4. Spazi Vettoriali e prodotti scalari

Ricordiamo il linguaggio astratto e discutiamo alcuni fatti fondamentali relativi alla *struttura lineare*.

4.1 Spazi vettoriali

La struttura lineare di $\mathbb{K}^n$ è condivisa da molti oggetti matematici. Ad esempio, abbiamo già notato come le matrici soddisfano le stesse leggi di somma e moltiplicazione per scalari. Conviene quindi partire da definizioni generali per mettere a punto un linguaggio utile in molte situazioni.

4.1.1 Spazi vettoriali: definizioni e fatti elementari

Sia $\mathbb{K}$ un corpo commutativo che per noi sarà $\mathbb{R}$ o $\mathbb{C}$, ma che potrebbe essere un qualunque altro corpo commutativo.

4.1 Definizione. *Si chiama* spazio vettoriale su $\mathbb{K}$ *ogni insieme X dotato di*

(i) *una operazione interna $+ : X \times X \to X$ detta* addizione *per la quale X sia un gruppo commutativo, i.e.,*
  a) $(x + y) + z = x + (y + z)$, $x + y = y + x$, $\forall x, y \in X$,
  b) *esiste un elemento $0 \in X$, detto zero, tale che $x + 0 = 0 + x = x$ $\forall x \in X$,*
  c) *per ogni $x \in X$ esiste $-x \in X$ tale che $x + (-x) = 0$,*

(ii) *una operazione di moltiplicazione $\mathbb{K} \times X \to X$ che associa ad ogni $\lambda \in \mathbb{K}$ ed $x \in X$ un elemento in X notato λx tale che*
  a) $\lambda(x + y) = \lambda x + \lambda y$, $(\lambda + \mu)x = \lambda x + \mu x$,
  b) $\lambda(\mu x) = (\lambda\mu)x$, $1 \cdot x = x$.

In particolare $(-1)x = -x$ $\forall x \in X$. Si scrive poi $x - y$ al posto di $x + (-y)$.

Se X ha una struttura di spazio vettoriale su $\mathbb{K}$, gli elementi di X si chiamano *vettori*, gli elementi di $\mathbb{K}$ *scalari* ed il prodotto λx si chiama *prodotto per scalari*. La moltiplicazione per scalari ci permette di guardare il vettore x in ogni scala.

4.2 Esempio $\mathbb{K}^n$, $n \geq 1$, è uno spazio vettoriale su $\mathbb{K}$. Anche $M_{m,n}(\mathbb{K})$ è uno spazio vettoriale su $\mathbb{K}$.

4.3 Esempio Questi esempi non esauriscono i possibili spazi con una struttura vettoriale. Molte classi di funzioni hanno una struttura vettoriale naturale. Ad esempio

- l'insieme delle funzioni reali su $[0, 1]$; l'insieme delle funzioni continue su $[0, 1]$, $C^0([0, 1])$; l'insieme delle funzioni con derivate continue su $[0, 1]$ fino all'ordine k, $C^k([0, 1])$,
- L'insieme delle funzioni derivabili di qualunque ordine su $[0, 1]$,
- l'insieme dei polinomi, o dei polinomi di grado al più n; l'insieme dei polinomi trigonometrici, o dei polinomi trigonometrici di grado al più n,
- l'insieme delle funzioni integrabili secondo Riemann su $(0, 1)$,

◦ l'insieme di tutte le successioni a valori in $[0,1]$.

Lo studio che qui sviluppiamo riguarda proprietà universali che dipendono dalla struttura di spazio vettoriale, indipendentemente dallo specifico esempio. Le seguenti definizioni sono modellate sulle analoghe per sottospazi di $\mathbb{K}^n$.

4.4 Definizione. *Un sottoinsieme W di uno spazio vettoriale X si chiama un* sottospazio vettoriale *di X se*

(i) $0 \in W$,
(ii) $\forall\, x, y \in W$ *si ha* $x + y \in W$,
(iii) $\forall\, x \in W\ \forall\, \lambda \in \mathbb{K}$ *si ha* $\lambda x \in W$.

a. Combinazioni lineari e generatori

In uno spazio vettoriale X su $\mathbb{K}$ hanno senso le *combinazioni lineari* finite di elementi di X a coefficienti in $\mathbb{K}$, i.e.,

$$\sum_{i=1}^{n} \lambda^i v_i \in X$$

se $\lambda^1, \lambda^2, \dots, \lambda^n \in \mathbb{K}$ e $v_1, v_2, \dots, v_n \in X$. Per comodità usiamo sistematicamente la convenzione di mettere in basso l'indice di una lista di vettori e in alto l'indice dei rispettivi coefficienti,

Si prova subito che $W \subset X$ *è un sottospazio vettoriale dello spazio vettoriale X su $\mathbb{K}$ se e solo se tutte le combinazioni lineari di elementi di W a coefficienti in $\mathbb{K}$ sono ancora elementi di W.*

Sia S un sottoinsieme di X. L'insieme di tutte le combinazioni lineari finite di elementi di S è un sottospazio vettoriale di X. Lo si chiama *sottospazio vettoriale generato da S in X* ed è spesso indicato con $\operatorname{Span} S$,

$$\operatorname{Span} S := \Big\{ \sum_{i=1}^{n} \lambda^i e_i \in X \,|\, \lambda^1, \lambda^2, \dots, \lambda^n \in \mathbb{K},\ e_1, e_2, \dots, e_n \in S, n \in \mathbb{N} \Big\}.$$

Si può anche provare che $\operatorname{Span} S$ è l'intersezione di tutti i sottospazi vettoriali di X che contengono S. $\operatorname{Span} S$ è dunque *il più piccolo sottospazio vettoriale contenente S*. Se $S, W \subset X$ e $W = \operatorname{Span} S$, allora W è un sottospazio vettoriale di X e si dice anche che S è un *insieme di generatori* di X o che S genera W (a coefficienti in $\mathbb{K}$). Dalla definizione di $\operatorname{Span} S$ segue subito che *S è un insieme di generatori di X se e solo se ogni $x \in \operatorname{Span} S$ è combinazione lineare di vettori in S*, i.e., se per ogni $x \in X$ esistono un intero $n \geq 1$, vettori $e_1, e_2, \dots, e_n$ in S e scalari $\lambda^1, \lambda^2, \dots, \lambda^n \in \mathbb{K}$ tali che $x = \sum_{i=1}^n \lambda^i e_i$.

b. Vettori linearmente indipendenti e basi

n vettori $v_1, v_2, \dots, v_n$ in uno spazio vettoriale X si dicono *linearmente dipendenti* se esistono scalari $\lambda^1, \dots, \lambda^n$ *non tutti nulli* tali che

$$\lambda^1 v_1 + \dots + \lambda^n v_n = 0,$$

o, in altri termini, se uno di essi è combinazione lineare degli altri. Se n vettori non sono linearmente dipendenti si dice che sono *linearmente indipendenti*. Più in generale un insieme S di vettori si dice *linearmente indipendente* se ogni scelta di un numero finito fra essi dà luogo ad un insieme di vettori linearmente indipendenti.

4.5 Definizione. *Un insieme di vettori linearmente indipendenti che genera X si chiama una* base *di X.*

4.6 Definizione. *Sia $\mathcal{C} \subset X$ un sottoinsieme di uno spazio vettoriale X. Un insieme di vettori $\mathcal{A} \subset \mathcal{C}$ si dice un* insieme massimale di vettori linearmente indipendenti in $\mathcal{C}$ *se $\mathcal{A}$ è costituito da vettori tra loro linearmente indipendenti e se aggiungendo ad $\mathcal{A}$ un vettore $w \in \mathcal{C} \setminus \mathcal{A}$, si ottiene un insieme di vettori linearmente dipendenti.*

Si ha

4.7 Proposizione. *Sia $\mathcal{C}$ un insieme di generatori di X. Un insieme $\mathcal{B} \subset \mathcal{C}$ è un insieme massimale di vettori linearmente indipendenti se e solo se $\mathcal{B}$ è una base di X.*

4.1.2 Basi e coordinate

a. Spazi vettoriali di dimensione finita

Si dice che uno spazio vettoriale X è *finitamente generato* se ha un insieme di generatori $\{v_1, v_2, \dots, v_n\}$ finito. Se X è finitamente generato da $\{v_1, v_2, \dots, v_n\}$, si può scegliere un insieme massimale di vettori linearmente indipendenti in $\{v_1, v_2, \dots, v_n\}$. X ha quindi una base costituita da un numero finito di elementi.

Sia X uno spazio vettoriale finitamente generato, sia $\{e_1, e_2, \dots, e_n\}$ una base di X e indichiamo con $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ la base standard di $\mathbb{K}^n$. Allora per ogni $x \in X$ esistono unici $x^1, x^2, \dots, x^n$ tali che $x = \sum_{i=1}^n x^i e_i$. Con altre parole la mappa $\mathcal{E} : \mathbb{K}^n \to X$ definita da $\mathcal{E}(\mathbf{e}_i) := e_i$ $\forall i = 1, \dots, n$, o equivalentemente da

$$\mathcal{E}(\mathbf{x}) := \sum_{i=1}^n x^i e_i, , \qquad \text{se } \mathbf{x} = (x^1, x^2, \dots, x^n),$$

è una mappa surgettiva e iniettiva. La mappa inversa $\mathcal{E}^{-1} : X \to \mathbb{K}^n$ associa ad ogni vettore $x \in X$, $x = \sum_{i=1}^n x^i e_i$, il vettore di $\mathbf{x} \in \mathbb{K}^n$, $\mathbf{x} := (x^1, x^2, \dots, x^n)$ delle sue coordinate. L'isomorfismo $\mathcal{E}$ e la sua inversa $\mathcal{E}^{-1}$ costituiscono il *sistema di coordinate in X determinato dalla base* $\{e_1, e_2, \dots, e_n\}$. Si noti che entrambe le mappe $\mathcal{E}, \mathcal{E}^{-1}$ sono lineari, cfr. Section 4.2, e che

4.8 Proposizione. *k vettori $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ in $\mathbb{K}^n$ sono linearmente indipendenti in $\mathbb{K}^n$ se e solo se $\mathcal{E}(\mathbf{v}_1), \dots, \mathcal{E}(\mathbf{v}_k)$ sono linearmente indipendenti su X.*

Come conseguenza dei risultati della Section **??** e della Proposizione 4.8 si trova

4.9 Proposizione. *Sia X uno spazio finitamente generato. Allora*

(i) *se $\{e_1, e_2, \dots, e_n\}$ è una base di X e se $v_1, v_2, \dots, v_p$, $p \le n$, sono p vettori linearmente indipendenti, allora è possibile completare $v_1, v_2, \dots, v_p$ con $n - p$ elementi della base $\{e_1, e_2, \dots, e_n\}$ in modo che assieme formino una nuova base di X,*
(ii) *se $\{v_1, v_2, \dots, v_n\}$ è un sistema di generatori di X, allora si può ottenere una base di X cancellando eventualmente alcuni dei v_i,*
(iii) *due qualsiasi basi hanno lo stesso numero di elementi.*

Il numero di elementi di ogni base è un elemento caratteristico di uno spazio vettoriale finitamente generato. Lo si chiama *dimensione* di X e lo si indica con $\dim X$. Come per $\mathbb{K}^n$ e per i suoi sottospazi

4.10 Corollario. *Sia X uno spazio vettoriale. X è finitamente generato se e solo se ha dimensione finita. Se X ha dimensione n, allora*

(i) *n vettori $v_1, v_2, \dots, v_n$ linearmente indipendenti sono una base di X,*
(ii) *se $k > n$, k vettori $v_1, v_2, \dots, v_k$ di X sono sempre linearmente dipendenti,*
(iii) *ogni sottospazio vettoriale W di X ha dimensione minore o uguale a n e $\dim W = n$ se e solo se $W = X$.*

4.2 Applicazioni lineari

Passiamo ora a considerare applicazioni lineari tra spazi vettoriali.

4.2.1 Definizioni

4.11 Definizione. *Siano X e Y due spazi vettoriali su $\mathbb{K}$. Una applicazione $\varphi : X \to Y$ si dice* $\mathbb{K}$-lineare, *o brevemente* lineare *se*

$$\varphi(x+y) = \varphi(x) + \varphi(y) \quad e \quad \varphi(\lambda x) = \lambda\varphi(x) \qquad \forall\, x, y \in X \text{ e } \forall\, \lambda \in \mathbb{K}.$$

Un'applicazione lineare tra spazi vettoriali si dice un isomorfismo *se è bigettiva.*

Ovviamente per ogni applicazione lineare $\varphi : X \to Y$ si ha $\varphi(0) = 0$ e

$$\varphi(\lambda^1 e_1 + \lambda^2 e_2 + \cdots + \lambda^n e_n) = \lambda^1\varphi(e_1) + \lambda^2\varphi(e_2) + \cdots + \lambda^n\varphi(e_n)$$

per ogni $\lambda^1, \lambda^2, \ldots, \lambda^n \in \mathbb{K}$, $e_1, e_2, \ldots, e_n \in X$. In particolare *ogni applicazione lineare è univocamente individuata dai valori che assume su una base.*

Lo spazio di tutte le applicazioni lineari $\varphi : X \to Y$ fra due spazi vettoriali si indica con $\mathcal{L}(X, Y)$. $\mathcal{L}(X, Y)$ è uno spazio vettoriale su $\mathbb{K}$. Si noti che la composizione di applicazioni lineari è lineare e che, se $\varphi : X \to Y$ è un isomorfismo, anche l'inversa è lineare e quindi un isomorfismo.

4.12 ¶ Sia $e_1, e_2, \ldots, e_n$ una base di uno spazio vettoriale X. Per ogni vettore v siano $v^1, v^2, \ldots, v^n$ le coordinate di v nella base $e_1, e_2, \ldots, e_n$. Per $i = 1, \ldots, n$, la *mappa coordinata* $e^i : X \to \mathbb{K}$ che ad ogni vettore $v \in X$ associa la sua i-esima coordinata, $e^i : X \to \mathbb{K}$, $e^i(v) := v^i$, è una applicazione lineare.

4.13 ¶ Sia V di dimensione finita ed $V = V_1 \oplus V_2$. Allora ogni elemento v di V si scrive in modo unico come $v = v_1 + v_2$ con $v_i \in V_i$. Le mappe coordinate $\pi_i : V \to V_i$, $i = 1, 2$, che a v associano v_i sono mappe lineari.

4.14 ¶ Sia X uno spazio vettoriale su $\mathbb{K}$. Dati n vettori $v_1, \ldots, v_n$ di X, l'applicazione $\varphi : \mathbb{K}^n \to X$ definita da $\varphi(\lambda^1, \ldots, \lambda^n) = \sum_1^n \lambda^i v_i$, è lineare. Inoltre φ è
(i) iniettiva se e solo se i vettori $v_1, \ldots, v_n$ sono linearmente indipendenti,
(ii) surgettiva se e solo se $v_1, \ldots, v_n$ generano X,
(iii) bigettiva se e solo se $v_1, \ldots, v_n$ è una base di X.

4.15 ¶ Se $\varphi : X \to Y$ è un isomorfismo tra spazi vettoriali di dimensione finita, allora $\dim X = \dim Y$.

4.16 ¶ Si osservi che le seguenti applicazioni sono lineari:
(i) l'applicazione derivazione $D : C^1([0,1]) \to C^0([0,1])$ che ad ogni funzione C^1 associa la sua derivata $f(x) \mapsto f'(x)$,
(ii) l'applicazione integrale da $C^0([0,1])$ in $\mathbb{R}$ che ad ogni funzione continua associa il suo integrale

$$f(x) \to \int_0^1 f(t)\,dt,$$

(iii) l'applicazione funzione integrale $C^0([0,1]) \to C^1([0,1])$ che ad ogni funzione continua associa la sua funzione integrale

$$f(x) \to F(x) := \int_0^x f(t)\,dt.$$

a. Immagine e nucleo

Sia $\varphi : X \to Y$ una applicazione lineare tra spazi vettoriali. Si verifica subito che, per ogni sottospazio $U \subset X$, l'immagine di U,

$$\varphi(U) := \{y \in Y \mid \exists\, x \in U : \varphi(x) = y\},$$

è un sottospazio vettoriale di Y, e che per ogni sottospazio $V \subset Y$ la controimmagine di V,

$$\varphi^{-1}(V) := \{x \in X \mid \varphi(v) \in V\},$$

è un sottospazio vettoriale di X. In particolare

$$\operatorname{Im}\varphi := \varphi(X), \qquad \ker\varphi := \varphi^{-1}(0) = \{x \in X \,|\, \varphi(x) = 0\},$$

detti rispettivamente *immagine di* φ e *nucleo* di φ, sono sottospazi vettoriali rispettivamente di Y e X. Si chiama *rango di* φ, e si indica con $\operatorname{Rank}\varphi$ la dimensione dell'immagine di φ, $\operatorname{Rank}\varphi = \dim \operatorname{Im}\varphi$.

4.17 ¶

Proposition. *Una applicazione lineare $\varphi : X \to Y$ è iniettiva se e solo se $\ker\varphi = \{0\}$.*

4.2.2 Matrice associata ad una applicazione lineare

Siano X, Y due spazi vettoriali di dimensione finita e siano $\mathcal{E} : \mathbb{K}^n \to X$, $\mathcal{F} : \mathbb{K}^m \to Y$ due sistemi di coordinate rispettivamente su X e su Y, e siano $\{e_1, e_2, \dots, e_n\}$, $\{f_1, f_2, \dots, f_m\}$ le relative basi. Ogni applicazione lineare $\ell : X \to Y$ induce un'applicazione lineare $L : \mathbb{K}^n \to \mathbb{K}^m$ definita da

$$L := \mathcal{F}^{-1} \circ \ell \circ \mathcal{E}, \tag{4.1}$$

o, se si vuole dal *diagramma commutativo*

$$\begin{array}{ccc} X & \xrightarrow{\ell} & Y \\ \uparrow \mathcal{E} & & \uparrow \mathcal{F}. \\ \mathbb{K}^n & \xrightarrow{L} & \mathbb{K}^m \end{array}$$

L manda le coordinate $\mathbf{x} \in \mathbb{K}^n$ di un vettore $x \in X$ nelle coordinate $\mathbf{y} \in \mathbb{K}^m$ del vettore $\ell(x) \in Y$. Se rappresentiamo L con la sua matrice associata $L \in M_{m,n}(\mathbb{K})$, allora

$$\mathbf{y} = L\mathbf{x}, \tag{4.2}$$

righe per colonne poiché da $\mathcal{E}(\mathbf{x}) = x$, $\mathcal{F}(\mathbf{y}) = \ell(x)$ segue che $\mathbf{y} = \mathcal{F}^{-1}(\ell(x)) = \mathcal{F}^{-1} \circ \ell \circ \mathcal{E}(\mathbf{x})$.

Un altro modo di leggere la (4.2) è dire che la colonna j-esima di L contiene le coordinate di $\ell(e_j)$ nella base $\{f_1, f_2, \dots, f_m\}$

$$\ell(e_j) = \sum_{i=1}^{n} L^i_j f_i. \tag{4.3}$$

Naturalmente la corrispondenza tra applicazioni lineari da X su Y e matrici dipende dalla scelta delle basi su X e Y. Quando è opportuna una notazione che specifichi le basi utilizzate, si scriverà

$$M^{\mathcal{F}}_{\mathcal{E}}(\ell)$$

per indicare la matrice associata ad ℓ nelle basi $\mathcal{E}$ e $\mathcal{F}$.

Essendo i sistemi di coordinate $\mathcal{E}$ e $\mathcal{F}$ isomorfismi, $\dim \ker L = \dim \ker \ell$ e $\dim \operatorname{Im} L = \dim \operatorname{Im} \ell$. La formula del rango in $\mathbb{K}^n$ dà allora

4.18 Proposizione (Formula del rango). *Se $\ell : X \to Y$ è lineare e X ha dimensione finita, allora*

$$\operatorname{Rank} \ell = \dim \operatorname{Im} \ell = \dim X - \dim \ker \ell.$$

4.2.3 Equazioni lineari

Siano X, Y spazi vettoriali sullo stesso corpo $\mathbb{K}$ e $\varphi : X \to Y$ un'applicazione lineare. Si chiama *equazione lineare* nell'incognita x una equazione del tipo

$$\varphi(x) = y, \tag{4.4}$$

dove y è un elemento di Y. L'equazione $\varphi(x) = 0$ si chiama l'*equazione omogenea associata* alla (4.4). Come abbiamo già visto nel Vol. I nel caso delle equazioni differenzali lineari, una conseguenza della linearità è il *principio di sovrapposizione*

4.19 Proposizione (Principio di sovrapposizione). *Si ha*

(i) *le soluzioni dell'equazione omogenea $\varphi(x) = 0$, i.e., gli elementi del nucleo di φ, sono uno spazio vettoriale,*

(ii) *Se $\overline{x}$ è una soluzione dell'equazione $\varphi(x) = y$, tutte le eventuali altre soluzioni di $\varphi(x) = y$ si ottengono aggiungendo ad $\overline{x}$ una qualunque soluzione dell'equazione omogenea associata $\varphi(x) = 0$, i.e., l'insieme delle soluzioni di $\varphi(x) = y$ è dato da*

$$\overline{x} + \ker \varphi := \{x = \overline{x} + z \,|\, z \in \ker \varphi\}.$$

Ovviamente l'equazione $\varphi(x) = y$ *ha una o più soluzioni se e solo se* $y \in \varphi(X)$, *e ha al più una soluzione se e solo se* $\ker \varphi = \{0\}$.

Si ricava quindi dalla formula del rango nella Proposizione 4.18

4.20 Corollario. *Siano* X, Y *spazi vettoriali di uguale dimensione finita, e sia* $\varphi : X \to Y$ *un'applicazione lineare. Allora* φ *è iniettiva se e solo se è surgettiva, i.e.,* $\varphi(x) = y$ *ha soluzione per ogni* $y \in Y$ *se e solo se* $\varphi(x) = 0$ *ha* 0 *come unica soluzione.*

Il Corollario 4.20 è una delle forme del *teorema dell'alternativa di Fredholm*: *o l'equazione non omogenea* $\varphi(x) = y$ *è risolubile per ogni secondo membro* y *o l'equazione omogenea* $\varphi(x) = 0$ *ha una soluzione non banale.*

4.3 Cambiamenti di coordinate

Le coordinate di un vettore come pure la matrice associata ad una applicazione lineare dipendono dalla scelta delle basi. Vediamo come cambiano cambiando base.

4.3.1 Cambiamento di coordinate in uno spazio vettoriale

Sia X uno spazio vettoriale di dimensione finita e siano $\{e_1, e_2, \dots, e_n\}$ una base di X, e $\{\epsilon_1, \epsilon_2, \dots, \epsilon_n\}$ un'altra base di X. Siano poi $\mathcal{E}, \mathcal{F} : \mathbb{K}^n \to X$ i rispettivi sistemi di coordinate

4.21 Definizione. *La mappa lineare* $L : \mathbb{K}^n \to \mathbb{K}^n$ *definita da* $L := \mathcal{F}^{-1} \circ \mathcal{E}$, *o se si vuole dal diagramma commutativo*

$$\begin{array}{ccc} X & \xrightarrow{\mathrm{Id}} & X \\ \uparrow \mathcal{E} & & \uparrow \mathcal{F} \\ \mathbb{K}^n & \xrightarrow{L} & \mathbb{K}^n \end{array}$$

e la matrice associata L, *definita da*

$$\epsilon_j = \sum_{i=1}^n L^i_j e_i, \tag{4.5}$$

sono rispettivamente l'applicazione e la matrice di cambiamento di base da $\mathcal{E}$ a $\mathcal{F}$.

Se ora $\mathbf{x}, \mathbf{y}$ sono le coordinate di un vettore $x \in X$ rispettivamente nelle basi $\mathcal{E}$ e $\mathcal{F}$, si ha, cfr. la (4.1)

$$\mathbf{x} = L\mathbf{y} \tag{4.6}$$

righe per colonne. In altre parole L manda le coordinate di un punto relative a $\mathcal{F}$ nelle coordinate dello stesso punto relative a $\mathcal{E}$, equivalentemente, L manda la base iniziale nella nuova e le coordinate nuove in quelle iniziali. Per questo si dice che *le coordinate dei vettori cambiano in modo controvariante*. Questo comportamento non è poi così misterioso. Se l'unità di misura ϵ_1 è 100 volte e_1, $\epsilon_1 := 100e_1$, ci si aspetta che la coordinata ξ^1 di un oggetto x rispetto a ϵ_1 sia $1/100$ della coordinata x^1 di x rispetto a e_1.

4.3.2 Cambiamenti di coordinate e applicazioni lineari

Uno spazio vettoriale di dimensione finita n si identifica dunque con le n-ple di numeri in $\mathbb{K}^n$ fissando una base $\{e_1, e_2, \dots, e_n\}$ di X e facendo corrispondere ad ogni vettore la n-upla delle sue coordinate, che si conviene di indicizzare con un indice in alto, e di ordinare, se necessario, in un vettore colonna.

Ogni funzione lineare definita sullo spazio vettoriale X (geometrico o fisico) può dunque essere vista come funzione delle coordinate; nel fare questo siamo liberi di scegliere il nostro sistema di riferimento: è quindi utile capire cosa succede alle differenti rappresentazioni in coordinate di una funzione su X cambiando coordinate su X. Non bisogna confondere una funzione su X, che è sempre la stessa, con la sua rappresentazione in un sistema di coordinate che varia da sistema a sistema. La proposizione seguente in Proposizione **??** esprime appunto il legame esistente fra le matrici associate ad una applicazione lineare su X in basi differenti.

Sia X di dimensione finita, siano $\{e_1, e_2, \dots, e_n\}$ e $\{\epsilon_1, \epsilon_2, \dots, \epsilon_n\}$ due basi su X e $\mathcal{E}, \mathcal{E}' : \mathbb{K}^n \to X$ i rispettivi sistemi di coordinate. Sia $R : X \to X$ l'applicazione lineare e la relativa matrice associata relative al cambiamento di base da $\mathcal{E}$ a $\mathcal{E}'$. Siano Y un altro spazio di dimensione finita, $\{f_1, f_2, \dots, f_m\}$ e $\{\varphi_1, \varphi_2, \dots, \varphi_m\}$ due basi di Y, $\mathcal{F}, \mathcal{F}' : \mathbb{K}^m \to Y$ i rispettivi sistemi di coordinate, e $S : Y \to Y$ l'applicazione lineare e la relativa matrice associata al cambiamento di base da $\mathcal{F}$ a $\mathcal{F}'$.

Se ora $\ell : X \to Y$ è un'applicazione lineare, indichiamo con A la matrice associata ad ℓ nelle basi $\mathcal{E}$ e $\mathcal{F}$ e con B la matrice associata a ℓ nelle basi $\mathcal{E}'$ e $\mathcal{F}'$. Allora $B = SAR^{-1}$. Infatti il diagramma

$$\begin{array}{ccccccc}
X & \overset{id}{\to} & X & \overset{\ell}{\to} & Y & \overset{id}{\to} & Y \\
\uparrow \mathcal{E}' & & \uparrow \mathcal{E} & & \uparrow \mathcal{F} & & \uparrow \mathcal{F}' \\
\mathbb{K}^n & \overset{R^{-1}}{\to} & \mathbb{K}^n & \overset{A}{\to} & \mathbb{K}^m & \overset{S}{\to} & \mathbb{K}^m
\end{array}$$

è commutativo. L'applicazione lineare associata a ℓ nelle basi $\mathcal{E}'$ e $\mathcal{F}$ è la seconda riga del diagramma, i.e. $B := S \circ A \circ R^{-1}$, o, in termini di matrici, $B = S^{-1}AR$ righe per colonne.

Più in generale si ha

4.22 Teorema. *Due matrici A, B rappresentano la stessa applicazione lineare da X in Y in basi diverse se esistono una matrice R $n \times n$ e una matrice S $m \times m$ tali che $B = SAR^{-1}$.*

4.23 ¶ Scrivere alcune matrici 3×3, interpretarle come applicazioni lineari da $\mathbb{R}^3$ su $\mathbb{R}^3$. Per ciascuna applicazione fissare una nuova base in $\mathbb{R}^3$ e calcolare nella nuova base la matrice associata all'applicazione.

4.24 ¶ Siano $V_1, V_2, \dots V_n$ spazi vettoriali su $\mathbb{K}$ e $f_0, f_1, \dots, f_n$ applicazioni lineari concatenati nel senso che

$$\{0\} \overset{f_0}{\to} V_1 \overset{f_1}{\to} V_2 \overset{f_3}{\to} \dots \overset{f_{n-2}}{\to} V_{n-1} \overset{f_{n-1}}{\to} V_n \overset{f_n}{\to} \{0\}.$$

Mostrare che se $\operatorname{Im}(f_i) = \ker(f_{i+1})$ $\forall i = 0, \dots, n-1$, allora $\sum_{i=1}^n (-1)^i \dim V_i = 0$.

4.25 ¶ Siano X, Y, Z spazi vettoriali su $\mathbb{K}$ e siano $f : X \to Y$, $g : X \to Z$ applicazioni lineari. Provare che $\ker f \subset \ker g$ se e solo se esiste una applicazione lineare $\ell : Y \to Z$ tale che $g := \ell \circ f$.

4.4 Struttura euclidea

La struttura lineare di uno spazio vettoriale si arricchisce dotandolo di un *prodotto scalare.*

4.4.1 La struttura metrica di $\mathbb{R}^n$

Dati due vettori in $\mathbb{R}^n$, si misurano le rispettive lunghezze e l'angolo fra essi mediante il *prodotto scalare.*

a. Prodotto scalare

Se $\mathbf{x} = (x_1, x_2, \dots, x_n)$ e $\mathbf{y} = (y_1, y_2, \dots, y_n)$ sono vettori in $\mathbb{R}^n$, il *prodotto scalare* di $\mathbf{x}$ e $\mathbf{y}$ è definito come il numero reale

$$(\mathbf{x}|\mathbf{y}) := \sum_{i=1}^{n} x_i y_i.$$

Due vettori $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ si dicono *ortogonali* se $(\mathbf{x}|\mathbf{y}) = 0$. Il prodotto scalare ha le seguenti proprietà

- $(\mathbf{x}+\mathbf{y}|\mathbf{z}) = (\mathbf{x}|\mathbf{z}) + (\mathbf{y}|\mathbf{z})$,
- $(\lambda\mathbf{x}|\mathbf{y}) = (\mathbf{x}|\lambda\mathbf{y}) = \lambda(\mathbf{x}|\mathbf{y})$,
- $(\mathbf{x}|\mathbf{z}) = (\mathbf{z}|\mathbf{x})$,
- $(\mathbf{x}|\mathbf{x}) \geq 0$ e $(\mathbf{x}|\mathbf{x}) = 0$ se e solo se $\mathbf{x} = 0$,

per ogni $\mathbf{x},\ \mathbf{y},\ \mathbf{z} \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$. Se $\mathbf{x} \in \mathbb{R}^n$, il numero reale non negativo

$$|\mathbf{x}| := \sqrt{(\mathbf{x}|\mathbf{x})} = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2},$$

si chiama *norma* di $\mathbf{x} \in \mathbb{R}^n$. Si ha

$$|\mathbf{x}-\mathbf{y}|^2 = (\mathbf{x}-\mathbf{y}|\mathbf{x}-\mathbf{y}) = \dots = |\mathbf{x}|^2 + |\mathbf{y}|^2 - 2(\mathbf{x}|\mathbf{y}),$$

in particolare vale il

4.26 Teorema (Pitagora). $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *sono ortogonali se e solo se* $|\mathbf{x}+\mathbf{y}|^2 = |\mathbf{x}|^2 + |\mathbf{y}|^2$.

4.27 Proposizione (Diseguaglianza di Cauchy-Schwarz). *Per ogni* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *si ha* $|(\mathbf{x}|\mathbf{y})| \leq |\mathbf{x}|\ |\mathbf{y}|$, *e* $(\mathbf{x}|\mathbf{y}) = |\mathbf{x}||\mathbf{y}|$ *se e solo se o* $\mathbf{y} = 0$, *o* $\mathbf{x} = \lambda\mathbf{y}$, $\lambda \geq 0$.

Dimostrazione. Se $\mathbf{y} = 0$, la tesi è ovvia. Per $\mathbf{y} \neq 0$, l'espressione $|\mathbf{x}+t\mathbf{y}|^2$ è un polinomio di secondo grado in t,

$$(\mathbf{x}+t\mathbf{y}|\mathbf{x}+t\mathbf{y}) = (\mathbf{x}+t\mathbf{y}|\mathbf{x}) + (\mathbf{x}+t\mathbf{y}|t\mathbf{y}) = |\mathbf{x}|^2 + 2(\mathbf{x}|\mathbf{y})t + |\mathbf{y}|^2 t^2,$$

non negativo; perciò il suo discriminante $(\mathbf{x}|\mathbf{y})^2 - |\mathbf{x}|^2|\mathbf{y}|^2$ è non positivo.

Infine, se $(\mathbf{x}|\mathbf{y}) = |\mathbf{x}|\ |\mathbf{y}|$, allora il discriminante del polinomio $t \to |\mathbf{x}+t\mathbf{y}|^2$ è nullo. Perciò per qualche $t \in \mathbb{R}$ $|\mathbf{x}+t\mathbf{y}|^2 = 0$ e dunque $\mathbf{x} = -t\mathbf{y}$, Che $-t$ sia non negativo segue infine dalla $-t = (\mathbf{x}|\mathbf{y}) = |\mathbf{x}|\ |\mathbf{y}| \geq 0$. □

Se $\mathbf{x}, \mathbf{y}$ sono vettori di $\mathbb{R}^2$, pensato come piano della geometria euclidea, e θ è l'angolo fra di essi, si vede facilmente che $(\mathbf{x}|\mathbf{y}) = |\mathbf{x}|\,|\mathbf{y}|\cos\theta$. In generale, essendo $\frac{|(\mathbf{x}|\mathbf{y})|}{|\mathbf{x}|\ |\mathbf{y}|} \leq 1$ per ogni $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, esiste $\theta \in \mathbb{R}$ tale che

$$\cos\theta := \frac{(\mathbf{x}|\mathbf{y})}{|\mathbf{x}|\,|\mathbf{y}|},$$

e θ si chiama *angolo tra i vettori* $\mathbf{x}$ *e* $\mathbf{y}$ e spesso si indica con $\widehat{\mathbf{x}\mathbf{y}}$. In particolare $\mathbf{x}, \mathbf{y}$ sono perpendicolari se e solo se $\theta = \pi/2 + k\pi$, $k \in \mathbb{Z}$, e vale il *teorema di Carnot*

$$|\mathbf{x}-\mathbf{y}|^2 = |\mathbf{x}|^2 + |\mathbf{y}|^2 - 2|\mathbf{x}|\,|\mathbf{y}|\cos\theta.$$

Si può quindi misurare l'angolo tra due vettori di $\mathbb{R}^n$ con tre misure di lunghezza.

b. Norma

La norma $|\mathbf{x}| = \sqrt{(x|x)}$ ha le seguenti proprietà

- $|\mathbf{x}| = 0$ se e solo se $\mathbf{x} = 0$,
- $|\lambda\mathbf{x}| = |\lambda|\ |\mathbf{x}|\ \forall\lambda \in \mathbb{R}, \forall\mathbf{x} \in \mathbb{R}^n$,
- (Diseguaglianza triangolare) $|\mathbf{x}+\mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|\ \forall\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

La diseguaglianza triangolare segue dalla diseguaglianza di Cauchy-Schwarz. Infatti

$$|\mathbf{x}+\mathbf{y}|^2 = |\mathbf{x}|^2 + |\mathbf{y}|^2 + 2(\mathbf{x}|\mathbf{y}) \leq |\mathbf{x}|^2 + |\mathbf{y}|^2 + 2|\mathbf{x}|\ |\mathbf{y}| = (|\mathbf{x}| + |\mathbf{y}|)^2.$$

c. Distanza

La *distanza euclidea* tra due punti di $\mathbb{R}^n$ è definita da $d(\mathbf{x},\mathbf{y}) := |\mathbf{x}-\mathbf{y}|$. Dalle proprietà della norma segue che la distanza tra due punti di $\mathbb{R}^n$ è un numero reale non negativo e che

- $d(\mathbf{x},\mathbf{y}) = 0$ se e solo se $\mathbf{x} = \mathbf{y}$,
- (SIMMETRIA) $d(\mathbf{x},\mathbf{y}) = d(\mathbf{y},\mathbf{x})$,
- (DISEGUAGLIANZA TRIANGOLARE) $d(\mathbf{x},\mathbf{y}) \leq d(\mathbf{x},\mathbf{z}) + d(\mathbf{z},\mathbf{x})$.

4.4.2 Spazi euclidei

a. Prodotti scalari

4.28 Definizione. *Sia X uno spazio vettoriale su $\mathbb{R}$. Si chiama* prodotto scalare *su X un'applicazione $(\ |\ ) : X \times X \to \mathbb{R}$,*

(i) (SIMMETRICA) $(x|y) = (y|x)$,
(ii) (BILINEARE) $(\lambda x + \mu y|z) = \lambda(x|z) + \mu(y|z)$, $(z|\lambda x + \mu y) = \lambda(z|x) + \mu(z|y)$,
(iii) (DEFINITA POSITIVA) $(x|x) > 0$ *se* $x \neq 0$.

per ogni $x, y, z \in X$ e $\forall \lambda, \mu \in \mathbb{R}$. Uno spazio euclideo *è uno spazio vettoriale su $\mathbb{R}$ di dimensione finita e con un prodotto scalare.*

4.29 Esempio $\mathbb{R}^n$ con il prodotto scalare *standard* dato da

$$(x|y) := \sum_{i=1}^n x^i y^i, \qquad x = \sum_{i=1}^n x^i e_i,\ y = \sum_{i=1}^n y^i e_i,$$

è il fondamentale modello di spazio euclideo. È anche in un certo senso l'unico spazio euclideo di dimensione n, cfr. l'esercizio in Esercizio 4.38.

4.30 Esempio Altri esempi si ottengono ad esempio pesando opportunamente le coordinate. Se $\lambda_1, \lambda_2, \dots, \lambda_n$ sono numeri positivi, la forma bilineare su $\mathbb{R}^n$ $b(x,y) := \sum_{i=1}^n \lambda_i x^i y^i$ è un prodotto scalare su $\mathbb{R}^n$.

4.31 Esempio Sia $\mathcal{M}_2$ lo spazio vettoriale delle variabili aleatorie discrete su uno spazio X con momento di ordine due finito. La *covarianza* di due variabili aleatorie, $\mathrm{Cov}\,[X,Y]$, è un prodotto scalare su $\mathcal{M}_2$, cfr. Sez. 2, Cap. 9, Vol. II.

In uno spazio X con prodotto scalare $(\ |\ )$, si possono misurare le lunghezze dei vettori e l'angolo fra due vettori. Si chiama *norma di $x \in X$* il numero reale non negativo

$$||x|| := \sqrt{(x|x)}.$$

Ripetendo alla lettera i conti fatti nella (**??**), si vede subito che valgono le seguenti formule

(i) (DISEGUALIANZA DI CAUCHY-SCHWARZ) per ogni $x, y \in X$ $|(x|y)| \leq ||x||\,||y||$, e $(x|y) = ||x||\,||y||$ se e solo se o $y = 0$, o $x = \lambda y$ per qualche $\lambda \geq 0$.
(ii) (TEOREMA DI CARNOT) $||x+y||^2 = ||x||^2 + ||y||^2 + 2(x|y)$.

In particolare si calcola il prodotto scalare di due vettori con tre misure di lunghezza,

$$(x|y) = \frac{1}{2}\Big(||x+y||^2 - ||x||^2 - ||y||^2\Big),$$

come in $\mathbb{R}^n$ e vale la *formula del parallelogramma*

$$||x+y||^2 + ||x-y||^2 = 2(||x||^2 + ||y||^2).$$

La norma ha le proprietà

(i) (NON DEGENERAZIONE) $||x|| \in \mathbb{R}_+$ $\forall x \in X$, $||x|| = 0$ se e solo se $x = 0$,
(ii) (1-OMOGENEITÀ) $||\lambda x|| = |\lambda|\,||x||$ $\forall \lambda \in \mathbb{R}$, $\forall x \in X$,
(iii) (PROPRIETÀ TRIANGOLARE) $||x+y|| \leq ||x|| + ||y||$.

b. Vettori ortonormali

Due vettori x, y di uno spazio euclideo si dicono *ortogonali* se $(x|y) = 0$. Dal teorema di Carnot segue che x, y sono ortogonali se e solo se vale per x, y il *teorema di Pitagora*, $||x + y||^2 = ||x||^2 + ||y||^2$.

4.32 Definizione. *Sia X uno spazio vettoriale con prodotto scalare. Un sistema di vettori $\{e_\alpha\} \subset X$ si dice* ortonormale *se $(e_\alpha|e_\beta) = 1$ se $\alpha = \beta$ e $(e_\alpha|e_\beta) = 0$ se $\alpha \neq \beta$, o come si scrive con notazione più compatta se e solo se $(\alpha|\beta) = \delta_{\alpha\beta}$.*

Vettori ortogonali sono linearmente indipendenti. In particolare n vettori ortonormali in uno spazio X di dimensione n costituiscono una base. Una base di vettori ortonormali si dice una *base ortonormale.*

Se $\{e_1, e_2, \dots, e_n\}$ è una base ortonormale di X, i numeri $(x|e_1), \dots, (x|e_n)$ sono detti i *coseni direttori* di x nella base $\{e_1, e_2, \dots, e_n\}$. *Se $\{e_1, e_2, \dots, e_n\}$ è una base ortonormale, allora i coseni direttori di x sono le sue coordinate.* Infatti se $x := \sum_{i=1}^n x^j e_j$, allora

$$(x|e_i) = \sum_{j=1}^n x^j (e_j|e_i) = \sum_{j=1}^n x^j \delta_{ji} = x^i$$

e quindi anche $x = \sum_{j=1}^n (x|e_j)e_j$. Inoltre vale il *teorema di Pitagora* nella forma

$$||x||^2 = (x|x) = \sum_{i,j=1}^n x^i x^j (e_i|e_j) = \sum_{j=1}^n |x^j|^2 = \sum_{j=1}^n |(x|e_j)|^2.$$

c. L'algoritmo di Gram-Schmidt

Se x_1, x_2 sono due vettori linearmente indipendenti, il vettore $x_2' := x_2 - \frac{(x_2|x_1)}{(x_1|x_1)} x_1$, è ortogonale a x_1 e dunque i vettori $y_1 := x_1/||x_1||$, $y_2 := x_2'/||x_2'||$ sono ortonormali e generano lo stesso sottospazio di x_1, x_2.

Si può estendere questa costruzione ad un numero arbitrario di vettori.

4.33 Teorema (Metodo di Gram-Schmidt). *Sia X uno spazio vettoriale con prodotto scalare e siano $v_1, v_2, \dots, v_n, \dots$ vettori linearmente indipendenti. Esistono allora vettori ortonormali $e_1, e_2, \dots, e_n, \dots$ tali che per ogni $k = 1, 2, \dots$*

$$\mathrm{Span}\,\{e_1, e_2, \dots, e_k\} = \mathrm{Span}\,\{v_1, v_2, \dots, v_k\}.$$

Dimostrazione. Procediamo per induzione su k. Se $k = 1$, si sceglie $e_1' := v_1/||v_1||$. Supponiamo di aver provato la base ortogonale per $k-1$ vettori, proviamolo per k. Per l'ipotesi di induzione abbiamo $k-1$ vettori $e_1, e_2, \dots . e_{k-1}$ ortonormali con $\mathrm{Span}\,\{e_1, e_2, \dots, e_{k-1}\} = \mathrm{Span}\,\{v_1, v_2, \dots, v_{k-1}\}$. Il vettore $x' := v_k - \sum_{i=1}^{k-1}(v_k|e_i)e_i$ è ortogonale ai vettori $v_1, v_2, \dots, v_{k-1}$. Perciò se si pone $e_k := x'/||x'||$, i vettori $e_1, e_2, \dots, e_k$ risultano ortonormali e $\mathrm{Span}\,\{e_1, e_2, \dots, e_k\} = \mathrm{Span}\,\{v_1, v_2, \dots, v_k\}$. □

Una conseguenza del metodo di ortonormalizzazione è che se $V \subset X$ è un sottospazio di uno spazio euclideo X, allora V ha una base ortonormale e la si può completare a formare una base ortonormale per X.

4.34 Osservazione. Osserviamo esplicitamente che il processo di ortonormalizzazione funziona anche nel caso di spazi di dimensione infinita, nel caso in cui si parta da una infinità numerabile di vettori. Inoltre conviene notare esplicitamente che il metodo di Gram-Schmidt è un algoritmo. Si calcola iterativamente una base ortogonale $\{e_1', e_2', \dots, e_n'\}$ ponendo $e_1' := v_1$ e per $j = 2, \dots$

$$e_j' := v_j - \sum_{i=1}^{j-1} \frac{(v_j|e_i')}{(e_i'|e_i')} e_i'$$

procedendo simultaneamente alla loro normalizzazione,

$$e_j := \frac{e_j'}{||e_j'||} \qquad \forall j = 1, \dots .$$

d. Prodotto scalare in coordinate

Sia X uno spazio euclideo e $\mathcal{E} = \{e_1, e_2, \dots, e_n\}$ una base di X. La matrice $G := [G_{ij}] \in M_{n,n}(\mathbb{R})$ data da

$$G_{ij} := (e_i|e_j).$$

si chiama *matrice metrica*, o *metrica*, del prodotto scalare nella base $\{e_1, e_2, \dots, e_n\}$.

Se ora $\mathbf{x} = (x^1, x^2, \dots, x^n), \mathbf{y} = (y^1, y^2, \dots, y^n) \in \mathbb{K}^n$ sono le coordinate rispettivamente di x e $y \in X$, allora per linearità

$$(x|y) = \Big(\sum_{i=1}^n x^i e_i \Big| \sum_{j=1}^n y^j e_j\Big) = \sum_{i,j=1}^n x^i y^j (e_i|e_j) = \mathbf{y}^T G \mathbf{x}$$

righe per colonne, se $\mathbf{x}$ e $\mathbf{y}$ sono stati listati in colonna. In altri termini $\mathbf{y}^T G\mathbf{x}$ è il prodotto scalare $(x|y)$ in coordinate.

La matrice metrica G è evidentemente simmetrica. G è anche invertibile, $\det G \neq 0$: se per assurdo non lo fosse, $\mathbf{x} \to G\mathbf{x}$ non sarebbe iniettiva. Esisterebbe allora $\mathbf{x} \neq 0$ per cui $G\mathbf{x} = 0$; se $x := \sum_{i=1}^n x^i e_i \in X$, si avrebbe quindi $x \neq 0$ e $(x|x) = \mathbf{x}^T G\mathbf{x} = 0$, una contraddizione.

Si noti che la base $\{e_1, e_2, \dots, e_n\}$ è ortonormale in X se e solo se la metrica nella base $\{e_1, e_2, \dots, e_n\}$ è la matrice identica. In questo caso il prodotto scalare indotto sulle coordinate è proprio il prodotto scalare usuale di $\mathbb{R}^n$.

Vediamo ora come cambia la matrice metrica associata ad un prodotto scalare al cambiare della base. Siano $\mathcal{E} := \{e_1, e_2, \dots, e_n\}$ e $\mathcal{F} := \{f_1, f_2, \dots, f_n\}$ due basi su uno spazio euclideo e $R \in M_{n,n}(\mathbb{R})$ la matrice di cambiamento di base da $\mathcal{E}$ a $\mathcal{F}$, i.e., definita da $f_j := \sum_{j=1}^n L^i_j e_i \ \forall j = 1, n$, cfr. Definizione 4.21. Siano G e H le metriche associate al prodotto scalare rispettivamente nelle basi $\mathcal{E}$ e $\mathcal{F}$. Allora non è difficile vedere che si ha

4.35 Proposizione. *$H = R^T G R$, righe per colonne.*

e. Applicazioni isometriche

Una applicazione $\ell : X \to Y$ tra spazi euclidei con prodotti scalari $(\ |\ )_X$, $(\ |\ )_Y$, si dice *isometrica* se conserva i prodotti scalari, i.e.,

$$(\ell(x)|\ell(y))_Y = (x|y)_X \qquad \forall x, y \in X.$$

Ogni applicazione isometrica è ovviamente iniettiva. Se una applicazione isometrica è anche surgettiva, si dice una *isometria*.

Si verifica facilmente

4.36 Proposizione. *Siano X uno spazio euclideo, $\mathcal{E} := \{e_1, e_2, \dots, e_n\}$ una base ortonormale di X. Sia $\mathcal{F} = \{f_1, f_2, \dots, f_n\}$ un'altra base di X, e $S \in M_{n,n}(\mathbb{R})$ la matrice che cambia base da $\mathcal{E}$ a $\mathcal{F}$, i.e., definita da $f_j = \sum_{i=1}^n S^i_j e_i$, $j = 1, \dots, n$. Sono fatti equivalenti*

(i) *$f_1, f_2, \dots, f_n$ sono ortonormali,*
(ii) *la mappa lineare $s : X \to X$, $s(e_i) := f_i$, $i = 1, \dots, n$, è una isometria,*
(iii) *le colonne di S sono ortonormali in $\mathbb{R}^n$,*
(iv) *$S^T S = \mathrm{Id}_n$,*
(v) *le righe di S sono ortonormali in $\mathbb{R}^n$.*

A parole, fissata una base ortonormale, c'è dunque un isomorfismo di X con $\mathbb{R}^n$ e quindi fra le isometrie di X in sé e il *gruppo ortogonale*

$$O(n) := \{S \in M_{n,n}(\mathbb{R}) \,|\, S^T S = \mathrm{Id}_n\}.$$

4.37 ¶ Il sistema di coordinate $\mathcal{E} : \mathbb{R}^n \to X$ individuato da una base ortonormale è un esempio di isometria tra $\mathbb{R}^n$ con il prodotto scalare euclideo e X. Più in generale, un'applicazione $\ell : X \to Y$ tra spazi euclidei è una applicazione isometrica se e solo se manda una base ortonormale in un sistema ortonormale di vettori. Mostrare che se tra X e Y vi è una applicazione isometrica, allora $\dim X \leq \dim Y$. Mostrare che ℓ è isometrica se e solo se la matrice associata L in basi ortonormali ha la proprietà $L^T L = \mathrm{Id}_n$, $n = \dim X$.

4.38 ¶ Se X è uno spazio euclideo di dimensione n, un sistema di coordinate $\ell : X \to \mathbb{R}^n$ fissato da una base ortonormale è una isometria su $\mathbb{R}^n$. Dedurne che $\mathbb{R}^n$ è l'unico spazio euclideo di dimensione n a meno di isometrie.

4.39 ¶ Mostrare che

Theorem. *Un'applicazione $\ell : X \to Y$ tra spazi reali con prodotto scalare è una mappa lineare isometrica se e solo $\ell(0) = 0$ e conserva le lunghezze, $||\ell(x) - \ell(y)||_Y = ||x - y||_X$ per ogni $x, y \in X$.*

f. Il teorema dell'alternativa

Sia $V \subset X$ un sottospazio di uno spazio euclideo X con prodotto scalare $(\ |\ )$. L'*ortogonale* a V è definito da
$$V^\perp := \{x \in X \,|\, (x|y) = 0 \ \forall y \in V\}.$$

4.40 Proposizione. *$V^\perp$ è un sottospazio vettoriale di X, V e $V^\perp$ sono supplementari in X, e $(V^\perp)^\perp = V$.*

Sia $\ell : X \to Y$ è una applicazione lineare tra spazi euclidei con prodotti scalari $(\ |\ )_X, (\ |\ )_Y$. La *trasposta* o *aggiunta* di ℓ è l'unica applicazione $\ell^* : Y \to X$ definita da
$$(x|\ell^*(y))_X := (\ell(x)|y)_Y, \qquad \forall x \in X,\ \forall y \in Y.$$
In coordinate, se G, H sono le matrici metriche associate ai prodotti scalari rispettivamente in X e Y, e L, M sono le matrici associate ad ℓ, ℓ^*, si ha
$$\mathbf{x}^T GM\mathbf{y} = (x|\ell^*(y))_X = (\ell(x)|y)_Y = (L\mathbf{x})^T H\mathbf{y}, \qquad \forall \mathbf{x}, \mathbf{y}$$
e quindi
$$M = G^{-1} L^T H, \tag{4.7}$$
In particolare, *se si sono scelte coordinate ortonormali su X e Y, la matrice associata a ℓ^* è L^T.*

Si ha poi

4.41 Proposizione. *Siano X, Y spazi euclidei e $\varphi : X \to Y$ una mappa lineare. Allora $(\operatorname{Im}\varphi)^\perp = \ker\varphi^*$, $\operatorname{Im}\varphi = (\ker\varphi^*)^\perp$, e, essendo $\varphi^{**} = \varphi$, anche $(\operatorname{Im}\varphi^*)^\perp = \ker\varphi$, $\operatorname{Im}\varphi^* = (\ker\varphi)^\perp$. Inoltre $\dim\operatorname{Im}\varphi = \dim\operatorname{Im}\varphi^*$.*

Dimostrazione. Si ha $(\varphi^*(y)|z) = (y|\varphi(z))$, $\forall z \in X$, $\forall y \in Y$. Perciò $y \in (\operatorname{Im}\varphi)^\perp$ se e solo se $y \in \ker\varphi^*$, il che prova che $(\operatorname{Im}\varphi)^\perp = \ker\varphi^*$ e passando ai supplementari, $\operatorname{Im}\varphi = (\ker\varphi^*)^\perp$. L'ultima parte della tesi segue dalla formula di Grassmann e dalla formula del rango,
$$\dim\operatorname{Im}\varphi = n - \dim\ker\varphi = \dim\ker\varphi^\perp = \dim\operatorname{Im}\varphi^*.$$
□

Come conseguenza,

4.42 Teorema (dell'alternativa). *Siano X e Y spazi euclidei, ed $\phi : X \to Y$ lineare. Allora $\phi(x) = y$ ha almeno una soluzione se e solo se y è perpendicolare a $\ker\phi^*$.*

4.43 ¶ Provare che
- (i) $(\ell \circ \varphi)^* = \varphi^* \circ \ell^*$, $(\varphi + \psi)^* = \varphi^* + \psi^*$, $(\lambda\varphi)^* = \lambda\varphi^*$, $\varphi^{**} = \varphi$,
- (ii) se φ è invertibile, allora $(\varphi^{-1})^* = (\varphi^*)^{-1}$,
- (iii) se φ è una isometria, allora anche φ^* è una isometria.

4.44 ¶ Sia $A \in M_{m,n}(\mathbb{R})$. Provare che le righe di A generano $\ker A^\perp$ e dedurne la proposizione in Proposizione 4.41.

g. Proiezioni ortogonali

4.45 Definizione. *Siano X uno spazio con prodotto scalare $(\ |\ )$, anche di dimensione infinita, $V \subset X$ un sottospazio di dimensione finita n, e $\{e_1, e_2, \dots, e_n\}$ una base ortonormale per V. Per ogni $x \in X$ si chiama* proiezione ortogonale di x su V *il vettore*

$$P(x) := \sum_{j=1}^{n} (x|e_j)e_j$$

e si chiama proiezione ortogonale *su V la mappa $P : X \to X$ così definita.*

Evidentemente $\mathrm{Im}\, P = V$, $||P(z)||^2 = \sum_{j=1}^{n} |(z|e_j)|^2$ e $P^2(z) = P(z)$. Si ha inoltre

4.46 Teorema (della proiezione ortogonale). *Siano X uno spazio con prodotto scalare $(\ |\ )$, $V \subset X$ un sottospazio di dimensione finita n, $e_1, e_2, \dots, e_n$ una base ortonormale per V e $x \in X$. Sono fatti equivalenti*

(i) *$z \in V$ è la proiezione ortogonale di x su V, $z = P(x)$,*
(ii) *$x - z$ è perpendicolare a V, i.e. $(x - z|w) = 0\ \forall w \in V$,*
(iii) *z è l'unico punto di minima distanza di x da V, i.e. $||z - x|| < ||w - x||$ per ogni $w \in V$, $w \neq z$.*

4.47 ¶ Se $P : X \to X$ è la proiezione ortogonale su V, allora $\mathrm{Id} - P : X \to X$ è la proiezione ortogonale su $V^\perp$.

4.48 ¶ Sia $P : X \to X$ un'applicazione lineare tra spazi euclidei e $V := \mathrm{Im}\,(P)$. Mostrare che P è la proiezione ortogonale su V se e solo se $P^2 = P$ e $P^T = P$.

4.49 ¶ Provare i teoremi non dimostrati in questa sottosezione.

4.4.3 Minimi quadrati

a. Regressione lineare

Supponiamo che una variabile y dipenda linearmente dalla variabile x

$$y = \alpha + \beta x$$

e che si voglia trovare i coefficienti α, β con esperimenti. Si procede quindi fissando alcuni valori certi $x_1, x_2, \dots x_n$ per la variabile x e effettuando misure delle corrispondenti quantità $y_1, \dots y_n$ della variabile y. Il problema è quello di ritrovare i *parametri attesi* α e β a partire dalle misure. In generale però queste misure sono affette da una incertezza, sono cioè variabili aleatorie. Non sarà dunque possibile ritrovare i valori veri α, β, ma solo trovare delle stime per essi, i migliori α, β possibili sulla base di un qualche criterio. Più precisamente ogni criterio di scelta fornirà in generale a partire dalle variabili aleatorie $y_1, y_2, \dots, y_n$ due distribuzioni di probabilità a, b ragionevolmente concentrate attorno ai *valori attesi* o veri α, β. Le distribuzioni α, β sono detti *stimatori* per i valori attesi, rispettivamente α e β, prodotti dal *predittore* scelto.

Come si è detto, il criterio di selezione, il *predittore*, non è evidentemente fissato a priori e va scelto sulla base delle necessità. Un modo classico, che produce distribuzioni di probabilità a e b per i coefficienti con buone proprietà statistiche, è quello di scegliere, fissata una serie di misure $y_1, y_2, \dots, y_n$, numeri $a, b \in \mathbb{R}$ in modo da minimizzare lo *scarto quadratico* tra i valori misurati y_i e quelli attesi $a + bx_i$, i.e., la quantità

$$S^2 := \sum_{i=1}^{n} (y_i - a - bx_i)^2.$$

al variare di $a, b \in \mathbb{R}$. È il cosidetto *metodo dei minimi quadrati* o della *regressione lineare*.

Trattandosi di un problema di minimizzazione per una funzione quadratica, è possibile ricondurre questo problema ad un problema di proiezione ortogonale nel modo seguente.

Siano dunque $x_1, x_2, \dots x_n \in \mathbb{R}$ assegnati, e $y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$ i valori misurati. Sia

$$V := \{y = (z_1, z_2, \dots, z_n) \in \mathbb{R}^n \mid z_i = a + bx_i \ \forall i = 1, n, \ \forall a, b \in \mathbb{R}\}$$

l'insieme dei valori attesi al variare di $a, b \in \mathbb{R}$. Si verifica che V è un sottospazio affine di $\mathbb{R}^n$ e se $z = (z_1, z_2, \dots, z_n) \in V$, allora $\forall i = 1, n$ $z_i = a + bx_i$ e $|y - z|^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 = S^2$. Perciò minimizzare S^2 corrisponde a minimizzare la distanza di $y \in \mathbb{R}^n$ da V. Il punto di minimo esiste dunque ed è la proiezione perpendicolare di y su V, i.e., il punto $\widehat{y} = (\widehat{y}_1, \widehat{y}_2, \dots, \widehat{y}_n)$, $\widehat{y}_i := a + bx_i$, tale che

$$\sum_{i=1}^n (y_i - \widehat{y}_i)(\lambda + \mu x_i) = 0 \qquad \forall \lambda, \mu \in \mathbb{R},$$

vale a dire

$$\begin{cases} \sum_{i=1}^n (y_i - \widehat{y}_i) = 0 \\ \sum_{i=1}^n x_i (y_i - \widehat{y}_i) = 0. \end{cases}$$

Ricordando che $\widehat{y}_i = a + bx_i$, si trova infine che

$$\begin{cases} b := \dfrac{\sum_{i=1}^n x_i (y_i - \overline{y})}{\sum_{i=1}^n x_i (x_i - \overline{x})} \\ a := \overline{y} - b_1 \overline{x} \end{cases},$$

dove $\overline{x} := \frac{1}{n} \sum_{i=1}^n x_i$, $\overline{y} := \frac{1}{n} \sum_{i=1}^n y_i$, sono le medie rispettivamente di $(x_1, x_2, \dots, x_n)$ e $(y_1, y_2, \dots, y_n)$.

Supponendo poi che le misure siano variabili aleatorie della forma

$$y_i := \alpha + \beta x_i + w_i$$

dove le incertezze attorno ai valori attesi sono rappresentate da distribuzioni di probabilità indipendenti con legge normale $N(0, \sigma^2)$ con $\sigma > 0$ incognito ma uguale per tutte le misure, si può dimostrare che le distribuzioni stimate, gli *stimatori di α e β* hanno anch'essi distribuzione normale, e precisamente

$$a \text{ ha legge } N(\alpha, \sigma^2 \frac{1/n + \overline{x}^2}{\sigma_x^2})$$

$$b \text{ ha legge } N(\beta_0, \frac{\sigma^2}{\sigma_x^2}),$$

dove $\sigma_x^2 := \sum_{i=1}^n (x_i - \overline{x})^2$.

b. Distanza da un iperpiano

Supponiamo di avere N dati sperimentali provenienti da un esperimento e di aver costruito un modello matematico dell'esperimento dipendente da $n << N$ parametri liberi. Il problema e' quello di determinare i parametri $x^1, x^2, \dots, x^n$ in modo che i dati calcolati dal modello siano i piu' vicini possibili a quello dell'esperimento.

Si puo' matematizzare la situazione raggruppando i dati sperimentali e i dati calcolati dal modello come vettori di uno spazio vettoriale V di dimensione N. I dati sperimentali sono ora un vettore $b \in \mathbb{R}^N$.

4.50 (Minimi quadrati lineare) Se il modello è funzione lineare dei parametri da minimizzare, il modello è una mappa $A : \mathbb{R}^n \to V$ che associa ai valori dei parametri $x^1, x^2, \dots, x^n$ un vettore $Ax \in V$ di dati calcolati. Se si introduce una norma $|| \ ||$ in V, e il problema diventa quello di trovare il parametro $x \in \mathbb{R}^n$ in modo da minimizzare la funzione

$$x \to ||Ax - b|| \tag{4.8}$$

Il problema ha una soluzione particolarmente semplice ed elegante quando la norma scelta deriva da un prodotto scalare $(a|b)$ su V. Si ha

4.51 Proposizione. *Sia V uno spazio con prodotto scalare e* $||a||^2 := (a|a)$. *Sia $A : \mathbb{R}^n \to V$ iniettiva. Allora il minimo di $||Ax - b||$ è unico ed e' l'unica soluzione della equazione*

$$A^*Au = A^*b, \tag{4.9}$$

detta equazione canonica. *L'applicazione $A^* : V \to \mathbb{R}^n$ è l'operatore aggiunto di A.*

Dimostrazione. Si tratta di trovare il punto $Ax \in \operatorname{Im} A$ di minima distanza da b. Il teorema della proiezione ortogonale assicura che il minimo Ax e' unico ed è il piede della prioiezione perpendicolare di b su $\operatorname{Im} A$, i.e. $Ax - b \perp \operatorname{Im} A$. D'altra parte $\operatorname{Im} A$ e $\ker A^*$ sono tra loro perpendicolari e supplementari e quindi $Ax - b \in \operatorname{Im} A^\perp$ se e solo se $Ax - b \in \ker A^*$, cioe' la (4.9).

Infine avendo A^*A ha tutti gli autovalori positivi perche' essendo A iniettiva,

$$A^*Au \bullet u = (Au|Au) = ||Au||^2.$$

Segue che A^*A è iniettiva. □

La risoluzione numerica dell'equazione canonica si opera riducendo, ad esempio mediante l'algoritmo SVD, la matrice simmetrica associata a A^*A a forma canonica,

$$A^*A = R^T \Lambda R, \qquad R^T = R^{-1}, \ \Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n),$$

e quindi calcolando per ogni $i = 1, \dots, n$ $y^i := (RA^*b)^i/\lambda_i$ e quindi $x = R^T y$, $y = (y^1, y^2, \dots, y^n)$.

4.52 (Minimi quadrati non lineare) Analogamente se il modello dipende in modo arbitrario ma C^1 dai parametri, il modello si schematizza come una immersione $\varphi : U \to \mathbb{R}^N$ e il problema diventa quello di minimizzare la funzione

$$x \to ||\phi(x) - b||^2.$$

Si ha

4.53 Proposizione. *Se x_0 è un minimo relativo per $||\phi(x) - b||$ allora x_0 soddisfa il sistema di equazioni non lineari*

$$[\mathbf{D}\varphi(x_0)]^*(\phi(x_0) - b) = 0. \tag{4.10}$$

Dimostrazione. Per il teorema di Fermat deve essere

$$0 = \frac{\partial ||\phi(x) - b||^2}{\partial x^i}(x_0) = (\phi(x_0) - b | D_i\varphi(x_0)) = 0 \qquad \forall i = 1, \dots n,$$

o, tenendo conto che le colonne della matrice $\mathbf{D}\phi(x_0)$ generano lo spazio tangente a $\varphi(U)$,

$$\phi(x_0) - b \in \operatorname{Im}(\mathbf{D}\varphi(x_0))^\perp = \ker([\mathbf{D}\varphi(x_0)]^*).$$

□

4.54 ¶ Supponiamo che un sistema sia descritto da una funzione lineare $y = mx + q$. Siano (x_i, y_i) n punti sperimentali in $\mathbb{R}^2$. Determinare m, q dai dati sperimentali in modo da minimizzare lo scarto quadratico medio, i.e. la funzione

$$S(m, q) := \sum_{i=1}^n (y_i - mx_i - q)^2.$$