ scelta quando abbiamo fissato il sistema di riferimento) che ha il vantaggio che la matrice associata alla rotazione f rispetto a B' è, come abbiamo visto nella (6.31),

$$M_{B'}(f) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{6.34}$$

in quanto, rispetto a tale sistema di riferimento, è la rotazione di angolo θ attorno all'asse z.
Tuttavia, noi vogliamo la matrice $M_B(f)$ che rappresenta la rotazione rispetto alla base B del sistema di riferimento che abbiamo fissato all'inizio: ma questa può essere ottenuta facilmente grazie a quanto visto nel Teorema 5.28, dove abbiamo dimostrato che $A = M_B(f)$ e $A' = M_{B'}(f)$ sono simili, ovvero

sono legate dalla relazione $A' = M^{-1}AM$, dove M è la matrice $M_{BB'}$ di cambiamento di coordinate dalla base B' alla base B:

$$M_{B'}(f) = M_{BB'}^{-1} M_B(f) M_{BB'}$$

ovvero, moltiplicando entrambi i membri di questa uguaglianza a sinistra per $M_{BB'}$ e a destra per $M_{BB'}^{-1}$,

$$M_B(f) = M_{BB'} M_{B'}(f) M_{BB'}^{-1} \tag{6.35}$$

Questa relazione ci consente di determinare la matrice $M_B(f)$ che rappresenta la rotazione f rispetto a B una volta costruita la base B' e trovate le matrici $M_{BB'}$ e la sua inversa.

Per quello che riguarda il primo punto, la base B' può essere trovata grazie al procedimento di ortogonalizzazione di Gram-Schmidt. Illustriamo il procedimento direttamente con un esempio: supponiamo di voler determinare la rotazione di asse con direzione data dal vettore di coordinate $(-1, 1, 1)$ e angolo $\frac{\pi}{2}$.

Come abbiamo detto, per costruire la base B' ci servono un vettore $\vec{OP'_3}$ sull'asse che abbia lunghezza 1 e due vettori $\vec{OP'_1}$, $\vec{OP'_2}$ perpendicolari all'asse e perpendicolari tra loro e di lunghezza 1.

Per trovare $\vec{OP'_3}$, prendiamo il vettore dato $(-1, 1, 1)$ e dividiamolo per la sua lunghezza, in modo che il vettore risultante abbia lunghezza 1. Poiché, sempre usando il prodotto scalare standard, si ha che la lunghezza di $(-1, 1, 1)$ è $\sqrt{(-1)^2 + 1^2 + 1^2} = \sqrt{3}$, possiamo porre

$$\vec{OP'_3} = \left(-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) \tag{6.36}$$

Per trovare i vettori $\vec{OP'_1}$ e $\vec{OP'_2}$, iniziamo con l'osservare che i vettori perpendicolari all'asse stanno sul piano che ha come normale l'asse stesso $(-1, 1, 1)$ e che quindi ha equazione cartesiana $-x + y + z = 0$.

Ponendo $y = t$ e $z = s$ si trova $x = t + s$, ovvero i vettori di tale piano sono tutti e soli quelli della forma $(t + s, t, s)$ al variare di $t, s \in \mathbb{R}$; inoltre, essendo $(t + s, t, s) = t(1, 1, 0) + s(1, 0, 1)$, una sua base è data dai vettori di coordinate $(1, 1, 0)$ e $(1, 0, 1)$, che denotiamo rispettivamente w_1 e w_2.

Essendo tali vettori non perpendicolari tra loro, non possiamo prenderli (neanche dopo averli normalizzati dividendoli per le loro lunghezze) come primi due vettori $\vec{OP'_1}$ e $\vec{OP'_2}$ della base B'.

Possiamo però applicare il procedimento di ortogonalizzazione di Gram-Schmidt, e ottenere una base w'_1, w'_2 ortogonale del piano perpendicolare all'asse di rotazione in base alle formule

$$w_1' = w_1$$
$$w_2' = w_2 - \frac{w_2 \cdot w_1'}{\|w_1'\|^2} w_1'$$

Applicando tali formule a $w_1 = (1,1,0)$ e $w_2 = (1,0,1)$, si ottiene

$$w_2' = (1,0,1) - \frac{1 \cdot 1 + 1 \cdot 0 + 0 \cdot 1}{1^2 + 1^2 + 0^2}(1,1,0) =$$
$$= (1,0,1) - \frac{1}{2}(1,1,0) = (\frac{1}{2}, -\frac{1}{2}, 1)$$

Ora abbiamo due vettori ($w_1' = (1,1,0)$ e $w_2' = (\frac{1}{2}, -\frac{1}{2}, 1)$) ortogonali all'asse e tra loro e di lunghezze rispettivamente $\|w_1\| = \sqrt{1^2 + 1^2 + 0^2} = \sqrt{2}$ e $\|w_2'\| = \sqrt{(\frac{1}{2})^2 + (-\frac{1}{2})^2 + 1^2} = \sqrt{\frac{3}{2}}$: dividendoli per la loro lunghezza possiamo quindi porre

$$\vec{OP_1'} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), \quad \vec{OP_2'} = \left(\frac{1}{2}\sqrt{\frac{2}{3}}, -\frac{1}{2}\sqrt{\frac{2}{3}}, \sqrt{\frac{2}{3}}\right) \tag{6.37}$$

La (6.37) e la (6.36) ci danno in effetti le coordinate dei vettori $\vec{OP_1'}, \vec{OP_2'}, \vec{OP_3'}$ rispetto alla base B scelta per fissare il nostro sistema di riferimento: quindi, per la definizione data a pagina 219, la matrice che ha queste coordinate come colonne è la matrice di cambiamento di coordinate da B' a B:

$$M_{BB'} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{2}\sqrt{\frac{2}{3}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{2}\sqrt{\frac{2}{3}} & \frac{1}{\sqrt{3}} \\ 0 & \sqrt{\frac{2}{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \tag{6.38}$$

Per applicare la (6.35) ci manca quindi solo da determinare l'inversa $M_{BB'}^{-1}$. Come sappiamo, in generale calcolare l'inversa di una matrice richiede o l'applicazione di operazioni elementari o il calcolo dei cofattori. In questo tipo di esercizio, però, non è necessario applicare questi metodi, in quanto la $M_{BB'}$ è di un tipo speciale che consente di determinare la sua inversa praticamente senza fare calcoli.

Per mostrare ciò, abbiamo bisogno di due definizioni preliminari: data una qualunque matrice $A \in M_{m,n}(\mathbb{K})$, si dice *trasposta di A* (e si denota ${}^TA$) la matrice che si ottiene da A scambiando le righe con le colonne. Ad esempio,

per la teoria sulle
matrici, fare riferimento
anche al file
matrici.pdf

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \quad {}^TA = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

OSSERVAZIONE: indichiamo la matrice trasposta
con la notazione A^T (apice T a destra).
In queste dispense è indicato con l'apice a sinistra

Osservazione 6.24. Si noti che quando si scambiano le righe con le colonne, l'entrata di A che si trova nella i-esima riga e j-esima colonna viene a trovarsi nella trasposta ${}^T\!A$ in corrispondenza della j-esima riga e i-esima colonna, con gli indici scambiati: quindi le entrate della trasposta soddisfano la relazione $({}^T\!A)_{ij} = A_{ji}$, uguaglianza che viene spesso usata per definire la trasposta.

Si ha la seguente, importante

Definizione 6.25. Una matrice $A \in M_n(\mathbb{R})$ si dice *ortogonale* se è invertibile e la sua inversa A^{-1} coincide con la trasposta ${}^T\!A$.

Mostreremo ora che la matrice $M_{BB'}$ che abbiamo costruito per determinare la rotazione mediante la (6.35) è ortogonale, quindi per calcolare la sua inversa $M_{BB'}^{-1}$ (che ci serve sempre nella (6.35)) basta scrivere la trasposta. Tale fatto è conseguenza del seguente risultato generale.

Teorema 6.26. Sia V uno spazio vettoriale euclideo e siano B e B' due sue basi ortonormali (ovvero costituite di vettori ortogonali tra loro e di norma 1). Allora la matrice di cambiamento di coordinate $M_{BB'}$ è una matrice ortogonale.

Proof. Per definizione, la matrice $M_{BB'} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ & \vdots & \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$ ha sulle colonne le coordinate dei vettori della base $B' = \{v'_1, \dots, v'_n\}$ rispetto alla base $B = \{v_1, \dots, v_n\}$, quindi dati due vettori v'_i e v'_j della base B' si ha $v'_i =$
NO
$a_{1i}v_1 + \dots + a_{ni}v_n = \sum_{k=1}^n a_{ki}v_k$ e $v'_j = a_{1j}v_1 + \dots + a_{nj}v_n = \sum_{l=1}^n a_{lj}v_l$.
Ma allora il prodotto scalare tra v'_i e v'_j è

$$v'_i \cdot v'_j = (\sum_{k=1}^n a_{ki}v_k) \cdot (\sum_{l=1}^n a_{lj}v_l) = \sum_{k,l=1}^n (a_{ki}v_k) \cdot (a_{lj}v_l) = \sum_{k,l=1}^n a_{ki}a_{lj}(v_k \cdot v_l) \tag{6.39}$$

(nella seconda uguaglianza abbiamo usato le proprietà (1) e (2) della Definizione 6.1 di forma bilineare, in base alle quali ogni addendo della prima sommatoria si moltiplica per ogni addendo della seconda, nella terza uguaglianza abbiamo invece usato la proprietà (3) della Definizione 6.1 di forma bilineare, grazie alla quale gli scalari a_{ki} e a_{lj} che in ogni addendo della somatoria moltiplicano v_k e v_l escono dal prodotto).
Ora, essendo per ipotesi la base $B' = \{v'_1, \dots, v'_n\}$ ortonormale, i suoi vettori sono tra loro ortogonali, cioè $v'_i \cdot v'_j = 0$ per $i \neq j$, e hanno norma 1, cioè $v'_i \cdot v'_i = \|v'_i\|^2 = 1$: ricordando che il delta di Kronecker δ_{ij} (Osservazione 4.16) vale 0 per $i \neq j$ e 1 per $i = j$, possiamo scrivere $v'_i \cdot v'_j = \delta_{ij}$.

NO

Ragionando analogamente con la base $B = \{v_1, \ldots, v_n\}$, anch'essa ortnormale, possiamo scrivere che $v_k \cdot v_l = \delta_{kl}$. Quindi la (6.39) può essere riscritta

$$\delta_{ij} = \sum_{k,l=1}^{n} a_{ki} a_{lj} \delta_{kl}$$

Dal momento che $\delta_{kl} = 0$ quando $k \neq l$, nella sommatoria a secondo membro rimangono solo i termini con $l = k$, quindi possiamo scrivere (ricordando che $\delta_{kk} = 1$)

$$\delta_{ij} = \sum_{k=1}^{n} a_{ki} a_{kj} \tag{6.40}$$

Ricordando che δ_{ij} non è nient'altro che l'entrata $(I_n)_{ij}$ di posto ij della matrice identica I_n, mentre a_{ki} può essere pensata come l'entrata $({}^T A)_{ik}$ di posto ik della matrice trasposta di A (Osservazione 6.24), la (6.40) può essere riscritta

$$(I_n)_{ij} = \sum_{k=1}^{n} ({}^T A)_{ik} A_{kj}$$

Ma la sommatoria a secondo membro è, per definizione di prodotto righe per colonne, l'entrata di posto ij del prodotto ${}^T AA$, quindi concludiamo che $I_n = {}^T AA$, che ci dice che la matrice A ha come inversa la sua trasposta ${}^T A$, cioè è ortogonale. □

Grazie al risultato appena dimostrato, possiamo finalmente completare il nostro esercizio: la matrice (6.38) è per costruzione la matrice di cambiamento di coordinate tra due basi ortonormali, quindi, in base al Teorema 6.26 è ortogonale, ovvero ha come inversa la sua trasposta:

$$M_{BB'}^{-1} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2}\sqrt{\frac{2}{3}} & -\frac{1}{2}\sqrt{\frac{2}{3}} & \sqrt{\frac{2}{3}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}$$

Possiamo allora finalmente calcolare la matrice $M_B(f)$ che rappresenta la rotazione rispetto alla base B usando la (6.35), e ponendo $\theta = \frac{\pi}{2}$ nella $M_{B'}(f)$ data dalla (6.34), come richiesto nell'esercizio:

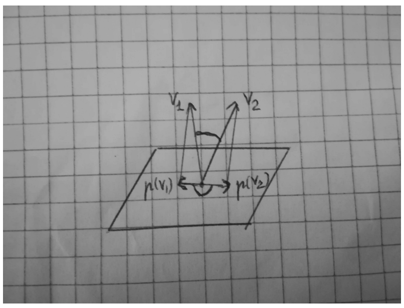
$$M_B(f) = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{2}\sqrt{\frac{2}{3}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{2}\sqrt{\frac{2}{3}} & \frac{1}{\sqrt{3}} \\ 0 & \sqrt{\frac{2}{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2}\sqrt{\frac{2}{3}} & -\frac{1}{2}\sqrt{\frac{2}{3}} & \sqrt{\frac{2}{3}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} =$$

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

$$\begin{pmatrix} \frac{1}{3} & \frac{\sqrt{3}-1}{3} & -\frac{\sqrt{3}+1}{3} \\ -\frac{\sqrt{3}+1}{3} & \frac{1}{3} & \frac{-\sqrt{3}+1}{3} \\ \frac{\sqrt{3}-1}{3} & \frac{\sqrt{3}+1}{3} & \frac{1}{3} \end{pmatrix}$$

6.5 Isometrie lineari nel piano e nello spazio

Le rotazioni e le riflessioni (sia nel piano che nello spazio) hanno l'importante proprietà che non modificano lunghezze e angoli tra vettori: un vettore e il suo vettore ruotato o riflesso hanno la stessa lunghezza, e se due vettori formano tra loro un certo angolo, anche le loro immagini mediante una rotazione o una riflessione formeranno lo stesso angolo.
Questa proprietà non è condivisa ad esempio dalle proiezioni: ad esempio, la proiezione ortogonale su un piano di un vettore che è già ortogonale al piano stesso è il vettore nullo, quindi la sua lunghezza diventa zero anche se il vettore di partenza non aveva lunghezza nulla; allo stesso modo, l'angolo tra due vettori prima e dopo la proiezione in generale cambia, come si vede nel seguente disegno



Ci poniamo ora l'obiettivo di determinare tutti gli endomorfismi di V_O^2 e di V_O^3 che preservano lunghezze e angoli, che chiameremo *isometrie lineari.*

A questo scopo, come già fatto per risolvere altri problemi di natura geometrica, trasformeremo il problema in un problema algebrico traducendo tutto in coordinate rispetto a una base ortonormale. Dal momento che la prima parte delle nostre considerazioni varrà sia per V_O^2 che per V_O^3, scriveremo V_O^k, con k che può essere uguale a 2 o a 3.
Come sappiamo, in coordinate rispetto a una base ortonormale la lunghezza di un vettore $v \in V_O^k$ di coordinate x e il coseno dell'angolo tra due vettori $v, w \in V_O^k$ di coordinate x e y sono dati rispettivamente dalle due formule

$$\sqrt{x \cdot x}, \quad \frac{x \cdot y}{\sqrt{x \cdot x}\sqrt{y \cdot y}} \tag{6.41}$$

dove $x \cdot y$ denota il prodotto scalare standard, ovvero $x \cdot y = x_1 y_1 + \cdots + x_k y_k$ (con $k = 2$ o 3 a seconda che siamo in V_O^2 o V_O^3).
D'altra parte, la nostra isometria lineare $f : V_O^k \to V_O^k$, in coordinate si traduce come una funzione $\mathbb{R}^k \to \mathbb{R}^k$, che manda le coordinate x e y dei vettori v e w nelle coordinate Ax e Ay delle loro immagini $f(v)$ e $f(w)$ (dove A è la matrice $M_B(f)$ associata a f rispetto alla base ortonormale B fissata). Quindi, in base alle (6.41), la lunghezza di $f(v)$ e il coseno dell'angolo tra $f(v)$ e $f(w)$ sono date da

$$\sqrt{Ax \cdot Ax}, \quad \frac{Ax \cdot Ay}{\sqrt{Ax \cdot Ax}\sqrt{Ay \cdot Ay}} \tag{6.42}$$

Confrontando la (6.41) con la (6.42) vediamo allora che la f preserva lunghezze e angoli se e solo se per ogni x e y sono soddisfatte le due condizioni

$$\sqrt{Ax \cdot Ax} = \sqrt{x \cdot x} \tag{6.43}$$

$$\frac{Ax \cdot Ay}{\sqrt{Ax \cdot Ax}\sqrt{Ay \cdot Ay}} = \frac{x \cdot y}{\sqrt{x \cdot x}\sqrt{y \cdot y}} \tag{6.44}$$

Ora, vediamo subito che la (6.43) e la (6.44) sono in realtà equivalenti all'unica condizione

$$Ax \cdot Ay = x \cdot y \tag{6.45}$$

per ogni x, y: infatti, se vale la (6.45) allora in particolare prendendo $y = x$ si ha $Ax \cdot Ax = x \cdot x$, e quindi è soddisfatta la (6.43); ma allora chiaramente, oltre ad essere uguali i numeratori di primo e secondo membro della (6.44) lo sono anche i denominatori. Viceversa, se valgono la (6.43) e la (6.44), allora tenuto conto che i denominatori della (6.44) sono uguali grazie alla (6.43), il fatto che valga la (6.44) ci dice che sono uguali i suoi numeratori, ovvero la (6.45).

Allo scopo di riscrivere la (6.45) in termini della sola A (senza far comparire x e y), osserviamo che se x e y denotano le k-uple scritte in colonna, ovvero $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}$ e $y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix}$, il prodotto scalare standard può essere riscritto come il prodotto righe per colonne tra x scritto in riga e y:

$$x \cdot y = x_1 y_1 + x_2 y_2 + \cdots + x_k y_k = \begin{pmatrix} x_1 & x_2 & \dots & x_k \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix}$$

Ma poichè, in base alla Definizione 6.24 di trasposta di una matrice (che scambia le righe con le colonne), x scritto in riga è la trasposta ${}^T x$ di x, possiamo scrivere che

$$x \cdot y = {}^T x y \tag{6.46}$$

(a secondo membro si tratta di un prodotto di matrici).
Quindi possiamo riscrivere la (6.45) come

$${}^T(Ax)(Ay) = {}^T x y \tag{6.47}$$

Ora, l'operazione di trasposta ha in generale la proprietà

$${}^T(AB) = {}^T B {}^T A \tag{6.48}$$

Infatti, per ogni entrata di posto $i\ j$ si ha

$$({}^T B {}^T A)_{ij} = \sum_{k=1}^{n} ({}^T B)_{ik} ({}^T A)_{kj} = \sum_{k=1}^{n} B_{ki} A_{jk} = \sum_{k=1}^{n} A_{jk} B_{ki} = (AB)_{ji} = {}^T(AB)_{ij}$$

Quindi a primo membro della (6.47) possiamo riscrivere ${}^T(Ax) = {}^T x {}^T A$ e scrivere

$${}^T x {}^T A A y = {}^T x y \tag{6.49}$$

Vediamo allora che perché questa condizione sia soddisfatta per ogni x e ogni y si deve avere

$$^T AA = I_n$$

che rappresenta finalmente la condizione che deve soddisfare la matrice A associata a un endomorfismo f perché questo sia un'isometria lineare, ovvero *A è una matrice ortogonale* (Definizione 6.25). Per determinare tutte le isometrie lineari del piano e dello spazio, basterà allora studiare e classificare le matrici ortogonali di ordine 2 e 3.

6.5.1 Isometrie lineari nel piano

Determiniamo ora esplicitamente tutte le matrici ortogonali di ordine 2. Sia $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Allora $^T AA = I_2$ significa

$$^T\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2+c^2 & ab+cd \\ ab+cd & b^2+d^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

che equivale alle tre condizioni

$$\begin{aligned} a^2 + c^2 &= 1 \\ b^2 + d^2 &= 1 \\ ab + cd &= 0 \end{aligned}$$

Per quanto noto dalla trigonometria, la prima di queste condizioni ci dice che deve essere $a = \cos\theta$, $c = \sin\theta$, per qualche $\theta \in \mathbb{R}$, e analogamente la seconda che deve essere $b = \sin\phi$, $d = \cos\phi$ per qualche $\phi \in \mathbb{R}$.
Quindi, la terza condizione pùo essere riscritta

$$ab + cd = \cos\theta \sin\phi + \sin\theta \cos\phi = 0$$

ovvero, ricordando la formula trigonometrica per il seno della somma,

$$\sin(\phi + \theta) = 0. \tag{6.50}$$

Questa equazione ci dice che sono possibili due casi[5]:

$$\phi + \theta = 0, \quad \phi + \theta = \pi$$

Nel primo caso, ovvero $\phi = -\theta$, si ottiene la matrice ortogonale

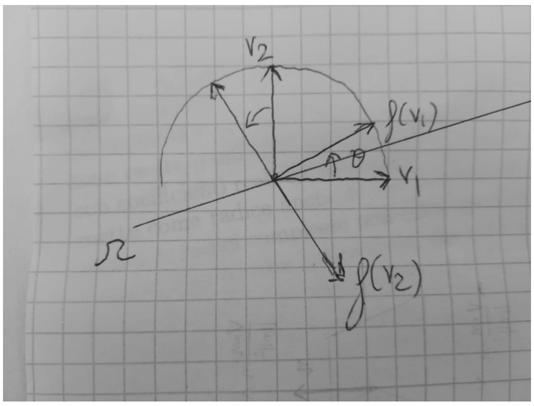
[5] Ovviamente se ci limitiamo all'intervallo $[0, 2\pi]$, altrimenti si intende che a ognuno dei due casi va aggiunto un multiplo di 2π e si ottengono ancora soluzioni della (6.50)

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin(-\theta) \\ \sin\theta & \cos(-\theta) \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{6.51}$$

Nel secondo caso, ovvero $\phi = \pi - \theta$, si ottiene la matrice ortogonale

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin(\pi-\theta) \\ \sin\theta & \cos(\pi-\theta) \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \tag{6.52}$$

Quindi concludiamo che una matrice $A \in M_2(\mathbb{R})$ è ortogonale se e solo se è di uno dei due tipi (6.51) o (6.52), per qualche $\theta \in \mathbb{R}$.
Ora, una matrice del tipo (6.51) è esattamente la matrice associata a una rotazione di angolo θ (attorno all'origine O) rispetto a una base ortonormale $B = \{\vec{OP_1}, \vec{OP_2}\}$ del piano, come visto nella (4.12).
Per capire invece che tipo di trasformazione sia quella rappresentata dalle matrici del tipo (6.52), ricordiamo che per definizione la matrice associata a un endomorfismo ha sulle colonne le coordinate delle immagini dei vettori di base trasformati, calcolate rispetto alla base stessa. Dal momento che la prima colonna di (6.52) è uguale alla prima colonna della matrice di rotazione, mentre la seconda colonna è la seconda colonna della matrice di rotazione cambiata di segno, significa che, se v_1 e v_2 sono i due vettori della base ortonormale fissata, allora il vettore v_1 ha la stessa immagine che avrebbe con una rotazione di angolo θ, mentre l'immagine di v_2 è il vettore opposto a quello che si avrebbe ruotandolo:



Come si vede dal disegno, si deduce allora che i due vettori sono riflessi rispetto alla retta passante per O che forma un angolo di $\frac{\theta}{2}$ con la direzione del vettore v_1, ovvero le matrici del tipo (6.52) rappresentano riflessioni.

Concludiamo quindi che le matrici ortogonali di ordine 2 rappresentano o rotazioni o riflessioni e quindi che *le isometrie lineari del piano sono o rotazioni o riflessioni.*

Osservazione 6.27. Notiamo che le matrici del tipo (6.51) hanno determinante uguale a $\cos^2\theta + \sin^2\theta = 1$, mentre le matrici del tipo (6.52) hanno determinante uguale a $-\cos^2\theta - \sin^2\theta = -1$. Questo non è un caso, in quanto in generale si ha

Proposizione 6.28. Se A è una matrice ortogonale, allora il determinante di A è uguale a $+1$ o -1.

Proof. La matrice A è ortogonale se e solo se ${}^TAA = I_n$. Ma allora, da una parte si ha $\det({}^TAA) = \det(I_n) = 1$, dall'altra, per il teorema di Binet $\det({}^TAA) = \det({}^TA)\det(A)$. Ora, usando il fatto, che non abbiamo dimostrato, che $\det({}^TA) = \det(A)$, si ha $1 = \det({}^TA)\det(A) = \det(A)^2$, da cui $\det(A) = \pm 1$, come volevamo. □

L'insieme delle matrici ortogonali di ordine n si denota $O(n)$; tra queste, il sottoinsieme costituito dalle matrici ortogonali di ordine n con determinante uguale a $+1$ si denota $SO(n)$ (e una matrice $A \in SO(n)$ si dice *ortogonale speciale*).
Nel prossimo paragrafo questa distinzione sarà fondamentale per poter classificare matrici ortogonali di ordine 3 e quindi le isometrie lineari nello spazio.

6.5.2 Isometrie lineari nello spazio

Cenni

Per la determinazione delle isometrie lineari nello spazio, o equivalentemente delle matrici ortogonali di ordine 3, seguiremo un metodo diverso da quello usato nel caso delle matrici di ordine 2: in tal caso, infatti, lo studio delle condizioni equivalenti alla ${}^TAA = I_3$ si rivela troppo complicato. Useremo quindi un altro metodo, che coinvolge sia considerazioni geometriche che considerazioni algebriche, e in particolar modo la teoria degli autovalori e autovettori.

Iniziamo con il dimostrare il seguente

Lemma 6.29. Sia $f : V_O^3 \to V_O^3$ un'isometria lineare e sia A la matrice ortogonale associata a f rispetto a una base ortonormale di V_O^3. Allora

(1) se $\det(A) = +1$, l'isometria f ha un autovettore relativo all'autovalore $+1$

(2) se $\det(A) = -1$, l'isometria f ha un autovettore relativo all'autovalore -1

NO

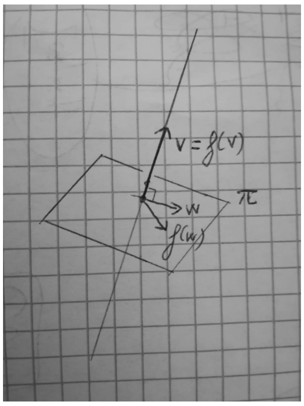
Proof. Dal momento che A ha ordine 3, il suo polinomio caratteristico ha grado 3 e quindi avrà 3 radici (eventualmente con ripetizioni, ovvero con molteplicità algebrica maggiore di 1): di esse, quelle complesse non reali possono essere o due o nessuna: infatti, dall'algebra è noto che se un polinomio con coefficienti reali ha una radice complessa $z = a+ib$, necessariamente ammette come radice anche la sua cosiddetta complessa coniugata $\bar{z} = a - ib$, quindi le radici complesse non reali appaiono sempre a coppie[6]. Quindi deve esistere almeno una radice reale, che sarà quindi un autovalore dell'isometria f: ma dal momento che f preserva le lunghezze, la relazione $f(v) = \lambda v$ può essere vera solo se λ è uguale a $+1$ o a -1 (per valori diversi di λ, la lunghezza di $f(v)$ sarebbe diversa dalla lunghezza di v).
Riassumendo, A ha sicuramente come autovalore o $\lambda = +1$ o $\lambda = -1$. Mostriamo ora che il primo caso si ha sempre quando $\det(A) = +1$, e il secondo quando $\det(A) = -1$.
Per assurdo, se si avesse $\det(A) = +1$ e $+1$ non fosse radice del polinomio caratteristico, le radici (contate con le loro molteplicità) potrebbero essere solo o $-1, -1, -1$, o $-1, z, \bar{z}$ (abbiamo detto che le radici complesse o sono due tra loro coniugate o nessuna). Ma sia nel primo che nel secondo caso si avrebbe una contraddizione col fatto che $\det(A) = +1 > 0$, in quanto come abbiamo visto a pagina 224 il determinante di una matrice è uguale al prodotto delle radici del suo polinomio caratteristico, e si avrebbe $\det(A) = (-1)(-1)(-1) = -1 < 0$ oppure $\det(A) = (-1)z\bar{z} = -(a+ib)(a-ib) = -a^2 - b^2 < 0$.
Analogamente, nel caso $\det(A) = -1$, se per assurdo -1 non fosse radice del polinomio caratteristico, le radici (contate con le loro molteplicità) potrebbero essere solo o $+1, +1, +1$, o $+1, z, \bar{z}$, ma di nuovo tenendo conto che il determinante di una matrice è uguale al prodotto delle radici del suo polinomio caratteristico, si avrebbe una contraddizione con $\det(A) = -1 < 0$, in quanto avremmo $\det(A) = (+1)(+1)(+1) = +1 > 0$ oppure $\det(A) = (+1)z\bar{z} = (a+ib)(a-ib) = a^2 + b^2 > 0$. $\square$

Ora, siamo pronti a determinare tutte le isometrie lineari $f : V_O^3 \to V_O^3$, distinguendo i casi $\det(A) = +1$ e $\det(A) = -1$.

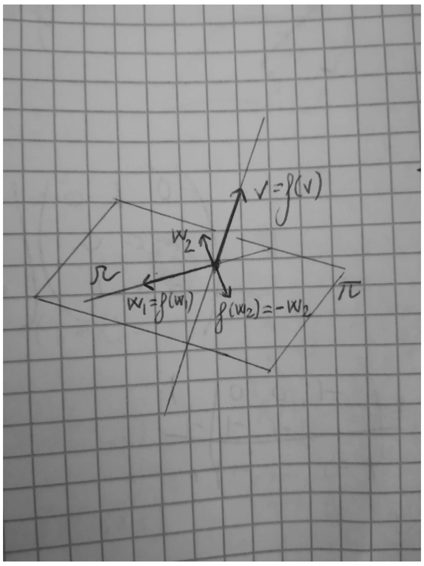
Nel primo caso, in base al Lemma 6.29 appena dimostrato, f ha necessariamente autovalore $+1$, cioè esiste un $v \in V_O^3$ non nullo per cui $f(v) = v$: in altre parole, esiste un vettore v fissato da f, e necessariamente saranno

[6] Ad esempio, per un polinomio di grado due $ax^2 + bx + c$ questo si vede dalla formula risolutiva $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a} = \frac{-b}{2a} \pm \frac{\sqrt{b^2-4ac}}{2a}$, dove nel caso in cui $\Delta = b^2 - 4ac < 0$ troviamo proprio due radici complesse coniugate grazie al $\pm$ della formula: ad esempio, $x^2 + x + 1$ ha come soluzioni $x = \frac{-1 \pm \sqrt{-3}}{2} = \frac{-1}{2} \pm \frac{\sqrt{-3}}{2} = \frac{-1}{2} \pm \frac{\sqrt{3}}{2} i$.

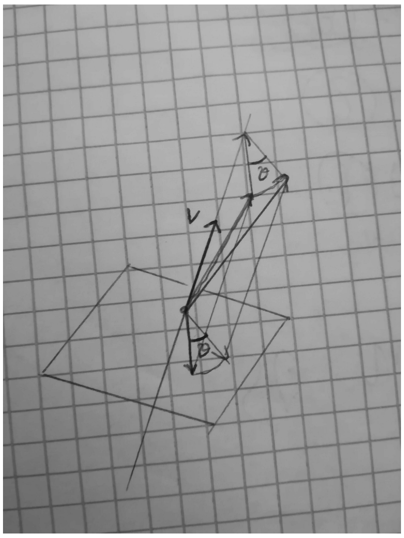
fissati anche tutti i suoi multipli tv (l'autospazio è un sottospazio vettoriale). Mostreremo ora che f è esattamente la rotazione (di un certo angolo θ) attorno alla retta per O che ha come direzione v.
A questo scopo, consideriamo come nel disegno seguente il piano π passante per O e perpendicolare a v:



Se un vettore w sta su π, significa che w è perpendicolare a v: ma poiché f preseva gli angoli, $f(w)$ deve essere perpendicolare a $f(v)$, cioè, essendo $f(v) = v$, concludiamo che $f(w)$ è ancora perpendicolare a v e quindi $f(w)$ sta ancora su π: in altre parole f manda i vettori di π in vettori di π e quindi, se ci restringiamo a π, abbiamo un endomorfismo di un piano che preserva lunghezze e angoli, ovvero un'isometria lineare del piano.
Nella sezione precedente abbiamo classificato tutte tali isometrie mostrando che esse sono o rotazioni attorno a O o riflessioni per rette passanti per O: ma se sul piano π f fosse una riflessione rispetto a una retta r contenuta in π, come illustrato nel disegno seguente dovrebbe esistere un vettore w_1 di π fissato da f (basta prendere un vettore che sta sulla retta di riflessione r) e un vettore w_2 di π mandato da f nel suo opposto, ovvero $f(w_2) = -w_2$ (basta prendere un vettore perpendicolare a r).

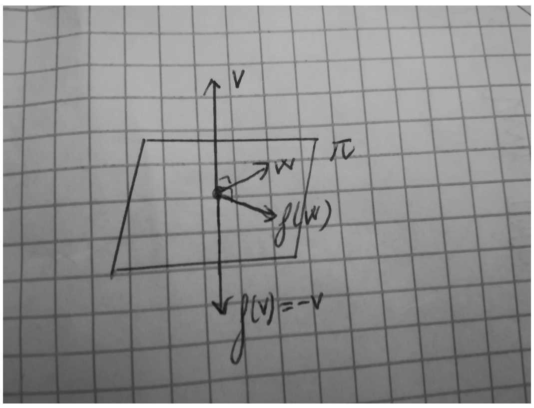


Quindi, oltre all'autovettore v (con autovalore associato $+1$) avremmo come autovettori w_1 (sempre associato all'autovalore $+1$, che quindi ha molteplicità geometrica, e quindi anche algebrica, uguale a 2) e l'autovettore w_2, con autovalore associato -1.
Gli autovalori, considerati con le loro molteplicità, sarebbero quindi $+1, +1, -1$, e il loro prodotto, che come sappiamo è uguale al determinante di A, sarebbe $(+1)(+1)(-1) = -1$, contro l'ipotesi che $\det(A) = +1$.
Concludiamo quindi che, ristretta al piano π, f non può che essere una rotazione di un certo angolo θ: ma assieme al fatto che f fissa la retta perpendicolare al piano questo implica che f agisca come una rotazione di angolo θ attorno a tale asse nell'intero spazio: infatti, come illustrato nel seguente disegno, ogni vettore v si decompone come somma di un vettore sull'asse e di uno sul piano π, e il fatto che la sua componente sull'asse rimanga fissa mentre quella sul piano ruoti di θ ci dice che il vettore v sta ruotando attorno all'asse di θ

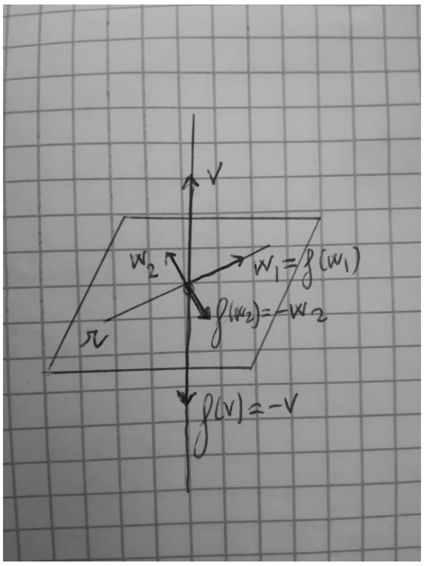


Riassumendo, abbiamo dimostrato che *le matrici $A \in SO(3)$ rappresentano sempre rotazioni attorno a un asse.*

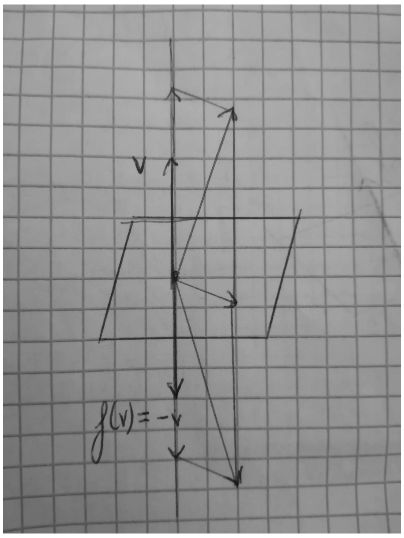
Per il caso in cui $\det(A) = -1$, possiamo applicare un'analisi del tutto parallela: in base al Lemma 6.29, f ha necessariamente autovalore -1, ovvero esiste un $v \in V_O^3$ non nullo per cui $f(v) = -v$. Consideriamo allora come nel disegno seguente il piano π passante per O e perpendicolare a v:



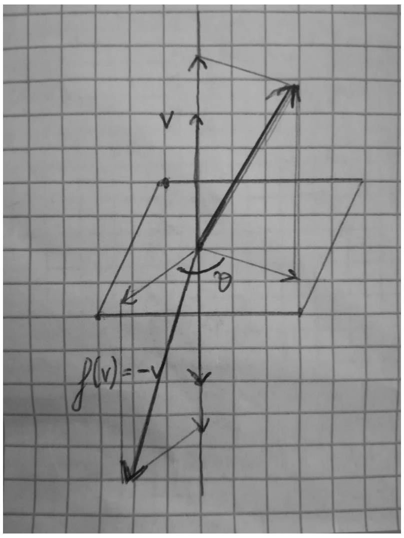
Come nel caso precedente, se un vettore w sta su π, significa che w è perpendicolare a v: ma poiché f preseva gli angoli, $f(w)$ deve essere perpendicolare a $f(v)$, cioè, essendo $f(v) = -v$, concludiamo che $f(w)$ è perpendicolare a $-v$ e quindi $f(w)$ sta ancora su π: allora f manda i vettori di π in vettori di π e quindi se ci restringiamo a π, abbiamo un endomorfismo di un piano che preserva lunghezze e angoli, ovvero un'isometria lineare del piano. Come sappiamo, allora la restrizione di f al piano deve essere o una rotazione attorno a O o una riflessione per una retta r passante per O: ma se sul piano π f fosse una riflessione, come illustrato nel disegno seguente dovrebbe esistere un vettore w_1 di π fissato da f (basta prendere un vettore che sta sulla retta di riflessione r) e un vettore w_2 di π mandato da f nel suo opposto, ovvero $f(w_2) = -w_2$ (basta prendere un vettore perpendicolare a r).



Quindi, oltre all'autovettore v (con autovalore associato -1) avremmo come autovettori w_2 (sempre associato all'autovalore -1, che quindi ha molteplicità geometrica, e quindi anche algebrica, uguale a 2) e l'autovettore w_1, con autovalore associato $+1$.
Gli autovalori, considerati con le loro molteplicità, sarebbero quindi $-1, -1, +1$, e il loro prodotto, che come sappiamo è uguale al determinante di A, sarebbe $(-1)(-1)(+1) = +1$, contro l'ipotesi che $\det(A) = -1$.
Concludiamo quindi che, ristretta al piano π, f non può che essere una rotazione di un certo angolo θ.
Ora, se $\theta = 0$ tale rotazione è in realtà l'identità, cioè fissa i vettori sul piano: ma assieme al fatto che f manda nel suo opposto ogni vettore della retta perpendicolare al piano π implica che f agisca come una riflessione rispetto a π: infatti, come illustrato nel seguente disegno, ogni vettore v si decompone come somma di un vettore sull'asse e di uno sul piano π, e il fatto che la sua componente sulla retta venga rovesciata mentre quella sul piano rimanga fissa ci dice che il vettore v viene riflesso rispetto al piano



Nel caso in cui invece $\theta \neq 0$, f non è né una riflessione né una rotazione, ma la composizione di una riflessione rispetto al piano con una rotazione attorno all'asse perpendicolare al piano: tali isometrie si chiamano *riflessioni rotatorie*.



Riassumendo, abbiamo dimostrato che *le matrici* $A \in O(3)$ *con determinante* -1 *rappresentano o riflessioni rispetto a un piano o riflessioni rotatorie.*

La classificazione appena dimostrata ha ad esempio le seguenti conseguenze:

(1) La composizione di due rotazioni nello spazio è una rotazione

(2) La composizione di due riflessioni nello spazio è una rotazione

Infatti, se f e g sono due rotazioni, con matrici associate A e B rispettivamente, la loro composizione $f \circ g$ (che è necessariamente ancora un'isometria perché se f e g preservano lunghezze e angoli lo farà anche la composizione che per definizione è ottenuta semplicemente applicando le due funzioni in successione) è rappresentata dal prodotto di matrici AB, e per il teorema di Binet $\det(AB) = \det(A)\det(B) = (+1)(+1) = +1$. Quindi $AB \in SO(3)$ e in base alla classificazione fatta è una rotazione.
Analogamente, se f e g sono due riflessioni, con matrici associate A e B rispettivamente, la loro composizione $f \circ g$ è rappresentata dal prodotto di matrici AB, e per il teorema di Binet stavolta si ha $\det(AB) = \det(A)\det(B) = (-1)(-1) = +1$. Di nuovo, $AB \in SO(3)$, ovvero è una rotazione.

Esempio 6.30. La composizione della rotazione di $\frac{\pi}{2}$ attorno all'asse x (ovvero attorno alla retta per O che ha come direzione il primo vettore $\vec{OP_1}$ della base ortonormale) e della rotazione di $\frac{\pi}{2}$ attorno all'asse y (ovvero attorno alla retta per O che ha come direzione il secondo vettore $\vec{OP_2}$ della base ortonormale) deve essere una rotazione, in quanto come abbiamo osservato sopra la composizione di due rotazioni è una rotazione: determiniamone asse e angolo di rotazione.

Come sappiamo, la matrice che rappresenta una rotazione di angolo θ attorno all'asse x è $\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}$, mentre la matrice che rappresenta una rotazione di angolo θ attorno all'asse y è $\begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}$: in particolare, per $\theta = \frac{\pi}{2}$ si ha rispettivamente $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$ e $\begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$.
La composizione di tali rotazioni è quindi data dalla matrice A che si ottiene eseguendo il prodotto

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & -1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{6.53}$$

Ora, l'asse di rotazione è l'insieme dei vettori fissi della trasformazione, ovvero tali che $f(v) = v$: quindi esso, in coordinate, coincide con l'autospazio di A relativo all'autovalore $+1$ della matrice A (si noti che sappiamo già dalla dimostrazione di sopra che A ammette $+1$ come autovalore, quindi non è necessario calcolare il polinomio caratteristico per verificarlo).

Gli autovettori relativi a $+1$ si ottengono, come sappiamo, risolvendo il sistema omogeneo $(A - I_3)x = 0$, ovvero

$$\begin{pmatrix} -1 & 0 & -1 \\ -1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Riducendo la matrice dei coefficienti a gradini si trova

$$\begin{pmatrix} -1 & 0 & -1 \\ -1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \underset{R_2 \to R_2 - R_1}{\longrightarrow} \begin{pmatrix} -1 & 0 & -1 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix} \underset{R_3 \to R_3 + R_2}{\longrightarrow} \begin{pmatrix} -1 & 0 & -1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

ovvero l'autospazio è dato dalle soluzioni del sistema ridotto $\begin{cases} -x_1 - x_3 = 0 \\ -x_2 + x_3 = 0 \end{cases}$
Posto $x_3 = t$, si trova $x_2 = x_3 = t$ e $x_1 = -x_3 = -t$: quindi gli autovettori relativi a $+1$ sono dati da tutti i vettori[7] $(-t, t, t)$, al variare di $t \in \mathbb{R}$: tali vettori stanno su una retta per O (di direzione $(-1, 1, 1)$) che è l'asse della rotazione.
Per determinare l'angolo di rotazione potremmo prendere un vettore sul piano perpendicolare all'asse e vedere di quanto viene ruotato (si noti che non è sufficiente prendere un vettore qualunque e verificare di quanto viene ruotato dalla rotazione, in quanto tale angolo è variabile: ad esempio, per i vettori dell'asse è zero in quanto questi sono fissati). Tuttavia, possiamo sfruttare un metodo algebrico più diretto, basato sulla seguente considerazione.
La nostra matrice (6.53) è in effetti la matrice $M_B(f)$ che rappresenta la rotazione f rispetto alla base ortonormale B del sistema di riferimento fissato; ma come abbiamo visto quando abbiamo spiegato il procedimento per determinare una rotazione noti asse e angolo, si può sempre costruire una nuova base ortonormale B' con asse z coincidente con l'asse di rotazione, e la cui matrice associata $M_{B'}(f)$ sarebbe quindi la matrice di una rotazione attorno all'asse z, ovvero $M_{B'}(f) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$, dove θ è l'angolo della rotazione.
Ora, il Teorema 5.28 afferma che le matrici $A = M_B(f)$ e $A' = M_{B'}(f)$ sono simili, e come abbiamo visto a pagina 223 esse devono avere lo stesso polinomio caratteristico e le stesse radici.

[7] Chiaramente, $(-t, t, t)$ sono le coordinate di tali vettori rispetto alla base fissata fin dall'inizio: il vettore corrispondente vero e proprio sarebbe $-t\vec{OP}_1 + t\vec{OP}_2 + t\vec{OP}_3$.

Ma il polinomio caratteristico di $M_{B'}(f) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$ è

$$\det\begin{pmatrix} \cos\theta - \lambda & -\sin\theta & 0 \\ \sin\theta & \cos\theta-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{pmatrix} = (1-\lambda)[(\cos\theta-\lambda)^2 + \sin^2\theta] =$$

$$= (1-\lambda)(\lambda^2 - 2\cos\theta + 1)$$

che, dalla formula di risoluzione delle equazioni di secondo grado, oltre a $\lambda = 1$ ha come radici

$$\cos\theta \pm \sqrt{\cos^2\theta - 1} = \cos\theta \pm \sqrt{-\sin^2\theta} = \cos\theta \pm i\sin\theta$$

Quindi anche $A = M_B(f)$ deve avere come soluzioni dell'equazione caratteristica $+1$ e $\cos\theta \pm i\sin\theta$, essendo θ l'angolo di rotazione. Riassumendo, per determinare l'angolo θ della rotazione rappresentata da una matrice $A \in SO(3)$ è sufficiente risolverne l'equazione caratteristica determinando le due radici complesse.
Nel nostro esempio, abbiamo

$$\det\begin{pmatrix} -\lambda & 0 & -1 \\ -1 & -\lambda & 0 \\ 0 & 1 & -\lambda \end{pmatrix} = -\lambda[\lambda^2] + (-1)(-1) = -\lambda^3 + 1$$

che, come si può verificare con il metodo di Ruffini, si decompone come $(\lambda - 1)(-\lambda^2 - \lambda - 1)$: risolvendo $-\lambda^2 - \lambda - 1 = 0$ mediante la formula risolutiva, si trova $\lambda = \frac{-1\pm\sqrt{-3}}{2} = -\frac{1}{2} \pm \frac{\sqrt{3}}{2}i$.
In base a quanto detto sopra, l'angolo di rotazione è allora dato da $\cos\theta = -\frac{1}{2}$ e $\sin\theta = \frac{\sqrt{3}}{2}$, ovvero $\theta = \frac{2}{3}\pi$.

6.6 Il teorema spettrale: diagonalizzazione di matrici simmetriche reali

Nell'Osservazione 5.25 abbiamo anticipato il fatto che le matrici simmetriche, ovvero quelle per cui $A_{ij} = A_{ji}$ o, equivalentemente[8], $A = {}^TA$, sono diagonalizzabili. Come sappiamo (Proposizione 5.23), questo significa che esiste

[8]Infatti, come abbiamo detto nell'Osservazione 6.24, per definizione di trasposta $({}^TA)_{ij} = A_{ji}$, quindi dire che $A_{ij} = A_{ji}$ equivale a scrivere $A_{ij} = ({}^TA)_{ij}$, ovvero $A = {}^TA$.

una matrice M invertibile di ordine n tale che il prodotto $M^{-1}AM$ è una matrice diagonale D, che ha sulla diagonale gli autovalori di A.
Ora, il *teorema spettrale* afferma che è sempre possibile far sì che la matrice M che diagonalizza A sia ortogonale, cioè non solo invertibile ma con inversa uguale alla sua trasposta (e quindi $M^{-1}AM = {}^TMAM$).
Non dimostreremo tale teorema, o meglio non dimostreremo che data una matrice simmetrica reale questa è sempre diagonalizzabile (che, come sappiamo, significa che gli autovalori sono tutti reali e le molteplicità geometriche sono uguali a quelle algebriche) ma, dando per buona questa prima parte, dimostreremo ora che M può essere scelta ortogonale, e vedremo come costruire esplicitamente M in alcuni esercizi.

La prima osservazione che facciamo è che una matrice M di ordine n è ortogonale se e solo se le sue colonne sono vettori di $\mathbb{R}^n$ ortogonali tra loro e di lunghezza unitaria rispetto al prodotto scalare standard $x \cdot y = x_1y_1 + \cdots + x_ny_n$.
Infatti, M è ortogonale se e solo se ${}^TMM = I_n$, ovvero, per definizione di prodotto di matrici, se e solo se moltiplicando la i-esima riga di ${}^TM$ per la j-esima colonna di M si ottiene l'elemento di posto $i\ j$ della matrice identica, che è il delta di Kronecker δ_{ij} (uguale a 0 se $i \neq j$ e uguale a 1 se $i = j$). Ma, per definizione di trasposta, sulla i-esima riga di ${}^TM$ c'è la i-esima colonna di M:

$$\begin{pmatrix} & \vdots & \\ M_{1i} & \dots & M_{ni} \\ & \vdots & \end{pmatrix} \begin{pmatrix} \dots & M_{1j} & \dots \\ & \vdots & \\ \dots & M_{nj} & \dots \end{pmatrix} = \begin{pmatrix} & \vdots & \\ \dots & \delta_{ij} & \dots \\ & \vdots & \end{pmatrix}$$

ovvero

$$M_{1i}M_{1j} + \cdots + M_{ni}M_{nj} = \delta_{ij}$$

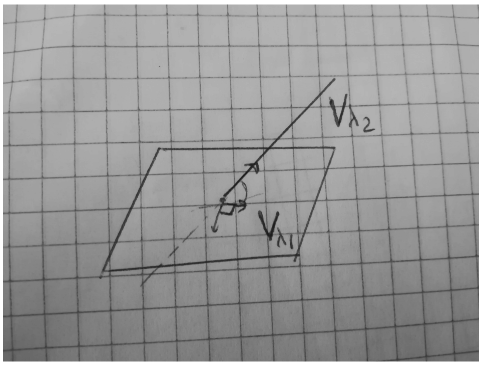
Questo ci dice che il prodotto scalare (standard) tra la i-esima colonna e la j-esima colonna di M è 0 se $i \neq j$ (quindi le colonne sono tra loro ortogonali) e 1 se $i = j$, cioè vale 1 il prodotto scalare tra la i-esima colonna e se stessa, che per definizione di norma significa che le colonne hanno norma 1.

Osservazione 6.31. La stessa osservazione appena fatta sulle colonne si può fare anche sulle righe di una matrice ortogonale, usando la $M^TM = I_n$ (ovvero la seconda delle due uguaglianze che ci dice che ${}^TM$ è l'inversa di M): per definizione di prodotto di matrici, infatti, questa significa che moltiplicando la i-esima riga di M per la j-esima colonna di ${}^TM$ si ottiene il delta di Kronecker δ_{ij}. Ma, per definizione di trasposta, sulla j-esima colonna di ${}^T M$ c'è la j-esima riga di M, e quindi

$$\begin{pmatrix} & \vdots & \\ M_{i1} & \dots & M_{in} \\ & \vdots & \end{pmatrix} \begin{pmatrix} \dots & M_{j1} & \dots \\ & \vdots & \\ \dots & M_{jn} & \dots \end{pmatrix} = \begin{pmatrix} & \vdots & \\ \dots & \delta_{ij} & \dots \\ & \vdots & \end{pmatrix}$$

ovvero $M_{i1}M_{j1}+\cdots+M_{in}M_{jn} = \delta_{ij}$ che, analogamente a quanto già visto per le colonne, ci dice che le righe sono tra loro ortogonali e di norma 1 (rispetto a prodotto scalare standard di $\mathbb{R}^n$).

Ora, dal momento che nella dimostrazione della Proposizione 5.23 abbiamo visto che la matrice M tale che $M^{-1}AM$ è diagonale si ottiene disponendo in colonna gli autovettori della matrice A, dire che si può scegliere M ortogonale significa, in base a quanto appena detto sulle sue colonne, che A ammette una base di autovettori fatta di vettori ortogonali tra loro e di norma 1 (cioè una base ortonormale di autovettori).
Perchè questo sia possibile, bisogna che gli autospazi $V_{\lambda_1}, \dots, V_{\lambda_k}$ siano tra loro ortogonali (rispetto al prodotto scalare standard): se non lo fossero, come nel seguente disegno illustrativo,



sarebbe impossibile trovare una base di tutto lo spazio fatta da autovettori e ortonormale: anche scegliendo i primi due autovettori ortogonali e di lunghezza 1 nel primo autospazio V_{λ_1}, il terzo autovettore preso dal secondo autospazio V_{λ_2} non sarebbe mai ortogonale ai primi due.
Questa condizione necessaria (e sufficiente) è garantita dalla seguente

Proposizione 6.32. Gli autospazi relativi a autovalori diversi di una matrice simmetrica reale sono tra loro ortogonali (rispetto al prodotto scalare standard)

Proof. Dobbiamo dimostrare che per ogni autovettore x relativo a un autovalore λ (ovvero $Ax = \lambda x$) e ogni autovettore y relativo a un autovalore μ (ovvero $Ay = \mu y$), con $\lambda \neq \mu$, si ha $x \cdot y = 0$.
Da una parte, essendo x autovettore, si ha

$$(Ax) \cdot y = (\lambda x) \cdot y = \lambda(x \cdot y) \tag{6.54}$$

Dall'altra, usando la (6.46) che ci dice che il prodotto scalare standard $x \cdot y$ di $\mathbb{R}^n$ può essere riscritto come ${}^T x y$,

$$(Ax) \cdot y = {}^T(Ax)y = {}^T x {}^T A y = {}^T x A y = x \cdot (Ay) = x \cdot (\mu y) = \mu(x \cdot y) \tag{6.55}$$

dove nella seconda uguaglianza abbiamo usato la proprietà (6.48) della trasposta e nella terza il fatto che A è simmetrica, cioè ${}^T A = A$.
Combinando la (6.54) con la (6.55) vediamo allora che deve essere $\lambda(x \cdot y) = \mu(x \cdot y)$, ovvero, portando a primo membro e mettendo in evidenza $x \cdot y$,

$$(\lambda - \mu)(x \cdot y) = 0.$$

Ma, essendo $\lambda \neq \mu$ e quindi $\lambda - \mu \neq 0$, questa uguaglianza ci dice che deve essere $x \cdot y = 0$, che è quello che volevamo dimostrare. $\square$

Esempio 6.33. Nell'Esempio 5.24 abbiamo calcolato autovalori e autovettori della matrice $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$, che è una matrice simmetrica (l'entrata a_{12} è uguale all'entrata a_{21}).

Come abbiamo visto, l'autospazio relativo all'autovalore $\lambda = 4$ è dato da tutti i vettori del tipo (t, t) al variare di $t \in \mathbb{R}$, mentre l'autospazio relativo a $\lambda = -2$ da tutti i vettori del tipo $(-t', t')$ (sempre al variare di $t' \in \mathbb{R}$. Come si vede subito, i vettori del primo autospazio sono ortogonali (rispetto al prodotto scalare standard $x \cdot y = x_1 y_1 + x_2 y_2$ di $\mathbb{R}^2$) ai vettori del secondo autospazio: $t \cdot (-t') + t \cdot t' = 0$, confermando quanto previsto dalla Proposizione 6.32.
Sempre nell'Esempio 5.24, abbiamo costruito la matrice $M = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ tale che $M^{-1}AM$ è diagonale mettendo nelle colonne le basi degli autospazi, in questo caso $(1, 1)$ e $(-1, 1)$. Questa matrice non è ortogonale perché nonostante le sue colonne siano ortogonali, esse non hanno lunghezza 1, come invece previsto dalla proprietà delle colonne delle matrici ortogonali vista sopra. Ma allora per ottenere una matrice ortogonale basta semplicemente normalizzare i due vettori dividendoli per la loro lunghezza, che è $\sqrt{2}$ per entrambi, e prendere come basi degli autospazi $\frac{1}{\sqrt{2}}(1, 1) = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ e $\frac{1}{\sqrt{2}}(-1, 1) = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$.

Mettendo questi vettori così normalizzati in colonna otteniamo la matrice ortogonale[9]

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

che ha ovviamente (avendo sempre sulle colonne una base di autovettori) ancora la proprietà che $M^{-1}AM = {}^{T}MAM$ è diagonale.

Esempio 6.34. Vediamo ora un esempio di ordine maggiore di 2: consideriamo la matrice simmetrica $A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$. Il suo polinomio caratteristico è

$$\det \begin{pmatrix} 2-\lambda & 1 & 1 \\ 1 & 2-\lambda & 1 \\ 1 & 1 & 2-\lambda \end{pmatrix} = (2-\lambda)[(2-\lambda)^2-1]-[(2-\lambda)-1]+[1-(2-\lambda)] =$$

$$= (2-\lambda)(\lambda^2-4\lambda+3)+2(\lambda-1) = (2-\lambda)(\lambda-3)(\lambda-1)+2(\lambda-1) =$$

$$(\lambda-1)[(2-\lambda)(\lambda-3)+2] = (\lambda-1)(-\lambda^2+5\lambda-4) = -(\lambda-1)^2(\lambda-4)$$

Quindi abbiamo due autovalori, $\lambda = 1$ con molteplicità algebrica 2, e $\lambda = 4$ con molteplicità algebrica 1. Sicuramente le molteplicità geometriche saranno uguali a quelle algebriche, perché il teorema spettrale ci dice che la matrice essendo simmetrica è sicuramente diagonalizzabile (questa è la parte che non abbiamo dimostrato), e lo verifichiamo calcolando gli autospazi: per $\lambda = 1$ si ha $A-\lambda I_3 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ e quindi l'insieme delle soluzioni di $(A-\lambda I_3)x = \bar{0}$ (l'autospazio relativo) si ottiene risolvendo l'unica equazione $x_1+x_2+x_3 = 0$ che, posto $x_2 = t$ e $x_3 = s$ ci dà $x_1 = -t-s$, e quindi l'autospazio è dato da tutti i vettori del tipo

$$(-t-s, t, s) = t(-1, 1, 0) + s(-1, 0, 1)$$

e i due vettori $v_1 = (-1, 1, 0)$ e $v_2 = (-1, 0, 1)$ ne costituiscono una base.

[9] Osserviamo che si tratta della matrice che rappresenta la rotazione di $\frac{\pi}{4}$ nel piano.

Per $\lambda = 4$ si ha

$$A - \lambda I_3 = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \underset{\substack{R_2 \to 2R_2 + R_1 \\ R_3 \to 2R_3 + R_1}}{\longrightarrow} \begin{pmatrix} -2 & 1 & 1 \\ 0 & -3 & 3 \\ 0 & 3 & -3 \end{pmatrix} \underset{R_3 \to R_3 + R_2}{\longrightarrow} \begin{pmatrix} -2 & 1 & 1 \\ 0 & -3 & 3 \\ 0 & 0 & 0 \end{pmatrix}$$

Dividendo anche la seconda riga per 3, si vede che il sistema $(A - \lambda I_3)x = 0$ si riduce a $\begin{cases} -2x_1 + x_2 + x_3 = 0 \\ -x_2 + x_3 = 0 \end{cases}$ posto $x_3 = u$, la seconda equazione dice che $x_2 = u$, e sostituendo nella prima si trova $-2x_1 + u + u = 0$, da cui $x_1 = u$. Le soluzioni del sistema, ovvero l'autospazio relativo all'autovalore $\lambda = 4$, è dato quindi tutte le terne del tipo

$$(u, u, u) = u(1, 1, 1)$$

e il vettore $(1, 1, 1)$ ne costituisce una base.

Come si vede subito, i vettori del primo autospazio sono ortogonali (rispetto al prodotto scalare standard $x \cdot y = x_1y_1 + x_2y_2 + x_3y_3$ di $\mathbb{R}^3$) ai vettori del secondo autospazio: $(-t-s)\cdot u + t \cdot u + s \cdot u = 0$, confermando quanto previsto dalla Proposizione 6.32.

Ora, la matrice $M = \begin{pmatrix} -1 & -1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$ che ha come colonne i vettori delle basi degli autospazi, ha sicuramente la proprietà che $M^{-1}AM$ è diagonale, ma non è ortogonale perché le sue colonne oltre a non avere lunghezza 1 non sono neanche ortogonali: la terza è ortogonale alle prime due in quanto stanno in autospazi diversi, ma le prime due, cioè la base del primo autospazio, non sono ortogonali tra loro. Per avere una matrice ortogonale dobbiamo quindi prendere per ogni autospazio una base ortonormale, che come sappiamo si può costruire dal procedimento di ortogonalizzazione di Gram-Schmidt.

Applicando tale procedimento ai vettori $v_1 = (-1, 1, 0)$ e $v_2 = (-1, 0, 1)$ del primo autospazio otteniamo

$$w_1 = v_1 = (-1, 1, 0)$$

$$w_2 = v_2 - \frac{v_2 \cdot w_1}{\|w_1\|^2} w_1 = (-1, 0, 1) - \frac{1}{2}(-1, 1, 0) = (-\frac{1}{2}, -\frac{1}{2}, 1)$$

Se, per facilità di calcolo, moltiplichiamo per 2 il secondo vettore, abbiamo quindi ottenuto per il primo autospazio la base ortogonale $(-1, 1, 0)$, $(-1, -1, 2)$ che rendiamo ortonormale dividendo i vettori per le rispettive norme: $\frac{1}{\sqrt{2}}(-1, 1, 0) = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)$ e $\frac{1}{\sqrt{6}}(-1, -1, 2) = (-\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}})$.

Usando questi vettori per costruire la matrice M, assieme al vettore $(1,1,1)$ del secondo autospazio anche lui normalizzato, cioè $\frac{1}{\sqrt{3}}(1,1,1) = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$, otteniamo

$$M = \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}$$

che è una matrice ortogonale con la proprietà che $M^{-1}AM = {}^TMAM$ è diagonale.

Osservazione 6.35. Quando costruiamo la matrice M mettendo in colonna i vettori delle basi degli autospazi, siamo liberi di metterli in qualunque ordine (dal loro ordine dipenderà solo l'ordine degli autovalori sulla diagonale della matrice $D = M^{-1}AM$). Questo implica che possiamo sempre fare in modo che la matrice ortogonale M che diagonalizza A sia ortogonale speciale, ovvero con determinante $+1$: infatti, se così non fosse (cioè se fosse $\det(M) = -1$), basterebbe cambiare l'ordine delle sue colonne scambiandone due, e come sappiamo in seguito a tale operazione il determinante cambia segno.

Come abbiamo già messo in evidenza nel precedente capitolo (Esempio 5.26) l'esistenza di una matrice per cui $M^{-1}AM$ è diagonale consente in alcune situazioni di semplificare il problema espresso in termini della matrice A mediante un opportuno cambio di coordinate, costruito tramite la matrice M. Grazie al teorema spettrale, se la matrice A del problema è simmetrica, il cambio di coordinate si può effettuare tramite una matrice ortogonale, il che ha dei vantaggi o è addirittura necessario in alcuni contesti. Ad esempio, si può vedere che all'equazione di una conica nel piano (ellissi, iperboli, parabole) o di una quadrica nello spazio (sfere, ellissoidi, paraboloidi, iperboloidi, cilindri etc.) è associata una matrice simmetrica A di ordine 2 o 3 rispettivamente, e semplificare questa matrice simmetrica diagonalizzandola mediante una matrice ortogonale speciale (che come sappiamo rappresenta una rotazione) significa semplificare l'equazione applicando sostanzialmente una rotazione della curva o della superficie stessa (o del sistema di riferimento in cui stiamo lavorando).

NO

Non vediamo i dettagli di questo fatto, ma per mostrare con un altro esempio come questo genere di procedimento si applichi per semplificare i problemi, vediamo la dimostrazione del fatto, che abbiamo visto nel Teorema 6.9, che data una forma bilineare simmetrica $f(x,y) = \sum_{i,j=1}^{n} a_{ij}x_iy_j$ di $\mathbb{R}^n$, se gli autovalori della matrice simmetrica A di entrate a_{ij} sono tutti positivi allora la forma f è definita positiva (ed è quindi un prodotto scalare).

NO

A questo scopo, mostriamo preliminarmente che vale l'uguaglianza

$$\sum_{i,j=1}^{n} a_{ij}x_iy_j = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & & \dots & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}. \tag{6.56}$$

Infatti, dalla definizione di prodotto righe per colonne si ha

$$\begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & & \dots & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix} =$$

$$= \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} a_{11}y_1 + a_{12}y_2 + \dots + a_{1n}y_n \\ a_{21}y_1 + a_{22}y_2 + \dots + a_{2n}y_n \\ . \\ . \\ a_{n1}y_1 + a_{n2}y_2 + \dots + a_{nn}y_n \end{pmatrix} =$$

(usando la notazione con la sommatoria)

$$= \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n} a_{1j}y_j \\ \sum_{j=1}^{n} a_{2j}y_j \\ . \\ . \\ \sum_{j=1}^{n} a_{nj}y_j \end{pmatrix} =$$

$$x_1 \sum_{j=1}^{n} a_{1j}y_j + x_2 \sum_{j=1}^{n} a_{2j}y_j + \dots + x_n \sum_{j=1}^{n} a_{nj}y_j =$$

(usando di nuovo la notazione con la sommatoria)

$$= \sum_{i=1}^{n} x_i \sum_{j=1}^{n} a_{ij}y_j = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}x_iy_j = \sum_{i,j=1}^{n} a_{ij}x_iy_j$$

come volevamo (nella prima uguaglianza abbiamo semplicemente moltiplicato x_i, che stava davanti alla sommatoria in j, per ogni addendo di tale sommatoria, come prevede la proprietà distributiva).
Usando la solita notazione

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & & \vdots & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

la (6.56) si può quindi anche scrivere

NO

$$\sum_{i,j=1}^{n} a_{ij} x_i y_j = {}^T x A y. \qquad (6.57)$$

In questa forma, possiamo dimostrare quanto detto: in base alla definizione di forma definita positiva, dobbiamo verificare se $f(x,x) = {}^T x A x \geq 0$ e $f(x,x) = 0$ solo per il vettore nullo (sotto l'ipotesi che gli autovalori di A siano tutti positivi).
Dal momento che la matrice A è simmetrica, esiste una matrice M ortogonale tale che $M^{-1}AM = {}^T MAM$ è la matrice diagonale $D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$
che ha sulle diagonale gli autovalori $\lambda_1, \dots, \lambda_n$ di A.
Introduciamo allora le nuove coordinate $x' = (x'_1, \dots, x'_n)$ ottenute dalle x tramite l'uguaglianza $x' = M^{-1}x$, ovvero, moltiplicando entrambi i membri per M, $x = Mx'$: sostituendo in ${}^T x A x$ troviamo

$$f(x,x) = {}^T x A x = {}^T (Mx') A (Mx') = {}^T x'^T M A M x' = {}^T x' D x'$$

ovvero

$$f(x,x) = \begin{pmatrix} x'_1 & x'_2 & \dots & x'_n \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix} =$$

$$= \begin{pmatrix} x'_1 & x'_2 & \dots & x'_n \end{pmatrix} \begin{pmatrix} \lambda_1 x'_1 \\ \lambda_2 x'_2 \\ \vdots \\ \lambda_n x'_n \end{pmatrix} = \lambda_1 x'^2_1 + \lambda_2 x'^2_2 + \cdots \lambda_n x'^2_n$$

Allora è chiaro che quest'ultima quantità, se gli autovalori sono positivi, è sempre maggiore o uguale a zero, e inoltre essa è uguale a zero solo se lo

NO

sono i singoli addendi, ovvero se e solo se $x'_1, x'_2, \dots, x'_n$ sono nulli: ma questo significa che $x' = \bar{0}$, e quindi $x = Mx' = M\bar{0} = \bar{0}$. Questo dimostra che f è definita positiva se gli autovalori sono tutti positivi[10].

6.7 Il caso complesso (cenni)

La maggior parte delle nozioni e dei risultati visti in questo corso si applicano a spazi vettoriali su qualunque campo $\mathbb{K}$, quindi allo stesso modo sia agli spazi vettoriali reali che a quelli complessi: la riduzione a gradini, il rango, il determinante, la teoria degli autovalori e autovettori, prevedono le stesse operazioni o gli stessi procedimenti sia che lavoriamo nel campo reale che in quello complesso.

L'unica nozione per cui abbiamo richiesto che lo spazio vettoriale fosse reale è stata quella di prodotto scalare, dove ad esempio già la definizione di forma definita positiva prevede che il campo su cui lavoriamo sia un campo ordinato (come il campo dei numeri reali, su cui, contrariamente a quello dei complessi, ha senso dire se un numero è maggiore o minore di un altro e in particolare maggiore di zero).

Tuttavia, nella matematica e nelle sue applicazioni può essere utile avere a disposizione nozioni come quella di lunghezza o quella di ortogonalità anche negli spazi vettoriali complessi. Per capire come questo possa essere fatto, prendiamo il prototipo[11] di spazio vettoriale complesso, ovvero lo spazio $\mathbb{C}^n$ delle n-uple $(z_1, \dots, z_n)$ di numeri complessi.

Come possiamo definire la norma o lunghezza di una tale n-upla?

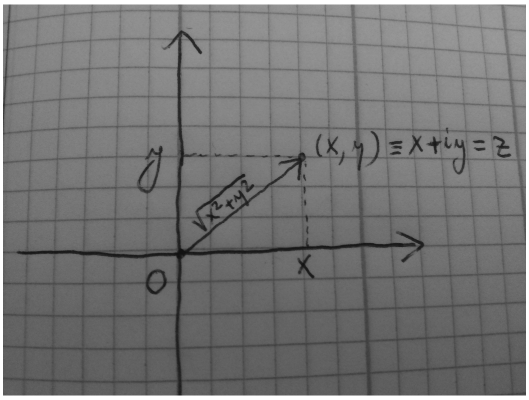
Per la teoria sui numeri complessi fare riferimento anche a complessi.pdf

Nel caso in cui $n = 1$, cioè di un solo numero complesso $z = a + ib$, il suo *modulo*, definito come $|z| = \sqrt{a^2 + b^2}$ è la quantità giusta per dare una misura della "lunghezza" di z: per giustificare questa affermazione, notiamo che ogni numero complesso $z = x + iy$ può essere identificato con la coppia (x, y) di numeri reali data dalla sua parte reale e dalla sua parte immaginaria; a sua volta, tale coppia (x, y) può essere identificata con un punto del piano cartesiano di ascissa x e ordinata y (una volta fissata un'origine e due assi orientati) e quindi in questo modo otteniamo un modo di vedere ogni numero complesso come un punto P del piano.

Allora, come si vede nel disegno seguente

[10] Non vediamo i dettagli del viceversa.

[11] Come sappiamo, in coordinate rispetto a una base fissata ogni $\mathbb{K}$-spazio vettoriale si rappresenta tramite $\mathbb{K}^n$, e quindi ogni spazio vettoriale complesso si rappresenta come $\mathbb{C}^n$.



il modulo $|z| = \sqrt{x^2 + y^2}$, per il teorema di Pitagora, è esattamente la lunghezza del segmento che congiunge l'origine O del piano cartesiano con il punto P. Questo giustifica geometricamente la scelta della quantità $|z| = \sqrt{x^2 + y^2}$ come misura della "lunghezza" di z.

Il modulo $|z|$ di $z = a + ib$ può essere espresso anche nel modo seguente: si definisce *complesso coniugato di* z il numero complesso $\bar{z} = a - ib$ che si ottiene da z cambiando di segno solo la sua parte immaginaria. Allora, si ha

$$z\bar{z} = (a + ib)(a - ib) = a^2 - (ib)^2 = a^2 - i^2b^2 = a^2 + b^2$$

e quindi $|z| = \sqrt{z\bar{z}}$.

A questo punto, se abbiamo invece una n-upla $z = (z_1, \dots, z_n)$ di numeri complessi, possiamo definire la sua lunghezza o norma nel modo seguente, tenendo conto del contributo di tutte le sue componenti:

$$\|z\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2} = \sqrt{z_1\bar{z}_1 + \cdots + z_n\bar{z}_n} \tag{6.58}$$

Questa norma è l'analogo della norma $\sqrt{x_1^2 + \cdots + x_n^2}$ di una n-upla reale $(x_1, \dots, x_n) \in \mathbb{R}^n$ rispetto al prodotto scalare standard: ora, così come quest'ultima si ottiene ponendo $y = x$ nel prodotto scalare $x \cdot y = x_1y_1 + \cdots + x_ny_n$, l'espressione sotto radice nella (6.58) si ottiene ponendo $w = z$ nella seguente espressione

$$z \cdot w = z_1\bar{w}_1 + \cdots + z_n\bar{w}_n \tag{6.59}$$

che rappresenta quindi l'analogo per $\mathbb{C}^n$ del prodotto scalare standard di $\mathbb{R}^n$. È facile (usando anche le proprietà del coniugato di un numero complesso) verificare che la (6.59) ha in comune con i prodotti scalari le seguenti proprietà:

(i) $z \cdot (w + w') = z \cdot w + z \cdot w'$

(ii) $(z+z')\cdot w = z\cdot w + z\cdot w'$

(iii) $(cz)\cdot w = c(z\cdot w)$

(le prime due sono le proprietà (1) e (2) della Definizione 6.1) ma *non* ha ad esempio la proprietà $z\cdot(cw) = c(z\cdot w)$: se nella (6.59) sostituiamo $w_1,\dots,w_n$ con $cw_1,\dots,cw_n$, il coniugato (in base alle sue proprietà, che non dimostriamo) si distribuisce anche su c, e quindi si ha

(iv) $z\cdot(cw) = \bar{c}(z\cdot w)$

ovvero lo scalare complesso c "esce" dal prodotto solo se moltiplica il primo vettore, mentre quando esce dal secondo vettore prende il coniugato. Questo ci dice che la (6.59) è lineare solo rispetto al primo argomento, mentre rispetto al secondo è lineare "a metà"[12].

Oltre a non essere bilineare, la (6.59) non è neanche simmetrica: infatti, non è difficile verificare che se scambiamo i vettori z e w il risultato del prodotto si trasforma nel suo coniugato, ovvero

(iv) $w\cdot z = \overline{z\cdot w}$

Per quello che riguarda infine quello che succede quando poniamo $w = z$, si ottiene invece esattamente la stessa proprietà che hanno anche i prodotti scalari, cioè di essere definiti positivi:

(v) $z\cdot z \geq 0$ e $z\cdot z = 0$ solo se $z = \bar{0}$.

In generale, in un generico spazio vettoriale complesso V, una funzione $V\times V \to \mathbb{C}$ che associa a due vettori un numero complesso e che ha le stesse proprietà (i)-(v) viste sopra per la (6.59) si chiama *prodotto hermitiano.* Ad esempio, la (6.59) si dice *prodotto hermitiano standard di* $\mathbb{C}^n$.

I prodotti hermitiani sono quindi per gli spazi vettoriali complessi quello che i prodotti scalari sono per gli spazi vettoriali reali e con essi si può svolgere una teoria parallela a quella vista per i prodotti scalari: ad esempio, abbiamo visto che le matrici ortogonali sono le matrici che preservano il prodotto scalare standard di $\mathbb{R}^n$. Analogamente, le matrici che preservano il prodotto hermitiano standard di $\mathbb{C}^n$ sono quelle che soddisfano la seguente

[12]Forme che si comportano in questo modo si chiamano *sesquilineari* (dove il prefisso "sesqui" significa proprio "una volta e mezzo", laddove nella parola bilineare il prefisso "bi" sta per "due volte").

Definizione 6.36. Una matrice $A \in M_n(\mathbb{C})$ si dice *unitaria* se si ha che $A \cdot {}^T\bar{A} = I_n$ (ovvero, equivalentemente, A è invertibile e ha come inversa A^{-1} la trasposta della sua coniugata ${}^T\bar{A}$).

L'insieme delle matrici unitarie di ordine n si denota $U(n)$.
Per le matrici ortogonali abbiamo dimostrato che il determinante può essere solo ± 1. Analogamente, si dimostra che il determinante di una matrice unitaria è sempre un numero complesso di modulo uguale a 1, ovvero $\det(A) = a + bi$ con $a^2 + b^2 = 1$ (questo ci dice che deve essere $a = cos\theta$ e $b = \sin\theta$, e quindi se A è unitaria $\det(A) = \cos\theta + i\sin\theta$).
Le matrici unitarie il cui determinante è in particolare uguale a $+1$ si dicono unitarie speciali, e l'insieme delle matrici unitarie speciali di ordine n si denota $SU(n)$.

Anche il teorema spettrale sulla diagonalizzazione delle matrici simmetriche mediante matrici ortogonali ha un suo analogo nel campo complesso: in tal caso, a svolgere il ruolo delle matrici simmetriche sono le matrici che soddisfano la seguente

Definizione 6.37. Una matrice $A \in M_n(\mathbb{C})$ si dice *hermitiana* se si ha che $A = {}^T\bar{A}$.

Si dimostra che per le matrici hermitiane vale la seguente

Proposizione 6.38. Sia $A \in M_n(\mathbb{C})$ una matrice hermitiana. Allora A è diagonalizzabile e esiste una matrice unitaria M tale che $M^{-1}AM$ è diagonale.

cioè proprio l'analogo del teorema spettrale reale.

Per la teoria sulle coniche e le quadriche, fare riferimento al CAPITOLO 7

Appendix A

Coniche e quadriche

Abbiamo più volte accennato (pagine 185 e 216) al fatto che i metodi che abbiamo imparato in questo corso possono essere utilizzati per studiare non solo gli insiemi di punti rappresentati da equazioni e sistemi di equazioni di primo grado (ovvero rette e piani), ma anche per studiare quelli rappresentati dalla generica equazione di secondo grado, ovvero nel piano (in coordinate rispetto a un riferimento dato da una base ortonormale)

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \tag{A.1}$$

e nello spazio (sempre in coordinate rispetto a un riferimento dato da una base ortonormale)

$$Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx + Hy + Iz + L = 0 \tag{A.2}$$

In questa appendice conclusiva daremo un'idea di come l'equazione (A.1) e la (A.2) corrispondano rispettivamente a 8 e 15 possibilità diverse di insiemi di punti: nel piano si parla di *coniche*, nello spazio di *quadriche*.

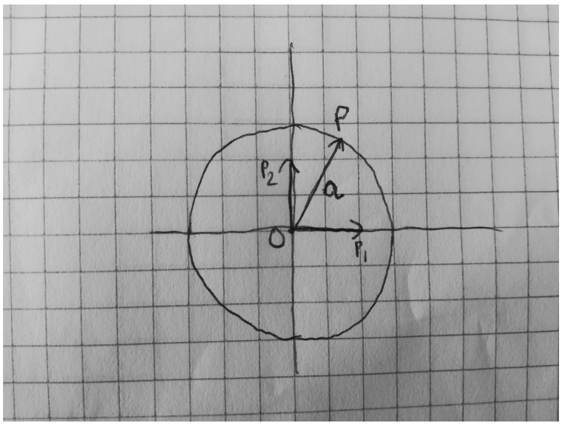
Sia per le prime che per le seconde, inizieremo con il vedere questi 8 e 15 casi rispettivamente, e poi spiegheremo perché qualunque sia l'equazione di partenza ci si può sempre ridurre a uno di essi.

A.1 Coniche

Consideriamo i seguenti casi particolari dell'equazione (A.1):

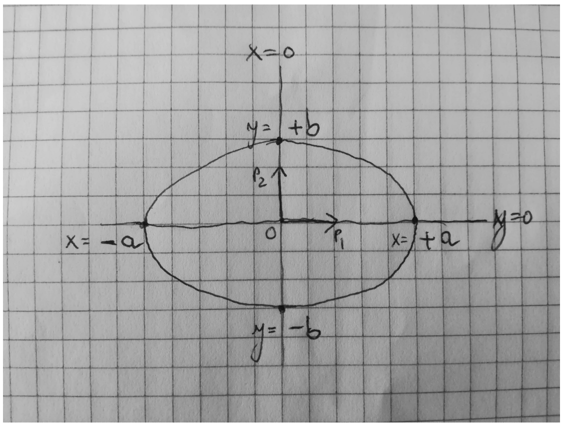
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{A.3}$$

L'insieme dei punti rappresentati dalla (A.3) è un'*ellisse*: nel caso particolare in cui $a = b$ l'equazione diventa $\frac{x^2}{a^2} + \frac{y^2}{a^2} = \frac{x^2+y^2}{a^2} = 1$, ovvero $x^2 + y^2 = a^2$, che dal momento che $x^2 + y^2$ è la lunghezza al quadrato del vettore $\vec{OP}$ rappresentato dalle coordinate (x, y), rappresenta l'insieme di punti tali che la lunghezza di $\vec{OP}$ (cioè la distanza di P da O) è costante e uguale a a, ed è quindi una circonferenza di raggio a



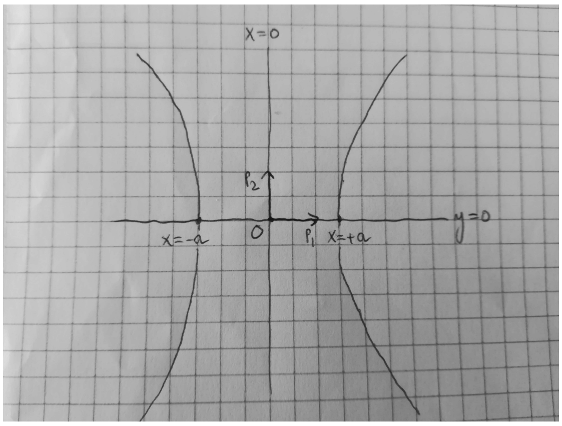
Per $a \neq b$, invece, l'ellisse rapresentata dalla (A.3) può essere pensata come una sorta di "circonferenza deformata", più allungata su un asse rispetto all'altro: questo può essere visto tenendo conto delle sue intersezioni con gli assi, ovvero le rette passanti per l'origine O del sistema di riferimento e aventi come direzioni i vettori $\vec{OP_1}$ e $\vec{OP_2}$ della base ortonormale scelta, che hanno equazioni[1] rispettivamente $y = 0$ e $x = 0$.
Ponendo $y = 0$ nella (A.3) si ottiene $\frac{x^2}{a^2} = 1$, ovvero $x^2 = a^2$ e quindi $x = \pm a$: quindi l'ellisse incontra l'asse orizzontale nei punti di coordinate $(+a, 0)$ e $(-a, 0)$; analogamente, ponendo $x = 0$ nella (A.3) si ottiene $\frac{y^2}{b^2} = 1$, ovvero $y^2 = b^2$ e quindi $y = \pm b$: quindi l'ellisse incontra l'asse verticale nei punti di coordinate $(0, +b)$ e $(0, -b)$

[1] Infatti, se un punto P sta sulla retta avente come direzione $\vec{OP_1}$, vuol dire che $\vec{OP}$ è multiplo di $\vec{OP_1}$ e quindi non ha componente in $\vec{OP_2}$, ovvero $y = 0$, e analogamente se un punto P sta sulla retta avente come direzione $\vec{OP_2}$, vuol dire che $\vec{OP}$ è multiplo di $\vec{OP_2}$ e quindi non ha componente in $\vec{OP_1}$, ovvero $x = 0$.



$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \tag{A.4}$$

l'insieme dei punti rappresentati dalla (A.4) è un'*iperbole*:



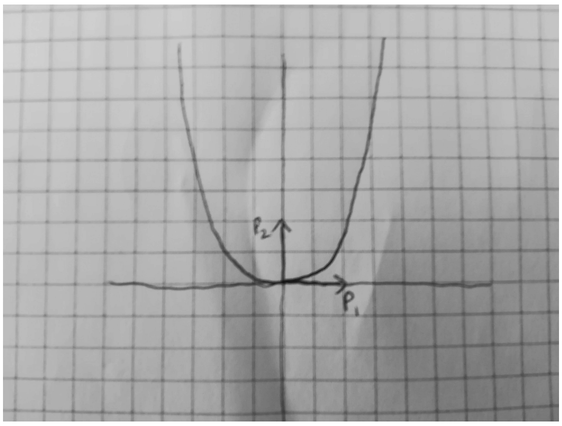
Il fatto che l'iperbole data dall'equazione (A.4) sia formata da due rami distinti che intersecano l'asse orizzontale ma non quello verticale si può di nuovo vedere calcolando le sue intersezioni con gli assi: ponendo $y = 0$ nella (A.4) si ottiene come per l'ellisse $\frac{x^2}{a^2} = 1$, e quindi anche l'iperbole incontra l'asse orizzontale nei punti di coordinate $(+a, 0)$ e $(-a, 0)$; ma stavolta, ponendo $x = 0$ nella (A.4) si ottiene $-\frac{y^2}{b^2} = 1$, ovvero $y^2 = -b^2$, che non ha soluzioni y reali, e quindi tale curva non incontra l'asse verticale, concordemente con il disegno. Si noti che lo stesso problema si ha in generale ponendo $x = k$, con k abbastanza piccolo (positivo o negativo), perché in tal caso l'equazione diventa $-\frac{y^2}{b^2} = 1 - \frac{k^2}{a^2}$, che di nuovo non ha soluzioni se k è troppo piccolo[2],

[2]Più precisamente, se k è in valore assoluto minore di a

in quanto il primo membro è negativo e il secondo positivo: poiché $x = k$ rappresenta una retta verticale parallela all'asse $x = 0$, otteniamo allora la conferma algebrica del fatto che l'iperbole non incontra tutta una fascia di rette verticali a destra e a sinistra dell'origine ed è quindi formata da due rami separati.

$$\frac{x^2}{a^2} - y = 0 \tag{A.5}$$

L'insieme dei punti rappresentati dalla (A.5) è una *parabola*:



Il disegno, e in particolare il fatto che la curva si estenda solo nella metà superiore del piano, è confermato dalla seguente osservazione di tipo algebrico: per ogni y fissato, l'equazione $\frac{x^2}{a^2} = y$ ha soluzioni se y è positivo mentre non ne ha se y è negativo (in quanto il primo membro $\frac{x^2}{a^2}$ è sempre maggiore o uguale a zero).

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0 \tag{A.6}$$

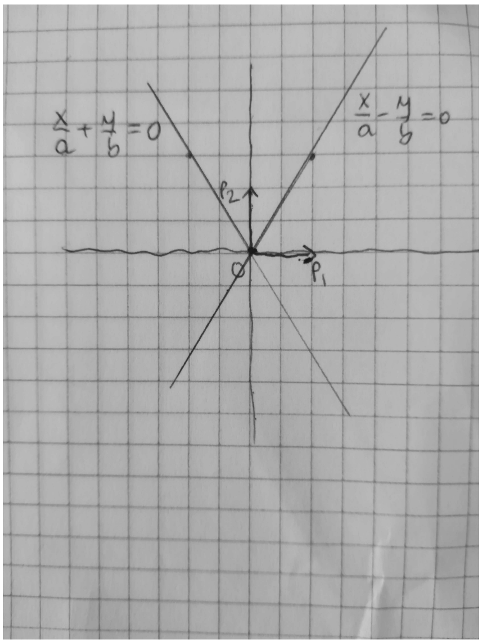
L'equazione (A.6) è soddisfatta solo se $x = 0$ e $y = 0$, quindi l'insieme dei punti da essa rappresentato contiene un *unico punto*, quello di coordinate $(0, 0)$ (ovvero l'origine O del sistema di riferimento)

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1 \tag{A.7}$$

L'equazione (A.7) non è soddisfatta da nessuna coppia di numeri reali x e y, quindi l'insieme dei punti da essa rappresentato è l'*insieme vuoto*.

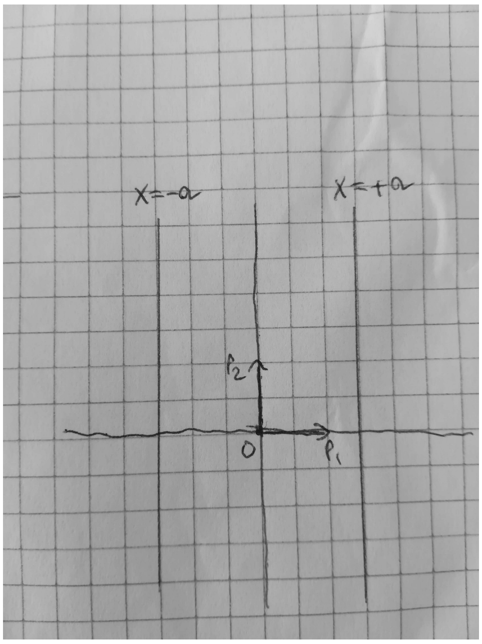
$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0 \tag{A.8}$$

L'equazione (A.8), in base alla nota formula $A^2 - B^2 = (A+B)(A-B)$, si spezza come $\left(\frac{x}{a} + \frac{y}{b}\right)\left(\frac{x}{a} - \frac{y}{b}\right) = 0$ ed è quindi soddisfatta da tutti i punti per cui $\frac{x}{a} + \frac{y}{b} = 0$ e da tutti quelli per cui $\frac{x}{a} - \frac{y}{b} = 0$: queste equazioni rappresentano *due rette incidenti nel piano* (che si incontrano nell'origine):



$$\frac{x^2}{a^2} = 1 \tag{A.9}$$

L'equazione (A.9) è soddisfatta da tutti e soli i punti per cui $x = a$ o $x = -a$, con y qualunque: queste equazioni rappresentano *due rette parallele tra loro* (e parallele all'asse verticale $x = 0$):



$$\frac{x^2}{a^2} = 0 \tag{A.10}$$

L'equazione (A.10) è soddisfatta da tutti e soli i punti per cui $x = 0$, con y qualunque: l'insieme rappresentato dalla (A.8) è allora *un'unica retta*, l'asse verticale $x = 0$.

Ora, il fatto importante è che gli insiemi geometrici rappresentati dalle equazioni (A.3)-(A.10) appena visti esauriscono tutti i possibili casi rappresentabili mediante la generica equazione (A.1): più precisamente, si ha che data una (A.1), l'insieme da essa rappresentato, se non è l'insieme vuoto, è sicuramente un'ellisse, una parabola, un'iperbole, un punto, una coppia di rette incidenti, una coppia di rette parallele o un' unica retta eventualmente ruotate, riflesse e traslate nel piano rispetto alla posizione in cui esse sono rappresentate nei disegni visti sopra.

Questo può essere spiegato mediante le nozioni e le tecniche viste nell'ultimo capitolo: infatti, nell'equazione (A.1) la parte di secondo grado, cioè $Ax^2 + Bxy + Cy^2$, rappresenta una forma quadratica (Osservazione 6.15), ovvero è uguale a $q(v) = f(v, v)$, con $v = (x, y)$, dove f è la forma bilineare simmetrica rappresentata dalla matrice $\begin{pmatrix} A & \frac{B}{2} \\ \frac{B}{2} & C \end{pmatrix}$: come abbiamo visto a pagina 282, con un cambio di coordinate dato da una matrice ortogonale questa matrice può essere diagonalizzata e nelle nuove coordinate x', y' la forma quadratica diventerebbe $\lambda_1 x'^2 + \lambda_2 y'^2$ (ultima formula a pagina 282). Dal momento

che, come sappiamo, una matrice ortogonale di ordine 2 rappresenta una rotazione o una riflessione, stiamo dicendo che si può ruotare o riflettere la curva in modo che la sua equazione inizi con $\lambda_1 x'^2 + \lambda_2 y'^2$, ovvero senza il termine misto xy.
A questo punto, una volta cioè che l'equazione è ridotta nella forma

$$\lambda_1 x^2 + \lambda_2 y^2 + Dx + Ey + F = 0 \tag{A.11}$$

(per facilità stiamo ridenotando x e y le nuove coordinate) si può far sparire almeno uno o entrambi gli addendi Dx e Ey di primo grado tramite delle opportune *traslazioni*. In questo corso non abbiamo studiato questo tipo di trasformazioni, ma basti qui sapere che esse in coordinate si realizzano mediante la trasformazione $(x, y) \mapsto (x + \alpha, y + \beta)$.
Allora, se ad esempio λ_1 è diverso da zero, nella (A.11) si può far sparire il termine Dx proprio con una traslazione: posto $x' = x + \frac{D}{2\lambda_1}$, ovvero sostituendo $x = x' - \frac{D}{2\lambda_1}$, la (A.11) diventa

$$\lambda_1 (x' - \frac{D}{2\lambda_1})^2 + \lambda_2 y^2 + D(x' - \frac{D}{2\lambda_1}) + Ey + F = 0$$

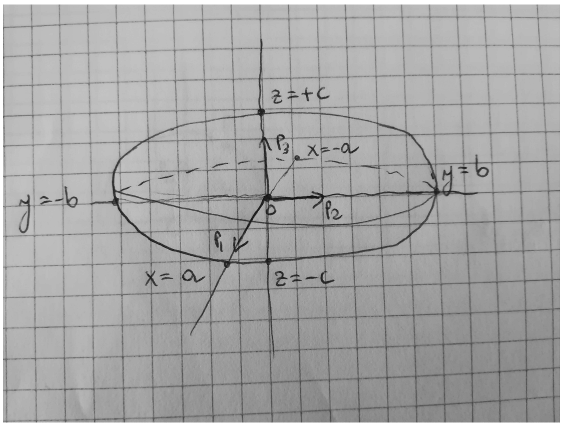
e come si vede sviluppando $\lambda_1(x' - \frac{D}{2\lambda_1})^2 = \lambda_1 x'^2 - Dx' + \frac{D^2}{4\lambda_1}$, il termine Dx' di primo grado scompare semplificandosi con il corrispondente termine dell'addendo $D(x' - \frac{D}{2\lambda_1}) = Dx' - \frac{D^2}{2\lambda_1}$.
Analizzando tutti i casi possibili, si può dimostrare che con passaggi algebrici di questo tipo l'equazione (A.1) si può sempre trasformare in una delle (A.3)-(A.10), ovvero, da un punto di vista geometrico, stiamo dicendo che, a meno di rotazioni, riflessioni e traslazioni, l'insieme di punti rappresentato dalla (A.1) è sicuramente uno degli insiemi visti discutendo le (A.3)-(A.10).

A.2 Quadriche

Il tipo di analisi appena applicata alle coniche nel piano può essere ripetuto anche per l'equazione (A.2). Iniziamo col vedere i seguenti casi particolari di tale equazione.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{A.12}$$

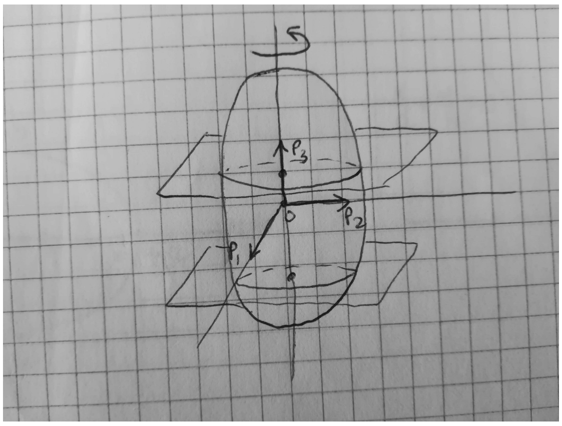
L'insieme di punti rappresentato da questa equazione è un cosiddetto *ellissoide*, rappresentato nel seguente disegno



Come l'ellisse poteva essere pensata come una sorta di circonferenza deformata, ovvero più allungata su un asse che su un altro, l'ellissoide può essere pensato come una sfera deformata, schiacciata lungo alcuni degli assi: infatti, analogamente a quanto visto per l'ellisse, se fossero $a = b = c$, l'equazione (A.12) diventerebbe $\frac{x^2}{a^2} + \frac{y^2}{a^2} + \frac{z^2}{a^2} = \frac{x^2+y^2+z^2}{a^2} = 1$, ovvero $x^2 + y^2 + z^2 = a^2$, che dal momento che $x^2 + y^2 + z^2$ è la lunghezza al quadrato del vettore $\vec{OP}$ dello spazio di coordinate (x, y, z), rappresenta l'insieme di punti tali che la lunghezza di $\vec{OP}$ (cioè la distanza di P da O) è costante e uguale a a, ed è quindi una sfera di centro O e raggio a.
Invece, se a, b e c non sono tutti e tre uguali, questo ci dà uno *schiacciamento* in certe direzioni, come si vede anche considerando le intersezioni dell'ellissoide con gli assi: intersecando con l'asse z, costituito da tutti i punti del tipo $(0, 0, t)$ e quindi ponendo $x = 0$ e $y = 0$ nell'equazione, si trova $\frac{z^2}{c^2} = 1$, ovvero $z = \pm c$, che ci dice dove l'ellissoide interseca l'asse verticale z (e quindi di quanto è "schiacciato" in questa direzione); analogamente, intersecando con l'asse x, costituito da tutti i punti del tipo $(t, 0, 0)$ e quindi ponendo $y = 0$ e $z = 0$ nell'equazione, si trova $\frac{x^2}{a^2} = 1$, ovvero $x = \pm a$, che ci dice invece dove l'ellissoide interseca l'asse x; infine, intersecando con l'asse y, costituito da tutti i punti del tipo $(0, t, 0)$ e quindi ponendo $x = 0$ e $z = 0$ nell'equazione, si trova $\frac{y^2}{b^2} = 1$, ovvero $y = \pm b$, che ci dice invece dove l'ellissoide interseca l'asse y.

Un'altra importante osservazione che si può fare per valutare la forma corretta dell'ellissoide guardando l'equazione (A.12), è che se i coefficienti a, b, c non sono tutti e tre uguali ma lo sono due di loro, allora l'ellissoide risulta essere invariante per rotazioni attorno a un asse: se ad esempio $a = b$, l'equazione si può riscrivere $\frac{x^2+y^2}{a^2} + \frac{z^2}{c^2} = 1$. Ora, se applichiamo una rotazione attorno all'asse z che, come sappiamo dalla (6.30), equivale ad applicare la trasformazione $(x, y, z) \mapsto (\cos\theta x - \sin\theta y, \sin\theta x + \cos\theta y, z)$, i punti ruotati $(x', y', z') = (\cos\theta x - \sin\theta y, \sin\theta x + \cos\theta y, z)$ verificano ancora l'equazione:

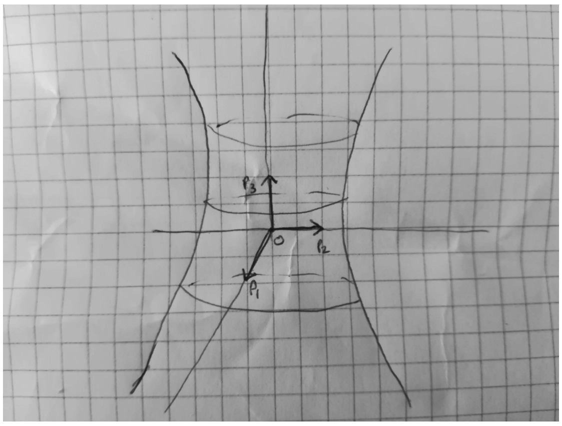
infatti, la trasformazione $(x, y) \mapsto (x', y') = (\cos\theta x - \sin\theta y, \sin\theta x + \cos\theta y)$ subita da x e y coincide con una rotazione nel piano orizzontale, che non cambia la lunghezza rispetto al prodotto scalare standard, ovvero $x'^2 + y'^2 = x^2 + y^2$. Assieme al fatto che $z' = z$, questo ci dice che $\frac{x'^2+y'^2}{a^2} + \frac{z'^2}{c^2} = \frac{x^2+y^2}{a^2} + \frac{z^2}{c^2} = 1$, ovvero (x', y', z') soddisfa ancora l'equazione, come affermato.
Quindi l'ellissoide rimane invariato se ruotiamo di qualunque angolo attorno all'asse z: questo equivale a dire che le sue sezioni con piani orizzontali sono circonferenze



Allo stesso modo, si può vedere che $b = c$ implica che l'ellissoide sia invariante per rotazioni attorno all'asse x e $a = c$ implica che sia invariante per rotazioni attorno all'asse y (in effetti la sfera, per cui $a = b = c$, è invariante per rotazioni attorno a qualunque asse).

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1 \tag{A.13}$$

L'insieme di punti rappresentato da questa equazione è un cosiddetto *iperboloide a una falda*, rappresentato nel seguente disegno

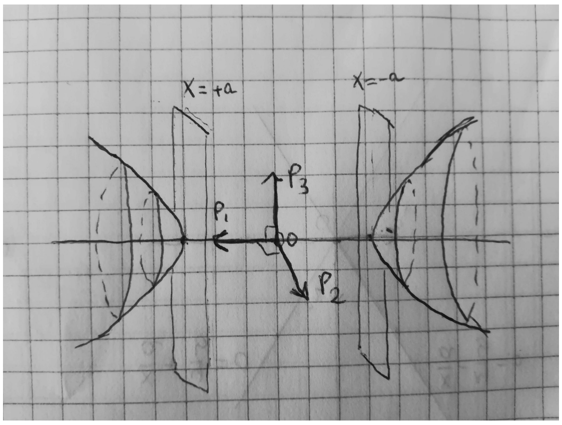


Per giustificare il disegno, notiamo ad esempio che ponendo $z = k$ nell'equazione (ovvero intersecando la superficie con un piano parallelo al piano orizzontale $z = 0$), si ottiene $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 + \frac{k^2}{c^2}$, ovvero, al variare di k, equazioni di ellissi (di "dimensioni" crescenti), mentre intersecando ad esempio con i piani $x = 0$ o $y = 0$ si ottiene rispettivamente $\frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$ e $\frac{x^2}{a^2} - \frac{z^2}{c^2} = 1$, ovvero iperboli.

Analogamente a quanto visto per l'ellissoide, anche in questo caso notiamo che se fosse $a = b$ allora l'equazione si scriverebbe $\frac{x^2+y^2}{a^2} - \frac{z^2}{c^2} = 1$, e come abbiamo spiegato sopra la presenza della quantità $x^2 + y^2$ ci dice che la superficie è invariante per rotazioni attorno all'asse z, ovvero equivalentemente le sue sezioni con piani orizzontali sono circonferenze (di raggio crescente): si parla in questo caso di iperboloide di rotazione.

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1 \tag{A.14}$$

L'insieme di punti rappresentato da questa equazione è un cosiddetto *iperboloide a due falde*, rappresentato nel seguente disegno

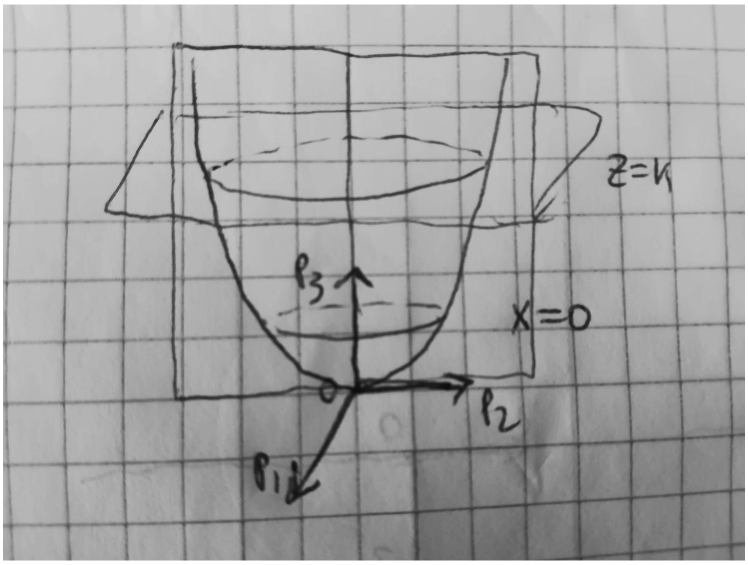


(nel disegno la visuale è ruotata rispetto alla solita rappresentazione degli assi per permettere di vedere meglio la superficie)
La verifica algebrica che questa superficie consiste di due parti sconnesse (da cui le due falde del nome) è analoga a quella fatta nel paragrafo precedente per l'iperbole per verificare che essa consiste di due rami sconnessi: ponendo $x = k$ nell'equazione (e cioè intersecando con piani perpendicolari all'asse x e paralleli al piano $x = 0$) si ottiene $-\frac{y^2}{b^2} - \frac{z^2}{c^2} = 1 - \frac{x^2}{a^2}$: dal momento che il primo membro di questa equazione è negativo o al più nullo, essa non ha soluzioni (ovvero non ci sono intersezioni) se il secondo membro è positivo, ovvero se $1 - \frac{x^2}{a^2} > 0$, cioè se $x^2 < a^2$, che risolto ci dà $-a < x < a$: quindi, la fascia di piani $x = k$, con $-a < k < a$, nella quale non ci sono punti

della superficie, separa le due falde dell'iperboloide, come si vede nel disegno sopra.
Si noti che per valori di k maggiori di a o minori di $-a$, per cui $-\frac{y^2}{b^2} - \frac{z^2}{c^2} = 1 - \frac{k^2}{a^2}$ ha soluzioni perché il piano $x = k$ interseca la superficie, questa equazione è quella di un'ellisse[3]; se invece sezioniamo la superficie con piani $y = k$ o $z = k$ otteniamo rispettivamente $\frac{x^2}{a^2} - \frac{z^2}{c^2} = 1 + \frac{k^2}{b^2}$ e $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 + \frac{k^2}{c^2}$, ovvero delle iperboli (che giustificano il nome di iperboloide dato a questa superficie).

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - z = 0 \tag{A.15}$$

L'insieme di punti rappresentato da questa equazione è un cosiddetto *paraboloide ellittico*:



Il motivo del nome di questa superficie è che sezionando con piani del tipo $z = k$, con k positivo[4], si ottengono le ellissi $\frac{x^2}{a^2} + \frac{y^2}{b^2} = k$, mentre sezionando con i piani $x = 0$ o $y = 0$ si ottiene rispettivamente $z = \frac{y^2}{b^2}$ e $z = \frac{x^2}{a^2}$, ovvero delle parabole[5].

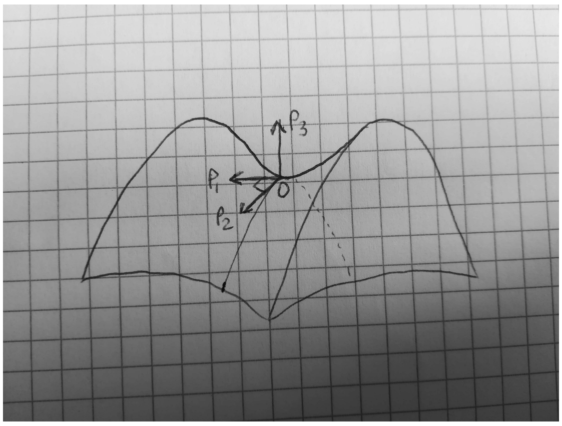
[3]Si cambino di segno entrambi i membri ottenendo $\frac{y^2}{b^2} + \frac{z^2}{c^2} = \frac{k^2}{a^2} - 1$

[4]Se k è negativo, l'equazione $\frac{x^2}{a^2} + \frac{y^2}{b^2} = k$ non ha soluzioni in quanto il primo membro è sempre positivo o nullo, il che spiega perché la superficie sia concentrata nella metà dello spazio corrispondente alla direzione positiva dell'asse z.

[5]Ci limitiamo a questi piani per facilità, ma in realtà è possibile vedere che si ottengono delle parabole sezionando con qualunque piano "verticale", ovvero contenente l'asse z.

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} - z = 0 \tag{A.16}$$

L'insieme di punti rappresentato da questa equazione è un cosiddetto *paraboloide iperbolico*:



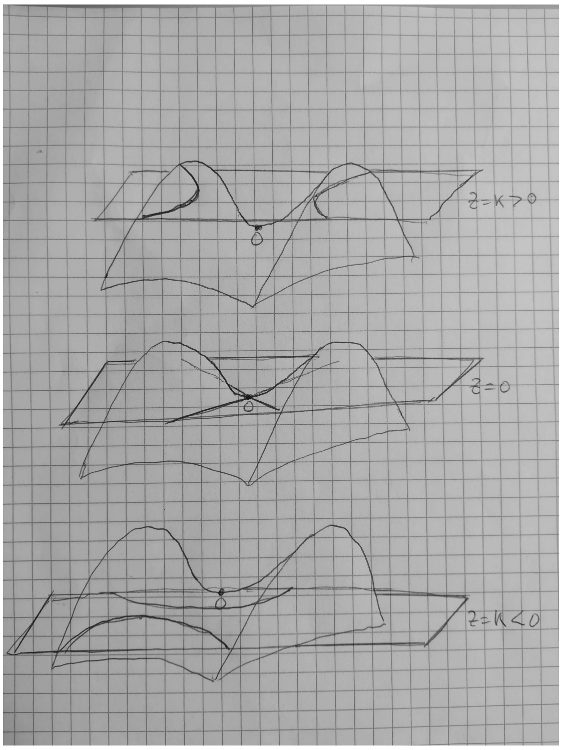
(La visuale è ruotata rispetto alla solita visuale degli assi per permettere di vedere meglio la superficie).

Per capire meglio il disegno e giustificarlo, sezioniamo la superficie con piani orizzontali del tipo $z = k$, ottenendo così l'equazione $\frac{x^2}{a^2} - \frac{y^2}{b^2} = k$. Ora, notiamo che quest'equazione ha soluzioni per ogni valore di k, e più precisamente abbiamo:

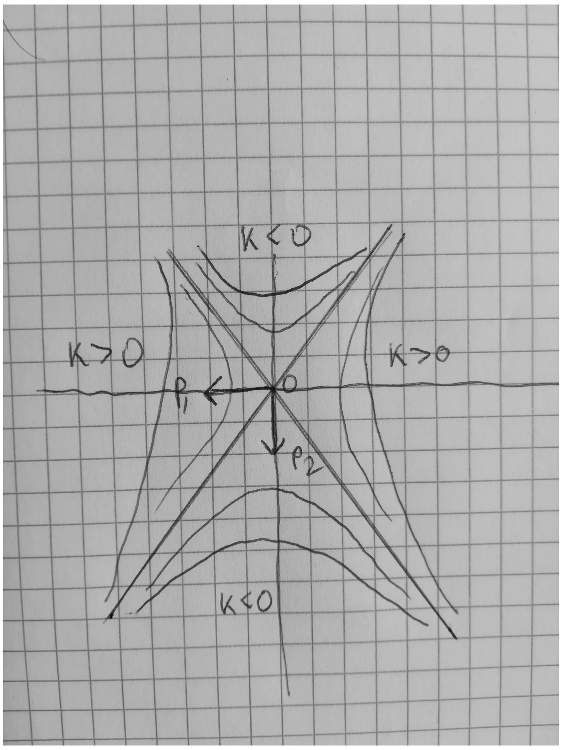
(1) se $k > 0$, l'equazione rappresenta un'iperbole nelle coordinate x, y tale che se $y = 0$ si ha $x = \pm a\sqrt{k}$, mentre per $x = 0$ non ha soluzioni

(2) se $k = 0$, l'equazione diventa $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$ e si spezza come prodotto $(\frac{x}{a} + \frac{y}{b})(\frac{x}{a} - \frac{y}{b}) = 0$, cioè, come abbiamo visto discutendo le coniche nel caso (A.8), una coppia di rette incidenti

(3) se $k < 0$, l'equazione rappresenta un'iperbole nelle coordinate x, y tale che stavolta se $y = 0$ non ha soluzioni, mentre se $x = 0$ si ha $y = \pm b\sqrt{-k}$



Guardando "dall'alto" queste sezioni, si ottiene la rappresentazione seguente, che mostra che sezionando con piani sotto l'origine si ottengono iperboli ruotate di 90 gradi rispetto alle iperboli che si ottengono sezionando piani sopra l'origine.



Si noti anche che sezionando il paraboloide ellittico con il piano $y = 0$ si ottiene la parabola $z = \frac{x^2}{a^2}$ (che essendo $z = \frac{x^2}{a^2} > 0$ si estende verso il semiasse positivo delle z), mentre sezionando con il piano $x = 0$ si ottiene la parabola $z = -\frac{y^2}{b^2}$ (che essendo $z = -\frac{y^2}{b^2} < 0$ si estende verso il semiasse negativo delle z): quindi l'origine, in cui queste due parabole si incontrano, è un punto di minimo lungo la prima parabola ma di massimo lungo la seconda, e quindi sulla superficie non è né di massimo né di minimo (è il tipico cosiddetto "punto di sella")

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = -1 \tag{A.17}$$

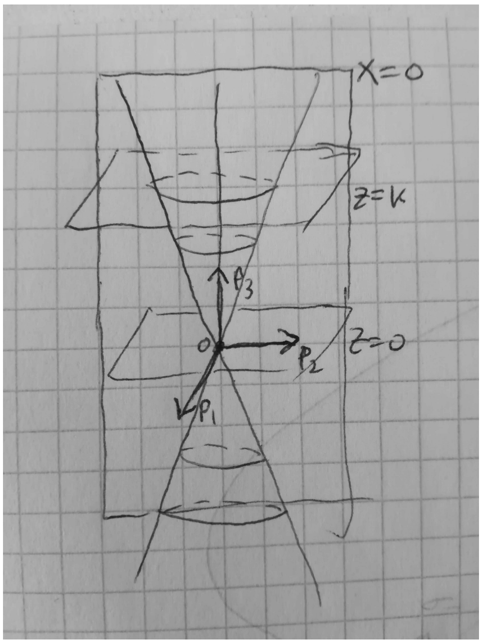
Quest'equazione non è mai soddisfatta per nessun valore di x, y, z, e quindi, ovvero l'insieme da lei rappresentato è l'*insieme vuoto*.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 0 \tag{A.18}$$

Quest'equazione è soddisfatta se e solo se $x = 0, y = 0, z = 0$, ovvero l'insieme dello spazio da lei rappresentata contiene *un solo punto* (l'origine).

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0 \tag{A.19}$$

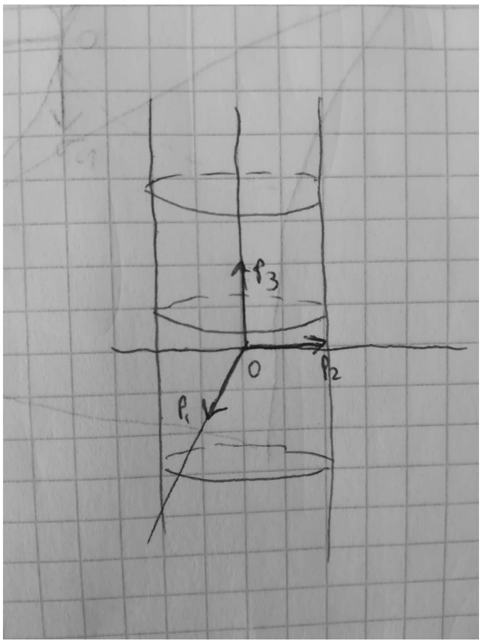
L'insieme di punti rappresentato da questa equazione è un *cono*:



Infatti, sezionando con i piani orizzontali $z=k$ si ottiene $\frac{x^2}{a^2}+\frac{y^2}{b^2}=\frac{k^2}{c^2}$, ovvero equazioni di ellissi, che per $k=0$ degenerano in $\frac{x^2}{a^2}+\frac{y^2}{b^2}=0$ (da cui $x=0$ e $y=0$), ovvero nell'unico punto O.
Il fatto che il profilo della superficie sia formato da rette, si vede sezionando tramite il piano $x=0$ o $y=0$, per cui si ottiene rispettivamente $\frac{y^2}{b^2}-\frac{z^2}{c^2}=(\frac{y}{b}+\frac{z}{c})(\frac{y}{b}-\frac{z}{c})=0$ e $\frac{x^2}{a^2}-\frac{z^2}{c^2}=(\frac{x}{a}+\frac{z}{c})(\frac{x}{a}-\frac{z}{c})=0$, ovvero coppie di rette.

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \tag{A.20}$$

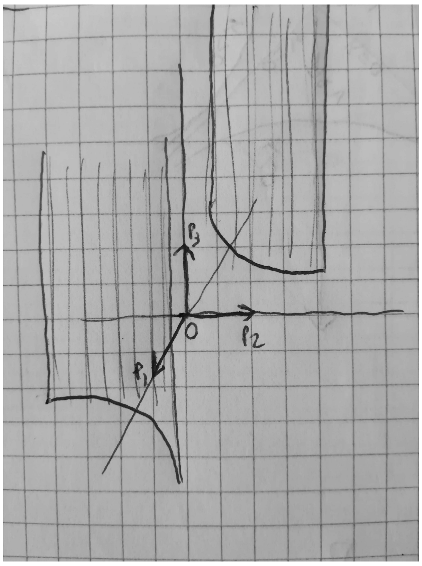
L'insieme di punti rappresentato da questa equazione è un *cilindro ellittico*: infatti, l'equazione non dipende da z e quindi su qualunque piano $z=k$ la sezione è data sempre dalla stessa ellisse $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$.



Analogamente, abbiamo il *cilindro iperbolico*, dato dall'equazione

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \tag{A.21}$$

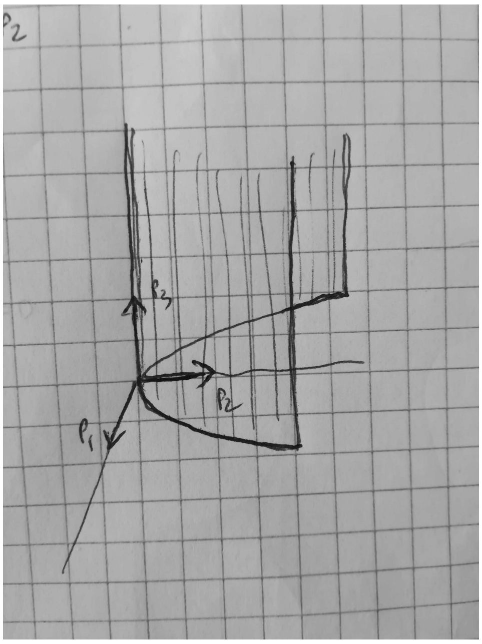
in cui di nuovo l'equazione non dipende da z e quindi su qualunque piano $z = k$ la sezione è data sempre dalla stessa iperbole $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$



e il *cilindro parabolico*, dato dall'equazione

$$\frac{x^2}{a^2} - y = 0 \tag{A.22}$$

in cui, analogamente, su qualunque piano $z = k$ la sezione è data sempre dalla stessa parabola $\frac{x^2}{a^2} - y = 0$

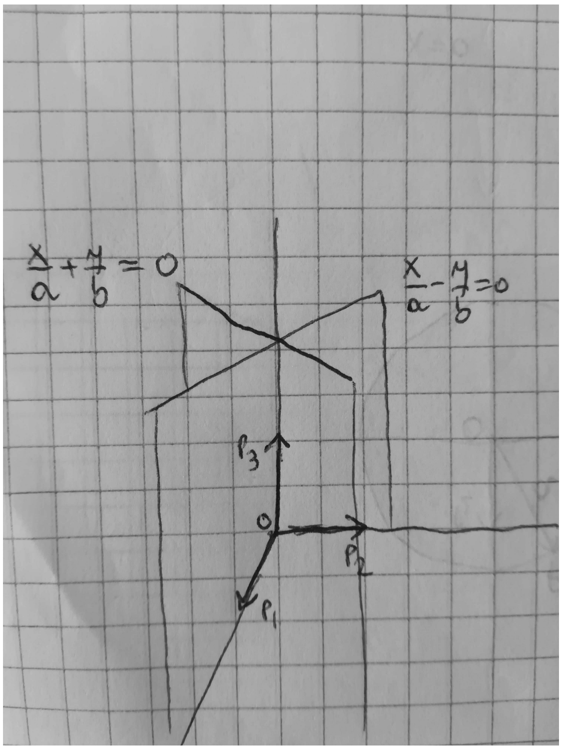


$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0 \tag{A.23}$$

Questa equazione è soddisfatta solo se $x = 0$ e $y = 0$, e da qualunque valore di z (dal momento che questa incognita non compare nell'equazione): quindi i punti che la verificano sono tutti e soli quelli del tipo $(0, 0, t)$, per $t \in \mathbb{R}$, che costituiscono *una retta* (l'asse z stesso).

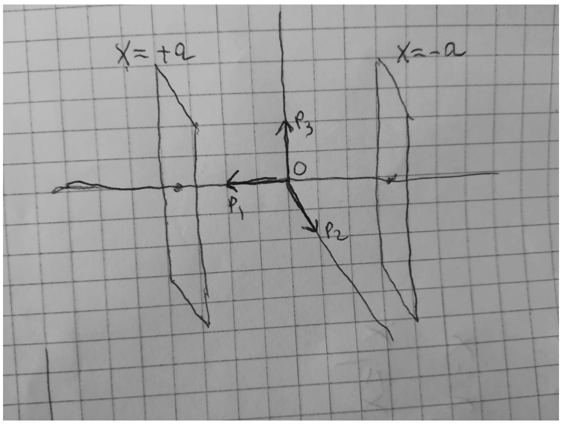
$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0 \tag{A.24}$$

Questa equazione si spezza come $\frac{x^2}{a^2} - \frac{y^2}{b^2} = (\frac{x}{a} + \frac{y}{b})(\frac{x}{a} - \frac{y}{b}) = 0$ ed è quindi soddisfatta sia dai punti tali che stanno sul piano $\frac{x}{a} + \frac{y}{b} = 0$ sia da quelli del piano $\frac{x}{a} - \frac{y}{b} = 0$ (che non sono paralleli): quindi l'equazione rappresenta in questo caso *l'unione di due piani incidenti*



$$\frac{x^2}{a^2} = 1 \tag{A.25}$$

Questa equazione equivale alle due equazioni $x = +a$ e $x = -a$, che rappresentano *due piani nello spazio paralleli tra loro* (e al piano coordinato $x = 0$):



$$\frac{x^2}{a^2} = 0 \tag{A.26}$$

Questa equazione equivale all'unica condizione $x = 0$: la superficie rappresentata è quindi semplicemente *un piano*.

Ora, analogamente a quanto visto per le coniche si ha che gli insiemi geometrici rappresentati dalle equazioni (A.12)-(A.26) appena visti esauriscono tutti i possibili casi rappresentabili mediante la generica equazione (A.2), e il motivo è lo stesso spiegato sopra per le coniche: data una (A.2), l'insieme da essa rappresentato, se non è l'insieme vuoto, è sicuramente uno dei casi visti eventualmente ruotato, riflesso o traslato nel piano rispetto alla posizione in cui essi sono rappresentati nei disegni visti sopra, perché nell'equazione (A.2) la parte di secondo grado $Ax^2+By^2+Cz^2+Dxy+Exz+Fyz$ rappresenta una forma quadratica la cui matrice in base a quanto visto nell'Osservazione 6.15 è la matrice simmetrica $\begin{pmatrix} A & \frac{D}{2} & \frac{E}{2} \\ \frac{D}{2} & B & \frac{F}{2} \\ \frac{E}{2} & \frac{F}{2} & C \end{pmatrix}$: questa matrice può essere diagonalizzata mediante una matrice ortogonale nello spazio (ovvero una rotazione, una riflessione o una riflessione rotatoria) e nelle nuove coordinate x', y', z' definite tramite tale trasformazione l'equazione inizierebbe con $\lambda_1 x'^2 + \lambda_2 y'^2 + \lambda_3 z'^2$ (ultima formula a pagina 282), facendo così sparire i termini misti di secondo grado dall'equazione.
A questo punto, esattamente come detto per le coniche, si possono far sparire uno o più addendi di primo grado tramite delle opportune *traslazioni*.

In questo modo si dimostra che 'equazione (A.2) si può sempre trasformare in una delle (A.12)-(A.26), ovvero, da un punto di vista geometrico, stiamo dicendo che, a meno di rotazioni, riflessioni e traslazioni, l'insieme di punti rappresentato dalla (A.2) è sicuramente uno degli insiemi visti discutendo le (A.12)-(A.26).