Fattorizzazione QR e matrici di Householder

22 ottobre 2009

In questa nota considereremo un tipo di fattorizzazione che esiste sempre nel caso di matrici quadrate non singolari ad entrate reali.

Definizione 1. L'insieme di tutte le matrici $n \times n$ invertibili a coefficienti su di un campo $\mathbb{K}$ è detto *gruppo generale lineare* di ordine n su $\mathbb{K}$ e indicato con il simbolo $GL_n(\mathbb{K})$. Il sottoinsieme $SL_n(\mathbb{K})$ di tutte le matrici di $GL_n(\mathbb{K})$ con determinante 1 è un sottogruppo di $GL_n(\mathbb{K})$ chiamato *gruppo speciale lineare* di ordine n su $\mathbb{K}$.

Definizione 2. Una matrice $Q \in GL_n(\mathbb{K})$ è detta *ortogonale* se $Q^T = Q^{-1}$.

Poiché

$$1 = \det I_n = \det(Q^T Q) = \det(Q^T)\det(Q) = \det(Q)^2,$$

abbiamo che ogni matrice ortogonale ha determinate pari a ± 1.

Teorema 1. *L'insieme delle matrici ortogonali* $n \times n$ *su* $\mathbb{K}$ *è un sottogruppo di* $GL_n(\mathbb{K})$.

Dimostrazione. Poiché $GL_n(\mathbb{K})$ è un gruppo, sicuramente il prodotto di matrici ortogonali è associativo. Inoltre, se M è ortogonale, allora

$$M^{-1}(M^{-1})^T = M^{-1}M = I,$$

per cui anche l'inversa M^{-1} è ortogonale. Basta dimostrare ora che se M, N sono due matrici ortogonali, allora MN è anche essa ortogonale. D'altro canto,

$$(MN)^T(MN) = (M^T N^T)NM = M^T(N^T N)M = M^T M = I.$$

La tesi segue. $\square$

Sia ora $Q = \begin{pmatrix} Q_1 & Q_2 & \dots & Q_n \end{pmatrix}$ la matrice reale $n \times n$ con colonne rispettivamente $Q_1, Q_2, \dots, Q_n$. Allora,

$$Q^T Q = \begin{pmatrix} Q_1^T \\ Q_2^T \\ \vdots \\ Q_n^T \end{pmatrix} \begin{pmatrix} Q_1 & Q_2 & \dots & Q_n \end{pmatrix} = \begin{pmatrix} \langle Q_1, Q_1 \rangle & \langle Q_1, Q_2 \rangle & \dots & \langle Q_1, Q_n \rangle \\ \langle Q_2, Q_1 \rangle & \langle Q_2, Q_2 \rangle & \dots & \langle Q_2, Q_n \rangle \\ \vdots & & & \vdots \\ \langle Q_n, Q_1 \rangle & \langle Q_n, Q_2 \rangle & & \langle Q_n, Q_n \rangle \end{pmatrix},$$

ove con $\langle \cdot, \cdot \rangle$ si è indicato il prodotto scalare standard di $\mathbb{R}^n$. In particolare, la matrice Q è ortogonale se, e solamente se,

$$\langle Q_i, Q_j \rangle = \delta_{ij} = \begin{cases} 1 & \text{se i=j} \\ 0 & \text{altrimenti.} \end{cases}$$

In altre parole una matrice Q è ortogonale se, e solamente se le sue colonne costituiscono un insieme *ortonormale* di vettori. Osserviamo che, poiché Q ortognonale implica Q^T ortogonale si ha automaticamente anche che le righe di Q sono un insieme ortonormale di vettori.

Osserviamo che, in generale se $\mathbf{x} \in \mathbb{R}^n$ e Q è una matrice ortogonale, si ha

$$||Q\mathbf{x}||_2 = \langle Q\mathbf{x}, Q\mathbf{x} \rangle = \mathbf{x}^T Q^T Q \mathbf{x} = \mathbf{x}^T \mathbf{x} = ||\mathbf{x}||_2.$$

In altre parole, le trasformazioni indotte da matrici ortogonali preservano la $|| \cdot ||_2$ dei vettori. Più in generale, date una norma $|| \cdot ||$ su $\mathbb{R}^n$ e una matrice $A \in \mathrm{Mat}_n(\mathbb{R})$ è possibile definire

$$||A|| = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{||A\mathbf{x}||}{||\mathbf{x}||} = \sup_{||\mathbf{x}|| \leq 1} ||A\mathbf{x}||.$$

Si dimostra che in questo modo si è introdotta una norma sul gruppo generale lineare e che $\forall \mathbf{x} \in \mathbb{R}^n$

$$||A\mathbf{x}|| \leq ||A|| \, ||\mathbf{x}||.$$

Le matrici ortogonali hanno tutte $|| \cdot ||_2$ pari ad 1

Algoritmo di Gram-Schmidt

Mostriamo ora come sia possibile, dato un insieme

$$\mathfrak{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_t\}$$

di vettori liberi in $\mathbb{R}^n$ costruire un insieme ortonormale

$$\mathfrak{C} = \{\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_t\}$$

con la proprietà aggiuntiva che, per ogni $j < t$, gli spazi vettoriali generati dai vettori $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_j$ e $\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_j$ coincidano. Premettiamo due teoremi.

Teorema 2. *Sia $\mathfrak{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_t\}$ un insieme ortonormale di vettori. Allora gli elementi di $\mathfrak{B}$ sono linearmente indipendenti.*

Dimostrazione. Supponiamo che esista una combinazione lineare

$$\alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2 + \ldots + \alpha_t \mathbf{b}_t = \mathbf{0}.$$

Allora, usando il fatto che

$$\langle \alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2 + \ldots + \alpha_t \mathbf{b}_t, \mathbf{b}_i \rangle = \alpha_i,$$

si ha per ogni $i = 1, \ldots, t$

$$\alpha_i = \langle \alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2 + \ldots + \alpha_t \mathbf{b}_t, \mathbf{b}_i \rangle = \langle \mathbf{0}, \mathbf{b}_i \rangle = 0.$$

La tesi segue. □

Teorema 3. *Sia $\mathbf{v} \in V$ un vettore non nullo e $\mathbf{x} \in V$. Allora, la proiezione ortogonale di $\mathbf{x}$ sullo spazio vettoriale generato da $\mathbf{v}$ è data da*

$$\Pi_{\mathbf{v}}(\mathbf{x}) = \langle \mathbf{v}, \mathbf{x} \rangle \frac{\mathbf{v}}{||\mathbf{v}||_2^2} = \frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}} \mathbf{x}.$$

Dimostrazione. Basta mostrare

$$\langle \mathbf{x} - \Pi_{\mathbf{v}}(\mathbf{x}), \mathbf{v} \rangle = 0.$$

Infatti,

$$\mathbf{v}^T \left(\mathbf{x} - \frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}} \mathbf{x} \right) = \mathbf{v}^T \mathbf{x} - \frac{\mathbf{v}^T\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}} \mathbf{x} = \mathbf{v}^T\mathbf{x} - \mathbf{v}^T\mathbf{x} = 0.$$

La tesi segue. □

Procediamo come segue:

1. poniamo $\mathbf{b}_1'' = \mathbf{b}_1$ e
$$\mathbf{b}_1' = \frac{1}{||\mathbf{b}_1''||_2} \mathbf{b}_1''.$$
Chiaramente $||\mathbf{b}_1'|| = 1$.

2. Per ogni $i > 1$ sottraiamo a $\mathbf{b}_i$ la sua proiezione ortogonale su ognuno degli spazi generati dai vettori $\mathbf{b}_j'$ ottenuti in precedenza:
$$\mathbf{b}_i'' = \mathbf{b}_i - \sum_{j=1}^{i-1} \left\langle \mathbf{b}_i, \mathbf{b}_j' \right\rangle \mathbf{b}_j'$$

3. infine otteniamo $\mathbf{b}'_i$, normalizzando il vettore $\mathbf{b}''_i$:

$$\mathbf{b}'_i = \frac{1}{||\mathbf{b}''_i||_2}\mathbf{b}'_i.$$

Osserviamo che

- per ogni $i = 2, \ldots, t$ il vettore $\mathbf{b}'_i$ ha norma

$$||\mathbf{b}'_i||_2 = \frac{1}{||\mathbf{b}''_i||_2}||\mathbf{b}''_i||_2 = 1$$

- prodediamo ora per induzione al fine di dimostrare la ortogonalità dei vettori considerati:

  1. $\mathbf{b}'_1$ e $\mathbf{b}'_2$ sono ortogonali, infatti

  $$\langle \mathbf{b}'_1, \mathbf{b}'_2 \rangle = ||\mathbf{b}''_2||_2 \langle \mathbf{b}'_1, \mathbf{b}_2 - \langle \mathbf{b}'_1, \mathbf{b}_2 \rangle \mathbf{b}'_1 \rangle = ||\mathbf{b}''_2||_2 \left(\langle \mathbf{b}'_1, \mathbf{b}_2 \rangle - \langle \mathbf{b}'_1, \mathbf{b}_2 \rangle \langle \mathbf{b}'_1, \mathbf{b}'_1 \rangle \right).$$

  Poiché $1 = ||\mathbf{b}'_1||_2^2 = \langle \mathbf{b}'_1, \mathbf{b}'_1 \rangle$, la quantità fra parentesi e nulla e l'ortogonalità segue.

  2. Supponiamo che i vettori $\mathbf{b}'_1, \mathbf{b}'_2, \ldots, \mathbf{b}'_{i-1}$ formino una sequenza ortonormale; mostriamo che anche $\mathbf{b}'_i$ è ortogonale a tutti questi. In particolare, per ogni $k < i$ fissato abbiamo

  $$\begin{aligned} \langle \mathbf{b}'_i, \mathbf{b}'_k \rangle = \ & ||\mathbf{b}''_i||_2 \left\langle \mathbf{b}_i - \textstyle\sum_{j=1}^{i-1} \langle \mathbf{b}_i, \mathbf{b}'_j \rangle \mathbf{b}'_j, \mathbf{b}'_k \right\rangle = \\ & ||\mathbf{b}''_i||_2 \left(\langle \mathbf{b}_i, \mathbf{b}'_k \rangle - \textstyle\sum_{j=1}^{i-1} \langle \mathbf{b}_i, \mathbf{b}'_j \rangle \langle \mathbf{b}'_j, \mathbf{b}'_k \rangle \right) = \\ & ||\mathbf{b}''_i||_2 \left(\langle \mathbf{b}_i, \mathbf{b}'_k \rangle - \textstyle\sum_{j=1}^{i-1} \delta_{jk} \langle \mathbf{b}_i, \mathbf{b}'_j \rangle \right) = \\ & ||\mathbf{b}''_i||_2 \left(\langle \mathbf{b}_i, \mathbf{b}'_k \rangle - \langle \mathbf{b}_i, \mathbf{b}'_k \rangle \right) = 0 \end{aligned}$$

  In particolare, si ha che tutta la sequenza $\mathbf{b}'_1, \ldots, \mathbf{b}'_i$ è ortonormale.

- È chiaro dalla costruzione che la sequenza $\mathfrak{B}_i = \{\mathbf{b}'_1, \ldots, \mathbf{b}'_i\}$ è contenuta nel sottospazio generato dai vettori $\mathfrak{C}_i = \{\mathbf{b}_1, \ldots, \mathbf{b}_i\}$. Per il Teorema 2, tale sequenza è indipendente e, pertanto genera un sottospazio di dimensione i; d'altro canto, i vettori in $\mathfrak{C}_i$ generano un sottospazio di dimensione al più i. Ne segue che il sottospazio generato da $\mathfrak{B}_i$ e quello generato da $\mathfrak{C}_i$ coincidono.

Data una matrice M possiamo utilizzare il procedimento di Gram–Schmidt per ottenere una matrice Q' ad essa associata che abbia colonne (o, equivalentemente, righe) ortonormali.

Osserviamo che:

1. moltiplicare una colonna di una matrice M per uno scalare α corrisponde a calcolare il prodotto a destra della matrice M per la matrice $L_i(\alpha)$ con entrate
$$l_{hk} = \begin{cases} 0 & \text{se } h \neq k \\ \alpha & \text{se } h = k = i \\ 1 & \text{se } h = k \neq i \end{cases}$$

2. Similmente, sostituire la colonna i–esima di M con una combinazione lineare del tipo
$$M_i' = M_i + \sum_{j=1}^{i-1} \beta_j M_j$$
equivale a moltiplicare M a destra per la matrice $U_i(\beta_1, \dots, \beta_{i-1})$ data da
$$U = \begin{pmatrix} 1 & 0 & \dots & 0 & \beta_1 & \dots & 0 \\ 0 & 1 & \ddots & 0 & \beta_2 & \dots & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots & & \\ & & 0 & 1 & \beta_{i-1} & & \\ & & & 0 & 1 & \ddots & \\ & & & & \ddots & \ddots & 0 \\ & & & & & 0 & 1 \end{pmatrix}$$

3. Pertanto le operazioni richieste nel procedimento di ortonormalizzazione di Gram-Schmidt possono essere rappresentate mediante prodotti a destra di matrici triangolari superiori con entrate sulla diagonale principale strettamente maggiori di 0.

4. Poiché l'insieme delle matrici triangolari superiori di tale fatta è chiuso rispetto il prodotto si ottiene che, comunque data una matrice $M \in GL_n(\mathbb{R})$ esistono una matrice triangolare superiore $\widetilde{R}$ e una matrice ortogonale Q tali che
$$Q = M\widetilde{R}.$$

5. Osserviamo che $\widetilde{R}$ è invertibile, e la sua inversa è ancora una matrice triangolare superiore con entrate strettamente positive sulla diagonale principale. Pertanto, posto $R = \widetilde{R}^{-1}$ si ha
$$M = QR.$$

Abbiamo dimostrato il seguente teorema.

Teorema 4. *Comunque assegnata una matrice quadrata* $M \in GL_n(\mathbb{R})$ *esistono sempre una matrice ortogonale* Q *e una matrice triangolare superiore* R *con entrate sulla diagonale principale strettamente positive tali che*

$$M = QR.$$

La decomposizione descritta nel Teorema 4 prende il nome di *fattorizzazione* QR. Dimostriamo ora che essa è unica.

Teorema 5. *La fattorizzazione* QR *di una matrice è unica.*

Dimostrazione. Supponiamo $Q_1R_1 = Q_2R_2$ siano due fattorizzazioni QR della medesima matrice M. Poichè sia Q_1 che Q_2 sono matrici ortogonali si ha

$$Q_2^T Q_1 = R_2 R_1^{-1}.$$

Rammentiamo che il prodotto $Q = Q_2^T Q_1$ di matrici ortogonali è esso stesso una matrice ortogonale, per cui $R = R_2R_1^{-1}$ deve essere ortogonale; in particolare $R^{-1} = R^T$ è una matrice triangolare inferiore. D'altro canto R è anche prodotto di due matrici triangolari superiori e, conseguentemente, la sua inversa è anche essa risultare triangolare superiore. Ne segue che R^{-1} deve essere contemporaneamente triangolare superiore e inferiore, ovvero R^{-1} è una matrice diagonale. Siccome $R = R^T = R^{-1}$, si ha $R^2 = I$. Il quadrato R^2 di una matrice diagonale R è la matrice diagonale che ha per elementi i quadrati degli elementi di R. Pertanto si ottiene che gli elementi non nulli r_{ii} di R devono tutti soddisfare $r_{ii}^2 = 1$. Poiché per ipotesi le entrate sulla diagonale principale di R_1 (e, conseguentemente R_1^{-1}) e di R_2 sono positive, si ha $r_{ii} = 1$ per ogni i, ovvero $R_1 = R_2$ e $Q_1 = Q_2$. □

Matrici di Householder

Nel presente paragrafo determineremo alcune matrici ortogonali di forma particolarmente semplice; queste saranno poi usate per il calcolo effettivo della fattorizzazione QR, incidentalmente fornendo una dimostrazione alternativa della sua esistenza.

Definizione 3. Sia $\mathbf{v}$ un vettore colonna non nullo di $\mathbb{R}^n$. Una matrice P della forma

$$P = I - \frac{2}{\mathbf{v}^T\mathbf{v}}\mathbf{v}\mathbf{v}^T$$

è detta *matrice di Householder* (alternativamente *riflessione di Householder*, *trasformazione di Householder*) associata a $\mathbf{v}$.

Osserviamo che, per ogni $\mathbf{x} \in \mathbb{R}^n$ abbiamo

$$\mathsf{P}\mathbf{x} = \mathbf{x} - \frac{2}{\|\mathbf{v}\|_2^2}\mathbf{v}\mathbf{v}^\mathsf{T}\mathbf{x} = \mathbf{x} - 2\frac{\mathbf{v}}{\|\mathbf{v}\|_2}\left\langle \frac{\mathbf{v}}{\|\mathbf{v}\|_2}, \mathbf{x}\right\rangle;$$

in particolare, l'immagine $\mathsf{P}\mathbf{x}$ dipende solamente dalla direzione di $\mathbf{v}$ (e non dal suo modulo). Se $\mathbf{x} \in \mathbf{v}^\perp$, allora $\mathsf{P}\mathbf{x} = \mathbf{x}$, in quanto $\mathbf{v}^\mathsf{T}\mathbf{x} = 0$. Inoltre,

$$\langle \mathsf{P}\mathbf{x}, \mathbf{v}\rangle = \langle \mathbf{x}, \mathbf{v}\rangle - 2\frac{1}{\|\mathbf{v}\|_2^2}\langle \mathbf{v}, \mathbf{x}\rangle \langle \mathbf{v}, \mathbf{v}\rangle = -\langle \mathbf{x}, \mathbf{v}\rangle;$$

pertanto $\langle \mathbf{x} + \mathsf{P}\mathbf{x}, \mathbf{v}\rangle = 0$, cioè

$$\mathbf{x} + \mathsf{P}\mathbf{x} \in \mathbf{v}^\perp.$$

Riassumendo, P corrisponde alla *riflessione* rispetto l'iperspazio $\mathbf{v}^\perp$. Infatti, sia $\Pi_\mathbf{v}$ la proiezione ortogonale nello spazio vettoriale generato da $\mathbf{v}$ e $\Pi_\mathbf{v}^\perp$ la proiezione sul suo ortogonale. Allora, osservato che

$$\mathbf{x} = \Pi_\mathbf{v}(\mathbf{x}) + \Pi_\mathbf{v}^\perp(\mathbf{x})$$

e posto $\mathbf{y} = \mathsf{P}\mathbf{x}$ si ha

$$\mathbf{x} + \mathbf{y} = \Pi_\mathbf{v}^\perp(\mathbf{x} + \mathbf{y}) = \Pi_\mathbf{v}^\perp(\mathbf{y}) + \Pi_\mathbf{v}^\perp(\mathbf{x}),$$

da cui segue

$$\Pi_\mathbf{v}(\mathbf{x}) = \mathbf{x} - \Pi_\mathbf{v}^\perp(\mathbf{x}) = \mathbf{y} - \Pi_\mathbf{v}^\perp(\mathbf{y}) = -\Pi_\mathbf{v}(\mathbf{y}),$$

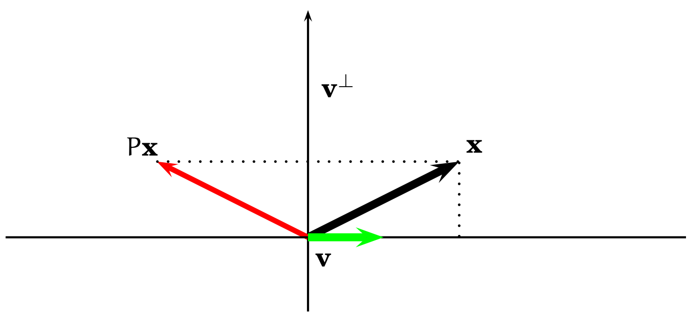
cioè la proiezione di $\mathbf{x}$ su $\mathbf{v}$ corrisponde alla proiezione di $\mathbf{y}$ su $\mathbf{v}$ cambiata di verso. Similmente, da $\Pi_\mathbf{v}^\perp(\mathbf{x}) \in \mathbf{v}^\perp$ segue che P agisce su $\Pi_\mathbf{v}(\mathbf{x})$ come l'identità. Pertanto,

$$\mathbf{y} = \Pi_\mathbf{v}(\mathbf{y}) + \Pi_\mathbf{v}^\perp(\mathbf{y}) = \mathsf{P}\Pi_\mathbf{v}(\mathbf{x}) + \mathsf{P}\Pi_\mathbf{v}^\perp(\mathbf{x}) = -\Pi_\mathbf{v}(\mathbf{x}) + \Pi_\mathbf{v}^\perp(\mathbf{x}).$$

Dall'unicità della decomposizione di un vettore in componenti ortogonali segue

$$\Pi_\mathbf{v}^\perp(\mathbf{y}) = \Pi_\mathbf{v}(\mathbf{x}),$$

cioè P fissa la proiezione di $\mathbf{x}$ su $\mathbf{v}^\perp$.



Teorema 6. *Ogni matrice di Householder* P *è una involuzione, cioè,* $\mathsf{P}^2 = \mathsf{I}$.

Dimostrazione. Abbiamo già visto che l'azione di P su un vettore $\mathbf{x}$ consiste nel fissare la proiezione di $\mathbf{x}$ su $\mathbf{v}^\perp$ e cambiare il segno della proiezione di $\mathbf{x}$ su $\mathbf{v}$. Pertanto, abbiamo che la proiezione di $\mathbf{y} = \mathsf{P}^2\mathbf{x}$ su $\mathbf{v}^\perp$ è uguale alla proiezione di $\mathbf{x}$, mentre la proiezione di $\mathbf{y}$ su $\mathbf{v}$ è uguale alla proiezione di $\mathbf{x}$ con il segno cambiato due volte (e quindi coincide con questa). Poiché $\mathbb{R}^n = \langle \mathbf{v} \rangle \oplus \mathbf{v}^\perp$, la tesi segue. □

Teorema 7. *Ogni matrice di Householder è ortogonale.*

Dimostrazione. Ogni matrice di Householder è simmetrica, infatti

$$\mathsf{P}^\mathsf{T} = (\mathsf{I} - \frac{2}{\|\mathbf{v}\|_2^2}\mathbf{v}\mathbf{v}^\mathsf{T})^\mathsf{T} = \mathsf{I} - \frac{2}{\|\mathbf{v}\|_2^2}(\mathbf{v}\mathbf{v}^\mathsf{T})^\mathsf{T} = \mathsf{I} - \frac{2}{\|\mathbf{v}\|_2^2}(\mathbf{v}\mathbf{v}^\mathsf{T}) = \mathsf{P}.$$

La tesi segue dal fatto che $\mathsf{P}^{-1} = \mathsf{P}$. □

Sia ora $\mathbf{f}$ un vettore non nullo di $\mathbb{R}^n$ e poniamo $\mathbf{e} = \frac{\mathbf{f}}{\|\mathbf{f}\|_2}$. Chiaramente, $\|\mathbf{e}\|_2 = 1$. Supponiamo che, assegnato $\mathbf{x} \in \mathbb{R}^n$, si voglia trovare una trasformazione di Householder tale che $\mathsf{P}\mathbf{x} = \alpha'\mathbf{f} = \alpha\mathbf{e}$ Consideriamo a tal proposito il vettore $\mathbf{v}^\pm = \mathbf{x} \pm \|\mathbf{x}\|_2\mathbf{e}$. Indichiamo con $\mathsf{P}^\pm$ le matrici di Householder associate a questi vettori. Si noti che, in generale,

$$\langle \mathbf{v}^-, \mathbf{v}^+ \rangle = \langle \mathbf{x} - \|\mathbf{x}\|_2\mathbf{e}, \mathbf{x} + \|\mathbf{x}\|_2\mathbf{e} \rangle = \|\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2\|\mathbf{e}\|_2 = 0.$$

Tenuto conto che $\|\mathbf{e}\|_2 = 1$, si ha

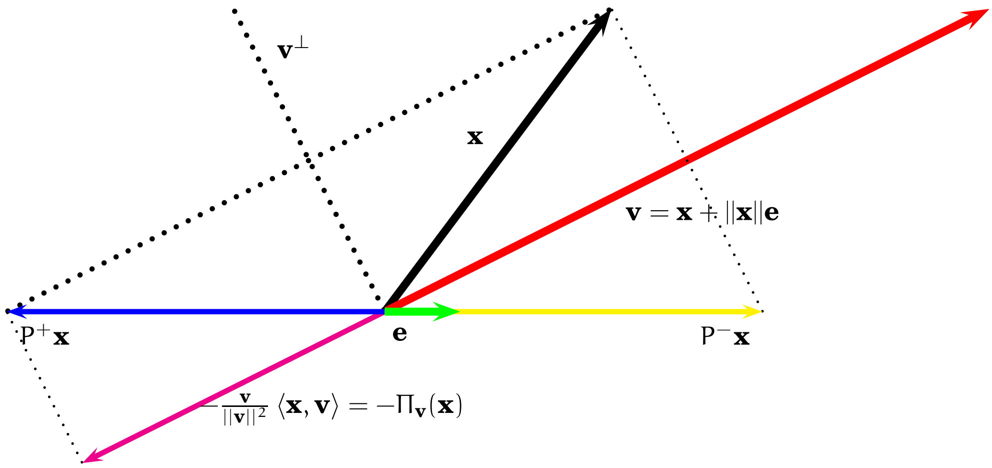
$$\|\mathbf{v}\|_2^2 = \langle \mathbf{x} \pm \|\mathbf{x}\|_2\mathbf{e}, \mathbf{x} \pm \|\mathbf{x}\|_2\mathbf{e} \rangle = 2\|\mathbf{x}\|_2^2 \pm 2\, \langle \mathbf{x}, \mathbf{e} \rangle\, \|\mathbf{x}\|_2;$$

$$\langle \mathbf{v}, \mathbf{x} \rangle = \|\mathbf{x}\|_2^2 \pm \|\mathbf{x}\|_2 \langle \mathbf{x}, \mathbf{e} \rangle .$$

Da questo si deduce

$$\begin{aligned}
\|\mathbf{v}\|_2^2\mathsf{P}^\pm\mathbf{x} = \quad & \|\mathbf{v}\|_2^2 \left(\mathbf{x} - 2\mathbf{v}\frac{\langle \mathbf{v}, \mathbf{x} \rangle}{\|\mathbf{v}\|_2^2} \right) = \|\mathbf{v}\|_2^2\mathbf{x} - 2\mathbf{v}\, \langle \mathbf{v}, \mathbf{x} \rangle = \\
& \left(\|\mathbf{v}\|_2^2 - 2\, \langle \mathbf{v}, \mathbf{x} \rangle \right) \mathbf{x} \mp 2\|\mathbf{x}\|_2 \langle \mathbf{v}, \mathbf{x} \rangle\, \mathbf{e} = \\
& \left(2\|\mathbf{x}\|_2^2 \pm 2\, \langle \mathbf{v}, \mathbf{x} \rangle\, \|\mathbf{x}\|_2 - 2\|\mathbf{x}\|_2^2 \mp 2\, \langle \mathbf{v}, \mathbf{x} \rangle \right) \mathbf{x} \mp 2\|\mathbf{x}\|_2 \langle \mathbf{v}, \mathbf{x} \rangle\, \mathbf{e} = \\
& \mp 2\|\mathbf{x}\|_2 \langle \mathbf{v}, \mathbf{x} \rangle\, \mathbf{e}.
\end{aligned}$$

In particolare, si vede che $\mathsf{P}\mathbf{x}$ è un multiplo scalare di $\mathbf{e}$ e, quindi, anche di $\mathbf{f}$.



Indicheremo la matrice di Householder sopra determinata, che definisce una trasformazione lineare che manda il vettore $\mathbf{x}$ in un multiplo scalare di un versore $\mathbf{e}$, con il simbolo $H^{\pm}(\mathbf{x}, \mathbf{e})$ a seconda che si sia scelto il segno positivo o negativo per la norma che compare nella scrittura del vettore $\mathbf{v}$. Per concludere, osserviamo che vale sempre

$$H^{-}(\mathbf{x}, \mathbf{e}) = -H^{+}(\mathbf{x}, \mathbf{e}).$$

Fattorizzazione QR con matrici di Householder

È possibile utilizzare le matrici di Householder per determinare la fattorizzazione QR di una matrice $M \in \mathrm{GL}_n(\mathbb{R})$. Questo metodo presenta alcuni vantaggi pratici rispetto l'ortonormalizzazione di Gram-Schmidt vista precedentemente; in particolare esso non richiede il calcolo dell'inversa di una matrice triangolare superiore.

Poiché ogni trasformazione di Householder è ortogonale, ci basta far vedere che esiste un prodotto di matrici del tipo $H^{\pm}(\mathbf{x_i}, \mathbf{e_j})$ che trasforma M in una matrice triangolare superiore con entrate sulla diagonale principale tutte positive. A questo punto, l'unicità della fattorizzazione garantisce che quanto abbiamo ottenuto sia effettivamente la decomposizione QR cercata.

Sia dunque $M = \begin{pmatrix} M_1 & M_2 & \dots & M_n \end{pmatrix}$.

1. Sia $\mathbf{e_1} = (1, 0, \dots, 0)$ il primo vettore della base canonica di $\mathbb{R}^n$ e consideriamo innanzi tutto la matrice di Householder $H_1 = H^{\pm}(M_1, \mathbf{e_1})$, ove la scelta del segno è fatta in modo da garantire che il primo coefficiente del vettore $H_1 M_1$ sia positivo. Allora, la prima colonna della matrice $H_1 M$ avrà un valore positivo nella prima riga, mentre sarà 0 altrove. Scriviamo $R_1 = H_1 M$. Osserviamo che mediante H_1 la matrice si modifica come

illustrato nel seguente schema

$$\begin{pmatrix} \square & \square & \dots & \square \\ \square & \square & \dots & \square \\ \vdots & & \vdots & \\ \square & \square & \dots & \square \end{pmatrix} \mapsto \begin{pmatrix} + & \lozenge & \dots & \lozenge \\ 0 & \lozenge & \dots & \lozenge \\ \vdots & & \vdots & \\ 0 & \lozenge & \dots & \lozenge \end{pmatrix},$$

ove con $+$ si è indicata una entrata sicuramente strettamente maggiore di 0, mentre con $\lozenge$ entrate da determinarsi, a priori differenti rispetto quelle originaria $\square$. Sottolineiamo che la trasformazione di Householder cambia, in generale, *tutte* le colonne della matrice M e non solamente la prima.

2. Fissiamo ora $i > 1$ e supponiamo che tutte le colonne con indice $j < i$ siano già in forma triangolare positiva, ovvero che per tutte le colonne M_j con $j < i$ si abbia:

   (a) $m_{jj} > 0$;

   (b) per $k > j$, $m_{kj} = 0$.

   Indichiamo con R_{i-1} questa matrice.

3. Vogliamo modificare M_i, la colonna i–esima di M, in modo che anche questa soddisfi le proprietà sopra enumerate. A tal fine, scriviamo la matrice di Householder $\widetilde{H_i}$ di dimensioni $(n-i+1) \times (n-i+1)$ che manda il vettore $\widetilde{\mathbf{M_i}} = (m_{ii}, m_{i+1,i}, \dots, m_{ni}) \in \mathbb{R}^{n-i+1}$ in un multiplo scalare positivo del vettore $\mathbf{e_i} = (1, 0, \dots, 0) \in \mathbb{R}^{n-i+1}$ e costruiamo ora la matrice diagonale a blocchi

$$H_i = \begin{pmatrix} I_{i-1} & 0 \\ 0 & \widetilde{H_i} \end{pmatrix}$$

4. Osserviamo che il prodotto di $H_i R_{i-1}$ non modifica le prime $i-1$ colonne di R_i (in quanto H_i agisce su questi vettori come la matrice identica, mentre manda la colonna i–esima in un vettore che ha tutti 0 nelle posizioni $i+1, i+2, \dots, n$ e un coefficiente maggiore di 0 in posizione i–esima. Pertanto la matrice

$$R_i = H_i R_{i-1}$$

soddisfa le ipotesi del punto precedente rispetto le prime i colonne. Rappresentando in modo schematico, accade quanto segue:

$$\begin{pmatrix} + & \square & \square & \square & \dots & \square \\ 0 & + & \square & \square & & \square \\ 0 & 0 & \square & \square & & \square \\ \vdots & & \vdots & & \vdots & \\ 0 & 0 & \square & \square & & \square \end{pmatrix} \mapsto \begin{pmatrix} + & \square & \Diamond & \Diamond & \dots & \Diamond \\ 0 & + & \Diamond & \Diamond & & \Diamond \\ 0 & 0 & + & \Diamond & & \Diamond \\ \vdots & & \vdots & & \vdots & \\ 0 & 0 & 0 & \Diamond & & \Diamond \end{pmatrix}.$$

5. Iterando la procedura sopra descritta $n-1$ volte si ottiene

$$(H_n H_{n-1} \cdots H_1) M = R_n$$

ove $H = (H_n H_{n-1} \cdots H_1)$ è una matrice ortogonale, mentre R_n è una matrice triangolare superiore con entrate positive sulla diagonale principale. Moltiplicando per $Q = H^T = H^{-1}$ l'uguaglianza di cui sopra si giunge ora alla fattorizzazione

$$M = QR$$

richiesta.