Complementi di Algebra e Fondamenti di Geometria

Capitolo 2

Lunghezza ed ortogonalità: Prodotto scalare in $\mathbb{R}^n$ e prodotto hermitiano in $\mathbb{C}^n$

L. Aceto, M. Ciampa

Ingegneria Elettrica, a.a. 2009/2010

Capitolo 2

Lunghezza ed ortogonalità: prodotto scalare in $\mathbb{R}^n$ e prodotto hermitiano in $\mathbb{C}^n$

In questo capitolo si definisce, nello spazio vettoriale $\mathbb{R}^n$, la funzione *prodotto scalare* e la si utilizza per estendere le nozioni di *lunghezza* di un vettore e di *ortogonalità* tra due vettori, già note in $\mathbb{R}^2$ ed $\mathbb{R}^3$, a tutti gli spazi $\mathbb{R}^n$. Si discutono poi i principali oggetti (*base ortonormale*, *sottospazi ortogonali*, *proiezione ortogonale*, *matrice ortogonale*) che tali nozioni consentono di definire. Infine, si mostra come le nozioni di lunghezza ed ortogonalità si possano introdurre anche negli spazi $\mathbb{C}^n$ definendovi una opportuna funzione — *prodotto hermitiano* — che svolge il ruolo della funzione prodotto scalare negli spazi $\mathbb{R}^n$.

Se $v_1, \ldots, v_k$ sono elementi di uno spazio vettorale V (in particolare: di $\mathbb{R}^n$ o di $\mathbb{C}^n$), useremo la sigla:

$$\langle v_1, \ldots, v_k \rangle$$

per indicare il sottospazio vettoriale di V costituito da tutte le combinazioni lineari di $v_1, \ldots, v_k$.[*]

2.1 Prodotto scalare, lunghezza ed ortogonalità

La funzione $\mathrm{PS} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ che ai vettori x di componenti $x_1, \ldots, x_n$ ed y di componenti $y_1, \ldots, y_n$ associa il numero reale:[†]

$$\mathrm{PS}(x, y) = y_1 x_1 + \cdots + y_n x_n = y^\mathsf{T} x$$

si chiama *prodotto scalare* (in $\mathbb{R}^n$). Non c'è una notazione universalmente adottata per la funzione prodotto scalare. In alternativa alla precedente, noi adotteremo la seguente: $\mathrm{PS}(x, y) = x \bullet y$.

2.1 Osservazione

Utilizzando le proprietà del prodotto riga per colonna si constata che la funzione prodotto scalare verifica le tre proprietà seguenti:

`PS1`: Per ogni $x, x', y \in \mathbb{R}^n$ ed $a, b \in \mathbb{R}$ si ha: $\mathrm{PS}(ax + bx', y) = a\,\mathrm{PS}(x, y) + b\,\mathrm{PS}(x', y)$, ovvero la funzione PS è *lineare rispetto al primo argomento.*

`PS2`: Per ogni $x, y \in \mathbb{R}^n$ si ha $\mathrm{PS}(x, y) = \mathrm{PS}(y, x)$, ovvero la funzione PS è *simmetrica.*

[*] Un'altra sigla comunemente usata per indicare lo stesso sottospazio è: $\mathrm{Span}(v_1, \ldots, v_k)$.

[†] Se x è la colonna di $\mathbb{R}^n$ di componenti $x_1, \ldots, x_n$, il simbolo x^T indica il *trasposto* di x, ovvero la *riga* di componenti $x_1, \ldots, x_n$.

PS3: Per ogni vettore non nullo $x \in \mathbb{R}^n$ si ha $\mathrm{PS}(x,x) > 0$, ovvero la funzione $x \to \mathrm{PS}(x,x)$ è *definita positiva.*

2.2 Problema

(1) Verificare l'asserto dell'osservazione precedente.

(2) Dedurre dalle proprietà PS1, PS2 e PS3 che la funzione PS è *lineare rispetto al secondo argomento.*

(3) Dedurre dalle proprietà PS1, PS2 e PS3 che se $u_1, \dots, u_k$ e v sono vettori di $\mathbb{R}^n$, allora:

$$\mathrm{PS}(u_1 + \cdots + u_k, v) = \mathrm{PS}(u_1, v) + \cdots + \mathrm{PS}(u_k, v)$$

La funzione $N : \mathbb{R}^n \to \mathbb{R}$ che al vettore x di componenti $x_1, \dots, x_n$ associa il numero reale[‡]

$$N(x) = \sqrt{\mathrm{PS}(x,x)} = \sqrt{x_1^2 + \cdots + x_n^2}$$

si chiama *norma* (in $\mathbb{R}^n$). La notazione usuale (che adotteremo salvo rare eccezioni) è $N(x) = \|x\|$.

L'esempio seguente mostra che la funzione norma estende la nozione di *lunghezza* di un vettore, già nota in $\mathbb{R}^2$ ed $\mathbb{R}^3$, a tutti gli spazi $\mathbb{R}^n$.

2.3 Esempio

Si consideri un piano cartesiano di origine Ω, ed un suo punto X di coordinate x_1, x_2. La usuale identificazione del vettore geometrico da Ω ad X con l'elemento x di $\mathbb{R}^2$ di componenti x_1, x_2 e l'uso del teorema di Pitagora rivelano che $\|x\|$ non è altro che la *lunghezza* del vettore geometrico.

Analogamente si constata che se $x \in \mathbb{R}^3$ allora $\|x\|$ non è altro che ...

Il significato geometrico consente, nei casi $n = 2$ e $n = 3$, di constatare che la funzione norma verifica le tre proprietà seguenti:

N1: Per ogni vettore non nullo $x \in \mathbb{R}^n$ si ha $\|x\| > 0$.

N2: Per ogni $x \in \mathbb{R}^n$ ed $\alpha \in \mathbb{R}$ si ha $\|\alpha x\| = |\alpha| \, \|x\|$.

N3: Per ogni $x, y \in \mathbb{R}^n$ si ha $\|x + y\| \leqslant \|x\| + \|y\|$.

La disuguaglianza che compare nella proprietà N3 si chiama *disuguaglianza triangolare.*

2.4 Problema

(1) Dimostrare, utilizzando la definizione, che per *ogni* n la funzione norma in $\mathbb{R}^n$ verifica le proprietà N1 ed N2.

La funzione norma verifica per ogni n *anche* la proprietà N3, ma per verificare questo occorre utilizzare un risultato – la *disuguaglianza di Schwarz* – che sarà illustrato in seguito.

(2) Dedurre, dalle proprietà N1, N2 ed N3, che $\|0\| = 0$.

[‡]... si ricordi che la funzione PS è definita positiva.

Se $\mathrm{PS}(x, y) = 0$, i vettori x ed y si dicono *ortogonali*. In tal caso si scrive:

$$x \perp y$$

L'esempio seguente mostra che la definizione introdotta estende la nozione di ortogonalità tra vettori, già nota in $\mathbb{R}^2$ ed $\mathbb{R}^3$, a tutti gli spazi $\mathbb{R}^n$.

2.5 Esempio

Si consideri un piano cartesiano di origine Ω, e due suoi punti: X di coordinate x_1, x_2 ed Y di coordinate y_1, y_2. Per il teorema di Pitagora, i vettori geometrici da Ω ad X e da Ω ad Y sono ortogonali se e solo se:

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2$$

Ma:

$$\|x\|^2 + \|y\|^2 = (x_1^2 + x_2^2) + (y_1^2 + y_2^2)$$

e:

$$\|x - y\|^2 = (x_1 - y_1)^2 + (x_2 - y_2)^2 = (x_1^2 + x_2^2) + (y_1^2 + y_2^2) - 2(x_1y_1 + x_2y_2)$$

dunque i vettori geometrici sono ortogonali se e solo se $x_1y_1 + x_2y_2 = \mathrm{PS}(x, y) = 0$.

Analogamente si constata che se $x, y \in \mathbb{R}^3$, i vettori geometrici sono ortogonali se e solo se ...

2.6 Problema

(1) Dedurre dalle proprietà `PS1`, `PS2` e `PS3` che il vettore nullo è ortogonale ad ogni elemento di $\mathbb{R}^n$.

(2) Dedurre dalle proprietà `PS1`, `PS2` e `PS3` che: se $a \bullet b = 0$ per ogni $b \in \mathbb{R}^n$, allora a è il vettore nullo.

(3) Dimostrare il Teorema di Pitagora in $\mathbb{R}^n$: Se a, b sono vettori ortogonali, allora:

$$\|a + b\|^2 = \|a\|^2 + \|b\|^2$$

Le nozioni di lunghezza e di ortogonalità consentono di definire le nozioni di *base ortogonale* e di *base ortonormale* di un sottospazio vettoriale di $\mathbb{R}^n$.

2.7 Osservazione

Siano $v_1, \ldots, v_k$ vettori di $\mathbb{R}^n$ *non nulli* e *a due a due ortogonali.* Allora $v_1, \ldots, v_k$ sono *linearmente indipendenti.*

Infatti: se $a_1v_1 + \cdots + a_kv_k = 0$ per qualche $a_1, \ldots, a_k \in \mathbb{R}$, allora per ogni $j = 1, \ldots, k$ si ha: $v_j \bullet (a_1v_1 + \cdots + a_kv_k) = a_j(v_j \bullet v_j) = 0$ e quindi, essendo v_j non nullo, $a_j = 0$.

Siano V un sottospazio vettoriale di $\mathbb{R}^n$ di dimensione k, e $v_1, \ldots, v_k$ vettori di V non nulli e a due a due ortogonali. L'osservazione precedente assicura che tali vettori costituiscono una base di V, precisamente costituiscono una *base ortogonale* di V. Se ciascuno dei vettori della base ha anche *norma uno* la base si dice *ortonormale.*

Basi ortonormali di $\mathbb{R}^n$ esistono certamente: la più comune è la *base canonica*:

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad , \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad , \quad \ldots \quad , \quad e_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

2.8 Problema

(1) Verificare che la base canonica di $\mathbb{R}^n$ è ortonormale.

(2) Verificare che per ogni $\theta \in [0, 2\pi)$ i vettori:

$$\begin{bmatrix} \cos\theta \\ \operatorname{sen}\theta \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} -\operatorname{sen}\theta \\ \cos\theta \end{bmatrix}$$

sono una base ortonormale di $\mathbb{R}^2$.

Avere a disposizione una base ortonormale di uno spazio vettoriale rende più semplici varie operazioni. Ad esempio: le componenti di un vettore $x \in \mathbb{R}^n$ rispetto ad una base ortonormale $q_1, \ldots, q_n$ di $\mathbb{R}^n$ si ottengono eseguendo i prodotti scalari $x \bullet q_1, \ldots, x \bullet q_n$, mentre, se la base non è ortonormale, è necessario risolvere un sistema di equazioni lineari. Una domanda interessante è se esistano basi ortonormali anche per un *qualsiasi* sottospazio vettoriale di $\mathbb{R}^n$ di dimensione almeno uno. Risponderemo (affermativamente!) a questa domanda in una prossima sezione.

2.2 Sottospazi ortogonali

In questa sezione si estende la nozione di ortogonalità introdotta per due vettori al caso dei *sottospazi vettoriali* e la si applica allo studio del nucleo e dell'immagine di un'applicazione lineare.

Siano x un vettore e V un sottospazio vettoriale di $\mathbb{R}^n$. *Il vettore x è ortogonale al sottospazio V* se x è ortogonale a ciascun elemento di V. In tal caso si scrive:

$$x \perp V$$

2.9 Esempio

In $\mathbb{R}^2$ si considerino il vettore:

$$v = \begin{bmatrix} 3 \\ 3 \end{bmatrix},$$

il sottospazio vettoriale V generato da v, ed i vettori:

$$f = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \text{e} \quad g = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Per ogni $\alpha \in \mathbb{R}$ si ha $(\alpha v) \bullet f = 0$ e quindi f *è* ortogonale a V, e $v \bullet g = 12 \neq 0$ e quindi g *non è* ortogonale a V.

(Esercizio: constatare il risultato disegnando, su un piano cartesiano, il sottospazio V ed i vettori f e g.)

Siano V e W sottospazi vettoriali di $\mathbb{R}^n$. *Il sottospazio V è ortogonale al sottospazio W* se ciascun elemento di V è ortogonale a ciascun elemento di W. In tal caso si scrive:

$$V \perp W$$

Verificare se due sottospazi vettoriali sono ortogonali è semplice se di ciascun sottospazio si ha un insieme (finito) di generatori:

2.10 Osservazione

Siano V il sottospazio vettoriale di $\mathbb{R}^n$ generato da $v_1, \ldots, v_k$ e W quello generato da $w_1, \ldots, w_j$. Allora:

$$V \perp W \text{ se e solo se } v_r \perp w_s \text{ per ogni } r, s$$

Infatti, se $V \perp W$ si ha certamente $v_r \perp w_s$ per ogni r, s; viceversa, utilizzando le proprietà del prodotto scalare si verifica che una qualsiasi combinazione lineare di $v_1, \ldots, v_k$ è ortogonale ad una qualsiasi combinazione lineare di $w_1, \ldots, w_j$.

2.11 Problema

(1) Si considerino, in $\mathbb{R}^4$, i seguenti sottospazi vettoriali:

$$U = \langle e_1, e_3 \rangle \quad , \quad V = \langle e_2 \rangle \quad , \quad W = \langle e_2, e_3, e_4 \rangle$$

Tra le nove coppie possibili ((U, U), (U, V), (U, W), (V, V), $\ldots$), individuare quelle costituite da sottospazi ortogonali.

(2) Siano v un vettore e W il sottospazio vettoriale di $\mathbb{R}^n$ generato da $w_1, \ldots, w_j$. Dimostrare che:

$$v \perp W \text{ se e solo se } v \perp w_s \text{ per ogni } s$$

Applicare il criterio di ortogonalità ottenuto ai dati dell'Esempio 2.9.

2.12 Osservazione (ortogonalità, nucleo ed immagine)

Sia $A \in \mathbb{R}^{n \times k}$ e si considerino l'applicazione lineare da $\mathbb{R}^k$ in $\mathbb{R}^n$ definita da A e quella da $\mathbb{R}^n$ in $\mathbb{R}^k$ definita da A^{T}. Coppie interessanti di sottospazi vettoriali ortogonali sono: $\operatorname{Ker} A$ e $\operatorname{Im} A^{\mathsf{T}}$, $\operatorname{Im} A$ e $\operatorname{Ker} A^{\mathsf{T}}$.

Infatti: Si ricordi che il nucleo di A è costituito da tutti gli elementi $x \in \mathbb{R}^k$ tali che $Ax = 0$ ovvero, indicando con $r_1, \ldots, r_n$ le *righe di* A, tali che $r_1 x = 0, \ldots, r_n x = 0$. Queste ultime n uguaglianze possono riscriversi utilizzando il prodotto scalare: $x \bullet r_1^{\mathsf{T}} = 0, \ldots, x \bullet r_n^{\mathsf{T}} = 0$. Poiché $r_1^{\mathsf{T}}, \ldots, r_n^{\mathsf{T}}$ sono le *colonne della matrice* A^{T}, che generano $\operatorname{Im} A^{\mathsf{T}}$, le n uguaglianze significano che:

$$\operatorname{Ker} A \perp \operatorname{Im} A^{\mathsf{T}}$$

Applicando lo stesso ragionamento a A^{T} si ottiene che:

$$\operatorname{Ker} A^{\mathsf{T}} \perp \operatorname{Im} A$$

L'ortogonalità di due sottospazi vettoriali V e W ha una immediata conseguenza: $V \cap W$ contiene solo il vettore nullo. Infatti, se $x \in V \cap W$ allora $x \perp x$ e quindi $x = 0$ (Esercizio: giustificare quest'ultimo passaggio). Se ne deduce che:

$$\text{Se } V \perp W \text{ allora la somma } V + W \text{ è diretta}$$

In tal caso, dalla formula di Grassmann si ottiene anche $\dim(V + W) = \dim V + \dim W$.

2.13 Osservazione (nucleo, immagine e decomposizione di $\mathbb{R}^n$)

Sia $A \in \mathbb{R}^{n \times k}$ e si consideri l'applicazione lineare definita da A. Per l'Osservazione 2.12 si ha:

$$\operatorname{Ker} A \perp \operatorname{Im} A^{\mathsf{T}}$$

Ne segue che la somma $\operatorname{Ker} A + \operatorname{Im} A^{\mathsf{T}}$ è diretta e:

$$\dim(\operatorname{Ker} A + \operatorname{Im} A^{\mathsf{T}}) = \dim(\operatorname{Ker} A) + \dim(\operatorname{Im} A^{\mathsf{T}})$$

ovvero, essendo $\dim(\operatorname{Im} A^{\mathsf{T}}) = \dim(\operatorname{Im} A)$:

$$\dim(\operatorname{Ker} A + \operatorname{Im} A^{\mathsf{T}}) = \dim(\operatorname{Ker} A) + \dim(\operatorname{Im} A)$$

Ma, per il teorema della dimensione:

$$\dim(\operatorname{Ker} A) + \dim(\operatorname{Im} A) = k$$

Se ne deduce la seguente *decomposizione di* $\mathbb{R}^k$:

$$\operatorname{Ker} A \oplus \operatorname{Im} A^\mathsf{T} = \mathbb{R}^k$$

Analogamente, a partire da:

$$\operatorname{Ker} A^\mathsf{T} \perp \operatorname{Im} A$$

si ottiene la seguente *decomposizione di* $\mathbb{R}^n$:

$$\operatorname{Im} A \oplus \operatorname{Ker} A^\mathsf{T} = \mathbb{R}^n$$

Le relazioni ottenute si sintetizzano nelle scritture:

$$\operatorname{Ker} A \overset{\perp}{\oplus} \operatorname{Im} A^\mathsf{T} = \mathbb{R}^k \quad \text{e} \quad \operatorname{Im} A \overset{\perp}{\oplus} \operatorname{Ker} A^\mathsf{T} = \mathbb{R}^n$$

dove il simbolo $\perp$ posto sopra al simbolo $\oplus$ indica che gli addendi della somma (diretta) sono ortogonali.

2.14 Osservazione
Sia $A \in \mathbb{R}^{n\times k}$ e si consideri l'applicazione lineare definita da A. Una conseguenza importante delle decomposizioni ottenute nelle osservazioni precedenti è la seguente: *la restrizione a* $\operatorname{Im} A^\mathsf{T}$ *dell'applicazione lineare definita da* A*, ovvero l'applicazione lineare da* $\operatorname{Im} A^\mathsf{T}$ *in* $\operatorname{Im} A$ *definita da* $x \to Ax$*, è invertibile.*
Infatti: (1) l'applicazione è surgettiva perché $y \in \operatorname{Im} A$ significa, per definizione, che esiste $x \in \mathbb{R}^k$ tale che $Ax = y$, ma $\mathbb{R}^k = \operatorname{Ker} A \oplus \operatorname{Im} A^\mathsf{T}$, e quindi x si scompone nella somma di un addendo in $\operatorname{Ker} A$ che ha immagine zero, ed uno in $\operatorname{Im} A^\mathsf{T}$ che ha immagine y; e: (2) l'applicazione è iniettiva perché se vi fossero due elementi $x, z \in \operatorname{Im} A^\mathsf{T}$ con la stessa immagine y, si avrebbe $x - z \in \operatorname{Im} A^\mathsf{T}$ e, poiché $A(x - z) = y - y = 0$, anche $x - z \in \operatorname{Ker} A$, dunque $x = z$.

Ritorneremo sugli argomenti delle ultime osservazioni nel capitolo riguardante la *decomposizione ai valori singolari.*

2.3 Proiezione ortogonale su una retta e angolo tra vettori

In questa sezione si definisce la nozione di *proiezione ortogonale di un vettore su un altro* e la si utilizza per dimostrare una disuguaglianza – la *disuguaglianza di Schwarz* – che permetterà, infine, di definire la nozione di *angolo* tra vettori di $\mathbb{R}^n$.

2.15 Osservazione (proiezione ortogonale e componente normale, I)
Siano $u, v \in \mathbb{R}^n$. Sono univocamente determinati un vettore p proporzionale ad u ed un vettore c ortogonale ad u tali che $p + c = v$. Il vettore p si chiama *proiezione ortogonale di* v *su* u ed il vettore c si chiama *componente normale di* v *rispetto ad* u.
Infatti: Se $u = 0$ allora $p = 0$ e $c = v$ sono gli unici vettori che soddisfano le richieste.
Se $u \neq 0$, si cerca $\alpha \in \mathbb{R}$ tale che il vettore $v - \alpha u$ sia ortogonale ad u, ovvero tale che:

$$(v - \alpha u) \bullet u = 0$$

L'equazione che si ottiene:

$$(u \bullet u)\, \alpha = v \bullet u \quad [\text{ovvero: } (u^\mathsf{T} u)\, \alpha = u^\mathsf{T} v]$$

ha una sola soluzione:

$$\alpha = \frac{v \bullet u}{u \bullet u}$$

dunque gli unici vettori che soddisfano le richieste sono:

$$p = \frac{v \bullet u}{u \bullet u}\, u \quad \text{e} \quad c = v - \frac{v \bullet u}{u \bullet u}\, u$$

2.16 Esempio

In $\mathbb{R}^3$ si considerino i vettori:

$$u = \begin{bmatrix} 3 \\ 3 \\ 0 \end{bmatrix} \quad , \quad v = \begin{bmatrix} 0 \\ 3 \\ 3 \end{bmatrix}$$

Poiché

$$v \bullet u = u^{\mathsf{T}} v = 9 \quad \text{e} \quad u \bullet u = \|u\|^2 = 18$$

la proiezione ortogonale di v su u e la componente normale di v rispetto ad u sono, rispettivamente:

$$p = \begin{bmatrix} \frac{3}{2} \\ \frac{3}{2} \\ 0 \end{bmatrix} \quad \text{e} \quad c = v - p = \begin{bmatrix} -\frac{3}{2} \\ \frac{3}{2} \\ 3 \end{bmatrix}$$

Si osservi che, geometricamente parlando, la proiezione ortogonale del vettore v sul vettore u è la proiezione ortogonale di v *sulla retta* generata da u.

2.17 Osservazione

Siano $u, v \in \mathbb{R}^n$. Un'interpretazione del vettore p proiezione ortogonale di v su u è la seguente.

Siano U lo spazio vettoriale generato da u, $F : U \to \mathbb{R}$ la funzione definita da:

$$F(x) = \|v - x\|^2$$

Allora:

$$\text{per ogni } x \in U \text{ tale che } x \neq p \text{ si ha: } F(p) < F(x)$$

ovvero p è *il* vettore che *rende minimo il valore di F*.

Infatti: $F(x) = \|v - x\|^2 = \|v - p + p - x\|^2$ e, ricordando che $v - p$ è la componente normale di v rispetto ad U (dunque $v - p \perp U$) ed osservando che $p - x \in U$, si ottiene: $\|v - p + p - x\|^2 = \|v - p\|^2 + \|p - x\|^2$. Se ne deduce che: se $x = p$ si ha $F(p) = \|v - p\|^2$, altrimenti $F(x) = \|v - p\|^2 + \|p - x\|^2 > \|v - p\|^2 = F(p)$.

Interpretando $\|v - x\|$ come *distanza tra x ed v* (si pensi al caso di $\mathbb{R}^2$), il vettore p identifica *l'elemento di U più vicino a v*.

2.18 Teorema (disuguaglianza di Schwarz)

Per ogni $u, v \in \mathbb{R}^n$ si ha:

$$|u \bullet v| \leqslant \|u\|\,\|v\|$$

e l'uguaglianza sussiste solo se u e v sono linearmente dipendenti.

Dimostrazione

Se $u = 0$ la relazione è vera (in forma di uguaglianza) per ogni v.

Se $u \neq 0$, detti p la proiezione ortogonale di v su u e c la componente normale, si ha:

$$\|c\|^2 = \|v - p\|^2 = \|v - \frac{v \bullet u}{u \bullet u}\, u\|^2 \geqslant 0$$

ovvero, ricordando che $u \bullet u = \|u\|^2$,

$$\|v\|^2 + \frac{|v \bullet u|^2}{\|u\|^4}\|u\|^2 - 2\,\frac{v \bullet u}{\|u\|^2}\,(v \bullet u) \geqslant 0$$

da cui si ottiene la disuguaglianza cercata. Si osservi che l'ultima relazione sussiste in forma di uguaglianza solo se $c = 0$, ovvero solo se $v = p$, nel qual caso v è un multiplo di u.
Fine Dimostrazione.

Siano u, v due vettori non nulli in $\mathbb{R}^n$. Per la disuguaglianza di Schwarz:

$$-1 \leqslant \frac{u \bullet v}{\|u\|\,\|v\|} \leqslant 1$$

e quindi *esiste un solo numero reale* $\theta \in [0, \pi]$ *tale che*:

$$\cos\theta = \frac{u \bullet v}{\|u\|\,\|v\|}$$

Il numero θ si chiama *angolo tra* u *e* v.

2.19 Esempio

Siano a, b vettori di $\mathbb{R}^2$ linearmente indipendenti e si considerino, in un piano cartesiano di origine Ω, il vettore geometrico da Ω ad A e quello da Ω a B ad essi corrispondenti.

Detto θ l'angolo tra i due vettori geometrici, si applica il Teorema di Carnot al triangolo di vertici Ω, A e B e si ottiene:

$$\|b - a\|^2 = \|a\|^2 + \|b\|^2 - 2\|a\|\,\|b\|\cos\theta$$

ma:

$$\|b - a\|^2 = \|a\|^2 + \|b\|^2 - 2(a \bullet b)$$

e quindi si ha l'espressione di $\cos\theta$ introdotta prima.

Si osservi che con la nozione di angolo introdotta, si ottiene la relazione, comunemente utilizzata per definire il prodotto scalare in $\mathbb{R}^2$ ed in $\mathbb{R}^3$:

$$a \bullet b = \|a\|\,\|b\|\cos\theta$$

2.20 Osservazione

Siano a, b due vettori di $\mathbb{R}^n$. Anche in questo caso si ha:

$$\|b - a\|^2 = \|a\|^2 + \|b\|^2 - 2(a \bullet b)$$

Dunque:

$$a \bullet b = \frac{1}{2}(\|a\|^2 + \|b\|^2 - \|b - a\|^2)$$

2.21 Problema

(1) Dimostrare, utilizzando la definizione e la disuguaglianza di Schwarz, che per *ogni* n la funzione norma in $\mathbb{R}^n$ verifica la proprietà `N3`.

(2) Siano u un vettore non nullo di $\mathbb{R}^n$ e

$$P = \frac{1}{u \bullet u}\, u u^{\mathsf{T}} \in \mathbb{R}^{n \times n}$$

Verificare che, per ogni $x \in \mathbb{R}^n$ si ha:

$$Px = \text{la proiezione ortogonale di } x \text{ su } u$$

2.4 Esistenza di basi ortonormali

In questa sezione si descrive la *procedura di Gram-Schmidt* che consente di determinare una base ortonormale di un sottospazio vettoriale di $\mathbb{R}^n$ di dimensione almeno uno, del quale sia nota una base. La procedura prova che *ogni sottospazio vettoriale di* $\mathbb{R}^n$ *di dimensione almeno uno ammette basi ortonormali.*

Siano V un sottospazio vettoriale di $\mathbb{R}^n$, di dimensione $k \geqslant 1$, e $v_1, \ldots, v_k$ una base di V. *Sono univocamente determinati coefficienti* $\tau_{ij} \in \mathbb{R}$ *tali che i vettori*:

$$\begin{aligned}
\omega_1 &= v_1 \\
\omega_2 &= v_2 - \tau_{2,1}\,\omega_1 \\
\omega_3 &= v_3 - (\,\tau_{3,1}\,\omega_1 + \tau_{3,2}\,\omega_2\,) \\
&\vdots \\
\omega_k &= v_k - (\,\tau_{k,1}\,\omega_1 + \cdots + \tau_{k,k-1}\,\omega_{k-1}\,)
\end{aligned}$$

sono k *elementi non nulli e ortogonali* e quindi, per l'Osservazione 2.7, *sono una base ortogonale di* V.

Infatti: Per $j = 1, \ldots, k$, il vettore ω_j è:

(0) una combinazione lineare a coefficienti non nulli dei vettori linearmente indipendenti $v_1, \ldots, v_j$ e quindi non nullo,

(1) ortogonale a ω_1 se e solo se:
$$\tau_{j,1} = \frac{v_j \bullet \omega_1}{\omega_1 \bullet \omega_1}$$

(2) ortogonale a ω_2 se e solo se:
$$\tau_{j,2} = \frac{v_j \bullet \omega_2}{\omega_2 \bullet \omega_2}$$

$\vdots$

$(j-1)$ ortogonale a ω_{j-1} se e solo se:
$$\tau_{j,j-1} = \frac{v_j \bullet \omega_{j-1}}{\omega_{j-1} \bullet \omega_{j-1}}$$

Dai vettori $\omega_1, \ldots, \omega_k$, base ortogonale di V, si ottiene immediatamente una *base ortonormale* $u_1, \ldots, u_k$ di V ponendo:

$$u_1 = \frac{1}{\|\omega_1\|}\,\omega_1, \ldots, u_k = \frac{1}{\|\omega_k\|}\,\omega_k$$

La procedura che a partire dalla base $v_1, \ldots, v_k$ calcola prima la base ortogonale $\omega_1, \ldots, \omega_k$ e poi quella ortonormale $u_1, \ldots, u_k$ prende il nome di *procedura di ortonormalizzazione di Gram-Schmidt.*

2.5 Proiezione ortogonale su un sottospazio vettoriale e problema dei minimi quadrati

In questa sezione si definisce la nozione di proiezione ortogonale di un vettore su un *sottospazio vettoriale* e la si utilizza per determinare la soluzione del *problema dei minimi quadrati in* $\mathbb{R}^n$.

2.22 Osservazione (proiezione ortogonale e componente normale, II)

Siano v un vettore e U un sottospazio vettoriale di $\mathbb{R}^n$. Sono univocamente determinati un vettore p di U ed un vettore c ortogonale ad U tali che $p + c = v$. Il vettore p si chiama *proiezione ortogonale di v su U* ed il vettore c si chiama *componente normale di v rispetto ad U.*

Infatti: Se U ha dimensione zero, gli unici vettori che verificano le richieste sono $p = 0$ e $c = v$.

Altrimenti, siano $k \geqslant 1$ la dimensione di U e $q_1, \dots, q_k$ una base ortonormale di U. Si cercano $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ tali che il vettore $v-(\alpha_1 q_1+\cdots+\alpha_k q_k)$ sia ortogonale ad U, ovvero tali che:

$$(v - (\alpha_1 q_1 + \cdots + \alpha_k q_k)) \bullet q_j = 0 \quad \text{per ogni } j$$

Per la ortonormalità degli elementi $q_1, \dots, q_k$, gli unici valori di $\alpha_1, \dots, \alpha_k$ che verificano le condizioni sono:

$$\alpha_1 = v \bullet q_1, \dots, \alpha_k = v \bullet q_k$$

dunque gli unici vettori che soddisfano le richieste sono:

$$p = (v \bullet q_1)\, q_1 + \cdots + (v \bullet q_k)\, q_k \quad \text{e} \quad c = v - p$$

2.23 Problema

Con riferimento ai dati dell'Osservazione 2.22:

(1) Si constati che nel caso $k = 1$ l'espressione data per la proiezione ortogonale coincide con quella data nell'Osservazione 2.15.

(2) Detto p_j il vettore proiezione ortogonale di v su q_j, si constati che

$$p = p_1 + \cdots + p_k$$

ovvero: *la proiezione ortogonale di v sul sottospazio vettoriale generato dagli elementi ortonormali $q_1, \dots, q_k$ si calcola sommando le proiezioni ortogonali di v sui singoli q_j.* Si rilegga la descrizione della procedura di Gram-Schmidt alla luce del risultato trovato.

(3) Ricordando che $q_1, \dots, q_k$ è una base ortonormale di U, detta $Q \in \mathbb{R}^{n \times k}$ la matrice di colonne $q_1, \dots, q_k$, dimostrare che per ogni $v \in \mathbb{R}^n$ si ha:

$$QQ^{\mathsf{T}} v = \text{la proiezione ortogonale di } v \text{ su } U$$

Siano v un vettore e U un sottospazio vettoriale di $\mathbb{R}^n$. Analogamente a quanto visto nel caso della proiezione ortogonale di un vettore su un altro, un'interpretazione del vettore p proiezione ortogonale di v su U è la seguente.

Sia $F : U \to \mathbb{R}$ la funzione definita da:

$$F(x) = \|v - x\|^2$$

Allora:

$$\text{per ogni } x \in U \text{ tale che } x \neq p \text{ si ha: } F(p) < F(x)$$

ovvero p è *il* vettore che *rende minimo il valore di F.*

Assegnati v vettore di componenti $v_1, \ldots, v_n$ ed U sottospazio vettoriale (di dimensione almeno uno) di $\mathbb{R}^n$ si chiama *problema dei minimi quadrati in* $\mathbb{R}^n$ il problema di determinare gli elementi

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in U$$

che rendono minima la quantità:

$$(v_1 - x_1)^2 + \cdots + (v_n - x_n)^2 = \|v - x\|^2$$

L'interpretazione appena data significa che *esiste una sola soluzione al problema dei minimi quadrati in* $\mathbb{R}^n$: *la proiezione ortogonale di* v *su* U.

Anche questo problema sarà ripreso nel capitolo riguardante la *decomposizione ai valori singolari.*

2.6 Matrici ortogonali e fattorizzazione QR

In questa sezione si introduce la nozione di *matrice ortogonale* e se ne mostrano alcune proprietà legate alle nozioni di lunghezza ed angolo. Infine, si mostra come la nozione di matrice ortogonale consenta di interpretare il risultato della procedura di Gram-Schmidt come una *fattorizzazione* di una opportuna matrice.

Una matrice $Q \in \mathbb{R}^{n \times k}$ si dice *ortogonale* se le sue colonne $q_1, \ldots, q_k$ sono elementi di norma uno di $\mathbb{R}^n$ a due a due ortogonali, ovvero sono *elementi ortonormali* di $\mathbb{R}^n$.

La definizione data è equivalente a:

$$Q^{\mathsf{T}} Q = I \in \mathbb{R}^{k \times k}$$

(<u>Esercizio</u>: verificare l'equivalenza.)

Se $n = k$, un'ulteriore formulazione equivalente alla definizione data è:

$$Q \text{ è invertibile e } Q^{-1} = Q^{\mathsf{T}}$$

2.24 Osservazione

Sia $Q \in \mathbb{R}^{n \times n}$ una matrice ortogonale, e si consideri l'applicazione lineare definita da Q. Allora:

(L) Per ogni $x \in \mathbb{R}^n$ si ha: $\|Qx\| = \|x\|$, ovvero l'applicazione *conserva le lunghezze*;

(PS) Per ogni $x, y \in \mathbb{R}^n$ si ha: $Qx \bullet Qy = x \bullet y$, ovvero l'applicazione *conserva il prodotto scalare.*

Per la definizione di norma e per l'Osservazione 2.20, le due proprietà sono equivalenti.

2.25 Problema

(1) Dimostrare le proprietà (L) e (PS) dell'asserto precedente.

(2) Dedurre dalle proprietà (L) e (PS) che l'applicazione lineare definita da una matrice ortogonale *conserva gli angoli*, ovvero: se x ed y sono due vettori non nulli di $\mathbb{R}^n$ e θ è l'angolo tra x ed y, allora l'angolo tra Qx e Qy vale anch'esso θ.

(3) Siano x ed y elementi di $\mathbb{R}^3$ e:

$$x' = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \text{e} \quad y' = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

i vettori delle *coordinate* di x ed y rispetto ad una base ortonormale di $\mathbb{R}^3$.

Determinare la lunghezza dei vettori x ed y e l'angolo da essi formato.

(4) Sia $Q \in \mathbb{R}^{n\times n}$ una matrice ortogonale. Dalla relazione $Q^\mathsf{T} Q = I$ segue che $QQ^\mathsf{T} = I$ e questo significa che *anche* Q^T *è una matrice ortogonale.*

Sia $Q \in \mathbb{R}^{3\times 2}$ una matrice ortogonale. Dimostrare che in questo caso Q^T *non è* una matrice ortogonale.

2.26 Osservazione

Siano $a_1, \dots, a_k$ elementi linearmente indipendenti di $\mathbb{R}^n$, e $u_1, \dots, u_k$ gli elementi ortonormali di $\mathbb{R}^n$ ottenuti applicando ad $a_1, \dots, a_k$ la procedura di Gram-Schmidt.

Rileggendo la procedura si constata che per $j = 1, \dots, k$ il vettore a_j è una combinazione lineare dei soli elementi $u_1, \dots, u_j$ e quindi esiste una matrice $T \in \mathbb{R}^{k\times k}$ triangolare superiore[§] tale che:

$$(a_1, \dots, a_k) = (u_1, \dots, u_k)\, T$$

Questa constatazione consente di rileggere il risultato della procedura di Gram-Schmidt come un metodo per ottenere, a partire da una matrice $A \in \mathbb{R}^{n\times k}$ a colonne linearmente indipendenti, una matrice $U \in \mathbb{R}^{n\times k}$ ortogonale ed una matrice $T \in \mathbb{R}^{k\times k}$ triangolare superiore tali che:

$$A = UT$$

La coppia U, T si chiama *fattorizzazione QR della matrice A.*

2.27 Problema

Siano $A, B \in \mathbb{R}^{n\times k}$ matrici a colonne linearmente indipendenti e $T \in \mathbb{R}^{k\times k}$ triangolare superiore tali che $A = BT$. Dimostrare che T è invertibile.

2.7 Prodotto hermitiano: norma ed ortogonalità in $\mathbb{C}^n$

In questa sezione si definisce, nello spazio vettoriale $\mathbb{C}^n$, la funzione *prodotto hermitiano* e si mostra come la si possa utilizzare per introdurre in $\mathbb{C}^n$ nozioni di lunghezza ed ortogonalità analoghe a quelle introdotte in $\mathbb{R}^n$ utilizzando il prodotto scalare.

La funzione $\mathrm{PH} : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ che ai vettori x di componenti $x_1, \dots, x_n$ ed y di componenti $y_1, \dots, y_n$ associa il numero complesso:[¶]

$$\mathrm{PH}(x, y) = \overline{y_1}\, x_1 + \cdots + \overline{y_n}\, x_n = y^\mathsf{H} x$$

si chiama *prodotto hermitiano* (in $\mathbb{C}^n$). Anche per la funzione prodotto hermitiano si adotta, in alternativa alla precedente, la notazione: $\mathrm{PH}(x, y) = x \bullet y$.

2.28 Osservazione

Utilizzando le proprietà del prodotto riga per colonna si constata che la funzione prodotto hermitiano verifica le tre proprietà seguenti:

[§] Una matrice di elementi m_{ij} si dice *triangolare superiore* (rispettivamente: triangolare inferiore) se $m_{ij} = 0$ per $i > j$ (rispettivamente: per $i < j$)

[¶] Se x è la colonna di $\mathbb{C}^n$ di componenti $x_1, \dots, x_n$, il simbolo x^H indica il *trasposto coniugato* di x, ovvero la *riga* di componenti $\overline{x_1}, \dots, \overline{x_n}$.

PH1: Per ogni $x, x', y \in \mathbb{C}^n$ ed $a, b \in \mathbb{C}$ si ha: $\mathrm{PH}(ax + bx', y) = a\,\mathrm{PH}(x, y) + b\,\mathrm{PH}(x', y)$, ovvero la funzione PH è *lineare rispetto al primo argomento.*

PH2: Per ogni $x, y \in \mathbb{C}^n$ si ha $\mathrm{PH}(y, x) = \overline{\mathrm{PH}(x, y)}$, ovvero la funzione PH ha la proprietà di *simmetria hermitiana.*

PH3: Per ogni vettore non nullo $x \in \mathbb{C}^n$, il numero $\mathrm{PH}(x, x)$ è reale e positivo, ovvero la funzione $x \to \mathrm{PH}(x, x)$ è *definita positiva.*

2.29 Problema

Verificare l'asserto dell'osservazione precedente e poi dedurre dalle proprietà PH1, PH2 e PH3 che: *Per ogni* $x, y, y' \in \mathbb{C}^n$ *ed* $a, b \in \mathbb{C}$ *si ha:* $\mathrm{PH}(x, ay + by') = \overline{a}\,\mathrm{PH}(x, y) + \overline{b}\,\mathrm{PH}(x, y')$.

2.30 Osservazione

Siano x e y vettori in $\mathbb{C}^n$. Se le componenti $x_1, \ldots, x_n$ e $y_1, \ldots, y_n$ sono numeri reali, allora $x, y \in \mathbb{R}^n$ e:

$$\mathrm{PH}(x, y) = \overline{y_1}\, x_1 + \cdots + \overline{y_n}\, x_n = y_1 x_1 + \cdots + y_n x_n = \mathrm{PS}(x, y)$$

In questo senso la funzione PH è una *estensione a* $\mathbb{C}^n$ *della funzione* PS.

La funzione $N : \mathbb{C}^n \to \mathbb{R}$ che al vettore x di componenti $x_1, \ldots, x_n$ associa il numero reale[‖]

$$N(x) = \sqrt{\mathrm{PH}(x, x)} = \sqrt{\overline{x_1}\, x_1 + \cdots + \overline{x_n}\, x_n} = \sqrt{|x_1|^2 + \cdots + |x_n|^2}$$

si chiama *norma* (in $\mathbb{C}^n$). La notazione usuale (che adotteremo salvo rare eccezioni) è, anche in questo caso: $N(x) = \|x\|$.

2.31 Osservazione

Sia x un vettore di $\mathbb{C}^n$. Se le componenti $x_1, \ldots, x_n$ sono numeri reali, allora $x \in \mathbb{R}^n$ e:

$$\mathrm{PH}(x, x) = \mathrm{PS}(x, x)$$

e quindi la norma di x come elemento di $\mathbb{C}^n$ è uguale a quella di x come elemento di $\mathbb{R}^n$. In questo senso la funzione norma in $\mathbb{C}^n$ è una *estensione a* $\mathbb{C}^n$ *della funzione norma in* $\mathbb{R}^n$.

Per la funzione norma in $\mathbb{C}^n$ sussistono proprietà analoghe a quelle della norma in $\mathbb{R}^n$:

N1: Per ogni vettore non nullo $x \in \mathbb{C}^n$ si ha $\|x\| > 0$.

N2: Per ogni $x \in \mathbb{C}^n$ ed $\alpha \in \mathbb{C}$ si ha $\|\alpha x\| = |\alpha|\, \|x\|$.

N3: Per ogni $x, y \in \mathbb{C}^n$ si ha $\|x + y\| \leqslant \|x\| + \|y\|$.

2.32 Problema

Dimostrare, utilizzando la definizione, che la funzione norma in $\mathbb{C}^n$ verifica le proprietà N1 ed N2 e poi dedurre, dalle proprietà N1, N2 ed N3, che $\|0\| = 0$.

Si osservi che nella definizione di norma in $\mathbb{C}^n$, la funzione prodotto hermitiano ha esattamente lo stesso ruolo della funzione prodotto scalare nella definizione di norma in $\mathbb{R}^n$. Questa osservazione suggerisce l'idea di *utilizzare in* $\mathbb{C}^n$ *le nozioni introdotte in* $\mathbb{R}^n$ *tramite il prodotto scalare, sostituendo a quest'ultimo il prodotto hermitiano.*

Si ha:

[‖]. . . si ricordi che la funzione $x \to \mathrm{PH}(x, x)$ è definita positiva.

- Vettori x ed y di $\mathbb{C}^n$ si dicono *ortogonali* se $\mathrm{PH}(x, y) = 0$. In tal caso si scrive:

$$x \perp y$$

- La nozione di ortogonalità consente di utilizzare in $\mathbb{C}^n$ quella di *proiezione ortogonale* di un vettore su una retta e di riprodurre la dimostrazione della *disuguaglianza di Schwarz*, che dunque sussiste nella stessa forma:

  Per ogni $u, v \in \mathbb{C}^n$ si ha:

$$|u \bullet v| \leqslant \|u\| \, \|v\|$$

  e l'uguaglianza sussiste solo se u e v sono linearmente dipendenti.

  La disuguaglianza di Schwarz consente, anche in questo caso, di dimostrare che la funzione norma in $\mathbb{C}^n$ verifica la proprietà `N3`.

- Le nozioni di *base ortogonale* e di *base ortonormale* di un sottospazio vettoriale di $\mathbb{C}^n$ sono identiche al caso di $\mathbb{R}^n$, ed anche l'Osservazione 2.7 sussiste inalterata:

  Elementi di $\mathbb{C}^n$ non nulli e a due a due ortogonali sono linearmente indipendenti.

  Basi ortonormali di $\mathbb{C}^n$ esistono certamente: la più comune è la *base canonica*:

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad , \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad , \quad \dots \quad , \quad e_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

  L'esistenza di basi ortonormali per un qualsiasi sottospazio vettoriale di $\mathbb{C}^n$ si ottiene rileggendo la Sezione 2.4 sostituendo $\mathbb{R}^n$ con $\mathbb{C}^n$. Anche in questo caso la costruzione della base ortonormale prende il nome di *procedura di Gram-Schmidt.*

- La nozione di matrice ortogonale nel caso complesso è identica a quella del caso reale, ma il termine usualmente adottato per indicarla è quello di *matrice unitaria*: Se le colonne $q_1, \dots, q_k$ di una matrice $Q \in \mathbb{C}^{n \times k}$ sono elementi ortonormali di $\mathbb{C}^n$, la matrice si dice unitaria. La definizione data è equivalente a:[**]

$$Q^{\mathsf{H}} Q = I \in \mathbb{C}^{k \times k}$$

  Analogamente al caso reale, anche nel caso complesso è possibile rileggere il risultato della procedura di Gram-Schmidt come un metodo per ottenere, a partire da una matrice $A \in \mathbb{C}^{n \times k}$ a colonne linearmente indipendenti, una matrice $U \in \mathbb{C}^{n \times k}$ unitaria ed una matrice $T \in \mathbb{C}^{k \times k}$ triangolare superiore tali che:

$$A = UT$$

  La coppia U, T è, anche in questo caso, una fattorizzazione QR della matrice A.

[**] Se $A \in \mathbb{C}^{n \times k}$ è la matrice di elementi a_{ij}, la sigla $\overline{A}$ indica la matrice $n \times k$ di elementi $\overline{a_{ij}}$ e la sigla A^{H} indica la matrice $\overline{A}^{\mathsf{T}} \in \mathbb{C}^{k \times n}$

2.33 Problema

(1) Dedurre dalle proprietà `PH1`, `PH2` e `PH3` che il vettore nullo è ortogonale ad ogni elemento di $\mathbb{C}^n$.

(2) Dedurre dalle proprietà `PH1`, `PH2` e `PH3` che: se $a \bullet b = 0$ per ogni $b \in \mathbb{C}^n$, allora a è il vettore nullo.

(3) Dimostrare il Teorema di Pitagora in $\mathbb{C}^n$: Se a, b sono vettori ortogonali, allora:

$$\|a + b\|^2 = \|a\|^2 + \|b\|^2$$

Si osservi che mentre in $\mathbb{R}^n$ sussiste anche l'implicazione inversa: *se a, b sono tali che $\|a + b\|^2 = \|a\|^2 + \|b\|^2$, allora sono ortogonali* ciò non vale in $\mathbb{C}^n$.