Ortogonalizzazione di Schmidt e fattorizzazione di Cholesky

Sia $\mathcal{U}$ uno spazio vettoriale definito su $\mathbb{R}$. In corrispondenza di una base $\{b_1 \dots b_n\}$ sia definito un prodotto interno assegnando i seguenti valori[1]

$$\langle b_i \, , b_j \rangle = g_{ij}$$

La matrice $\mathbf{G}$ i cui elementi siano gli scalari g_{ij} deve essere, per gli assiomi di definizione di un prodotto interno, simmetrica e definita positiva.
Si vuole costruire una base $\{\acute{b}_1 \dots \acute{b}_n\}$ ortonormale, cioè tale che sia

$$\langle \acute{b}_i \, , \acute{b}_j \rangle = \delta_{ij}$$

Si considerino i vettori $\acute{b}_i$ definiti nel modo seguente

$$\begin{aligned}
&\acute{b}_1 := b_1 / u_{11} \\
&\acute{b}_2 := (b_2 - u_{12} \acute{b}_1)/u_{22} \\
&\acute{b}_3 := (b_3 - u_{13} \acute{b}_1 - u_{23} \acute{b}_2)/u_{33} \\
&\dots
\end{aligned}$$

con $u_{ii} > 0$. Essendo i vettori $\{b_1 \dots b_n\}$ indipendenti, è semplice dimostrare che i vettori $\{\acute{b}_1 \dots \acute{b}_n\}$ risultano anch'essi indipendenti, comunque si scelgano i coefficienti u_{ij}.
Si dimostra ora che è possibile determinare i coefficienti u_{ij} in modo tale che i vettori $\{\acute{b}_1 \dots \acute{b}_n\}$ risultino ortonormali.
Procedendo nella costruzione dei nuovi vettori da $\acute{b}_1$ a $\acute{b}_n$ si ha per i primi due

$$\begin{aligned}
\langle \acute{b}_1 \, , \acute{b}_1 \rangle = 1 \quad &\Rightarrow \quad \langle b_1 \, , b_1 \rangle / u_{11}^2 = 1 \\
&\Rightarrow \quad u_{11} = \langle b_1 \, , b_1 \rangle^{1/2} \\
\langle \acute{b}_2 \, , \acute{b}_1 \rangle = 0 \quad &\Rightarrow \quad \langle b_2 \, , \acute{b}_1 \rangle - u_{12} = 0 \\
&\Rightarrow \quad u_{12} = \langle b_2 \, , \acute{b}_1 \rangle = \langle b_2 \, , b_1 \rangle / u_{11} = g_{21}/u_{11} \\
\langle \acute{b}_2 \, , \acute{b}_2 \rangle = 1 \quad &\Rightarrow \quad \langle b_2 - u_{12} \acute{b}_1 \, , b_2 - u_{12} \acute{b}_1 \rangle / u_{22}^2 = 1 \\
&\Rightarrow \quad u_{22}^2 = \langle b_2 - u_{12} \acute{b}_1 \, , b_2 - u_{12} \acute{b}_1 \rangle > 0 \\
&\Rightarrow \quad u_{22} = \left(\langle b_2 \, , b_2 \rangle - u_{12}^2 \right)^{1/2} = (g_{22} - u_{12}^2)^{1/2}
\end{aligned}$$

Supponendo poi di aver costruito i vettori $\{\acute{b}_1 \dots \acute{b}_{i-1}\}$ in modo tale che risultino ortonormali, si consideri il vettore

$$\acute{b}_i := \left(b_i - \sum_{k=1}^{i-1} u_{ki} \acute{b}_k \right) / u_{ii}$$

[1] Un punto di vista alternativo consiste nell'assumere che sia stato già assegnato un prodotto interno in altro modo, risultando in tal caso definiti gli scalari $g_{ij} := \langle b_i \, , b_j \rangle$.

La condizione di ortogonalità
$$\langle \acute{b}_i \,, \acute{b}_j \rangle = 0 \quad j < i$$
implica, sostituendo la espressione di $\acute{b}_i$,
$$j < i \quad \langle \acute{b}_i \,, \acute{b}_j \rangle = \left(\langle b_i \,, \acute{b}_j \rangle - \sum_{k=1}^{i-1} u_{ki} \langle \acute{b}_k \,, \acute{b}_j \rangle \right) / u_{ii} = \Big(\langle b_i \,, \acute{b}_j \rangle - u_{ji} \Big) / u_{ii} = 0$$
e quindi
$$j < i \quad u_{ji} = \langle b_i \,, \acute{b}_j \rangle$$
La condizione
$$\langle \acute{b}_i \,, \acute{b}_i \rangle = 1$$
implica poi
$$\langle \acute{b}_i \,, \acute{b}_i \rangle = \| b_i - \sum_{k=1}^{i-1} u_{ki} \acute{b}_k \|^2 / u_{ii}^2 = 1$$
$$u_{ii}^2 = \| b_i - \sum_{k=1}^{i-1} u_{ki} \acute{b}_k \|^2$$
Da tali relazioni si ottengono infine le espressioni esplicite dei coefficienti u_{ji}
$$j < i \quad u_{ji} = \langle b_i \,, \acute{b}_j \rangle = \left(\langle b_i \,, b_j \rangle - \sum_{k=1}^{j-1} u_{kj} \langle b_i \,, \acute{b}_k \rangle \right) / u_{jj} = \left(g_{ij} - \sum_{k=1}^{j-1} u_{kj} u_{ki} \right) / u_{jj}$$
$$u_{ii} = \| b_i - \sum_{k=1}^{i-1} u_{ki} \acute{b}_k \| = \left(g_{ii} - \sum_{k=1}^{i-1} u_{ki}^2 \right)^{1/2}$$
Si è così descritta in maniera completa la costruzione della base ortonormale $\{ \acute{b}_1 \dots \acute{b}_n \}$. Il procedimento seguito ha il nome di *ortogonalizzazione di Schmidt*.

Si osservi ora che dalle definizioni dei nuovi vettori si ha
$$\begin{aligned} &b_1 = u_{11} \acute{b}_1 \\ &b_2 = u_{12} \acute{b}_1 + u_{22} \acute{b}_2 \\ &b_3 = u_{13} \acute{b}_1 + u_{23} \acute{b}_2 + u_{33} \acute{b}_3 \\ &\dots \end{aligned}$$
e in generale
$$b_i = \sum_{j=1}^{i} u_{ji} \acute{b}_j$$
Da tali espressioni deriva che
$$\langle b_i \,, b_j \rangle = \sum_{k=1}^{i} u_{ki} \left(\sum_{h=1}^{j} u_{hj} \langle \acute{b}_k \,, \acute{b}_h \rangle \right) = \sum_{k=1}^{i} u_{ki} \left(\sum_{h=1}^{j} u_{hj} \delta_{hk} \right) = \sum_{k=1}^{\min i,j} u_{ki} u_{kj}$$

Risulta dunque che i coefficienti u_{ji} sono legati agli scalari g_{ij} dalla semplice relazione

$$\sum_{k=1}^{\min i,j} u_{ki} u_{kj} = g_{ij}$$

Indicando con $\mathbf{U}$ la matrice (*triangolare superiore*) che ha come colonne le componenti dei vettori $\{b_1 \dots b_n\}$ nella base $\{\acute{b}_1 \dots \acute{b}_n\}$, la relazione precedente si scrive[2]

$$\mathbf{U}^T \mathbf{U} = \mathbf{G}$$

L'algoritmo che produce, in corrispondenza di una matrice $\mathbf{G}$ definita positiva, gli elementi della matrice $\mathbf{U}$ nel modo indicato si chiama *fattorizzazione di Cholesky*.[3]
Si noti che la condizione che la matrice $\mathbf{G}$ sia definita positiva implica che $u_{ii}^2 > 0$. Infatti u_{ii}^2 risulta essere il quadrato della norma, nel prodotto interno definito dal $\mathbf{G}$, di un particolare vettore, diverso dal vettore nullo. Questo permette di utilizzare la fattorizzazione di Cholesky come test per decidere se una assegnata matrice reale simmetrica $\mathbf{G}$ sia o no definita positiva. Infatti, se nella valutazione dei termini u_{ii} risulta $u_{ii}^2 \leq 0$, questo implica che la matrice $\mathbf{G}$ non è definita positiva. Se invece questo non accade allora la matrice $\mathbf{G}$ è definita positiva poiché, essendo $\mathbf{U}$ invertibile, risulta

$$\mathbf{u} \neq \mathbf{o} \quad \Rightarrow \quad \mathbf{u}^T \mathbf{G} \mathbf{u} = \mathbf{u}^T \mathbf{U}^T \mathbf{U} \mathbf{u} = (\mathbf{U}\mathbf{u})^T (\mathbf{U}\mathbf{u}) > 0$$
$$\mathbf{u} = \mathbf{o} \iff (\mathbf{U}\mathbf{u})^T (\mathbf{U}\mathbf{u}) = 0$$

[2] Spesso la espressione $\mathbf{U}^T\mathbf{U}$ è scritta nella forma $\mathbf{L}\mathbf{L}^T$, essendo $\mathbf{L} = \mathbf{U}^T$. In ogni caso tali simboli non devono essere confusi con quelli utilizzati nella *eliminazione di Gauss* e nella *fattorizzazione LU*.
[3] L'algoritmo descritto può essere utilizzato per risolvere un sistema di equazioni lineari $\mathbf{G}\mathbf{x} = \mathbf{b}$ facendolo seguire dalla valutazione in successione dei vettori $\mathbf{y} := (\mathbf{U}^T)^{-1}\mathbf{b}$, $\mathbf{x} := \mathbf{U}^{-1}\mathbf{y}$, con una *sostituzione in avanti* nel primo caso e una *sostituzione all'indietro* nel secondo.