Chapter 4

Applicazioni lineari e prodotto di matrici

4.1 Applicazioni lineari: definizione e esempi

Inizieremo ora a parlare di funzioni tra spazi vettoriali. Ricordiamo che una funzione $f : X \to Y$ tra due insiemi X (detto *dominio*) e Y (detto *codominio*) è una legge che associa a ogni $x \in X$ un ben preciso elemento di Y, detto *immagine di* x e denotato $f(x)$.
Noi studieremo le funzioni $f : V \to W$ in cui dominio e codominio V e W sono spazi vettoriali su un certo campo $\mathbb{K}$, e in particolare studieremo quelle che soddisfano la seguente

Definizione 4.1. Una funzione $f : V \to W$ tra spazi vettoriali si dice *funzione lineare* (o *applicazione lineare*) se verifica le due seguenti proprietà:

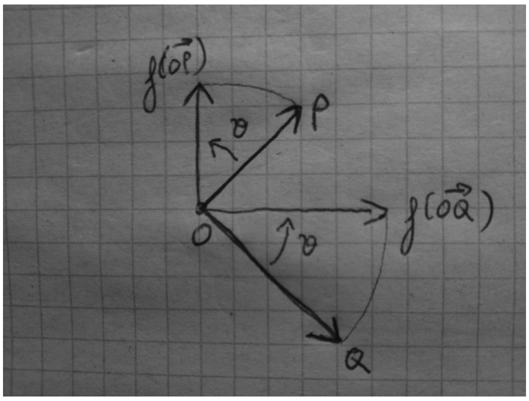
$$f(v + v') = f(v) + f(v') \text{ per ogni } v, v' \in V \tag{4.1}$$

$$f(cv) = cf(v) \text{ per ogni } v \in V \text{ e ogni scalare } c \in \mathbb{K} \tag{4.2}$$

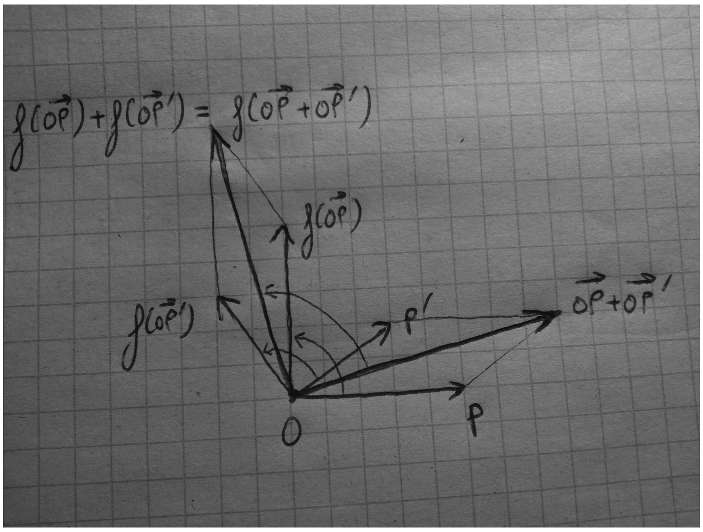
Limitarsi alle funzioni lineari può sembrare molto restrittivo: ad esempio, si può vedere che se[1] $V = W = \mathbb{R}$, le uniche funzioni lineari $f : \mathbb{R} \to \mathbb{R}$ sono quelle del tipo $f(x) = ax$, con $a \in \mathbb{R}$ fissato.
Tuttavia, vediamo subito che tra le applicazioni lineari vi sono funzioni di grande importanza e utilità in geometria e nelle sue applicazioni:

[1] Sappiamo che $\mathbb{R}^n$ è uno spazio vettoriale di dimensione n, in particolare per $n = 1$ si ottiene $\mathbb{R}^1 = \mathbb{R}$ (che risulta quindi uno spazio vettoriale di dimensione 1).

Esempio 4.2. Dato lo spazio V_O^2 dei vettori geometrici applicati nel piano, consideriamo la funzione $f : V_O^2 \to V_O^2$ che associa a ogni vettore $\vec{OP}$ il vettore che si ottiene ruotando $\vec{OP}$ di un angolo θ fissato in senso antiorario attorno all'origine O, come nel disegno seguente



Ora, come si vede nel disegno seguente, dati due vettori $\vec{OP}$ e $\vec{OP'}$, sommarli e poi ruotare il vettore risultante oppure prima ruotarli e poi sommare i vettori ruotati è equivalente, ovvero

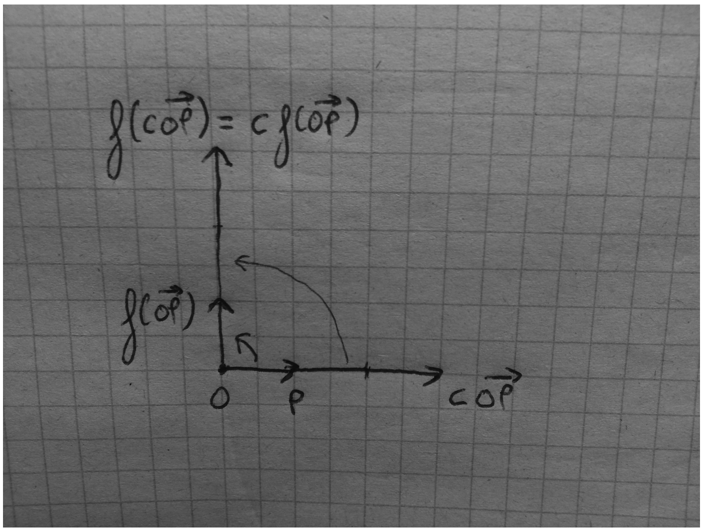


Quindi vale la

$$f(\vec{OP} + \vec{OP'}) = f(\vec{OP}) + f(\vec{OP'}) \tag{4.3}$$

che ci dice che questa funzione soddisfa la proprietà (4.1) della Definizione 4.1

Analogamente, dato un vettore $\vec{OP}$ e un numero reale c, moltiplicare il vettore per c e poi ruotarlo oppure prima ruotarlo e poi moltiplicarlo per c è equivalente:

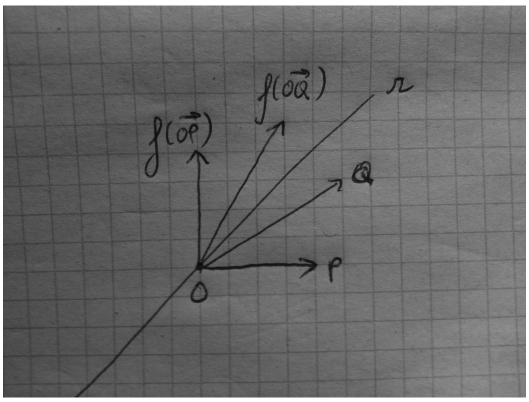


Quindi si ha

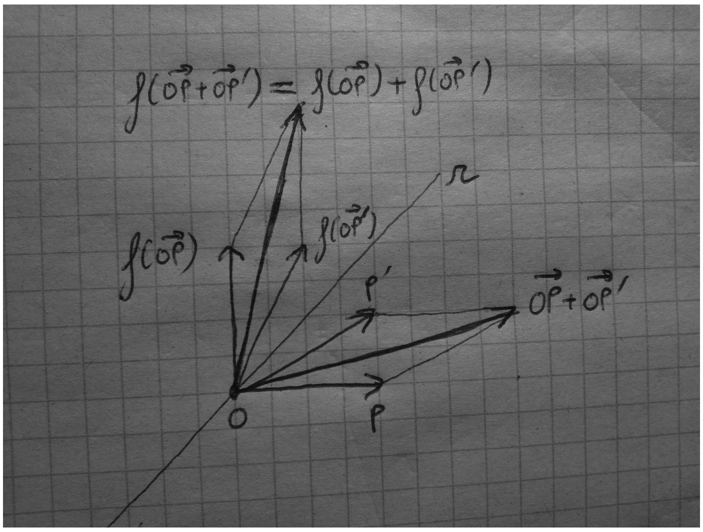
$$f(c\vec{OP}) = cf(\vec{OP}) \tag{4.4}$$

che ci dice che questa funzione soddisfa anche la proprietà (4.2) della Definizione 4.1
Concludiamo quindi che le rotazioni attorno a O sono applicazioni lineari dallo spazio vettoriale V_O^2 in se stesso.

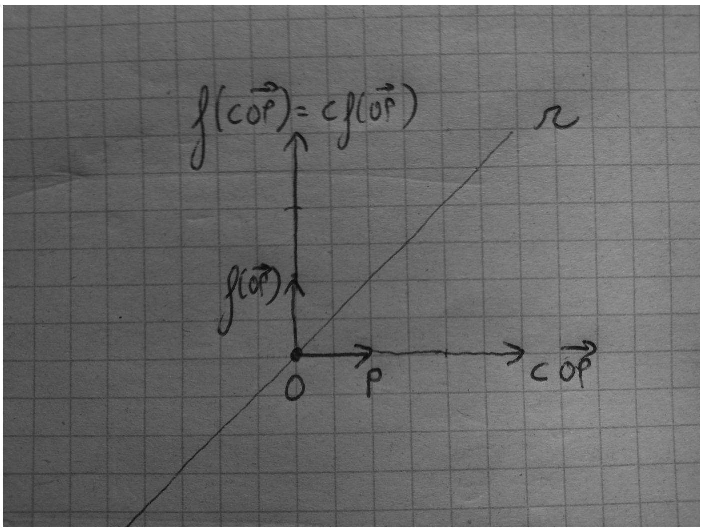
Possiamo arrivare alla stessa conclusione anche per altre importanti trasformazioni geometriche: ad esempio, si consideri la riflessione rispetto a una retta r passante per O, che manda ogni vettore $\vec{OP} \in V_O^2$ nel vettore simmetrico rispetto alla retta, come nel seguente disegno



Allora, come già fatto per le rotazioni, notiamo che, dati due vettori $\vec{OP}$ e $\vec{OP'}$, sommarli e poi riflettere il vettore risultante oppure prima rifletterli e poi sommare i vettori riflessi è equivalente

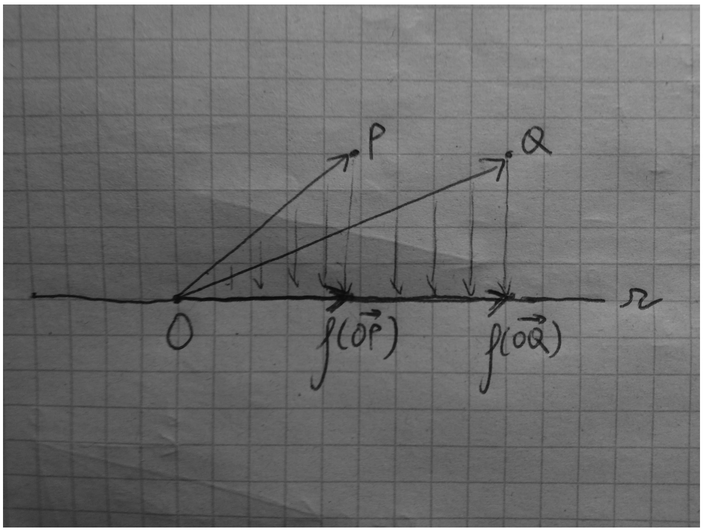


(cioè $f(\vec{OP} + \vec{OP'}) = f(\vec{OP}) + f(\vec{OP'})$) e dato un vettore $\vec{OP}$ e un numero reale c, moltiplicare il vettore per c e poi rifletterlo oppure prima rifletterlo e poi moltiplicarlo per c è equivalente



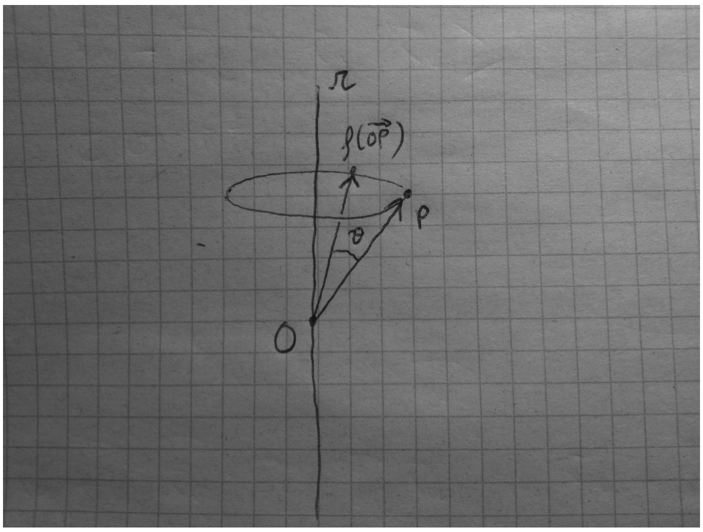
(cioè $f(c\vec{OP}) = cf(\vec{OP})$): quindi concludiamo che anche la riflessione rispetto a una retta che passa per O, avendo le proprietà (4.1) e (4.2) richieste nella Definizione 4.1, è un'applicazione lineare $f : V_O^2 \to V_O^2$.

Come terzo esempio di applicazione lineare $V_O^2 \to V_O^2$ citiamo la proiezione ortogonale, che proietta ortogonalmente i vettori su una retta fissata passante per O.

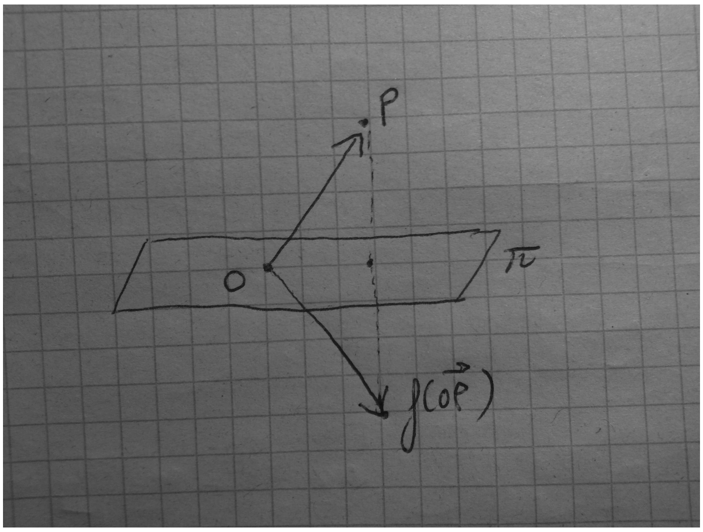


per la quale non è difficile vedere che valgono ancora le proprietà (4.1) e (4.2) (omettiamo ulteriori dettagli).

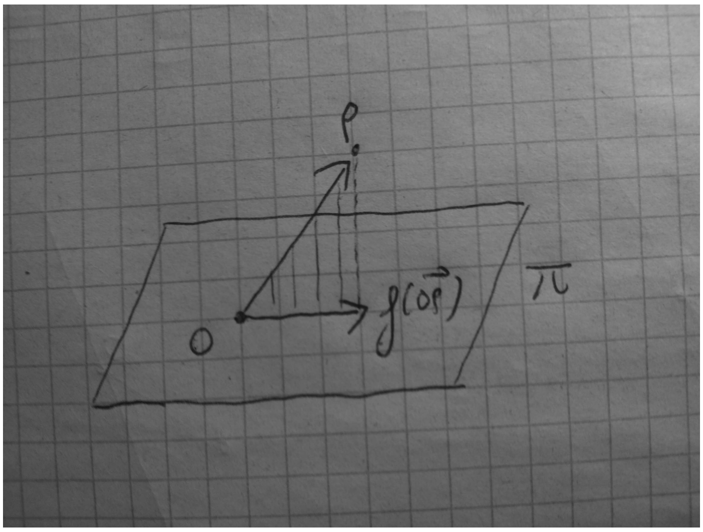
Analogamente a quanto visto per rotazioni, riflessioni e proiezioni nel piano, anche le corrispondenti trasformazioni $V_O^3 \to V_O^3$ dello spazio tridimensionale V_O^3 sono applicazioni lineari: più precisamente, si può mostrare che soddisfano le proprietà (4.1) e (4.2) della Definizione 4.1 la rotazione di un angolo fissato θ attorno a una retta data passante per O (detta *asse della rotazione*)



la riflessione rispetto a un piano passante per O



e la proiezione ortogonale su un piano passante per l'origine O:



(Quest'ultima può anche essere vista come applicazione $f : V_O^3 \to V_O^2$ se vediamo il piano su cui proiettiamo come spazio vettoriale a se stante).

4.2 Matrice associata a un'applicazione lineare

Una delle caratteristiche fondamentali di un'applicazione lineare $f : V \to W$ è che, se gli spazi V e W hanno dimensione finita, allora f può essere rappresentata da una matrice.

Vediamo i dettagli: sia $f : V \to W$ un'applicazione lineare, e siano $B_V = \{v_1, v_2, \ldots, v_n\}$ e $B_W = \{w_1, w_2, \ldots, w_m\}$ basi di V e W rispettivamente.
Allora, come sappiamo ogni vettore $v \in V$ può essere identificato con una n-upla $(x_1, x_2, \ldots, x_n)$, quella delle sue coordinate rispetto alla base B_V (ovvero $v = x_1v_1 + x_2v_2 + \cdots + x_nv_n$), e analogamente ogni vettore $w \in W$ può essere identificato con una m-upla $(y_1, y_2, \ldots, y_m)$, quella delle sue coordinate rispetto alla base B_W (ovvero $w = y_1w_1 + y_2w_2 + \cdots + y_mw_m$).
Con queste identificazioni, la f può essere pensata come una funzione $\mathbb{K}^n \to \mathbb{K}^m$ che associa a ogni n-upla una m-upla. Ci proponiamo di trovare l'espressione esplicita di tale funzione.
A questo scopo, sia $(x_1, \ldots, x_n)$ la n-upla delle coordinate di un vettore $v \in V$, ovvero come abbiamo ricordato sopra $v = x_1v_1 + \cdots + x_nv_n$. Allora la sua immagine $f(v)$ sarà

$$f(v) = f(x_1v_1 + \cdots + x_nv_n) =$$

(usando le proprietà di un'applicazione lineare)

$$= f(x_1v_1) + \cdots + f(x_nv_n) =$$

$$= x_1f(v_1) + \cdots + x_nf(v_n). \tag{4.5}$$

Ora, ciascuno dei vettori $f(v_1), f(v_2), \ldots, f(v_n)$ che compare nella (4.5) appartiene al codominio W della funzione, e quindi potrà essere espresso come combinazione lineare dei vettori $w_1, w_2, \ldots, w_m$ della base B_W fissata per W:

$$f(v_1) = a_{11}w_1 + a_{21}w_2 + \cdots + a_{m1}w_m \tag{4.6}$$

$$\vdots$$

$$f(v_n) = a_{1n}w_1 + a_{2n}w_2 + \cdots + a_{mn}w_m \tag{4.7}$$

Ora, sostituendo queste espressioni nella (4.5) si ottiene

$$f(v) = x_1(a_{11}w_1 + a_{21}w_2 + \cdots + a_{m1}w_m) + \cdots + x_n(a_{1n}w_1 + a_{2n}w_2 + \cdots + a_{mn}w_m) =$$

(svolgendo i conti e mettendo in evidenza i vettori)

$$= (a_{11}x_1 + \cdots + a_{1n}x_n)w_1 + \cdots + (a_{m1}x_1 + \cdots + a_{mn}x_n)w_m \tag{4.8}$$

Questa uguaglianza ci sta dicendo che le coordinate del vettore $f(v)$ rispetto alla base $B_W = \{w_1, w_2, \ldots, w_m\}$ sono date da $a_{11}x_1 + \cdots + a_{1n}x_n, \ldots, a_{m1}x_1 + \cdots + a_{mn}x_n$ e quindi che, tradotta in coordinate, la nostra applicazione lineare può essere identificata con la funzione $\mathbb{K}^n \to \mathbb{K}^m$ che associa a ogni n-upla $(x_1, \ldots, x_n)$ la m-upla formata dai coefficienti che appaiono nella (4.8), ovvero

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdot \\ \cdot \\ \cdot \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}. \tag{4.9}$$

(stiamo scrivendo le n-uple in verticale per questioni di spazio).
I coefficienti che compaiono nella (4.9) formano una matrice con m righe e n colonne

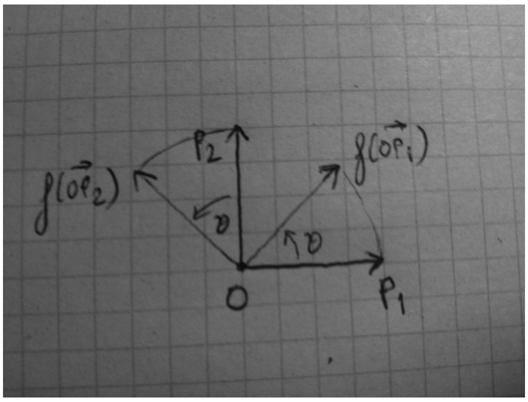
$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

che chiameremo la *matrice associata all'applicazione lineare rispetto alle basi B_V e B_W*. In base alle (4.6), ..., (4.7) tale matrice può essere definita come *la matrice che ha sulle colonne le coordinate dei vettori $f(v_1), \ldots, f(v_n)$ (ovvero le immagini dei vettori della base B_V fissata nel dominio) rispetto alla base B_W fissata nel codominio.*
Denoteremo $M_{B_W B_V}(f)$ la matrice associata a un'applicazione lineare $f : V \to W$ rispetto alle basi B_V e B_W.
Se dominio e codominio dell'applicazione coincidono, ovvero si ha una funzione lineare $f : V \to V$ (tali applicazioni si dicono *endomorfismi*), allora è possibile fissare la stessa base B_V sia nel dominio che nel codominio, e calcolare la matrice associata $M_{B_V B_V}(f)$: in tal caso, per brevità la si denota semplicemente $M_{B_V}(f)$.
In generale, data A, chiameremo *funzione determinata da A* la funzione definita dalla (4.9), e la denoteremo L_A.

Come vedremo, la matrice associata a un'applicazione lineare ci dà tutte le informazioni di cui abbiamo bisogno sull'applicazione, e usando gli strumenti imparati nei capitoli precedenti (rango, determinante) saremo in grado di capire molte proprietà della funzione data. Prima di fare ciò, vediamo subito alcuni esempi di calcolo della matrice associata.

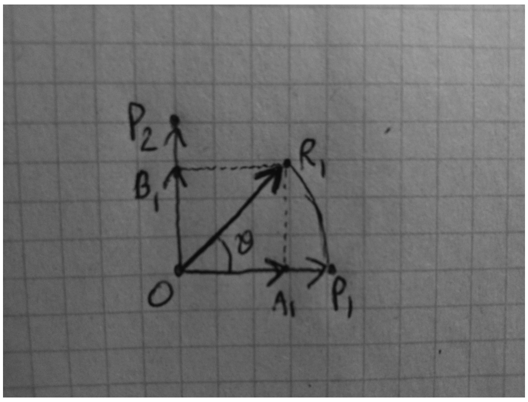
Esempio 4.3. (1) Sia $f : V_O^2 \to V_O^2$ la rotazione attorno a O di un angolo fissato θ (abbiamo visto nel paragrafo precedente che si tratta di un'applicazione lineare).

Per calcolarne la matrice associata, fissiamo una base B formata da due vettori $= v_1 = \vec{OP_1}$ e $v_2 = \vec{OP_2}$ della stessa lunghezza e perpendicolari tra loro, come nel disegno



e determiniamo la matrice $M_B(f)$.

A questo scopo, come afferma la definizione di matrice associata dobbiamo trovare le coordinate di $f(v_1)$ e $f(v_2)$ rispetto a B, ovvero esprimere $f(\vec{OP_1})$ e $f(\vec{OP_2})$ come combinazione lineare $x_1\vec{OP_1}+x_2\vec{OP_2}$ dei vettori della base B. Iniziamo con $f(\vec{OP_1})$: come si vede nel disegno seguente



(nel quale stiamo denotando R_1 il punto finale del vettore ruotato $f(\vec{OP_1})$) si ha $f(\vec{OP_1}) = \vec{OR_1} = \vec{OA_1} + \vec{OB_1}$, essendo A_1 e B_1 le proiezioni ortogonali di R_1 sui vettori di base. Ora, chiaramente $\vec{OA_1} = x_1\vec{OP_1}$ e $\vec{OB_1} = x_2\vec{OP_2}$, dove x_1 è dato dal rapporto $\frac{|\vec{OA_1}|}{|\vec{OP_1}|}$ tra la lunghezza di $\vec{OA_1}$ e quella di $\vec{OP_1}$, mentre x_2 è dato dal rapporto $\frac{|\vec{OB_1}|}{|\vec{OP_2}|}$ tra la lunghezza di $\vec{OB_1}$ e quella di $\vec{OP_2}$. Ma essendo la lunghezza di

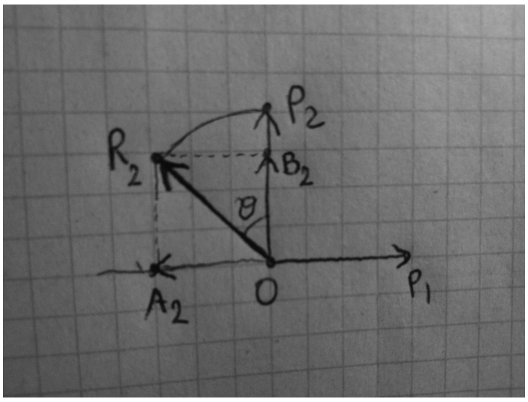
$\vec{OP_1}$ uguale alla lunghezza di $\vec{OR_1} = f(\vec{OP_1})$ (la rotazione non modifica le lunghezze dei vettori) possiamo dire che x_1 è uguale al rapporto tra la lunghezza del cateto $\vec{OA_1}$ e quella dell'ipotenusa $\vec{OR_1}$ del triangolo rettangolo OR_1A_1, ovvero (essendo il cateto in questione adiacente all'angolo), $x_1 = \cos\theta$.

Analogamente, poichè $\vec{OP_2}$ ha la stessa lunghezza di $\vec{OP_1}$ e quindi di $\vec{OR_1}$, mentre $\vec{OB_1}$ ha la stessa lunghezza del segmento A_1R_1, si ha che $x_2 = \frac{|\vec{OB_1}|}{|\vec{OP_2}|} = \frac{|A_1R_1|}{|OR_1|}$, ovvero (essendo il rapporto tra il cateto A_1R_1, opposto all'angolo θ, e l'ipotenusa OR_1 del triangolo rettangolo OR_1A_1), $x_2 = \sin\theta$.

Riassumendo,

$$f(\vec{OP_1}) = \vec{OR_1} = \vec{OA_1} + \vec{OB_1} = \cos\theta\ \vec{OP_1} + \sin\theta\ \vec{OP_2} \tag{4.10}$$

Ora, facciamo un ragionamento analogo per determinare $f(\vec{OP_2})$: come si vede nel disegno seguente



(nel quale stiamo denotando R_2 il punto finale del vettore ruotato $f(\vec{OP_2})$) si ha $f(\vec{OP_2}) = \vec{OR_2} = \vec{OA_2} + \vec{OB_2}$. Ora, chiaramente $\vec{OA_2} = -x_1\vec{OP_1}$, dove x_1 è dato dal rapporto $\frac{|\vec{OA_2}|}{|\vec{OP_1}|}$ tra la lunghezza di $\vec{OA_2}$ e quella di $\vec{OP_1}$ (il segno meno è dovuto al fatto che $\vec{OA_2}$ ha verso opposto rispetto a $\vec{OP_1}$) e $\vec{OB_2} = x_2\vec{OP_2}$, dove x_2 è dato dal rapporto $\frac{|\vec{OB_2}|}{|\vec{OP_2}|}$ tra la lunghezza di $\vec{OB_2}$ e quella di $\vec{OP_2}$. Ma essendo la lunghezza di $\vec{OP_1}$ uguale alla lunghezza di $\vec{OP_2}$ e quindi di $\vec{OR_2} = f(\vec{OP_2})$ (sempre perchè la rotazione non modifica le lunghezze dei vettori) mentre la lunghezza di $\vec{OA_2}$ è uguale alla lunghezza del segmento R_2B_2, possiamo dire che x_1 è uguale al rapporto tra la lunghezza del cateto R_2B_2 e quella dell'ipotenusa OR_2 del triangolo rettangolo OR_2B_2, ovvero (essendo il cateto in questione opposto all'angolo), $x_1 = \sin\theta$.

Analogamente, poichè $\vec{OP_2}$ ha la stessa lunghezza di $\vec{OR_2}$, si ha che $x_2 = \frac{|\vec{OB_2}|}{|\vec{OP_2}|} = \frac{|OB_2|}{|OR_2|}$, che è il rapporto tra il cateto e l'ipotenusa del triangolo rettangolo OB_2R_2, ovvero (essendo il cateto in questione adiacente all'angolo), $x_2 = \cos\theta$.

Riassumendo,

$$f(\vec{OP_2}) = \vec{OR_2} = \vec{OA_2} + \vec{OB_2} = -\sin\theta\ \vec{OP_1} + \cos\theta\ \vec{OP_2} \qquad (4.11)$$

Quindi, (4.10) e (4.11) ci dicono che la matrice associata a f rispetto a B avrà sulla prima colonna $(\cos\theta, \sin\theta)$ (le coordinate di $f(\vec{OP_1})$ rispetto a B) e sulla seconda colonna $(-\sin\theta, \cos\theta)$ (le coordinate di $f(\vec{OP_2})$ rispetto a B), ovvero

$$M_B(f) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \qquad (4.12)$$

In base a quanto visto nella (4.9), abbiamo allora che la rotazione, in coordinate, si traduce nella funzione $f : \mathbb{R}^2 \to \mathbb{R}^2$ data da

$$(x_1, x_2) \mapsto (\cos\theta x_1 - \sin\theta x_2, \sin\theta x_1 + \cos\theta x_2) \qquad (4.13)$$

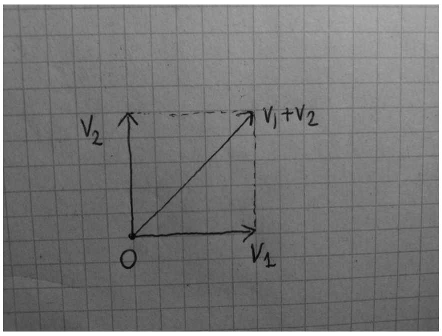
Ad esempio, scegliamo $\theta = \frac{\pi}{4}$ e sostituiamo in (4.12) e (4.13), ottenendo rispettivamente (si ricordi che $\cos\frac{\pi}{4} = \sin\frac{\pi}{4} = \frac{\sqrt{2}}{2}$)

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

e

$$(x_1, x_2) \mapsto \left(\frac{\sqrt{2}}{2}x_1 - \frac{\sqrt{2}}{2}x_2, \frac{\sqrt{2}}{2}x_1 + \frac{\sqrt{2}}{2}x_2\right) \qquad (4.14)$$

Per illustrare come questa semplice funzione $\mathbb{R}^2 \to \mathbb{R}^2$ rappresenti effettivamente la rotazione di $\frac{\pi}{4}$, prendiamo ad esempio il vettore $v = v_1 + v_2$, che come si vede nel disegno seguente coincide con la diagonale del quadrato che ha v_1 e v_2 come lati:

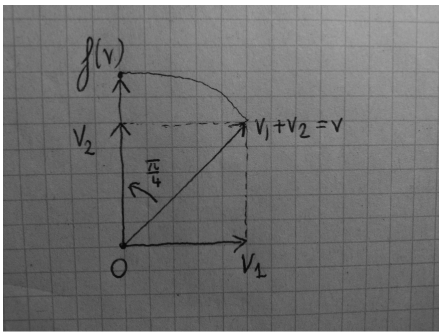


Tale vettore ha quindi come coordinate rispetto a B la coppia $(x_1, x_2) = (1, 1)$. In base alla (4.14), le coordinate di $f(v)$ rispetto a B sono date da

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

ovvero deve essere $f(v) = 0v_1 + \sqrt{2}v_2 = \sqrt{2}v_2$.

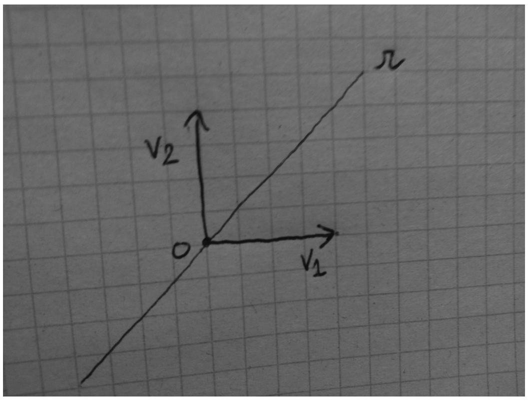
In effetti, tale risultato ottenuto analiticamente in coordinate è confermato dall'analisi grafica, che ci dice che il vettore $f(v)$ che si ottiene ruotando la diagonale v del quadrato di $\frac{\pi}{4}$ è proprio proporzionale al vettore v_2 e la sua lunghezza è proprio $\sqrt{2}$ volte la lunghezza di v_2 (in quanto v coincideva con la diagonale del quadrato di lati v_1 e v_2):



(2) Sia $V = V_O^2$ lo spazio vettoriale dei vettori applicati in un punto O nel piano e sia $f : V_O^2 \to V_O^2$ la proiezione ortogonale su una retta fissata

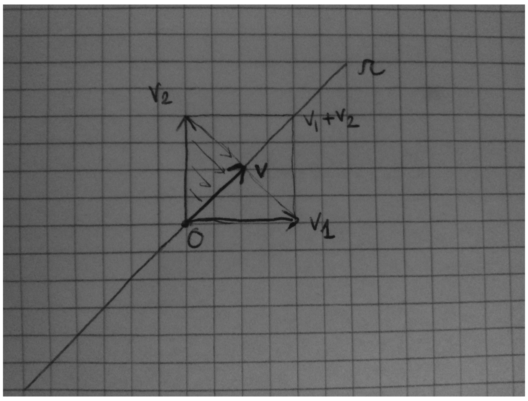
r passante per O (abbiamo visto nel paragrafo precedente che si tratta di un'applicazione lineare).

Per calcolarne la matrice associata $M_B(f)$, fissiamo una base $B = \{v_1, v_2\}$ di V_O^2 come nel disegno seguente



(i vettori v_1 e v_2 sono vettori della stessa lunghezza, perpendicolari tra loro e disposti in modo che la retta r su cui proiettiamo sia la bisettrice dell'angolo da essi formato).

Ora, notiamo che quando proiettiamo v_1 ortogonalmente su r, otteniamo un vettore v che sta sulla retta ed è lungo come metà della diagonale del quadrato i cui lati sono v_1 e v_2: essendo tale diagonale, per definizione di somma tra vettori, coincidente con $v_1 + v_2$, abbiamo quindi che $f(v_1) = \frac{1}{2}(v_1 + v_2) = \frac{1}{2}v_1 + \frac{1}{2}v_2$; analogamente, come si vede dal disegno, anche proiettando v_2 sulla retta di ottiene lo stesso vettore v, quindi si ha anche $f(v_2) = \frac{1}{2}(v_1 + v_2) = \frac{1}{2}v_1 + \frac{1}{2}v_2$.



Si vede quindi che le coordinate di $f(v_1)$ rispetto a B sono $(\frac{1}{2}, \frac{1}{2})$, e anche le le coordinate di $f(v_2)$ rispetto a B sono $(\frac{1}{2}, \frac{1}{2})$: disponendo tali coordinate rispettivamente sulla prima e sulla seconda colonna, come previsto dalla definizione di matrice associata, si ottiene

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

e la funzione $\mathbb{R}^2 \to \mathbb{R}^2$ corrispondente

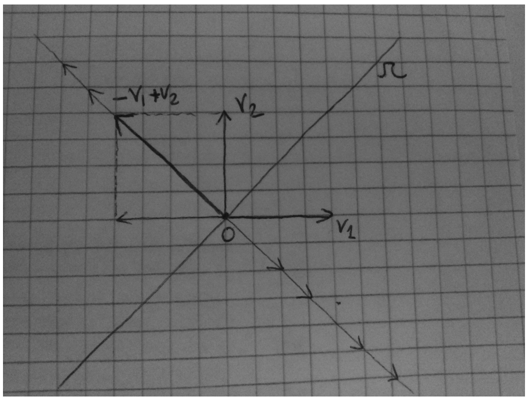
$$(x_1, x_2) \mapsto \left(\frac{1}{2}x_1 + \frac{1}{2}x_2, \frac{1}{2}x_1 + \frac{1}{2}x_2\right) \tag{4.15}$$

ci dà una rappresentazione in coordinate della proiezione.

Ad esempio, il vettore $v = -v_1 + v_2$, che ha coordinate $(-1, 1)$ rispetto a B, viene mandato in base alla (4.15) nel vettore di coordinate

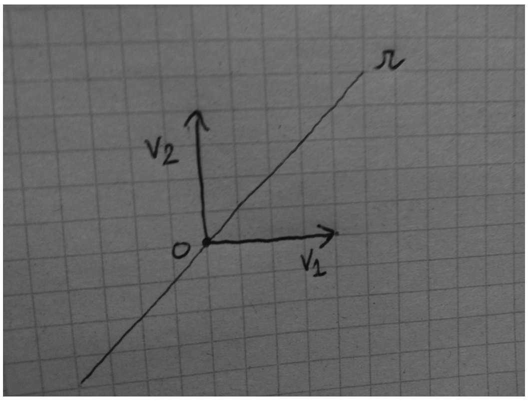
$$\left(\frac{1}{2} \cdot (-1) + \frac{1}{2} \cdot 1, \frac{1}{2} \cdot (-1) + \frac{1}{2} \cdot 1\right) = (0, 0)$$

ovvero nel vettore nullo $\vec{OO}$. Infatti, come si vede dal seguente disegno, tale vettore appartiene alla retta passante per O e ortogonale a r, e i vettori che giacciono su questa retta vengono chiaramente proiettati sul vettore nullo $\vec{OO}$.



(3) Come ultimo esempio, prendiamo ora come $f : V_O^2 \to V_O^2$ la riflessione rispetto a una retta fissata r passante per O, ovvero l'endomorfismo che associa a ogni vettore $\vec{OP}$ il suo simmetrico rispetto a r (sappiamo dal primo paragrafo che si tratta di un'applicazione lineare).

Per calcolarne la matrice associata $M_B(f)$, consideriamo la stessa base $B = \{v_1, v_2\}$ usata nell'esempio precedente

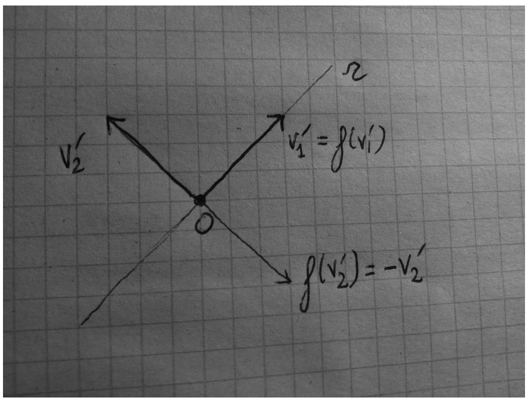


e notiamo che quando riflettiamo v_1 rispetto a r otteniamo v_2, ovvero $f(v_1) = v_2$, e analogamente quando riflettiamo v_2 rispetto a r, otteniamo v_1, ovvero $f(v_2) = v_1$.

Quindi, riscrivendo $f(v_1) = v_2$ come $f(v_1) = 0v_1 + 1v_2$ vediamo che le coordinate di $f(v_1)$ rispetto a B sono $(0, 1)$, e analogamente riscrivendo $f(v_2) = v_1$ come $f(v_2) = 1v_1 + 0v_2$ vediamo che le coordinate di $f(v_2)$ rispetto a B sono $(1, 0)$: disponendo tali coordinate in colonna, come previsto dalla definizione di matrice associata, si ottiene

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

Se, dato sempre lo stesso endomorfismo, consideriamo invece la base $B' = \{v'_1, v'_2\}$ come nel disegno seguente



allora si ha che $f(v_1') = v_1'$ (il vettore v_1' sta sulla retta e quindi la riflessione rispetto alla retta lo lascia invariato) e $f(v_2') = -v_2'$ (il vettore v_2' è perpendicolare alla retta quindi riflettendolo esso cambia verso), cioè $f(v_1') = 1v_1' + 0v_2'$ e $f(v_2) = 0v_1' + (-1)v_2'$ e quindi

$$M_{B'}(f) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

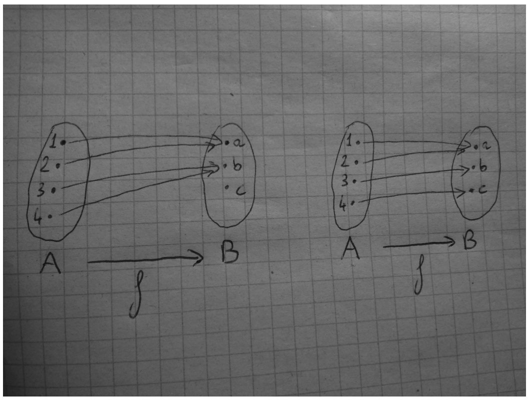
Questo esempio illustra il fatto ovvio che la matrice associata dipende dalla scelta delle basi.

4.3 Iniettività e suriettività di applicazioni lineari

Il primo problema che affronteremo sulle applicazioni lineari è determinare quando una tale funzione è iniettiva, suriettiva o biiettiva.

4.3.1 Richiami generali

Ricordiamo che una funzione $f : A \to B$ tra due insiemi A e B si dice *suriettiva* se ogni elemento del codominio B risulta essere immagine di qualche elemento di A (ovvero, se per ogni $b \in B$ esiste un $a \in A$ tale che $f(a) = b$). Ad esempio, delle funzioni rappresentate nel seguente disegno, la prima non è suriettiva (l'elemento $c \in B$ non è immagine di nessun elemento di A), la seconda sì.

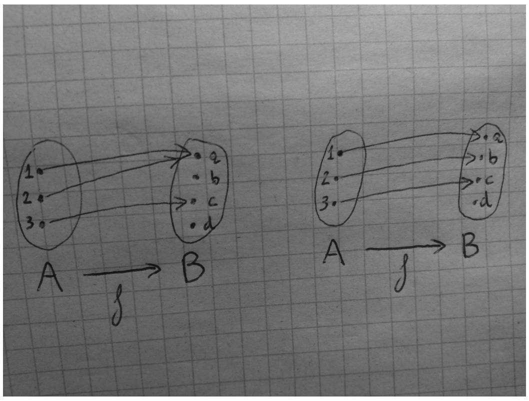


Un modo alternativo di dire che una funzione è suriettiva è fare riferimento alla cosiddetta *immagine* $Im(f)$ di f: per definizione, l'immagine di una funzione $f : A \to B$ è il sottoinsieme di B costituito da tutti gli elementi che sono immagine di qualche elemento di A (in riferimento ai disegni, quegli elementi "raggiunti da una freccia che proviene da A"), ovvero

$$Im(f) = \{b \in B \mid b = f(a) \text{ per qualche } a \in A\}$$

Ad esempio, la funzione a sinistra nel disegno precedente ha $Im(f) = \{a, b\}$, mentre la funzione a destra $Im(f) = \{a, b, c\}$: una funzione è suriettiva esattamente quando $Im(f) = B$, ovvero l'immagine coincide con tutto il codominio (dire $Im(f) = B$ significa in effetti dire che ogni elemento di B è immagine di qualche elemento di A).

Una funzione $f : A \to B$ si dice invece *iniettiva* se non succede che due elementi diversi di A abbiano la stessa immagine (in formule, f è iniettiva se $a_1 \neq a_2 \Rightarrow f(a_1) \neq f(a_2)$ o equivalentemente $f(a_1) = f(a_2) \Rightarrow a_1 = a_2$).
Ad esempio, delle funzioni rappresentate nel seguente disegno, la prima non è iniettiva (in quanto nonostante $1 \neq 2$ si ha $f(1) = f(2) = a$), la seconda sì.



La nozione di iniettività può essere riformulata tramite il concetto di *controimmagine*: dato un elemento b del codominio B, la sua controimmagine, denotata $f^{-1}(b)$, è l'insieme di tutti gli elementi di A che hanno b come immagine, ovvero

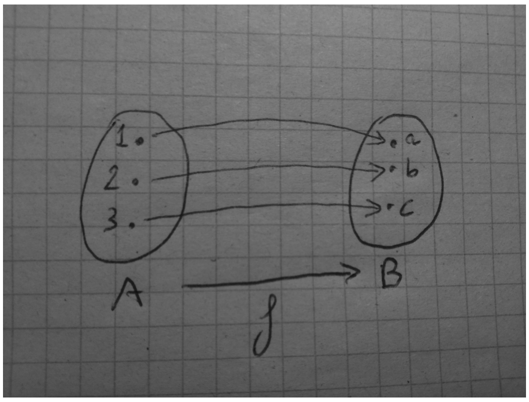
$$f^{-1}(b) = \{a \in A \mid f(a) = b\}$$

Dal momento che una funzione è iniettiva quando non esistono due elementi diversi che hanno la stessa immagine, dire che una funzione è iniettiva equivale a dire che tutte le controimmagini che non siano vuote[2] hanno un solo elemento.

[2]se la funzione non è suriettiva, ci saranno elementi $b \in B$ tali che non esiste nessun $a \in A$ con $f(a) = b$, e quindi la cui controimmagine $f^{-1}(b)$ non ha elementi.

Ad esempio, per la funzione a sinistra nel disegno precedente si ha $f^{-1}(a) = \{1, 2\}$, $f^{-1}(b) = f^{-1}(d) = \emptyset$, $f^{-1}(c) = \{3\}$: essa non è iniettiva in quanto la controimmagine di a ha due elementi.
Per la funzione a destra, invece, si ha $f^{-1}(a) = \{1\}$, $f^{-1}(b) = \{2\}$, $f^{-1}(c) = \{3\}$, $f^{-1}(d) = \emptyset$: essa è iniettiva in quanto le controimmagini non vuote hanno tutte un solo elemento.

Infine, una funzione si dice *biiettiva* se è sia iniettiva che suriettiva. Ad esempio, la funzione rappresentata nel seguente disegno è biiettiva.



4.3.2 Suriettività di applicazioni lineari

Abbiamo detto che una funzione $f : A \to B$ è suriettiva se e solo se per ogni elemento $b \in B$ esiste un $a \in A$ tale che $f(a) = b$, ovvero equivalentemente se e solo se la sua immagine $Im(f)$ coincide con tutto il codominio.
Quindi, per capire se un'applicazione è suriettiva, dobbiamo determinare l'insieme $Im(f)$.
Nel caso di un'applicazione lineare $f : V \to W$, vale la seguente importante

Proposizione 4.4. Sia $f : V \to W$ un'applicazione lineare. Allora $Im(f)$ è un sottospazio vettoriale di W.

Proof. Dobbiamo verificare che $Im(f)$ è chiuso rispetto alla somma e al prodotto per scalari. Per la prima proprietà dobbiamo prendere $w, w' \in Im(f)$ e vedere se $w + w' \in Im(f)$. Ora, se $w, w' \in Im(f)$, per definizione di $Im(f)$ significa che esistono un vettore $v \in V$ tale che $w = f(v)$ e un vettore $v' \in V$ tale che $w' = f(v')$. Ma allora, sfruttando il fatto che f è lineare, si ha

$$w + w' = f(v) + f(v') = f(v + v')$$

che ci dice che anche $w + w'$ è immagine di un elemento del dominio (cioè $v + v'$) e quindi $w + w' \in Im(f)$.

Per la chiusura rispetto al prodotto per scalari, dobbiamo verificare che se $w \in Im(f)$ e $c \in \mathbb{K}$, allora $cw \in Im(f)$. Ma, come prima, se $w \in Im(f)$ allora per definizione di $Im(f)$ esiste un vettore $v \in V$ tale che $w = f(v)$, e quindi, usando sempre il fatto che f è lineare,

$$cw = cf(v) = f(cv)$$

che ci dice che anche cw è immagine di un elemento del dominio (cioè cv) e quindi $cw \in Im(f)$. □

Il fatto che $Im(f)$ sia un sottospazio vettoriale ci dice che per determinarla possiamo trovarne un sistema di generatori o una base. Questo si fa facilmente grazie alla seguente

Proposizione 4.5. Sia $f : V \to W$ un'applicazione lineare e siano $v_1, \dots, v_n$ generatori di V. Allora le immagini $f(v_1), \dots, f(v_n)$ generano $Im(f)$ (in simboli, $Im(f) = \langle f(v_1), \dots, f(v_n) \rangle$)

Proof. Per definizione di generatori, dobbiamo verificare che ogni vettore $w \in Im(f)$ si può scrivere come combinazione lineare dei vettori $f(v_1), \dots, f(v_n)$. Ora, sappiamo che un vettore $w \in Im(f)$ è tale che $w = f(v)$ per qualche vettore $v \in V$. Ma essendo per ipotesi $v_1, \dots, v_n$ generatori di V, il vettore v potrà essere scritto come loro combinazione lineare $v = x_1v_1 + \cdots + x_nv_n$. Quindi, sfruttando la linearità di f si ha

$$w = f(v) = f(x_1v_1 + \cdots + x_nv_n) = x_1f(v_1) + \cdots + x_nf(v_n)$$

che dimostra proprio che w si scrive come combinazione lineare di $f(v_1), \dots, f(v_n)$, come volevamo. □

A questo punto, per determinare se un'applicazione lineare $f : V \to W$ è suriettiva, basta scegliere dei generatori $v_1, \dots, v_n$ di V, prendere le loro immagini $f(v_1), \dots, f(v_n)$ che, come abbiamo appena visto, formano un insieme di generatori di $Im(f)$, e poi estrarre da quest'ultimo insieme una base eliminando gli eventuali vettori che sono dipendenti dai rimanenti: contando i vettori della base ottenuta, sapremo la dimensione di $Im(f)$ e quindi f sarà suriettiva se e solo se[3] $\dim(Im(f)) = \dim(W)$.

Vediamo come tale criterio di suriettività si rivela particolarmente utile e di semplice applicazione nel caso dell'applicazione lineare $L_A : \mathbb{K}^n \to \mathbb{K}^m$ determinata da una matrice A, ovvero della forma data dalla (4.9) (come

[3] Questa affermazione è giustificata dal fatto, che non abbiamo dimostrato, che se S è un sottospazio vettoriale di uno spazio vettoriale V, allora $\dim(S) \leq \dim(V)$ e le dimensioni coincidono se e solo se $S = V$.

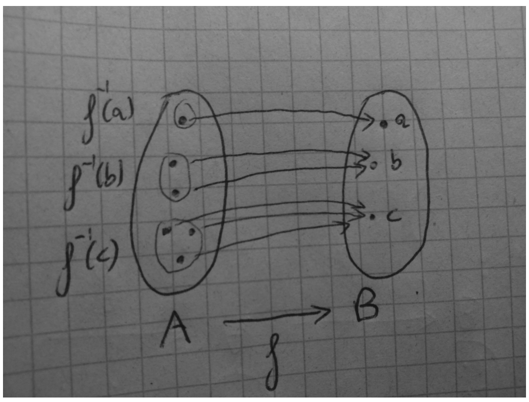
sappiamo, ogni applicazione lineare può essere identificata con una tale funzione grazie alla matrice associata).
Ora, se come generatori del dominio $V = \mathbb{K}^n$ scegliamo i vettori della base canonica $v_1 = (1, 0, \dots, 0), v_2 = (0, 1, \dots, 0) \dots, v_n = (0, 0, \dots, 1)$, dalla (4.9) si ha

$$f(v_1) = L_A \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \ f(v_2) = L_A \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \dots, f(v_n) = L_A \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

cioè $f(v_1), f(v_2), \dots, f(v_n)$ sono le colonne $C_1, C_2, \dots, C_n$ della matrice A che determina l'applicazione. Quindi, in base alla Proposizione 4.5, si ha $Im(f) = \langle C_1, C_2, \dots, C_n \rangle$ e la funzione è suriettiva se e solo se $\dim\langle C_1, C_2, \dots, C_n \rangle = \dim(\mathbb{K}^m) = m$. Ma poichè la dimensione del sottospazio generato dalle colonne di una matrice è per definizione il suo rango, concludiamo che *un'applicazione del tipo (4.9) è suriettiva se e solo se il rango di A è uguale a m.*

4.3.3 Iniettività di applicazioni lineari

Mentre nel caso della suriettività abbiamo visto che per verificare se una funzione f è suriettiva basta controllare un solo sottoinsieme del codominio (l'immagine $Im(f)$), in generale per verificare se f è iniettiva dobbiamo a priori controllare tutte le controimmagini degli elementi del codominio e verificare che queste, quando non sono vuote, hanno un solo elemento.
Ora, per una generica funzione $f : A \to B$ le controimmagini degli elementi di B sono sottoinsiemi del tutto indipendenti tra loro: come nel seguente disegno



può accadere che un elemento abbia controimmagine costituita da un solo elemento ma altri abbiano controimmagine costituita da più elementi.

Vedremo ora invece che le applicazioni lineari hanno il particolare comportamento per cui le controimmagini degli elementi di B, se non sono vuote, o sono *tutte* costituite da un solo elemento (nel qual caso la funzione è iniettiva) o hanno *tutte* più di un elemento (nel qual caso la funzione non è iniettiva): quindi basta controllare una sola controimmagine non vuota per capire come sono fatte tutte le altre.
Più precisamente abbiamo la seguente

Proposizione 4.6. Sia $f : V \to W$ un'applicazione lineare. Allora valgono i seguenti fatti:

(i) la controimmagine $f^{-1}(\bar{0}) = \{v \in V \mid f(v) = \bar{0}\}$ del vettore nullo di W è un sottospazio vettoriale di V (detto *nucleo di* f e denotato $N(f)$)

(ii) per ogni $w_0 \in W$, la controimmagine $f^{-1}(w_0)$ di w_0, se non è vuota, è un sottospazio affine[4] di V, e più precisamente

$$f^{-1}(w_0) = v_0 + N(f) = \{v_0 + n \mid n \in N(f)\}$$

dove v_0 è un qualunque elemento fissato di $f^{-1}(w_0)$.

Proof. Per dimostrare (i), iniziamo con l'osservare che il nucleo di f non è mai vuoto, in quanto il vettore nullo di V (che, con un abuso di notazione, denotiamo ancora $\bar{0}$) è sicuramente tale che $f(\bar{0}) = \bar{0}$: infatti, possiamo pensare il vettore nullo $\bar{0}$ di V come $0v$ (dove v è un qualunque vettore di V) e quindi, sfruttando la linearità di f, si ha $f(\bar{0}) = f(0v) = 0f(v) = \bar{0}$.
Dimostriamo allora che $N(f)$ è chiuso rispetto alla somma e al prodotto per scalari.
Siano v, v' due vettori di $N(f)$, cioè $f(v) = \bar{0}$ e $f(v') = \bar{0}$. Allora, essendo f lineare,

$$f(v + v') = f(v) + f(v') = \bar{0} + \bar{0} = \bar{0}$$

e quindi anche $v + v' \in N(f)$: questo ci dice che $N(f)$ è chiuso rispetto alla somma.
Dati invece un vettore v del nucleo (quindi $f(v) = \bar{0}$) e uno scalare $c \in \mathbb{K}$, allora, sempre per la linearità di f,

$$f(cv) = cf(v) = c\bar{0} = \bar{0}$$

[4] Si veda pagina 95 per la definizione.

ovvero $cv \in N(f)$: questo ci dice che $N(f)$ è chiuso rispetto al prodotto per scalari.
La (i) è dimostrata.

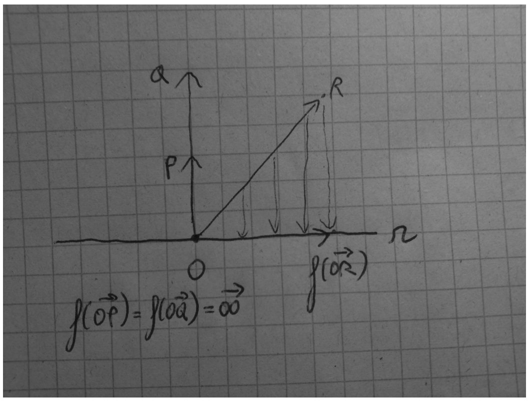
Per dimostrare la (ii), ovvero l'uguaglianza $v_0 + N(f) = f^{-1}(w_0)$, dobbiamo dimostrare che ogni elemento di $v_0+N(f)$ sta nella controimmagine $f^{-1}(w_0)$ di w_0 (ovvero $v_0+N(f) \subseteq f^{-1}(w_0)$), e viceversa che ogni elemento di $f^{-1}(w_0)$ appartiene a $v_0 + N(f)$ (ovvero l'inclusione opposta $f^{-1}(w_0) \subseteq v_0 + N(f)$).
Per dimostrare la prima inclusione, consideriamo il generico elemento di $v_0+N(f)$, cioè, per definizione di sottospazio affine, un vettore v del tipo $v = v_0 + n$, con $n \in N(f)$. Allora

$$f(v) = f(v_0 + n) = f(v_0) + f(n) = f(v_0) + \bar{0} = f(v_0) = w_0$$

(nella seconda uguaglianza abbiamo usato il fatto che f è lineare, nella terza il fatto che n appartiene al nucleo di f, e quindi $f(n) = \bar{0}$). Abbiamo quindi dimostrato che $f(v) = w_0$, cioè v appartiene alla controimmagine $f^{-1}(w_0)$ di w_0, come volevamo.

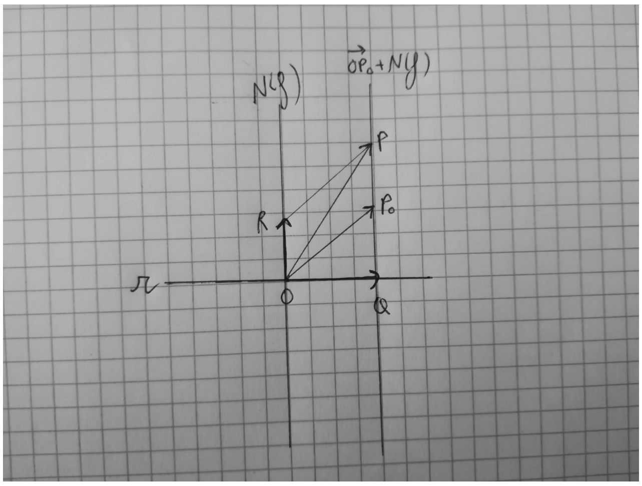
Per dimostrare la seconda inclusione, consideriamo un qualunque elemento v della controimmagine di w_0, cioè $f(v) = w_0$. Essendo $w_0 = f(v_0)$, si ha quindi $f(v) = f(v_0)$, da cui, portando a primo membro, $f(v) - f(v_0) = \bar{0}$. Essendo f lineare, quest'ultima uguaglianza può essere riscritta $f(v-v_0) = \bar{0}$, il che ci dice che il vettore $v-v_0$ appartiene al nucleo $N(f)$ di f. Ma allora, osservando che chiaramente $v = v_0 + (v - v_0)$, vediamo che v si decompone proprio come somma di v_0 e di un elemento del nucleo $N(f)$, cioè $v \in v_0 + N(f)$, come volevamo. $\square$

La proposizione appena dimostrata afferma in pratica che tutte le controimmagini non vuote di un'applicazione lineare f sono "copie" o traslati del nucleo $N(f)$: per illustrare ciò, consideriamo ad esempio lo spazio V_O^2 dei vettori nel piano applicati in O e l'applicazione lineare $f : V_O^2 \to V_O^2$ data dalla proiezione ortogonale su una retta r fissata.
Come si vede, un vettore viene proiettato sul vettore nullo $\vec{OO}$ (cioè appartiene al nucleo $N(f)$ della funzione) se e solo se appartiene alla retta passante per O e ortogonale a r, come i vettori $\vec{OP}$, $\vec{OQ}$ del disegno seguente



Ma, come sappiamo, i vettori che stanno su una retta per O formano un sottospazio vettoriale: questo conferma che il nucleo $N(f)$ è un sottospazio vettoriale.

Per verificare ora che le controimmagini non vuote sono copie o traslati del nucleo, consideriamo come nel disegno seguente



un qualunque vettore $\vec{OQ}$ che stia nell'immagine di f, diciamo $\vec{OQ} = f(\vec{OP_0})$: la sua controimmagine, oltre che da $\vec{OP_0}$, è data da tutti i i vettori $\vec{OP}$ che vengono proiettati su $\vec{OQ}$, cioè, come si vede nel disegno, tutti i vettori che hanno secondo estremo sulla retta ortogonale a r e passante per Q.
Ognuno di tali vettori $\vec{OP}$ si decompone come somma di $\vec{OP_0}$ più un vettore $\vec{OR}$ appartenente al nucleo $N(f)$: quindi, come previsto dalla Proposizione 4.6, la controimmagine $f^{-1}(\vec{OQ})$ è data dal sottospazio affine $\vec{OP_0} + N(f)$, traslato della retta ortogonale a r e passante per O che rappresenta il nucleo.

La Proposizione 4.6 ha come immediato corollario il seguente criterio necessario e sufficiente di iniettività per un'applicazione lineare:

Corollario 4.7. Un'applicazione lineare $f : V \to W$ è iniettiva se e solo se $N(f) = \{\bar{0}\}$.

Proof. Un'applicazione lineare è iniettiva se e solo se la controimmagine di ogni elemento $w_0 \in W$, se non è vuota (cosa che succede se w_0 non sta nell'immagine dell'applicazione), contiene un solo elemento. Ma poichè, come abbiamo visto nella proposizione, la controimmagine di ogni elemento w_0 dell'immagine è del tipo $v_0 + N(f)$, allora questa conterrà un solo elemento v_0 esattamente quando il nucleo contiene il solo vettore nullo $\bar{0}$, cioè $N(f) = \{\bar{0}\}$. □

Notiamo che il fatto che il nucleo sia un sottospazio vettoriale ci dice che possiamo parlare di base e dimensione del nucleo, e che possiamo riformulare il criterio di iniettività per un'applicazione lineare f come segue: *f è iniettiva se e solo se* $\dim(N(f)) = 0$ (infatti, un sottospazio ha dimensione 0 se e solo se è $\{\bar{0}\}$).

Riassumendo quanto visto finora, abbiamo dimostrato che per verificare l'iniettività di un'applicazione lineare f si deve guardare la dimensione $\dim(N(f))$ del nucleo $N(f)$ di f, mentre per verificare la suriettività di f si deve guardare la dimensione $\dim(Im(f))$ dell'immagine $Im(f)$ di f.
Vedremo ora che in realtà è sufficiente calcolare una sola di queste dimensioni per determinare automaticamente anche l'altra. Infatti, queste dimensioni sono collegate dalla formula del seguente risultato, detto anche *teorema della dimensione* o *teorema nullità più rango*.

Teorema 4.8. Sia $f : V \to W$ un'applicazione lineare, con $\dim(V)$ finita[5]. Allora

$$\dim(N(f)) + \dim(Im(f)) = \dim(V). \tag{4.16}$$

Proof. Supponiamo che $\dim(N(f)) = s$ e sia $v_1, \dots, v_s$ una base di $N(f)$. Se $\dim(V) = n$, allora possiamo[6] aggiungere alla base del nucleo $n - s$ vettori $v_{s+1}, \dots, v_n$ in modo che $\{v_1, \dots, v_s, v_{s+1}, \dots, v_n\}$ sia una base di V.
Ora, dimostreremo che le immagini $f(v_{s+1}), \dots, f(v_n)$ di questi ultimi $n - s$ vettori formano una base di $Im(f)$: questo implicherà che $\dim(Im(f)) = n - s$, che assieme a $\dim(N(f)) = s$ e $\dim(V) = n$ ci dà $\dim(N(f)) + \dim(Im(f)) = s + (n - s) = n = \dim(V)$, cioè la formula.
Per dimostrare che $\{f(v_{s+1}), \dots, f(v_n)\}$ è una base di $Im(f)$, dobbiamo dimostrare che i vettori generano $Im(f)$ e sono linearmente indipendenti.

[5] Si veda l'Osservazione 1.32.

[6] Il fatto che data una base di un sottospazio, questa possa sempre essere completata a una base di tutto lo spazio aggiungendo dei vettori è in effetti un teorema (che non abbiamo dimostrato) detto *teorema del completamento*.

In effetti, noi sappiamo che essendo $v_1, \dots, v_s, v_{s+1}, \dots, v_n$ una base e quindi un insieme di generatori di V, le immagini $f(v_1), \dots, f(v_s), f(v_{s+1}), \dots, f(v_n)$ generano $Im(f)$ (lo abbiamo dimostrato nella Proposizione 4.5); ma in questi generatori i primi $f(v_1), \dots, f(v_s)$ sono uguali al vettore nullo $\bar{0}$, in quanto $v_1, \dots, v_s$ sono i vettori della base del nucleo fissata inizialmente. Quindi possiamo eliminarli dalla lista $f(v_1), \dots, f(v_s), f(v_{s+1}), \dots, f(v_n)$ dei generatori, concludendo che per generare $Im(f)$ bastano $f(v_{s+1}), \dots, f(v_n)$, che è quello che volevamo.

Ora dimostriamo che $f(v_{s+1}), \dots, f(v_n)$ sono linearmente indipendenti. Dobbiamo dimostrare che se $c_{s+1}f(v_{s+1}) + \cdots + c_n f(v_n) = \bar{0}$ allora $c_{s+1} = 0, \dots, c_n = 0$.

In effetti, sfruttando il fatto che f è lineare possiamo riscrivere l'uguaglianza $c_{s+1}f(v_{s+1}) + \cdots + c_n f(v_n) = \bar{0}$ come $f(c_{s+1}v_{s+1} + \cdots + c_n v_n) = \bar{0}$: ma questa uguaglianza ci dice che il vettore $c_{s+1}v_{s+1} + \cdots + c_n v_n$ appartiene al nucleo $N(f)$, e quindi esso può essere scritto come combinazione lineare di $v_1, \dots, v_s$ (che del nucleo costituiscono una base):

$$c_{s+1}v_{s+1} + \cdots + c_n v_n = c_1 v_1 + \cdots + c_s v_s.$$

Portando tutto a primo membro in questa uguaglianza si ottiene

$$c_{s+1}v_{s+1} + \cdots + c_n v_n - c_1 v_1 - \cdots - c_s v_s = \bar{0}$$

ovvero una combinazione lineare uguale al vettore nullo dei vettori $v_1, \dots, v_s, v_{s+1}, \dots, v_n$: essendo questi vettori indipendenti (sono i vettori che formano la base completata di V) necessariamente tutti i coefficienti $c_1, \dots, c_s, c_{s+1}, \dots, c_n$ sono uguali a zero, e in particolare $c_{s+1} = 0, \dots, c_n = 0$, che è quello che ci restava da dimostrare. □

Vediamo ora alcune notevoli conseguenze della formula (4.16):

Corollario 4.9. Sia $f : V \to W$ un'applicazione lineare, con $\dim(V)$ finita. Allora valgono le tre seguenti:

(1) se $\dim(V) > \dim(W)$ allora f non è iniettiva

(2) se $\dim(V) < \dim(W)$ allora f non è suriettiva

(3) se $\dim(V) = \dim(W)$ allora f è iniettiva se e solo se è suriettiva

Proof. Dimostriamo la (1). Per assurdo, se la funzione f fosse iniettiva, il suo nucleo, come abbiamo visto nel Corollario 4.7, sarebbe nullo, ovvero $\dim(N(f)) = 0$. Sostituendo questo nella formula (4.16), avremmo $\dim(V) = \dim(Im(f))$. Ma essendo $Im(f)$ un sottospazio di W, la sua dimensione è sicuramente minore o uguale a $\dim(W)$, e quindi concluderemmo $\dim(V) \leq \dim(W)$, contro l'ipotesi che $\dim(V) > \dim(W)$. Quindi f non può essere iniettiva.

Per dimostrare la (2), supponiamo per assurdo che la funzione f sia suriettiva: allora, essendo $Im(f) = W$, avremmo $\dim(Im(f)) = \dim(W)$ che, sostituita nella formula (4.16), ci dà $\dim(V) = \dim(W) + \dim(N(f))$. Poichè $\dim(N(f))$ è un numero reale positivo o nullo, questa uguaglianza implica che $\dim(V) \geq \dim(W)$, contro l'ipotesi che $\dim(V) < \dim(W)$. Quindi f non può essere suriettiva.

Infine, dimostriamo la (3). Se f è iniettiva, $\dim(N(f)) = 0$ e quindi la formula (4.16) si riduce a $\dim(Im(f)) = \dim(V)$. Essendo per ipotesi la dimensione di V uguale a quella di W, questo significa che $\dim(Im(f)) = \dim(W)$, che implica che f è suriettiva.
Viceversa, se f è suriettiva, $\dim(Im(f)) = \dim(W)$ e quindi la formula (4.16) si riduce a $\dim(V) = \dim(N(f)) + \dim(W)$. Essendo per ipotesi la dimensione di V uguale a quella di W, il primo membro $\dim(V)$ si semplifica con l'addendo $\dim(W)$ del secondo membro, e quindi rimane $0 = \dim(N(f))$, che implica che f è iniettiva. □

Ad esempio, un'applicazione lineare $f : \mathbb{R}^3 \to \mathbb{R}^2$ non può mai essere iniettiva (ma può essere suriettiva); un'applicazione lineare $f : \mathbb{R}^3 \to \mathbb{R}^4$ non è sicuramente suriettiva (ma può essere iniettiva); un'applicazione lineare $f : \mathbb{R}^3 \to \mathbb{R}^3$ o è contemporaneamente iniettiva e suriettiva o nessuna delle due: basta mostrare che vale una delle due proprietà e automaticamente varrà anche l'altra.

Ora, supponiamo di aver determinato grazie ai risultati precedenti che un'applicazione lineare f data è biiettiva: nel prossimo paragrafo ci porremo (e risolveremo) il problema di determinare la sua *inversa*.

4.4 Composizione di applicazioni, inversa e prodotto di matrici

Ricordiamo che, data una funzione $f : X \to Y$ tra due insiemi, questa si dice *invertibile* se esiste una funzione $g : Y \to X$ (detta appunto l'inversa di f) tale che

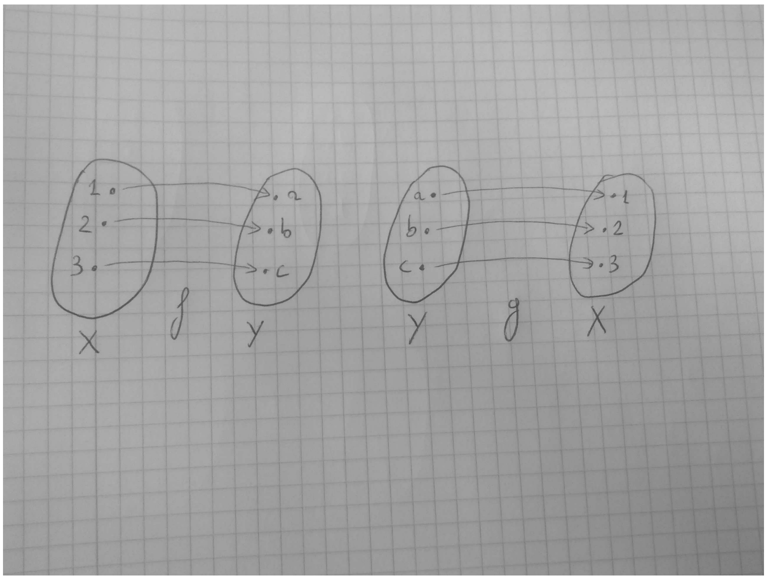
$$f \circ g = id_Y, \quad g \circ f = id_X \tag{4.17}$$

dove stiamo denotando con id la funzione che manda ogni elemento in se stesso[7] e con il simbolo $\circ$ la composizione di funzioni, ovvero l'operazione che consiste nell'applicare prima una funzione e poi l'altra: più precisamente, ricordiamo che ogni volta che si hanno due funzioni $f : X \to Y$ e $g : Y \to Z$, tali che *il codominio della prima coincida con il dominio della seconda*, allora, per ogni elemento $x \in X$, possiamo applicare prima f ottenendo $f(x) \in Y$, e poi dal momento che Y è anche il dominio della g possiamo applicare la g a $f(x)$, ottenendo $g(f(x))$. In questo modo otteniamo una nuova funzione che associa a ogni elemento di X un elemento di Z:

$$f : X \underset{x \mapsto g(f(x))}{\longrightarrow} Z$$

(si noti che a causa dell'ordine in cui appaiono le funzioni in $g(f(x))$, la nuova funzione ottenuta si denota $g \circ f$ e si legge "g composto f", benché si intenda che stiamo applicando prima la f e poi la g).
Quindi, le (4.17) significano che una funzione $f : X \to Y$ è invertibile se esiste una funzione $g : Y \to X$ tale $g(f(x)) = x$ per ogni $x \in X$ e $f(g(y)) = y$ per ogni $y \in Y$.

Ora, si può vedere che le uniche funzioni f invertibili sono quelle biiettive. Ad esempio, consideriamo $X = \{1, 2, 3\}$, $Y = \{a, b, c\}$ e la funzione $f : X \to Y$ tale che $f(1) = a$, $f(2) = b$ e $f(3) = c$: si tratta chiaramente di una funzione biiettiva, che ha come inversa la funzione $g : Y \to X$ rappresentata nel disegno seguente:



[7] Più precisamente, id_X è la funzione $X \to X$ che manda ogni elemento di x in se stesso e id_Y denota la funzione $Y \to Y$ che manda ogni elemento di Y in se stesso.

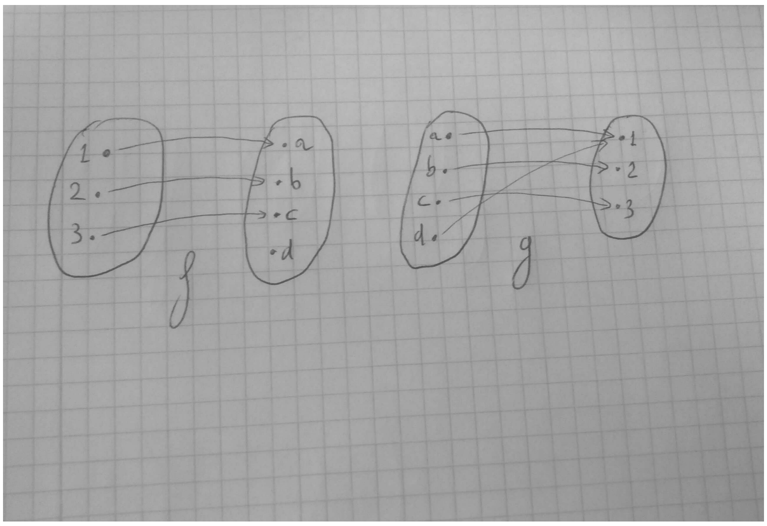
Infatti, come si vede subito, si ha

$$(g \circ f)(1) = g(f(1)) = g(a) = 1$$
$$(g \circ f)(2) = g(f(2)) = g(b) = 2$$
$$(g \circ f)(3) = g(f(3)) = g(c) = 3$$

ovvero $g \circ f : \{1, 2, 3\} \to \{1, 2, 3\}$ è la funzione che manda ogni elemento dell'insieme $X = \{1, 2, 3\}$ in se stesso (ovvero la funzione identica id_X di X) e analogamente

$$(f \circ g)(a) = f(g(a)) = f(1) = a$$
$$(f \circ g)(b) = f(g(b)) = f(2) = b$$
$$(f \circ g)(c) = f(g(c)) = f(3) = c$$

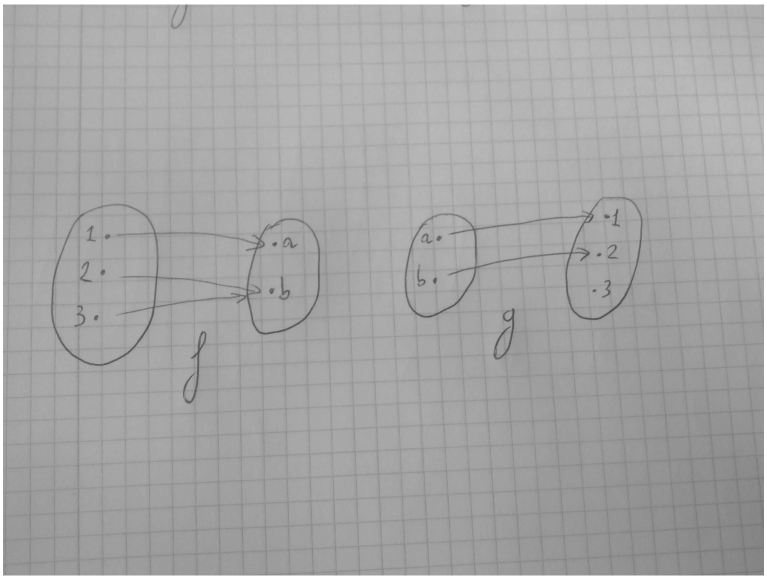
ovvero $f \circ g : \{a, b, c\} \to \{a, b, c\}$ è la funzione che manda ogni elemento dell'insieme $Y = \{a, b, c\}$ in se stesso (cioè è la funzione identica id_Y).
Per giustificare l'affermazione che le funzioni biiettive sono *le sole* a essere invertibili, consideriamo ad esempio la funzione f rappresentata nel seguente disegno, che è iniettiva ma non suriettiva (e quindi non è biiettiva)



Si vede subito che $g \circ f$ è la funzione identica di $\{1, 2, 3\}$, ma $f \circ g$ non è la funzione identica di $\{a, b, c, d\}$ in quanto pur essendo $f(g(a)) = f(1) = a$, $f(g(b)) = f(2) = b$ e $f(g(c)) = f(3) = c$, si ha $f(g(d)) = f(1) = a$, ovvero $f \circ g$ non manda d in se stesso.

Si osservi che non c'è alcun modo di modificare g in modo che $f \circ g$ mandi ogni elemento di $\{a, b, c, d\}$ in se stesso: qualunque valore assegniamo a d non sarà mai $f(g(d)) = d$, perchè $g(d)$ dovrebbe essere un elemento mandato da f in d, ma non esiste nessun elemento di $\{1, 2, 3\}$ che viene mandato da f in d.
In altre parole, il motivo per cui l'uguaglianza $f \circ g = id$ non può mai essere verificata è la non suriettività di f. Si dice che f ha un'*inversa sinistra* ($g \circ f = id$ è verificato) ma non ammette un'*inversa destra* (cioè la $f \circ g = id$ non può valere per nessuna g).

Analogamente, si prendano $f : \{1, 2, 3\} \to \{a, b\}$ e $g : \{a, b\} \to \{1, 2, 3\}$ definite come nel seguente disegno



Allora, si vede subito che $f \circ g$ è la funzione identica di $\{a, b\}$, ma $g \circ f$ non è la funzione identica di $\{1, 2, 3\}$ in quanto pur essendo $g(f(1)) = g(a) = 1$ e $g(f(2)) = g(b) = 2$, si ha $g(f(3)) = g(b) = 2$, ovvero $g \circ f$ non manda 3 in se stesso.
Si osservi che anche qui non c'è alcun modo di modificare g : se avessimo posto $g(b) = 3$ avremmo sì ottenuto $g(f(3)) = g(b) = 3$ ma stavolta sarebbe stato $g(f(2)) = g(b) = 3$, ovvero $g \circ f$ non avrebbe mandato 2 in se stesso: come si vede, il problema stavolta è che f manda i due elementi 2 e 3 entrambi in b, quindi necessariamente $g(f(2))$ e $g(f(3))$ saranno uguali e non potranno mai essere il primo 2 e il secondo 3.

In altre parole, il motivo per cui non esiste un'inversa sinistra di f è dovuto alla non iniettività di f: la f ha cioè un'inversa destra ma non un'inversa sinistra.

Esempio 4.10. Ad esempio, la funzione $f : \mathbb{R} \to \mathbb{R}$ data da $f(x) = x^2$, non essendo nè iniettiva (due numeri uno l'opposto dell'altro hanno lo stesso quadrato) nè suriettiva (i numeri negativi non sono quadrati di nessun numero reale) non ha nè inversa sinistra nè inversa destra. In effetti, la candidata ad essere inversa di f, la radice quadrata $g(x) = \sqrt{x}$, da una parte non soddisfa $g(f(x)) = x$ se x è negativo perché in tal caso $g(f(x)) = g(x^2) = \sqrt{x^2} = |x|$, e $|x| \neq x$ se x è negativo ($|-2| = +2 \neq -2$); dall'altra parte $f(g(y)) = y$ non ha senso se y è negativo in quando in tal caso $g(y) = \sqrt{y}$ non è neanche un numero reale. Per eliminare i due problemi (e rendere g l'inversa di f) bisogna togliere i numeri negativi sia dal dominio che dal codominio di f, ovvero restringerla a una funzione $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, dove $\mathbb{R}_{\geq 0}$ denota l'insieme dei numeri non negativi: ma così facendo in effetti facciamo proprio in modo che diventi una funzione biiettiva (ora non ci sono due elementi del dominio con lo stesso quadrato, e ogni elemento del codominio è quadrato di qualcosa).

Ora, come abbiamo detto sopra ci poniamo il problema di calcolare, se esiste, l'inversa di un'applicazione lineare data. Prima di fare ciò, dal momento che l'inversa è definita tramite la composizione, dobbiamo vedere cosa succede quando si compongono due applicazioni lineari. In particolare, poiché ogni applicazione lineare può essere sempre tradotta (lavorando in coordinate) in una funzione del tipo (4.9), vedremo cosa succede quando si compongono due applicazioni di questo tipo.
Più precisamente, consideriamo $f = L_A : \mathbb{K}^n \to \mathbb{K}^m$,

$$f\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} \tag{4.18}$$

e $g = L_B : \mathbb{K}^p \to \mathbb{K}^n$:

$$g\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} b_{11}y_1 + b_{12}y_2 + \cdots + b_{1p}y_p \\ b_{21}y_1 + b_{22}y_2 + \cdots + b_{2p}y_p \\ \vdots \\ b_{n1}y_1 + b_{n2}y_2 + \cdots + b_{np}y_p \end{pmatrix} \tag{4.19}$$

In base a quanto appena visto, la composizione $f \circ g$ può essere calcolata in quanto il codominio di g, ovvero $\mathbb{K}^n$, è anche il dominio di f. Si ha

$$(f \circ g)\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} = f\left(g\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}\right) = f\begin{pmatrix} b_{11}y_1 + b_{12}y_2 + \cdots + b_{1p}y_p \\ b_{21}y_1 + b_{22}y_2 + \cdots + b_{2p}y_p \\ \vdots \\ b_{n1}y_1 + b_{n2}y_2 + \cdots + b_{np}y_p \end{pmatrix} =$$

$$= \begin{pmatrix} a_{11}(b_{11}y_1 + b_{12}y_2 + \cdots + b_{1p}y_p) + \cdots + a_{1n}(b_{n1}y_1 + b_{n2}y_2 + \cdots + b_{np}y_p) \\ a_{21}(b_{11}y_1 + b_{12}y_2 + \cdots + b_{1p}y_p) + \cdots + a_{2n}(b_{n1}y_1 + b_{n2}y_2 + \cdots + b_{np}y_p) \\ \vdots \\ a_{m1}(b_{11}y_1 + b_{12}y_2 + \cdots + b_{1p}y_p) + \cdots + a_{mn}(b_{n1}y_1 + b_{n2}y_2 + \cdots + b_{np}y_p) \end{pmatrix} = \tag{4.20}$$

(raccogliendo $y_1, y_2, \dots, y_p$)

$$= \begin{pmatrix} (a_{11}b_{11} + \cdots + a_{1n}b_{n1})y_1 + (a_{11}b_{12} + \cdots + a_{1n}b_{n2})y_2 + \cdots + (a_{11}b_{1p} + \cdots + a_{1n}b_{np})y_p \\ (a_{21}b_{11} + \cdots + a_{2n}b_{n1})y_1 + (a_{21}b_{12} + \cdots + a_{2n}b_{n2})y_2 + \cdots + (a_{21}b_{1p} + \cdots + a_{2n}b_{np})y_p \\ \vdots \\ (a_{m1}b_{11} + \cdots + a_{mn}b_{n1})y_1 + (a_{m1}b_{12} + \cdots + a_{mn}b_{n2})y_2 + \cdots + (a_{m1}b_{1n} + \cdots + a_{mn}b_{np})y_p \end{pmatrix} \tag{4.21}$$

Da quest'ultima espressione vediamo che la composizione $f \circ g$ è ancora una funzione del tipo (4.9), cioè è determinata da una matrice C, e più precisamente

$$C = \begin{pmatrix} a_{11}b_{11} + \cdots + a_{1n}b_{n1} & a_{11}b_{12} + \cdots + a_{1n}b_{n2} & \cdots & a_{11}b_{1p} + \cdots + a_{1n}b_{np} \\ a_{21}b_{11} + \cdots + a_{2n}b_{n1} & a_{21}b_{12} + \cdots + a_{2n}b_{n2} & \cdots & a_{21}b_{1p} + \cdots + a_{2n}b_{np} \\ & & \vdots & \\ a_{m1}b_{11} + \cdots + a_{mn}b_{n1} & a_{m1}b_{12} + \cdots + a_{mn}b_{n2} & \cdots & a_{m1}b_{1n} + \cdots + a_{mn}b_{np} \end{pmatrix} \tag{4.22}$$

Ora, notiamo che le entrate c_{ij} di tale matrice sono tutte espressioni del tipo

$$c_{ij} = a_{i1}b_{1j} + \cdots + a_{in}b_{nj} \tag{4.23}$$

in cui il primo indice dell'entrata di A è fisso e sempre uguale a i, il secondo indice dell'entrata di B è fisso e sempre uguale a j mentre gli indici "interni" (il secondo dell'entrata di A e il primo dell'entrata di B) variano da 1 a n (usando la notazione di sommatoria, potremmo quindi scrivere $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$).
Diamo allora la seguente importante

Definizione 4.11. La matrice C le cui entrate sono date dalla (4.23) si chiama *prodotto di A per B*, e si scrive $C = AB$.

In pratica, abbiamo definito il prodotto di due matrici A e B in modo che la matrice $C = AB$ ottenuta sia la matrice che determina la composizione $L_A \circ L_B$ delle funzioni determinate da A e B.

Ricordando che la composizione $L_A \circ L_B$ delle due funzioni L_A e L_B può essere fatta solo sotto opportune condizioni (il codominio di $L_B : \mathbb{K}^p \to \mathbb{K}^n$ deve essere uguale al dominio di $L_A : \mathbb{K}^n \to \mathbb{K}^m$) si osserva di conseguenza che anche il prodotto di due matrici può essere fatto solo sotto opportune condizioni: più precisamente, dal momento che la matrice di L_A ha m righe e n colonne, mentre la matrice B di L_B ha n righe e p colonne, vediamo che si possono moltiplicare tra loro due matrici A e B (in quest'ordine) se e solo se *il numero di colonne di A è uguale al numero di righe di B*: in tal caso, il risultato AB è una matrice con m righe e p colonne. Se denotiamo $M_{m,n}(\mathbb{K})$ l'insieme delle matrici con m righe e n colonne (e entrate appartenenti al campo $\mathbb{K}$), possiamo allora scrivere che

$$A \in M_{m,n}(\mathbb{K}),\ B \in M_{n,p}(\mathbb{K}) \Rightarrow AB \in M_{m,p}(\mathbb{K})$$

Il prodotto di matrici dato dalla Definizione 4.11 si chiama anche *prodotto righe per colonne*: il motivo è che nell'espressione (4.23) della generica entrata di posto $i\ j$ appaiono tutte e sole le entrate con primo indice i di A e tutte e sole le entrate con secondo indice j di B: dal momento che il primo indice ci dice in che riga siamo, e il secondo in quale colonna, questo significa che per calcolare l'entrata di posto $i\ j$ di AB dobbiamo prendere la riga i-esima di A, la j-esima colonna di B, moltiplicare tra di loro le entrate corrispondenti e sommare il tutto

$$\begin{pmatrix} \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \\ \end{pmatrix} \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{pmatrix} = \begin{pmatrix} & \dots & \\ \dots & a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} & \dots \\ & \dots & \end{pmatrix}$$

Esempio 4.12. Consideriamo le due matrici

$$A = \begin{pmatrix} 1 & 2 \\ 4 & 5 \\ 3 & 6 \end{pmatrix} \in M_{3,2}(\mathbb{R}), \quad B = \begin{pmatrix} 5 & 2 & 0 & 1 \\ 3 & 1 & -1 & 2 \end{pmatrix} M_{2,4}(\mathbb{R})$$

In base a quello che abbiamo detto, le due matrici possono essere moltiplicate e il risultato AB sarà una matrice 3 per 4.
Per trovare l'entrata di AB che sta *nella prima riga* e *prima colonna* si prendono *la prima riga di A* $\begin{pmatrix} 1 & 2 \end{pmatrix}$, la *prima colonna* di B $\begin{pmatrix} 5 \\ 3 \end{pmatrix}$, se ne moltiplicano gli elementi corrispondenti (il primo con il primo, il secondo con il secondo) e si somma il risultato: $1 \cdot 5 + 2 \cdot 3 = 5 + 6 = 11$. Quindi nella prima riga, prima colonna di AB dobbiamo scrivere 11.
Analogamente, per trovare l'entrata di AB che sta *nella prima riga* e *seconda colonna* si prendono *la prima riga di A* $\begin{pmatrix} 1 & 2 \end{pmatrix}$, la *seconda colonna* di B $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$, se ne moltiplicano gli elementi corrispondenti e si somma il risultato: $1 \cdot 2 + 2 \cdot 1 = 2 + 2 = 4$. Quindi nella prima riga, seconda colonna di AB c'è 4.
Facendo questo calcolo per tutte le entrate, si vede che

$$AB = \begin{pmatrix} 11 & 4 & -2 & 4 \\ 35 & 13 & -5 & 14 \\ 33 & 12 & -6 & 15 \end{pmatrix}$$

Esempio 4.13. Come abbiamo visto nella (4.12), la matrice associata alla rotazione f di angolo θ in senso antiorario attorno a O rispetto a una base ortonormale di V_O^2 è $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$: la funzione $L_A : \mathbb{R}^2 \to \mathbb{R}^2$ determinata da A, che manda (x_1, x_2) in $(\cos\theta x_1 - \sin\theta x_2, \sin\theta x_1 + \cos\theta x_2)$, è la traduzione in coordinate di f.
Analogamente, se g denota la rotazione di angolo ϕ, la traduzione in coordinate di g sarà data dalla funzione L_B determinata dalla matrice $B = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$. Dal momento che il prodotto di matrici ci dà la composizione delle applicazioni corrispondenti, se moltiplichiamo A e B otteniamo la matrice che rappresenta in coordinate la composizione delle due rotazioni: infatti

$$AB = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} =$$

$$= \begin{pmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi & -\cos\theta\sin\phi - \sin\theta\cos\phi \\ \sin\theta\cos\phi + \cos\theta\sin\phi & -\sin\theta\sin\phi + \cos\theta\cos\phi \end{pmatrix} =$$

(in base alla note formule di trigonometria per il coseno e il seno della somma e della differenza tra due angoli)

$$= \begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix}$$

che è ancora una matrice di rotazione ma relativa all'angolo $\theta+\phi$: questo era prevedibile in quanto la composizione non fa altro che applicare una rotazione di angolo ϕ seguita da una rotazione di angolo θ.
In un capitolo successivo faremo lo stesso per calcolare la composizione di due rotazioni nello spazio, dove il risultato è molto meno prevedibile e intuitivo che nel caso del piano.

Vediamo ora le proprietà fondamentali del prodotto di matrici.

In primo luogo, è importante osservare che tale prodotto *non gode della proprietà commutativa*, ovvero, date due matrici A e B e ammesso che si possano[8] eseguire entrambi i prodotti AB e BA, questi non saranno in generale uguali. Ad esempio, siano $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ e $B = \begin{pmatrix} -1 & 0 \\ 5 & 2 \end{pmatrix}$. Allora, per definizione di prodotto righe per colonne vediamo che

$$AB = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 5 & 2 \end{pmatrix} = \begin{pmatrix} 9 & 4 \\ 17 & 8 \end{pmatrix}$$

$$BA = \begin{pmatrix} -1 & 0 \\ 5 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} -1 & -2 \\ 11 & 18 \end{pmatrix}$$

ovvero $AB \neq BA$.

Osservazione 4.14. Il motivo della non commutatività in generale del prodotto di due matrici A e B è che, come sappiamo, tale prodotto rappresenta la composizione delle applicazioni L_A e L_B corrispondenti, e la composizione di funzioni non gode in generale della proprietà commutativa: ad esempio, date le due funzioni

$$f : \mathbb{R} \to \mathbb{R},\ f(x) = x+1, \qquad g : \mathbb{R} \to \mathbb{R},\ g(x) = x^2$$

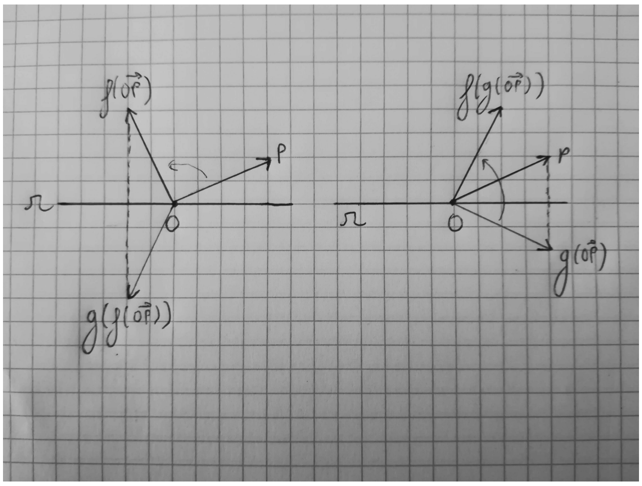
si ha $f(g(x)) = f(x^2) = x^2+1$ mentre $g(f(x)) = g(x+1) = (x+1)^2$, e quindi $f \circ g \neq g \neq f$.

Per un esempio geometrico, si considerino la funzione $f : V_O^2 \to V_O^2$ che ruota ogni vettore del piano applicato in O di 90 gradi in senso antiorario,

---

[8]Ad esempio, se $A \in M_{3,2}(\mathbb{R})$ e $B \in M_{2,4}(\mathbb{R})$, il prodotto AB si può eseguire mentre BA no; si noti anche che a volte si possono eseguire sia AB che BA ma si tratta di due matrici di tipo diverso. Ad esempio, se $A \in M_{3,2}(\mathbb{R})$ e $B \in M_{2,3}(\mathbb{R})$ si possono eseguire entrambi i prodotti, ma $AB \in M_{3,3}(\mathbb{R})$ mentre $BA \in M_{2,2}(\mathbb{R})$.

e la funzione $g : V_O^2 \to V_O^2$ che riflette ogni vettore attorno a una retta r passante per O: allora, come si vede nel seguente disegno, applicare prima la rotazione f e poi la riflessione g oppure viceversa porta in generale a risultati diversi (ovvero $f \circ g \neq g \circ f$)



Per quello che riguarda invece la proprietà associativa, si può dimostrare che questa vale, ovvero $(AB)C = A(BC)$. Ad esempio, siano $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $B = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix}$ e $C = \begin{pmatrix} 1 & 5 \\ 2 & 0 \end{pmatrix}$. Allora

$$(AB)C = \left(\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix} \right) \begin{pmatrix} 1 & 5 \\ 2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 3 \\ 2 & 7 \end{pmatrix} \begin{pmatrix} 1 & 5 \\ 2 & 0 \end{pmatrix} = \begin{pmatrix} 6 & 0 \\ 16 & 10 \end{pmatrix}$$

$$A(BC) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \left(\begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 5 \\ 2 & 0 \end{pmatrix} \right) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 4 & 10 \\ 1 & -5 \end{pmatrix} = \begin{pmatrix} 6 & 0 \\ 16 & 10 \end{pmatrix}$$

Osservazione 4.15. Di nuovo, l'associatività del prodotto di matrici si spiega facilmente ricordando che tale prodotto rappresenta la composizione delle funzioni corrispondenti, ovvero $(AB)C$ rappresenta $(L_A \circ L_B) \circ L_C$ mentre $A(BC)$ rappresenta $L_A \circ (L_B \circ L_C)$. Ma è facile vedere che la composizione di funzioni gode della proprietà associativa: infatti, applicando $(f \circ g) \circ h$ a x otteniamo $(f \circ g)(h(x)) = f(g(h(x)))$, e analogamente applicando $f \circ (g \circ h)$ a x otteniamo sempre $f((g \circ h)(x)) = f(g(h(x)))$, ovvero $(f \circ g) \circ h = f \circ (g \circ h)$.

Ci chiediamo ora se il prodotto di matrici ammetta un elemento neutro che svolga lo stesso ruolo che svolge il numero 1 per il prodotto tra numeri, per cui si ha $a \cdot 1 = 1 \cdot a = a$ per ogni numero a. La risposta è affermativa: più precisamente, per ogni n consideriamo la matrice con n righe e n colonne seguente

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & 1 \end{pmatrix} \tag{4.24}$$

ovvero la matrice che ha 1 nelle entrate con stesso indice di riga e di colonna (a_{11}, a_{22} etc.) e 0 in tutte le altre entrate.
Tale matrice si chiama *matrice identica di ordine* n e si denota I_n. Ad esempio,

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Osservazione 4.16. In generale, data una matrice quadrata di ordine n, le entrate $a_{11}, a_{22}, \dots, a_{nn}$ che hanno stesso indice di riga e di colonna formano la cosiddetta *diagonale della matrice*. La matrice identica I_n può essere quindi descritta come la matrice che ha 1 sulla diagonale e 0 nelle altre entrate. Le entrate di I_n si denotano solitamente con il simbolo δ_{ij}, detto *delta di Kronecker*, che vale quindi 1 se $i = j$, e 0 se $i \neq j$.

Ora, si può verificare che, per ogni $A \in M_{m,n}(\mathbb{K})$ si ha

$$AI_n = A, \quad I_m A = A$$

(l'ordine della matrice identica cambia perché deve essere tale che si possa svolgere il prodotto) e quindi la matrice identica svolge esattamente il ruolo di elemento neutro per il prodotto righe per colonne.
Ad esempio,

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

Osservazione 4.17. Si noti che la matrice identica I_n non è nient'altro che la matrice che determina la funzione identica $id_{\mathbb{K}^n} : \mathbb{K}^n \to \mathbb{K}^n$ che manda ogni elemento in se stesso: infatti,

$$id_{\mathbb{K}^n}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 1x_1 + 0x_2 + \cdots + 0x_n \\ 0x_1 + 1x_2 + \cdots + 0x_n \\ \vdots \\ 0x_1 + 0x_2 + \cdots + 1x_n \end{pmatrix} \tag{4.25}$$

Questo spiega perché tale matrice sia l'elemento neutro per il prodotto, in quanto il prodotto tra matrici rappresenta la composizione delle applicazioni corrispondenti e la funzione identica è esattamente l'elemento neutro per la composizione.

Continuando con l'analogia con il prodotto tra numeri, vogliamo ora esaminare la questione *dell'esistenza dell'inverso*: come sappiamo, nell'insieme dei numeri reali $\mathbb{R}$ per ogni numero a diverso da zero esiste un numero b tale che $ab = ba = 1$, detto appunto inverso di a (e denotato a^{-1}).

Nel caso delle matrici, vorremmo quindi capire se data una matrice $A \in M_{m,n}(\mathbb{K})$ esiste una matrice[9] $B \in M_{n,m}(\mathbb{K})$ tale che $AB = I_m$ e $BA = I_n$. Tuttavia, se una tale B esiste, dal momento che il prodotto di A e B corrisponde alla composizione delle applicazioni lineari $L_A : \mathbb{K}^n \to \mathbb{K}^m$ e $L_B : \mathbb{K}^m \to \mathbb{K}^n$ mentre la matrice identica corrisponde alla funzione identica, si ha $L_A \circ L_B = id$ e $L_B \circ L_A = id$, ovvero la funzione L_A deve essere invertibile. Ma allora, come abbiamo ricordato a pagina 155, $L_A : \mathbb{K}^n \to \mathbb{K}^m$ deve essere biiettiva, il che in base al Corollario 4.9 è possibile solo se dominio e codominio hanno la stessa dimensione, ovvero solo se $m = n$.

Riassumendo, abbiamo mostrato che il problema dell'inverso si pone solo per matrici che hanno stesso numero di righe e colonne, ovvero solo se $A \in M_{n,n}(\mathbb{K})$. Tali matrici si dicono *quadrate* e il numero n comune di righe e colonne si dice *ordine della matrice*. Per semplicità, l'insieme $M_{n,n}(\mathbb{K})$ si denota $M_n(\mathbb{K})$.

Diamo allora la seguente

Definizione 4.18. Una matrice $A \in M_n(\mathbb{K})$ quadrata di ordine n si dice *invertibile* se esiste una matrice $B \in M_n(\mathbb{K})$ tale che

$$AB = I_n, \quad BA = I_n \tag{4.26}$$

In tal caso, B si chiama *matrice inversa di* A e si denota A^{-1}.

[9] La scelta del numero di righe e colonne di B è obbligata se vogliamo poter eseguire sia il prodotto AB che quello BA.

Dimostriamo ora il seguente risultato, che ci dice che, contrariamente a quello che accade nel campo dei numeri reali dove l'unico numero non invertibile è lo zero, nell'insieme delle matrici, anche limitandosi alle sole matrici quadrate, ci sono molte matrici non invertibili:

Teorema 4.19. Una matrice $A \in M_n(\mathbb{K})$ è invertibile se e solo se il rango di A è uguale a n.

Proof. Come abbiamo ricordato poco prima della Definizione 4.18, l'invertibilità di A, ovvero l'esistenza di una matrice B tale che $AB = BA = I_n$, equivale a dire $L_A \circ L_B = L_B \circ L_A = id$, ovvero che L_A è invertibile[10] e quindi biiettiva. Quindi ci basta dimostrare che L_A è biiettiva se e solo se il rango di A è n. Ora, come abbiamo visto nel Corollario 4.9, una funzione lineare in cui dominio e codominio abbiano la stessa dimensione (e $L_A : \mathbb{K}^n \to \mathbb{K}^n$ verifica questa condizione) è iniettiva se e solo se è suriettiva, ovvero basta una delle due proprietà per avere anche l'altra: in altre parole, per avere entrambe le proprietà (e quindi la biiettività) è necessaria e sufficiente una delle due proprietà. Quindi, possiamo dire che L_A è biiettiva se e solo se è suriettiva. Ma, come abbiamo visto a pagina 148, L_A è suriettiva se e solo se il rango di A è n. Il teorema è dimostrato. $\square$

Vediamo ora come si calcola l'inversa di una matrice A, supponendo che questa esista (cioè che A sia invertibile).
Facciamo prima la seguente osservazione preliminare: data una matrice A quadrata, se B è tale che $AB = I_n$ allora automaticamente anche $BA = I_n$ (e quindi basta $AB = I_n$ per dire che A è invertibile e che B è l'inversa di A).
Infatti, se $AB = I_n$ allora $L_A \circ L_B = id$, che significa che L_A ha inversa destra (pagina 157): come sappiamo, questo implica che L_A è suriettiva (una funzione non suriettiva non può avere inversa destra). Ma per la (iii) del Corollario 4.9 L_A, essendo un'applicazione tra due spazi della stessa dimensione (in tal caso dominio e codiminio coincidono), è automaticamente anche iniettiva, e quindi (si veda sempre pagina 157) ha inversa sinistra, cioè esiste una funzione g tale che $g \circ L_A = id$. Ora mostriamo che $g = L_B$: infatti,

$$g = g \circ id = g \circ (L_A \circ L_B) = (g \circ L_A) \circ L_B = id \circ L_B = L_B$$

[10] Volendo essere rigorosi, $L_A \circ L_B = L_B \circ L_A = id$ implica chiaramente che L_A sia invertibile, ma viceversa il fatto che L_A sia invertibile implica solo che esista una funzione g tale che $g \circ L_A = L_A \circ g = id$, che a priori non sappiamo se è della forma $g = L_B$ per qualche matrice B. In realtà, si può dimostrare che g deve necessariamente essere di quella forma, e quindi otteniamo anche l'implicazione opposta per cui L_A invertibile implica che esista una matrice B per cui $L_A \circ L_B = L_B \circ L_A = id$ (e quindi $AB = BA = I_n$), ma omettiamo i dettagli.

(nella terza uguaglianza abbiamo sfruttato il fatto che la composizione di funzioni gode della proprietà associativa). In conclusione, possiamo scrivere $L_B \circ L_A = id$ che equivale a dire che $BA = I_n$, come volevamo.

A questo punto siamo pronti a descrivere il primo metodo per la determinazione dell'inversa di una matrice: grazie all'osservazione preliminare appena fatta, sappiamo che trovare l'inversa di A equivale a trovare una matrice B tale che $AB = I_n$ (senza dover verificare anche $BA = I_n$), ovvero

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & & \dots & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ & & \dots & \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

Tenuto conto della definizione di prodotto righe per colonne, vediamo che moltiplicando le righe di A per la prima colonna di B deve essere

$$\begin{cases} a_{11}b_{11} + a_{12}b_{21} + \dots + a_{1n}b_{n1} = 1 \\ a_{21}b_{11} + a_{22}b_{21} + \dots + a_{2n}b_{n1} = 0 \\ \dots \\ a_{n1}b_{11} + a_{n2}b_{21} + \dots + a_{nn}b_{n1} = 0 \end{cases}$$

ovvero la prima colonna di B soddisfa il sistema

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = 1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = 0 \\ \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = 0 \end{cases} \tag{4.27}$$

con matrice dei coefficienti uguale a A e termini noti uguali alla prima colonna della matrice identica.

Analogamente, moltiplicando le righe di A per la seconda colonna di B si vede che devono essere soddisfatte le seguenti

$$\begin{cases} a_{11}b_{12} + a_{12}b_{22} + \dots + a_{1n}b_{n2} = 0 \\ a_{21}b_{12} + a_{22}b_{22} + \dots + a_{2n}b_{n2} = 1 \\ \dots \\ a_{n1}b_{12} + a_{n2}b_{22} + \dots + a_{nn}b_{n2} = 0 \end{cases}$$

ovvero la seconda colonna di B deve soddisfare il sistema

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = 1 \\ \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = 0 \end{cases} \tag{4.28}$$

sempre con matrice dei coefficienti uguale a A ma stavolta con termini noti uguali alla seconda colonna della matrice identica, e così via possiamo ragionare allo stesso modo fino all'ultima colonna di B che dovrà soddisfare

$$\begin{cases} a_{11}b_{1n} + a_{12}b_{2n} + \cdots + a_{1n}b_{nn} = 0 \\ a_{21}b_{1n} + a_{22}b_{2n} + \cdots + a_{2n}b_{nn} = 0 \\ \qquad\qquad\qquad \cdots \\ a_{n1}b_{1n} + a_{n2}b_{2n} + \cdots + a_{nn}b_{nn} = 1 \end{cases}$$

ovvero il sistema

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ \qquad\qquad\qquad \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = 1 \end{cases} \tag{4.29}$$

Riassumendo, si ha $AB = I_n$ se e solo se le colonne $\begin{pmatrix} b_{11} \\ b_{21} \\ \cdots \\ b_{n1} \end{pmatrix}$, $\begin{pmatrix} b_{12} \\ b_{22} \\ \cdots \\ b_{n2} \end{pmatrix}$, ..., $\begin{pmatrix} b_{1n} \\ b_{2n} \\ \cdots \\ b_{nn} \end{pmatrix}$ sono soluzione rispettivamente degli n sistemi (4.27), (4.28), ..., (4.29).

Come sappiamo, per risolvere un sistema basta scriverne la matrice completa, composta da matrice dei coefficienti delle incognite e colonna dei termini noti, e ridurla a gradini mediante operazioni elementari sulle righe. Poiché i sistemi (4.27), (4.28), ..., (4.29) hanno tutti la stessa matrice dei coefficienti, cioè A, e differiscono solo per i termini noti, possiamo risolverli tutti contemporaneamente eseguendo le stesse operazioni elementari: a questo scopo, basta scrivere la matrice

$$(A|I_n) = \left(\begin{array}{cccc|cccc} a_{11} & a_{12} & \dots & a_{1n} & 1 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2n} & 0 & 1 & \dots & 0 \\ & & \dots & & & & \dots & \\ a_{n1} & a_{n2} & \dots & a_{nn} & 0 & 0 & \dots & 1 \end{array}\right) \tag{4.30}$$

ottenuta affiancando tutti i termini noti dei sistemi (4.27), (4.28), ..., (4.29) e risolverli contemporaneamente con una sola riduzione (chiaramente, in base al Teorema 4.19 la matrice A è invertibile se e solo se in seguito a tale riduzione non si annullerà nessuna delle sue righe).

Vediamo subito un esempio: data la matrice $A = \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix}$, verifichiamo se essa è invertibile e, in caso affermativo, calcoliamone l'inversa. Come detto sopra, affianchiamo a tale matrice la matrice identica dello stesso ordine

$$(A|I_n) = \left(\begin{array}{cc|cc} 1 & -1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{array}\right)$$

che rappresenta i due sistemi le cui soluzioni sono le colonne della matrice inversa, e iniziamo con l'applicare il procedimento di riduzione a gradini: a questo scopo basta il singolo passaggio

$$(A|I_n) = \left(\begin{array}{cc|cc} 1 & -1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{array}\right) \underset{R_2 \to R_2 - R_1}{\longrightarrow} \left(\begin{array}{cc|cc} 1 & -1 & 1 & 0 \\ 0 & 3 & -1 & 1 \end{array}\right) \tag{4.31}$$

da cui vediamo che, dopo la riduzione a gradini, ad A non si annulla nessuna riga e quindi, come abbiamo detto sopra, A è invertibile.

La prima colonna dell'inversa B di A è data dalla soluzione del sistema ridotto

$$\begin{cases} x_1 - x_2 = 1 \\ 3x_2 = -1 \end{cases} \tag{4.32}$$

ovvero, come si vede risolvendo dal basso, la coppia $(\frac{2}{3}, -\frac{1}{3})$, che è quindi la prima colonna della matrice inversa. Analogamente, la seconda colonna dell'inversa B di A è data dalla soluzione del sistema ridotto

$$\begin{cases} x_1 - x_2 = 0 \\ 3x_2 = 1 \end{cases} \tag{4.33}$$

ovvero, come si vede risolvendo dal basso, la coppia $(\frac{1}{3}, \frac{1}{3})$, che è quindi la seconda colonna della matrice inversa.
In conclusione, l'inversa della matrice A è

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

Infatti, si verifica subito con un calcolo che

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

concordemente con la definizione di inversa.

Per evitare di scrivere e risolvere separatamente i sistemi (4.32) e (4.33) e trovare invece in modo più diretto la matrice inversa, si può procedere come segue. Dopo aver effettuato la riduzione a gradini in (4.31), si applicano ulteriori operazioni elementari fino a trasformare la matrice A del blocco di sinistra nella matrice identica: a questo punto nel blocco di destra si legge direttamente l'inversa. Per vederlo, riprendiamo da (4.31) e facciamo comparire prima uno zero in posizione 1 2 eseguendo

$$\left(\begin{array}{cc|cc} 1 & -1 & 1 & 0 \\ 0 & 3 & -1 & 1 \end{array}\right) \underset{R_1 \to 3R_1 + R_2}{\longrightarrow} \left(\begin{array}{cc|cc} 3 & 0 & 2 & 1 \\ 0 & 3 & -1 & 1 \end{array}\right)$$

e applichiamo poi a ogni riga l'operazione elementare del secondo tipo che consiste nel dividerla per l'elemento che si trova sulla diagonale:

$$\left(\begin{array}{cc|cc} 3 & 0 & 2 & 1 \\ 0 & 3 & -1 & 1 \end{array}\right) \underset{\substack{R_1 \to (1/3)R_1 \\ R_2 \to (1/3)R_2}}{\longrightarrow} \left(\begin{array}{cc|cc} 1 & 0 & 2/3 & 1/3 \\ 0 & 1 & -1/3 & 1/3 \end{array}\right). \tag{4.34}$$

Come si vede la matrice identica che avevamo affiancato ad A si è trasformata nella matrice inversa di A già trovata sopra.

Per capire perché, ricordiamo che le operazioni elementari che stiamo eseguendo servono a risolvere contemporaneamente i sistemi che poi ci danno come soluzione le colonne della matrice inversa: ma allora, riducendo la matrice A (cioè la matrice dei coefficienti di tali sistemi) alla matrice identica come in (4.34) non stiamo facendo altro che ridurre i due sistemi alla forma

$$\begin{cases} x_1 = \frac{2}{3} \\ x_2 = -\frac{1}{3} \end{cases}, \quad \begin{cases} x_1 = \frac{1}{3} \\ x_2 = \frac{1}{3} \end{cases}$$

e cioè far comparire direttamente le soluzioni cercate (che sono proprio le colonne della matrice inversa).

Vediamo un altro esempio:

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

Iniziamo con il trasformare la matrice $(A|I_n)$ in una matrice a gradini

$$(A|I_n) = \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 2 & 1 & 1 & 0 & 0 & 1 \end{array}\right) \underset{\substack{R_2 \to R_2 + R_1 \\ R_3 \to R_3 - 2R_1}}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 2 & 1 & 1 & 0 \\ 0 & -1 & -3 & -2 & 0 & 1 \end{array}\right)$$

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 2 & 1 & 1 & 0 \\ 0 & -1 & -3 & -2 & 0 & 1 \end{array}\right) \underset{R_3 \to 2R_3 + R_2}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 2 & 1 & 1 & 0 \\ 0 & 0 & -4 & -3 & 1 & 2 \end{array}\right)$$

Il fatto che non si sia annullata nessuna riga nel blocco di sinistra ci dice che la matrice A è invertibile.
Ora, come spiegato sopra, facciamo comparire zeri sopra la diagonale, effettuando una sorta di riduzione a gradini "all'incontrario", dal basso verso l'altro e da destra verso sinistra (in ogni passaggio, mettiamo in evidenza in grassetto i nuovi zeri che facciamo comparire)

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 2 & 1 & 1 & 0 \\ 0 & 0 & -4 & -3 & 1 & 2 \end{array}\right) \underset{\substack{R_2 \to 2R_2 + R_3 \\ R_1 \to 2R_1 + R_3}}{\longrightarrow} \left(\begin{array}{ccc|ccc} 2 & 2 & \mathbf{0} & -1 & 1 & 2 \\ 0 & 4 & \mathbf{0} & -1 & 3 & 2 \\ 0 & 0 & -4 & -3 & 1 & 2 \end{array}\right)$$

$$\left(\begin{array}{ccc|ccc} 2 & 2 & 0 & -1 & 1 & 2 \\ 0 & 4 & 0 & -1 & 3 & 2 \\ 0 & 0 & -4 & -3 & 1 & 2 \end{array}\right) \underset{R_1 \to 2R_1 - R_2}{\longrightarrow} \left(\begin{array}{ccc|ccc} 4 & \mathbf{0} & 0 & -1 & -1 & 2 \\ 0 & 4 & 0 & -1 & 3 & 2 \\ 0 & 0 & -4 & -3 & 1 & 2 \end{array}\right)$$

Infine, dividiamo ogni riga per l'elemento sulla diagonale applicando operazioni elementari del secondo tipo

$$\left(\begin{array}{ccc|ccc} 4 & 0 & 0 & -1 & -1 & 2 \\ 0 & 4 & 0 & -1 & 3 & 2 \\ 0 & 0 & -4 & -3 & 1 & 2 \end{array}\right) \underset{\substack{R_1 \to (1/4)R_1 \\ R_1 \to (1/4)R_2 \\ R_3 \to -(1/4)R_3}}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1/4 & -1/4 & 1/2 \\ 0 & 1 & 0 & -1/4 & 3/4 & 1/2 \\ 0 & 0 & 1 & 3/4 & -1/4 & -1/2 \end{array}\right)$$

Quindi

$$A^{-1} = \begin{pmatrix} -1/4 & -1/4 & 1/2 \\ -1/4 & 3/4 & 1/2 \\ 3/4 & -1/4 & -1/2 \end{pmatrix}.$$

Vediamo ora un modo alternativo per calcolare l'inversa di una matrice invertibile, basato sul determinante e sulla nozione di cofattore, vista nel precedente capitolo.
Abbiamo visto nel Teorema 4.19 che una matrice A di ordine n è invertibile se e solo se il suo rango è n. Ma come sappiamo dal Teorema 3.6, questo equivale ad avere $\det(A) \neq 0$. Dimostriamo ora la seguente

Proposizione 4.20. Sia $A \in M_n(\mathbb{K})$ una matrice invertibile (ovvero con $\det(A) \neq 0$). Allora la sua inversa A^{-1} è data da

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ & \vdots & & \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{pmatrix} \tag{4.35}$$

dove C_{ij} indica il cofattore di a_{ij}, e $\frac{1}{\det(A)}$ davanti alla matrice dei cofattori significa che ogni entrata di tale matrice deve essere moltiplicata per $\frac{1}{\det(A)}$.

Osservazione 4.21. A proposito della disposizione dei cofattori nella (4.35), si noti che i cofattori delle entrate della prima *riga* di A sono nella prima *colonna* della (4.35), i cofattori delle entrate della seconda *riga* di A sono nella seconda *colonna* della (4.35), e così via.

Proof. Dimostrare che la matrice data in (4.35) è l'inversa di A significa dimostrare che il prodotto righe per colonne tra A e tale matrice è uguale alla matrice identica I_n

$$\begin{pmatrix} & & \dots & \\ a_{i1} & a_{i2} & \dots & a_{in} \\ & & \dots & \end{pmatrix} \begin{pmatrix} & \frac{C_{j1}}{\det(A)} & \\ & \frac{C_{j2}}{\det(A)} & \\ \dots & \vdots & \dots \\ & \frac{C_{jn}}{\det(A)} & \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & 1 \end{pmatrix} \tag{4.36}$$

ovvero che

$$a_{i1}\frac{C_{j1}}{\det(A)} + a_{i2}\frac{C_{j2}}{\det(A)} + \dots + a_{in}\frac{C_{jn}}{\det(A)} \tag{4.37}$$

è uguale a 1 se $i = j$ e uguale a 0 se $i \neq j$. Ora, riscriviamo la (4.37) con un unico denominatore come segue

$$\frac{a_{i1}C_{j1} + a_{i2}C_{j2} + \dots + a_{in}C_{jn}}{\det(A)} \tag{4.38}$$

e osserviamo che il numeratore di tale espressione, quando $i = j$, non è nient'altro che lo sviluppo di Laplace $a_{j1}C_{j1} + a_{j2}C_{j2} + \dots + a_{jn}C_{jn}$ del determinante di A rispetto alla j-esima riga, e quindi per $i = j$ la (4.38) è uguale a $\frac{\det(A)}{\det(A)} = 1$, come volevamo.

Ci rimane da dimostrare che quando $i \neq j$, la (4.38) si annulla: questo equivale chiaramente all'annullarsi del suo numeratore

$$a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn} \tag{4.39}$$

Confrontando tale espressione con quella dello sviluppo di Laplace del determinante di A rispetto alla j-esima riga, ovvero $a_{j1}C_{j1} + a_{j2}C_{j2} + \cdots + a_{jn}C_{jn}$, notiamo che esse differiscono solamente per il fatto che al posto di $a_{j1}, a_{j2}, \ldots, a_{jn}$ (ovvero le entrate della j-esima riga di A) ci sono $a_{i1}, a_{i2}, \ldots, a_{in}$, ovvero le entrate della i-esima riga di A: quindi, la (4.39) coincide con il determinante di A se $a_{j1} = a_{i1}$, $a_{j2} = a_{i2}$, ..., $a_{jn} = a_{in}$, ovvero se la sua j-riga e la sua i-esima riga sono uguali. Ma come sappiamo, il determinante di una matrice con due righe uguali è nullo (pagina 122), il che dimostra che la (4.39) è nulla, come volevamo.

□

Esempio 4.22. Calcoliamo l'inversa della matrice $A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & 2 \end{pmatrix}$.

I cofattori sono

$$C_{11} = (-1)^{1+1} \det \begin{pmatrix} -1 & 1 \\ 0 & 2 \end{pmatrix} = -2, \quad C_{12} = (-1)^{1+2} \det \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = -1$$

$$C_{13} = (-1)^{1+3} \det \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} = 1, \quad C_{21} = (-1)^{2+1} \det \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} = -4$$

$$C_{22} = (-1)^{2+2} \det \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = 1, \quad C_{23} = (-1)^{2+3} \det \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix} = 2$$

$$C_{31} = (-1)^{3+1} \det \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix} = 3, \quad C_{32} = (-1)^{3+2} \det \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = 0$$

$$C_{33} = (-1)^{3+3} \det \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix} = -3.$$

Quindi

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{2n} & C_{33} \end{pmatrix} = -\frac{1}{3} \begin{pmatrix} -2 & -4 & 3 \\ -1 & 1 & 0 \\ 1 & 2 & -3 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & \frac{4}{3} & -1 \\ \frac{1}{3} & -\frac{1}{3} & 0 \\ -\frac{1}{3} & -\frac{2}{3} & 1 \end{pmatrix}.$$

dove il determinante di A è stato calcolato sviluppandolo secondo Laplace rispetto alla terza riga, usando i cofattori già calcolati:

$$\det(A) = a_{31}C_{31} + a_{32}C_{32} + a_{33}C_{33} = 1 \cdot 3 + 0 \cdot 0 + 2 \cdot (-3) = -3$$

Osservazione 4.23. Osserviamo che la (4.35), nel caso $n = 2$, diventa la semplice formula

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}. \tag{4.40}$$

in quanto i cofattori sono semplicemente $C_{11} = (-1)^{1+1}a_{22}, C_{12} = (-1)^{1+2}a_{21}, C_{21} = (-1)^{2+1}a_{12}, C_{22} = (-1)^{2+2}a_{11}$ (ricordiamo che il determinante di una matrice di ordine 1 si definisce come il valore della sua unica entrata). In pratica, a parte dividere per il determinante, la matrice inversa si ottiene scambiando tra loro i due elementi sulla diagonale e cambiando di segno le restanti entrate.

Esempio 4.24. Applichiamo la (4.40) al caso della matrice che rappresenta una rotazione di angolo θ nel piano, ovvero $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$: si ha

$$A^{-1} = \frac{1}{\cos^2\theta + \sin^2\theta} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} =$$

(per l'identità $\cos^2\theta + \sin^2\theta = 1$ e le formule trigonometriche per il seno e il coseno dell'opposto di un angolo)

$$= \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix}$$

che è ancora una matrice di rotazione, quella associata alla rotazione di angolo $-\theta$, che è l'inversa della rotazione di angolo θ (se applichiamo prima una e poi l'altra il vettore torna alla posizione iniziale).

In effetti, in generale, la matrice A^{-1} inversa di una matrice A data rappresenta la funzione inversa della funzione $L_A : \mathbb{K}^n \to \mathbb{K}^n$ determinata da A: infatti, come sappiamo la composizione $L_A \circ L_{A^{-1}}$ è la funzione determinata dal prodotto $AA^{-1} = I_n$, ovvero la funzione identica, cioè $L_A \circ L_{A^{-1}} = id$ (e analogamente $L_{A^{-1}} \circ L_A = id$).

Vediamo ora che la formula (4.35) per il calcolo dell'inversa può essere utilizzata per ricavare un modo alternativo alla riduzione a gradini per risolvere certi sistemi di equazioni lineari.

A tale scopo, osserviamo prima che un generico sistema lineare di m equazioni in n incognite

$$
\left\{
\begin{array}{c}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\
\cdot \\
\cdot \\
\cdot \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m
\end{array}
\right.
\tag{4.41}
$$

può essere riscritto in forma molto più concisa grazie al prodotto di matrici. Più precisamente, se denotiamo con $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$ la matrice dei coefficienti del sistema, con $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ la matrice (costituita da una sola colonna) che ha come entrate le incognite, e con $b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$ la matrice (sempre costituita da una sola colonna) che ha come entrate i termini noti del sistema si ha, svolgendo il prodotto righe per colonne, che

$$
\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}
\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}
=
\begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}
$$

E quindi il sistema (4.41) può essere riscritto

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

ovvero nella semplice forma

$$
Ax = b. \tag{4.42}
$$

Ora, supponiamo che la matrice A dei coefficienti del sistema sia quadrata di ordine n (quindi il sistema ha n equazioni e n incognite) e che abbia determinante diverso da zero. Allora A è invertibile e possiamo moltiplicare entrambi i membri dell'uguaglianza (4.42) a sinistra per l'inversa A^{-1}:

$$A^{-1}(Ax) = A^{-1}b$$

Tenendo conto che il prodotto di matrici gode della proprietà associativa, questo può essere riscritto come

$$(A^{-1}A)x = A^{-1}b$$

ovvero, visto che per definizione di inversa $A^{-1}A = I_n$,

$$I_n x = A^{-1}b$$

e quindi, essendo I_n elemento neutro per il prodotto,

$$x = A^{-1}b \tag{4.43}$$

Quindi la soluzione x di un sistema con stesso numero di equazioni e di incognite e matrice dei coefficienti con determinante diverso da zero può essere trovata mediante la (4.43).

Osservazione 4.25. Si noti che non abbiamo fatto altro che applicare gli stessi passaggi, normalmente sottointesi, che si applicano quando si vuole risolvere una semplice equazione di primo grado in una sola incognita $ax = b$. Infatti, ad esempio, se dobbiamo risolvere $2x = 3$ dividiamo entrambi i membri per 2, ovvero equivalentemente moltiplichiamo per l'inverso $\frac{1}{2}$ di 2 ottenendo $\frac{1}{2}(2x) = \frac{1}{2}3 = \frac{3}{2}$; per la proprietà associativa a primo membro si ha $(\frac{1}{2}2)x = \frac{3}{2}$ ovvero, essendo $\frac{1}{2}2 = 1$, si ha $1x = \frac{3}{2}$ cioè, essendo 1 elemento neutro per la moltiplicazione tra numeri, $x = \frac{3}{2}$, che è la soluzione dell'equazione.
L'unica differenza con il caso dei sistemi $Ax = b$ è che non sempre A è invertibile, mentre, a meno che a non sia zero, nell'equazione $ax = b$ tale ipotesi è sempre garantita.

Ora, se combiniamo la (4.43) con la formula per l'inversa (4.35), vediamo che la soluzione di un sistema con n equazioni e n incognite e matrice dei coefficienti invertibile è data da

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \frac{1}{\det(A)} \begin{pmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ & & \vdots & \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}. \tag{4.44}$$

Quindi, come si vede, le componenti x_i della soluzione del sistema si ottengono moltiplicando le righe della matrice dei cofattori (divisa per il determinante di A) per la colonna dei termini noti:

$$x_1 = \frac{b_1 C_{11} + b_2 C_{21} + \cdots + b_n C_{n1}}{\det(A)}$$

$$x_2 = \frac{b_1 C_{12} + b_2 C_{22} + \cdots + b_n C_{n2}}{\det(A)}$$

$$\vdots$$

e in generale

$$x_i = \frac{b_1 C_{1i} + b_2 C_{2i} + \cdots + b_n C_{ni}}{\det(A)} \tag{4.45}$$

Ora, se confrontiamo il numeratore della (4.45) con l'espressione del determinante di A calcolato con lo sviluppo di Laplace rispetto alla i-esima colonna

$$\det(A) = a_{1i} C_{1i} + a_{2i} C_{2i} + \cdots + a_{ni} C_{ni}$$

vediamo che l'unica differenza è che al posto degli elementi $a_{1i}, a_{2i}, \ldots, a_{ni}$ della colonna i-esima di A abbiamo i termini noti $b_1, b_2, \ldots, b_n$: in altre parole, tale numeratore coincide con lo sviluppo di Laplace del determinante della matrice che si ottiene da A sostituendo i termini noti al posto della i-esima colonna.
Riassumendo, abbiamo dimostrato il

Teorema 4.26. (Teorema di Cramer) Sia $Ax = b$ un sistema di n equazioni lineari in n incognite, con $\det(A) \neq 0$ (cioè A è invertibile). Allora tale sistema ha un'unica soluzione $(x_1, x_2, \ldots, x_n)$ le cui componenti sono date da

$$x_i = \frac{\det(B_i)}{\det(A)}, \quad i = 1, \ldots, n$$

dove B_i è la matrice che si ottiene da A sostituendo la colonna dei termini noti b al posto della i-esima colonna di A.

Esempio 4.27. Consideriamo il sistema

$$\begin{cases} 2x_1 + x_2 = 5 \\ x_1 - x_2 = 3 \end{cases}$$

La matrice $A = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}$ dei coefficienti del sistema ha determinante $\det(A) = -3$, quindi è invertibile e possiamo applicare il metodo di Cramer, ovvero abbiamo le componenti dell'unica soluzione date da

$$x_1 = \frac{\det(B_1)}{\det(A)} = \frac{\det\begin{pmatrix} 5 & 1 \\ 3 & -1 \end{pmatrix}}{\det(A)} = \frac{-8}{-3} = \frac{8}{3}$$

$$x_2 = \frac{\det(B_2)}{\det(A)} = \frac{\det\begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix}}{\det(A)} = \frac{1}{-3} = -\frac{1}{3}$$

Per concludere questa parte sull'inversa, mostriamo che, data una matrice invertibile A, vale la seguente

$$\det(A^{-1}) = \frac{1}{\det(A)} \tag{4.46}$$

Questa uguaglianza si dimostra come corollario del seguente, importante risultato, detto *teorema di Binet*:

Teorema 4.28. Siano $A, B \in M_n(\mathbb{K})$ due matrici quadrate. Allora

$$\det(AB) = \det(A)\det(B)$$

Non dimostriamo il teorema di Binet, ma illustriamolo con un esempio: se $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ e $B = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$, allora $AB = \begin{pmatrix} 4 & 7 \\ 10 & 15 \end{pmatrix}$, e si vede che $\det(A) = -2$, $\det(B) = 5$ e $\det(AB) = -10 = (-2) \cdot 5$, concordemente con il teorema di Binet.

Per dimostrare la (4.46), basta applicare il teorema di Binet al caso $B = A^{-1}$: allora abbiamo

$$\det(AA^{-1}) = \det(A)\det(A^{-1})$$

ovvero, tenuto conto che $AA^{-1} = I_n$,

$$\det(I_n) = \det(A)\det(A^{-1}) \tag{4.47}$$

Ora, è facile vedere che il determinante della matrice identica I_n, per qualunque ordine n, è uguale a 1: infatti, basta applicare lo sviluppo di Laplace rispetto alla prima riga:

$$\det \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & & \dots & \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} = 1 \cdot (-1)^{1+1} \det \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & 1 \end{pmatrix} =$$

(essendo l'unico cofattore ancora una matrice identica, di ordine $n-1$, riapplichiamo lo sviluppo di Laplace rispetto alla prima riga, e così via)

$$= \det \begin{pmatrix} 1 & \dots & 0 \\ & \dots & \\ 0 & \dots & 1 \end{pmatrix} = \dots = \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1.$$

Quindi la (4.47) diventa

$$1 = \det(A) \det(A^{-1})$$

che implica subito la (4.46).

Concludiamo questo capitolo vedendo come, oltre al prodotto, tra matrici può essere definita anche un'operazione di somma: definiamo la *somma di due matrici* $A, B \in M_{m,n}(\mathbb{K})$, denotata $A + B$, come la matrice di $M_{m,n}(\mathbb{K})$ la cui entrata di posto ij è ottenuta sommando le entrate di posto ij di A e B, ovvero

$$(A + B)_{ij} = A_{ij} + B_{ij}$$

Ad esempio, siano $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ e $B = \begin{pmatrix} 1 & 0 & -2 \\ 3 & 1 & 1 \end{pmatrix}$. Allora

$$A + B = \begin{pmatrix} 1+1 & 2+0 & 3+(-2) \\ 4+3 & 5+1 & 6+1 \end{pmatrix} = \begin{pmatrix} 2 & 2 & 1 \\ 7 & 6 & 7 \end{pmatrix}.$$

La somma di matrici così definita ha le proprietà seguenti

(1) vale la proprietà commutativa, ovvero $A + B = B + A$

(2) vale la proprietà associativa, ovvero $(A + B) + C = A + (B + C)$

(3) esiste un elemento neutro, dato dalla matrice che ha tutte le entrate nulle

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

In ogni insieme $M_{m,n}(\mathbb{K})$, chiameremo *matrice nulla*, e la denoteremo O, la matrice che ha tutte le entrate uguali a zero.

(4) ogni matrice A ha un'inversa additiva o opposta, data dalla matrice che ha come entrata di posto ij l'opposto dell'entrata di posto ij di A:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} -1 & -2 & -3 \\ -4 & -5 & -6 \end{pmatrix} = \begin{pmatrix} -1 & -2 & -3 \\ -4 & -5 & -6 \end{pmatrix} + \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Denoteremo $-A$ l'opposta di A.

(5) vale la proprietà distributiva rispetto al prodotto, ovvero se $A \in M_{m,n}(\mathbb{K})$ e $B, C \in M_{n,p}(\mathbb{K})$, si ha $A(B + C) = AB + AC$ e analogamente se $A, B \in M_{m,n}(\mathbb{K})$ e $C \in M_{n,p}(\mathbb{K})$, si ha $(A + B)C = AC + BC$.

Come abbiamo già visto ad esempio quando abbiamo parlato delle proprietà della somma di vettori nel Capitolo 1, sono proprio queste proprietà a garantirci che in un'uguaglianza tra matrici siamo autorizzati a fare alcune delle manipolazioni che facciamo normalmente con le espressioni tra numeri, ad esempio spostare da un membro all'altro gli addendi di un'uguaglianza. Infatti (si confronti con quanto già visto nell'Osservazione 1.1 a proposito dei vettori), se ad esempio ho $A + B = C$, posso sommare a entrambi i membri di questa uguaglianza la matrice $-B$ opposta di B, ovvero $(A + B) + (-B) = C + (-B)$, sfruttare la proprietà associativa a primo membro e scrivere $A + (B + (-B)) = C + (-B)$, sfruttare il fatto che $B + (-B) = O$, da cui $A + O = C + (-B)$ e quindi, per il fatto che la matrice nulla O è elemento neutro per la somma, $A = C - B$ (dove stiamo scrivendo $C - B$ come forma breve per $C + (-B)$).

Si presti comunque attenzione che la non commutatività del prodotto di matrici rende non valide alcune identità classiche dell'algebra numerica, come la formula per il quadrato di un binomio ($(a+b) = a^2+2ab+b^2$) o il prodotto notevole $(a + b)(a - b) = a^2 - b^2$. Infatti, per matrici si ha

$$(A + B)^2 = (A + B)(A + B) = A^2 + AB + BA + B^2$$

Nella seconda uguaglianza abbiamo usato la proprietà distributiva, che vale anche per matrici, ma non valendo la commutatività del prodotto non possiamo scrivere $AB + BA = 2AB$.
Analogamente,

$$(A + B)(A - B) = A^2 - AB + BA - B^2$$

e di nuovo non essendo valida la commutatività del prodotto di matrici non possiamo semplificare $-AB+BA$.

Come ultimo esempio di proprietà che vale nell'algebra numerica ma non in quella delle matrici citiamo la non validità della cosiddetta *legge di annullamento del prodotto*, ovvero può essere $AB=O$ pur non essendo nè $A=O$ nè $B=O$ (dove O denota la matrice nulla), ad esempio $\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 2 \\ -1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

Osservazione 4.29. Nella formula (4.35) per l'inversa di una matrice, abbiamo introdotto la notazione per cui uno scalare davanti a una matrice (in quel caso era $\frac{1}{\det(A)}$ davanti alla matrice dei cofattori) significa che dobbiamo moltiplicare ogni entrata della matrice per quello scalare.
In generale, data una matrice A e uno scalare $c \in \mathbb{K}$, scriveremo cA per indicare la matrice che si ottiene moltiplicando tutte le entrate di A per c, ovvero $(cA)_{ij} = c \cdot A_{ij}$.
In pratica, questo definisce una nuova operazione, di cui segnaliamo, senza dimostrarle, le seguenti proprietà, che ci saranno utili nei capitoli successivi:

$$c(A+B) = cA + cB \tag{4.48}$$

$$(cA)B = A(cB) = c(AB) \tag{4.49}$$