Matrici ortogonali, simmetriche, semi-definite e definite; la SVD di una matrice

S. Maset
Dipartimento di Matematica e Geoscienze
Università di Trieste
maset@units.it

June 15, 2018

Nel seguito considereremo $\mathbb{R}^n$ dotato del prodotto scalare canonico $\langle \cdot , \cdot \rangle$.

1 Matrici ortogonali

Definizione 1 *Una matrice $Q \in \mathbb{R}^{n\times n}$ si dice* ortogonale *se le sue colonne $q^{(1)}, \ldots, q^{(n)}$ sono ortonormali e quindi costituiscono una base ortonormale di $\mathbb{R}^n$.*

Un esempio di matrice ortogonale è I_n le cui colonne sono i vettori della base canonica di $\mathbb{R}^n$.

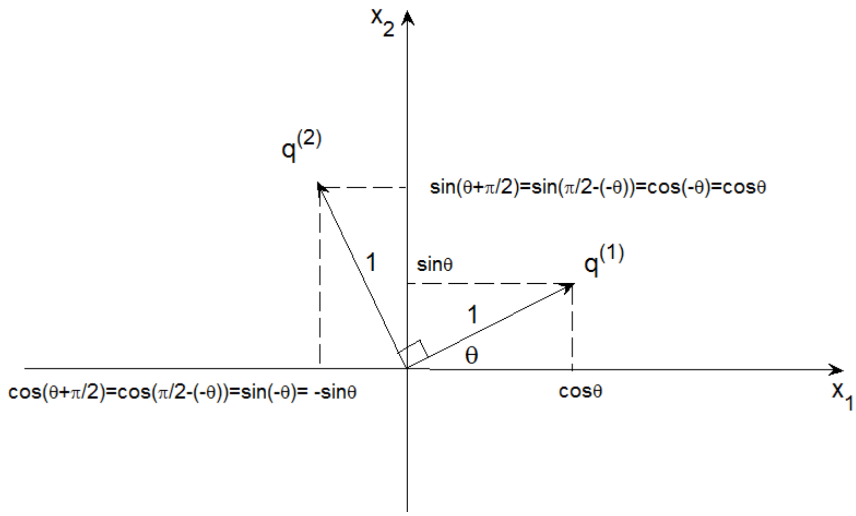
Le matrici ortogonali 2×2 hanno la forma

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \text{ per qualche } \theta \in [0, 2\pi)$$

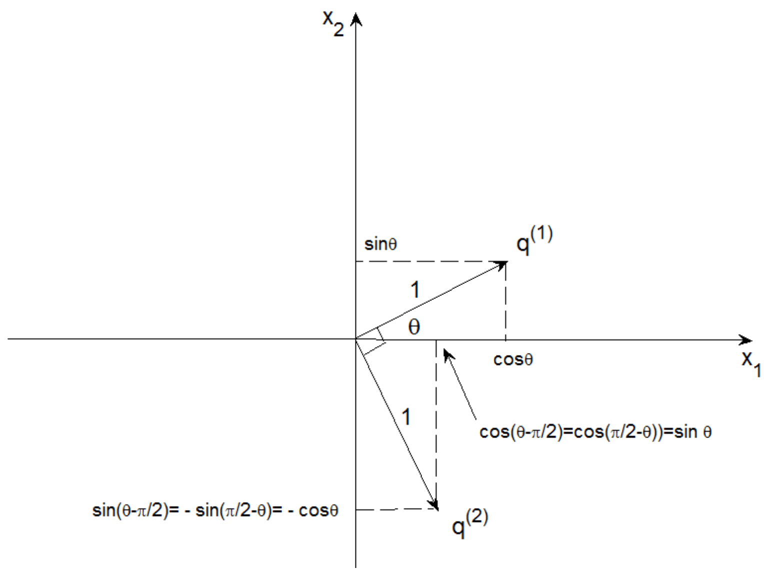
oppure la forma

$$\begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} \text{ per qualche } \theta \in [0, 2\pi).$$

La prima forma si ottiene quando $q^{(2)}$ è ruotato di $\frac{\pi}{2}$ in senso anti-orario rispetto a $q^{(1)}$:



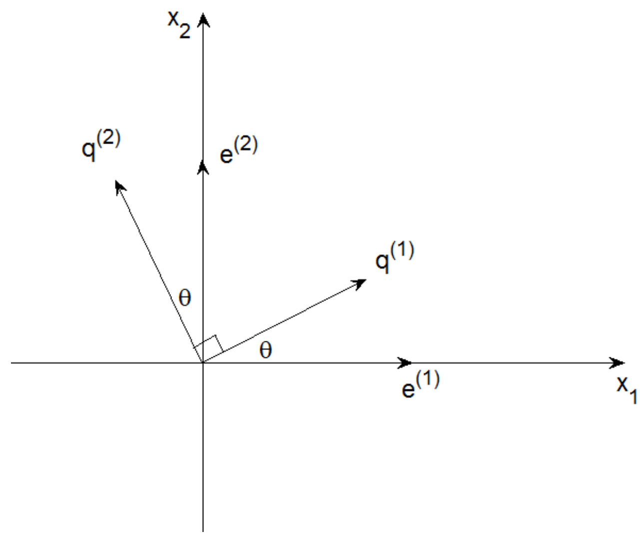
Invece, la seconda forma si ottiene quando $q^{(2)}$ è ruotato di $\frac{\pi}{2}$ in senso orario rispetto a $q^{(1)}$:



La prima forma di matrice ortogonale 2×2 corrisponde a una rotazione attorno all'origine.

Infatti, sia $Q \in \mathbb{R}^{2\times 2}$ una matrice ortogonale della prima forma. Allora

$$Qe^{(1)} = q^{(1)} = \text{“}e^{(1)} \text{ ruotato di } \theta \text{ in senso anti-orario attorno all'origine”}$$
$$Qe^{(2)} = q^{(2)} = \text{“}e^{(2)} \text{ ruotato di } \theta \text{ in senso anti-orario attorno all'origine”}.$$

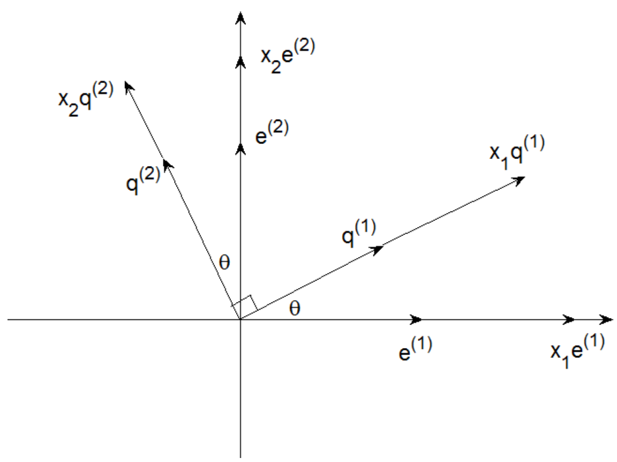


Per $x = x_1 e^{(1)} + x_2 e^{(2)} \in \mathbb{R}^2$, si ha

$$Qx = Q(x_1 e^{(1)} + x_2 e^{(2)}) = x_1 Q e^{(1)} + x_2 Q e^{(2)} = x_1 q^{(1)} + x_2 q^{(2)}.$$

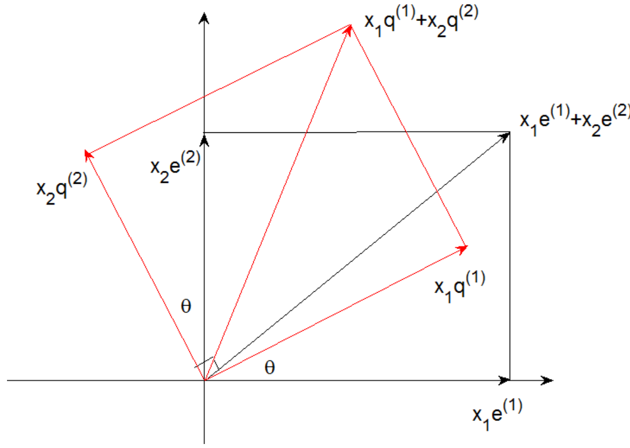
Ma

$$\begin{aligned} x_1 q^{(1)} &= \text{“}x_1 e^{(1)} \text{ ruotato di } \theta \text{ in senso anti-orario attorno all'origine”} \\ x_2 q^{(2)} &= \text{“}x_2 e^{(2)} \text{ ruotato di } \theta \text{ in senso anti-orario attorno all'origine”}. \end{aligned}$$



Inoltre

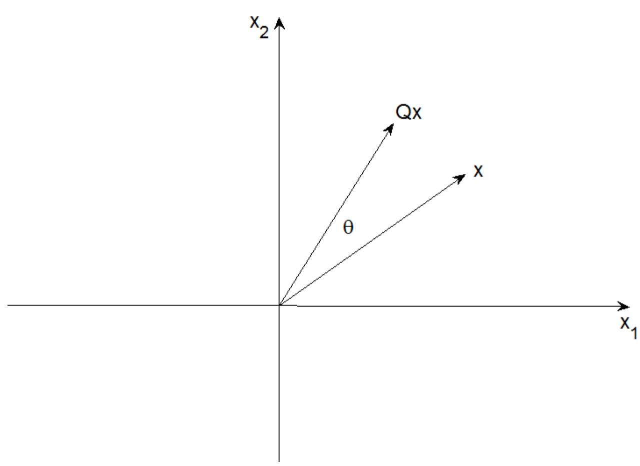
$$x_1 q^{(1)} + x_2 q^{(2)} = \text{“}x_1 e^{(1)} + x_2 e^{(2)} \text{ ruotato di } \theta \text{ in senso anti-orario attorno all'origine”}.$$



Il rettangolo nero ruotando si trasforma nel rettangolo rosso.

Si conclude che l'operatore lineare $L_Q : \mathbb{R}^2 \to \mathbb{R}^2$ associato alla matrice Q è un operatore di rotazione: per $x \in \mathbb{R}^2$,

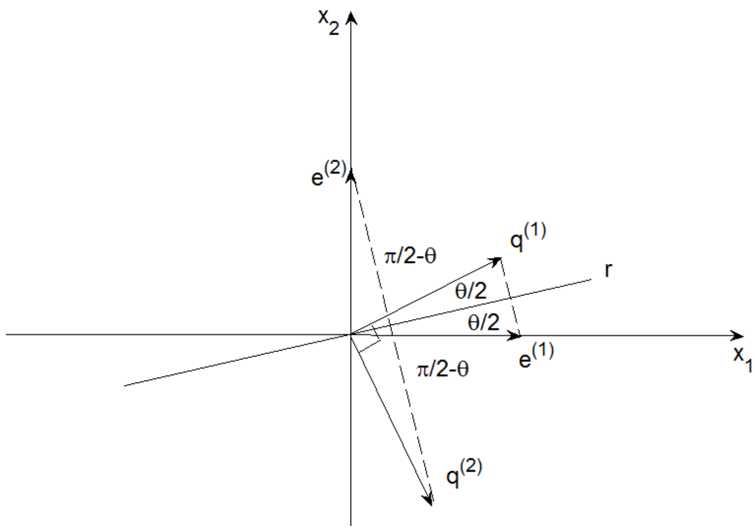
$$\begin{aligned} L_Q(x) &= Qx = x_1 q^{(1)} + x_2 q^{(2)} \\ &= \text{“}x = x_1 e^{(1)} + x_2 e^{(2)} \text{ ruotato di } \theta \text{ in senso anti-orario} \\ &\quad\ \text{attorno all'origine”.} \end{aligned}$$



La seconda forma di matrice ortogonale 2×2 corrisponde a una riflessione rispetto a una retta passante per l'origine, ovvero a una rotazione nello spazio di un angolo π con asse di rotazione la retta.

Infatti, sia $Q \in \mathbb{R}^{2\times 2}$ una matrice ortogonale della seconda forma e sia r la retta passante per l'origine formante un angolo $\frac{\theta}{2}$ con la direzione positiva del primo asse. Allora

$$Qe^{(1)} = q^{(1)} = \text{ “}e^{(1)} \text{ riflesso rispetto a } r\text{”},$$
$$Qe^{(2)} = q^{(2)} = \text{ “}e^{(2)} \text{ riflesso rispetto a } r\text{”}.$$

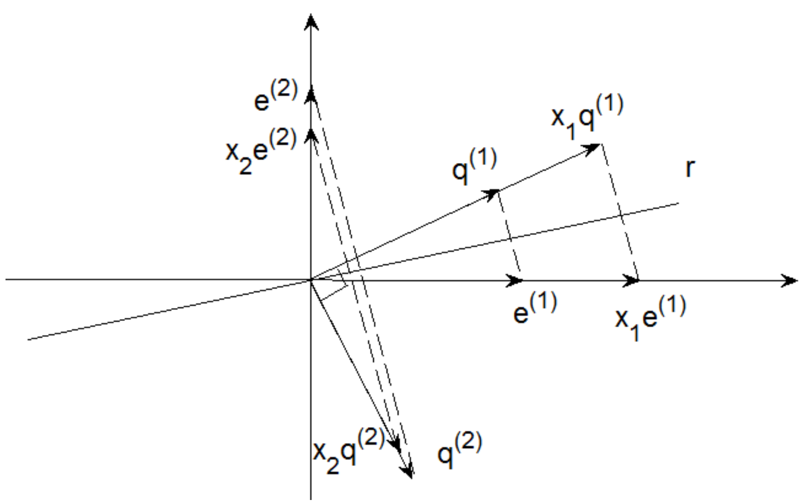


Per $x = x_1 e^{(1)} + x_2 e^{(2)} \in \mathbb{R}^2$, si ha

$$Qx = Q(x_1 e^{(1)} + x_2 e^{(2)}) = x_1 Q e^{(1)} + x_2 Q e^{(2)} = x_1 q^{(1)} + x_2 q^{(2)}.$$

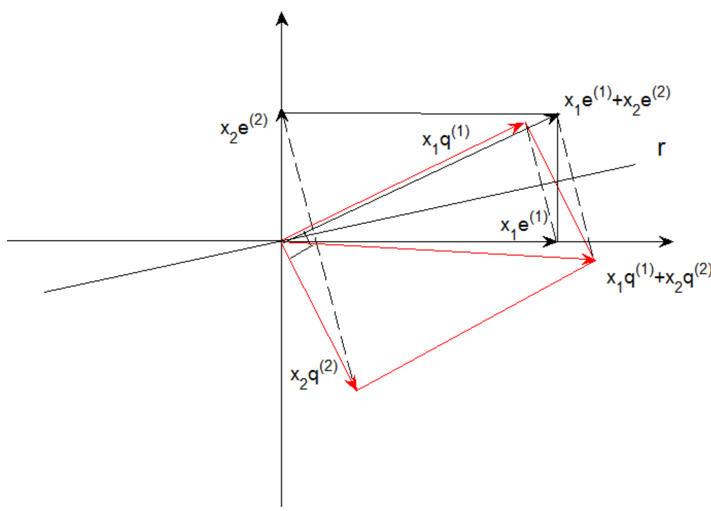
Ma

$$x_1 q^{(1)} = \text{ “}x_1 e^{(1)} \text{ riflesso rispetto a } r\text{”}$$
$$x_2 q^{(2)} = \text{ “}x_2 e^{(2)} \text{ riflesso rispetto a } r\text{”}.$$



Inoltre

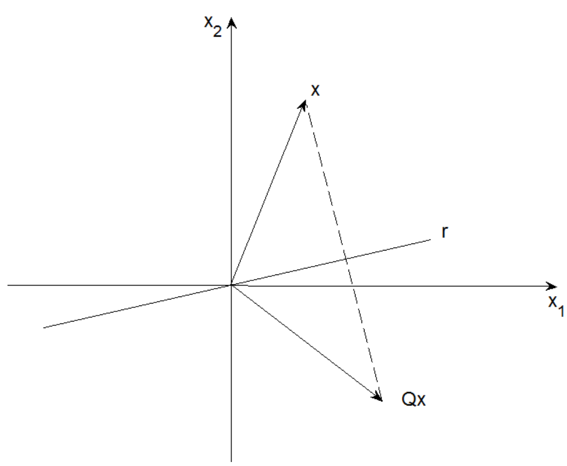
$$x_1 q^{(1)} + x_2 q^{(2)} = \text{“}x_1 e^{(1)} + x_2 e^{(2)} \text{ riflesso rispetto a } r\text{”}.$$



Il rettangolo nero riflettendosi sulla retta r si trasforma nel rettangolo rosso.

Si conclude che l'operatore lineare $L_Q : \mathbb{R}^2 \to \mathbb{R}^2$ associato alla matrice Q è un operatore di riflessione: per $x \in \mathbb{R}^2$,

$$\begin{aligned} L_Q(x) &= Qx = x_1 q^{(1)} + x_2 q^{(2)} \\ &= \text{“}x = x_1 e^{(1)} + x_2 e^{(2)} \text{ riflesso rispesso a } r\text{”}. \end{aligned}$$



1.1 Proprietà delle matrici ortogonali

Sia $Q \in \mathbb{R}^{n\times n}$. Si ha che Q è ortogonale se e solo se $Q^T Q = I_n$.

Infatti, più in generale, per una matrice $B \in \mathbb{R}^{n \times k}$ di colonne $b^{(1)}, \ldots, b^{(k)}$ e quindi B^T di righe $\left(b^{(1)}\right)^T, \ldots, \left(b^{(k)}\right)^T$ si ha

$$
\begin{aligned}
& B^T B = I_k \\
\Leftrightarrow\ & \forall i,j \in \{1,\ldots,k\} : \ \left(B^T B\right)(i,j) = \left(b^{(i)}\right)^T b^{(j)} = \left\langle b^{(i)}, b^{(j)} \right\rangle \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad = I_k\,(i,j) = \begin{cases} 0 \text{ se } i \neq j \\ 1 \text{ se } i = j \end{cases} \\
\Leftrightarrow\ & b^{(1)}, \ldots, b^{(k)} \text{ sono ortonormali.}
\end{aligned}
$$

Se Q è ortogonale, allora Q è non singolare e $Q^{-1} = Q^T$.
Infatti, da $Q^T Q = I_n$ segue che Q è non singolare con $Q^{-1} = Q^T$.

Si ha poi che Q è ortogonale se e solo se $QQ^T = I_n$.
Infatti, se Q è ortogonale, $QQ^T = QQ^{-1} = I_n$. Viceversa, da $QQ^T = I_n$ segue che Q è non singolare con $Q^{-1} = Q^T$ da cui $Q^T Q = Q^{-1} Q = I_n$ e quindi Q è ortogonale.

La moltiplicazione per una matrice ortogonale preserva il prodotto scalare canonico $\langle \cdot\, , \cdot \rangle$:

$$\forall Q \in \mathbb{R}^{n \times n} \text{ ortogonale } \forall x,y \in \mathbb{R}^n : \langle Qx, Qy \rangle = \langle x, y \rangle .$$

Infatti, sia Q ortogonale e siano $x, y \in \mathbb{R}^n$. Si ha

$$\langle Qx, Qy \rangle = \langle x, Q^T Q y \rangle = \langle x, I_n y \rangle = \langle x, y \rangle .$$

Come conseguenza si ha che la moltiplicazione per una matrice ortogonale preserva anche la norma euclidea:

$$\forall Q \in \mathbb{R}^{n \times n} \text{ ortogonale } \forall x \in \mathbb{R}^n : \|Qx\|_2 = \|x\|_2 .$$

Infatti, per Q ortogonale e $x \in \mathbb{R}^n$, si ha

$$\|Qx\|_2^2 = \langle Qx, Qx \rangle = \langle x, x \rangle = \|x\|_2^2 .$$

Infine, come ulteriore conseguenza, si ha che la moltiplicazione per una matrice ortogonale preserva la metrica euclidea:

$$\forall Q \in \mathbb{R}^{n \times n} \text{ ortogonale } \forall x,y \in \mathbb{R}^n : \rho_2(Qx, Qy) = \rho_2(x, y).$$

Infatti, per Q ortogonale e $x, y \in \mathbb{R}^n$, si ha

$$\rho_2(Qx, Qy) = \|Qy - Qx\|_2 = \|Q(y - x)\|_2 = \|y - x\|_2 = \rho_2(x, y).$$

Nel caso delle matrici ortogonali 2×2, la preservazione del prodotto scalare canonico implica che in una rotazione e in una riflessione si preservano le distanze tra punti e anche gli angoli: dati tre punti $x, y, z \in \mathbb{R}^2$, l'angolo $y\widehat{x}z$ di vertice x è dato da

$$\begin{aligned}
\arccos \frac{\langle y-x, z-x\rangle}{\|y-x\|_2 \, \|z-x\|_2} &= \arccos \frac{\langle Q(y-x), Q(z-x)\rangle}{\|Q(y-x)\|_2 \, \|Q(z-x)\|_2} \\
&= \arccos \frac{\langle Qy-Qx, Qz-Qx)\rangle}{\|Qy-Qx\|_2 \, \|Qz-Qx)\|_2}.
\end{aligned}$$

ed è uguale all'angolo $Qy\widehat{Qx}Qz$ di vertice Qx.

Vediamo come si comportano le matrici ortogonali rispetto ad alcune operazioni matriciali:

1) $\forall Q, R \in \mathbb{R}^{n\times n}$ ortogonali $: QR$ è ortogonale.

2) $\forall Q \in \mathbb{R}^{n\times n}$ ortogonale $: Q^{-1} = Q^T$ è ortogonale.

Dimostrazione di 1). Siano $Q, R \in \mathbb{R}^{n\times n}$ ortogonali. Si ha

$$(QR)^T QR = R^T Q^T QR = R^T I_n R = R^T R = I_n$$

e quindi QR è ortogonale.

Dimostrazione di 2). Sia $Q \in \mathbb{R}^{n\times n}$ ortogonale. Si ha

$$\left(Q^T\right)^T Q^T = QQ^T = QQ^{-1} = I_n$$

e quindi Q^T è ortogonale.

La proprietà 2) dice che se una matrice $Q \in \mathbb{R}^{n\times n}$ ha colonne ortonormali, allora anche le righe (che sono le colonne di Q^T) sono ortonormali.

Esercizio. Dato nel piano un sistema di riferimento cartesiano S, si consideri il nuovo sistema di riferimento cartesiano S' ottenuto ruotando gli assi attorno all'origine di un angolo θ in senso anti-orario. Trovare come si trasformano le coordinate di un punto passando da S a S'. In altri termini, detti $q^{(1)}$ e $q^{(2)}$ i nuovi versori degli assi e dato un punto $x = (x_1, x_2) \in \mathbb{R}^2$, dove x_1 e x_2 sono le coordinate di x in S, si tratta di trovare la coppia $x' = (x'_1, x'_2) \in \mathbb{R}^2$, dove x'_1 e x'_2 sono le coordinate di x in S', tale che

$$x = x'_1 q^{(1)} + x'_2 q^{(2)} = Qx',$$

dove Q è la matrice di colonne $q^{(1)}$ e $q^{(2)}$. Esprimere x'_1 e x'_2 in termini di x_1 e x_2.

Esercizio. Fare lo stesso quando il nuovo sistema di riferimento cartesiano S' è invece ottenuto con una riflessione degli assi rispetto a una retta passante per l'origine.

Esercizio. Si consideri nel piano $\mathbb{R}^2$ la retta r passante per l'origine di equazione $x_2 = mx_1$, dove $m \in \mathbb{R}$ è la pendenza della retta. Scrivere la matrice di riflessione rispetto a r in termini di m.

Esercizio. Provare che se $Q \in \mathbb{R}^{n\times n}$ è ortogonale, allora $-Q$ è ortogonale.

Esercizio. Quali sono le matrici ortogonali 1×1?

Esercizio. Mostrare che non valgono le seguenti proprietà

- $\forall Q, R \in \mathbb{R}^{n\times n}$ ortogonali : $Q + R$ è ortogonale;
- $\forall \alpha \in \mathbb{R}\ \forall Q \in \mathbb{R}^{n\times n}$ ortogonale : αQ è ortogonale.

Suggerimento: considerare la matrice ortogonale I_n.

Esercizio. Si è visto che le matrici ortogonali di $\mathbb{R}^{n\times n}$ preservano il prodotto scalare canonico $\langle\ \cdot, \cdot\ \rangle$ di $\mathbb{R}^n$. Provare ora che solo le matrici ortogonali possono farlo. Provare cioè che, data $Q \in \mathbb{R}^{n\times n}$, se

$$\forall x, y \in \mathbb{R}^n : \langle Qx, Qy \rangle = \langle x, y \rangle\,,$$

allora Q è ortogonale. Suggerimento: siano $q^{(1)} = Qe^{(1)}, \ldots,\ q^{(n)} = Qe^{(n)}$ le colonne di Q, dove $e^{(1)}, \ldots, e^{(n)}$ è la base canonica di $\mathbb{R}^n$; determinare $< q^{(i)}, q^{(j)} >$ per $i, j \in \{1, \ldots, n\}$.

Esercizio. Verificare la condizione $Q^T Q = I_2$ caratterizzante l'ortogonalità per le matrici 2×2 di rotazione e riflessione.

Esercizio. Sia $Q \in \mathbb{R}^{n\times n}$ ortogonale. Provare che $\det(Q) = 1$ o $\det(Q) = -1$. Suggerimento: usare il fatto che $Q^T Q = I_n$.

Esercizio. Qual è il determinante delle matrici 2×2 di rotazione e riflessione?

Esercizio. Verificare che la trasposta di una matrice $Q \in \mathbb{R}^{2\times 2}$ di rotazione di un angolo θ è la matrice di rotazione dell'angolo $-\theta$. Spiegare questo fatto pensando all'operatore inverso dell'operatore L_Q.

Esercizio. Verificare che la trasposta di una matrice $Q \in \mathbb{R}^{2\times 2}$ di riflessione è la matrice Q. Spiegare questo fatto pensando all'operatore inverso dell'operatore L_Q.

Esercizio. Sia $Q \in \mathbb{R}^{n\times n}$ ortogonale. Provare che per ogni autovalore $\lambda \in \mathbb{R}$ di Q si ha $|\lambda| = 1$. Suggerimento: determinare $\|Qx\|_2$ per un autovettore $x \in \mathbb{R}^n$ relativo a λ.

2 Matrici simmetriche

Definizione 2 *Una matrice $A \in \mathbb{R}^{n\times n}$ si dice* simmetrica *se $A^T = A$, cioè*

$$\forall i, j \in \{1, \ldots, n\} : A(i,j) = A^T(i,j) = A(j,i)\,.$$

Chiaramente, le matrici simmetriche 2×2 hanno la forma

$$\begin{bmatrix} a & b \\ b & d \end{bmatrix}.$$

Notare che, per $B \in \mathbb{R}^{m\times n}$, le matrici $B^TB \in \mathbb{R}^{n\times n}$ e $BB^T \in \mathbb{R}^{m\times m}$ sono simmetriche.

Infatti, si ha

$$\left(B^TB\right)^T = B^T\left(B^T\right)^T = B^TB \text{ e } \left(BB^T\right)^T = \left(B^T\right)^T B^T = BB^T.$$

2.1 Proprietà delle matrici simmetriche

Un'importante proprietà delle matrici simmetriche è la seguente.

Teorema 3 *Sia $A \in \mathbb{R}^{n\times n}$ simmetrica. Se*

$$\forall x \in \mathbb{R}^n : \langle x, Ax\rangle = x^TAx = 0 \tag{1}$$

allora $A = 0$.

Dimostrazione. Assumiamo (1). Siano $k, l \in \{1, \ldots, n\}$. Proviamo che $a_{kl} = 0$.

Prendendo $x = e^{(k)} + e^{(l)}$ in (1) si ha

$$\left\langle e^{(k)} + e^{(l)}, A\left(e^{(k)} + e^{(l)}\right)\right\rangle = 0.$$

Ma

$$\begin{aligned}
&\left\langle e^{(k)} + e^{(l)}, A\left(e^{(k)} + e^{(l)}\right)\right\rangle = \left\langle e^{(k)} + e^{(l)}, Ae^{(k)} + Ae^{(l)}\right\rangle \\
&= \underbrace{\left\langle e^{(k)}, Ae^{(k)}\right\rangle}_{=0} + \left\langle e^{(k)}, Ae^{(l)}\right\rangle + \left\langle e^{(l)}, Ae^{(k)}\right\rangle + \underbrace{\left\langle e^{(l)}, Ae^{(l)}\right\rangle}_{=0} \\
&= \underbrace{\left\langle e^{(k)}, Ae^{(l)}\right\rangle}_{=\left\langle e^{(k)}, A(:,l)\right\rangle = a_{kl}} + \underbrace{\left\langle e^{(l)}, Ae^{(k)}\right\rangle}_{=\left\langle e^{(l)}, A(:,k)\right\rangle = a_{lk}} \\
&= a_{kl} + a_{lk} \\
&= 2a_{kl} \text{ poichè } A \text{ è simmetrica}
\end{aligned}$$

Pertanto $2a_{kl} = 0$, da cui $a_{kl} = 0$. ■

Esercizio. Il precedente teorema non è più vero se si lascia cadere l'ipotesi che A sia simmetrica. Trovare una matrice $A \in \mathbb{R}^{n\times n}$ non nulla del tipo

$$A = \begin{bmatrix} 0 & b & 0 & \cdots & 0 \\ c & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

tale che

$$\forall x \in \mathbb{R}^n : \langle x, Ax \rangle = x^T Ax = 0.$$

Vediamo ora come si comportano le matrici simmetriche rispetto ad alcune operazioni matriciali.

1) $\forall A, B \in \mathbb{R}^{n\times n}$ simmetriche $: A + B$ è simmetrica.

2) $\forall \alpha \in \mathbb{R}\ \forall A \in \mathbb{R}^{n\times n}$ simmetrica $: \alpha A$ è simmetrica.

3) $\forall A \in \mathbb{R}^{n\times n}$ simmetrica non singolare $: A^{-1}$ è simmetrica.

Dimostrazione di 1). Siano $A, B \in \mathbb{R}^{n\times n}$ simmetriche. Si ha

$$(A + B)^T = A^T + B^T = A + B$$

e quindi $A + B$ è simmetrica.

Dimostrazione di 2). Siano $\alpha \in \mathbb{R}$ e $A \in \mathbb{R}^{n\times n}$ simmetrica. Si ha

$$(\alpha A)^T = \alpha A^T = \alpha A$$

e quindi αA è simmetrica.

Dimostrazione di 3). Sia $A \in \mathbb{R}^{n\times n}$ simmetrica non singolare. Si ha

$$(A^{-1})^T = (A^T)^{-1} = A^{-1}.$$

e quindi A^{-1} è simmetrica.

Esercizio. Mostrare che, per $n > 1$, non vale la seguente proprietà

- $\forall A, B \in \mathbb{R}^{n\times n}$ simmetriche : AB è simmetrica.

Suggerimento: considerare una matrice simmetrica A e una matrice simmetrica $B = \mathrm{diag}(b_1, \ldots, b_n)$ e poi calcolare $AB(n, 1)$ e $AB(1, n)$.

Esercizio. Siano $A, B \in \mathbb{R}^{n\times n}$ simmetriche. Provare che AB è simmetrica se e solo se $AB = BA$.

Esercizio. Si è visto che le matrici ortogonali di $\mathbb{R}^{n\times n}$ preservano la norma euclidea $\|\ \cdot\ \|_2$ di $\mathbb{R}^n$. Provare ora che solo le matrici ortogonali possono farlo. Provare cioè che, data $Q \in \mathbb{R}^{n\times n}$, se

$$\forall x \in \mathbb{R}^n : \|Qx\|_2 = \|x\|_2\,, \tag{2}$$

allora Q è ortogonale. Suggerimento: da (2) espressa con le norme al quadrato, ottenere

$$\forall x \in \mathbb{R}^n : \langle x, Ax \rangle = 0$$

per un'opportuna matrice simmetrica A; seguirà allora che $A = \ldots$ e quindi che Q è ortogonale.

3 Diagonalizzazione con trasformazione ortogonale

Ricordiamo che, data $A \in \mathbb{R}^{n\times n}$, si dice che A è *diagonalizzabile* se A è simile ad una matrice diagonale, vale a dire esiste una matrice non singolare $V \in \mathbb{R}^{n\times n}$, detta *matrice di trasformazione*, tale che $D = V^{-1}AV$ è diagonale.

Questa è la diagonalizzabilità di A in $\mathbb{R}$, che si contrappone alla diagonalizzabilità di A in $\mathbb{C}$, vale a dire esiste una matrice non singolare $V \in \mathbb{C}^{n\times n}$ tale che $D = V^{-1}AV \in \mathbb{C}^{n\times n}$ è diagonale.

Come è noto dal corso di Geometria, esistono matrici non diagonalizzabili in $\mathbb{R}$ che lo diventano in $\mathbb{C}$, come esistono matrici che non sono diagonalizzabili nemmeno in $\mathbb{C}$.

Ad esempio,

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

non è diagonalizzabile in $\mathbb{R}$ ma lo è in $\mathbb{C}$, mentre

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

non è diagonalizzabile né in $\mathbb{R}$ né in $\mathbb{C}$.

Dire che A è simile a $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ con matrice di trasformazione V di colonne $v^{(1)}, \ldots, v^{(n)}$, equivale a dire che $\lambda_1, \ldots, \lambda_n$ sono gli autovalori di A e $v^{(1)}, \ldots, v^{(n)}$ è una base di $\mathbb{R}^n$ tale che, per $i \in \{1, \ldots, n\}$, $v^{(i)}$ è autovettore di A relativo all'autovalore λ_i, cioè

$$Av^{(i)} = \lambda_i v^{(i)}.$$

Osservare che una matrice diagonalizzabile in $\mathbb{R}$ ha tutti gli autovalori reali.

Per una matrice diagonalizzabile A, la molteplicità geometrica di ogni autovalore distinto λ è uguale alla molteplicità algebrica. Per cui, per tali matrici si parlerà semplicemente di molteplicità. Essa è uguale al numero di volte che λ appare nella diagonale $\lambda_1, \ldots, \lambda_n$ della matrice diagonale simile a A.

Una matrice $A \in \mathbb{R}^{n\times n}$ si dice *diagonalizzabile con trasformazione ortogonale* se A è simile a una matrice diagonale con matrice di trasformazione ortogonale, cioè esiste $Q \in \mathbb{R}^{n\times n}$ ortogonale tale che $Q^{-1}AQ = Q^TAQ$ è diagonale.

Per cui, dire che A è simile a $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ con matrice di trasformazione ortogonale V di colonne $v^{(1)}, \ldots, v^{(n)}$, equivale a dire che $\lambda_1, \ldots, \lambda_n$ sono gli autovalori di A e $v^{(1)}, \ldots, v^{(n)}$ è una base ortonormale di $\mathbb{R}^n$ tale che, per $i \in \{1, \ldots, n\}$, $v^{(i)}$ è autovettore di A relativo all'autovalore λ_i.

Teorema 4 (Teorema di diagonalizzazione con trasformazione ortogonale o Teorema Spettrale). *Sia $A \in \mathbb{R}^{n\times n}$. Si ha che A è diagonalizzabile con trasformazione ortogonale, cioè esiste $Q \in \mathbb{R}^{n\times n}$ ortogonale tale che Q^TAQ è diagonale, se e solo se A è simmetrica.*

La dimostrazione di questo teorema è stata vista nel corso di Geometria.

Esercizio. Diagonalizzare con trasformazione ortogonale la matrice simmetrica

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix},$$

vale a dire determinare $Q \in \mathbb{R}^{2\times 2}$ ortogonale e $D \in \mathbb{R}^{2\times 2}$ diagonale tali che $D = Q^T AQ$.

Se $A \in \mathbb{R}^{n\times n}$ è simmetrica, allora autovettori di A relativi a autovalori distinti sono ortogonali.

Infatti, siano $x, y \in \mathbb{R}^n \setminus \{\underline{0}\}$ autovettori di A relativi agli autovalori distinti $\lambda, \mu \in \mathbb{R}$, rispettivamente. Si ha

$$\langle Ax, y\rangle = \langle \lambda x, y\rangle = \lambda \langle x, y\rangle$$

e

$$\langle Ax, y\rangle = \langle x, A^T y\rangle = \langle x, Ay\rangle = \langle x, \mu y\rangle = \mu \langle x, y\rangle .$$

Da cui, $\lambda \langle x, y\rangle = \mu \langle x, y\rangle$, vale a dire

$$(\lambda - \mu) \langle x, y\rangle = 0.$$

Essendo $\lambda \neq \mu$, si ottiene $\langle x, y\rangle = 0$.

Il fatto che abbiamo appena dimostrato è anche conseguenza immediata del Teorema di diagonalizzazione con trasformazione ortogonale.

Infatti, siano $v^{(1)}, \ldots, v^{(n)}$ le colonne della matrice ortogonale $Q \in \mathbb{R}^{n\times n}$ tale che $Q^T AQ = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$.

Se λ compare nella lista $\lambda_1, \ldots, \lambda_n$ nelle posizioni $\lambda_{i_1}, \ldots, \lambda_{i_K}$ e μ compare nella posizioni (diverse dalle precedenti) $\lambda_{j_1}, \ldots, \lambda_{j_L}$, dove K e L sono le molteplicità di λ e μ rispettivamente, allora

$$x = \sum_{k=1}^{K} \alpha_k v^{(i_k)} \quad \text{e} \quad y = \sum_{l=1}^{L} \beta_l v^{(j_l)}$$

per degli scalari $\alpha_k, \beta_l,\ k \in \{1, \ldots, n\}$ e $l \in \{1, \ldots, L\}$, dal momento che $v^{(i_1)}, \ldots, v^{(i_K)}$ è una base dell'autospazio relativo all'autovalore λ e $v^{(j_1)}, \ldots, v^{(j_L)}$ è una base dell'autospazio relativo all'autovalore μ.

Quindi, essendo $v^{(1)}, \ldots, v^{(n)}$ ortonormali, si vede che

$$\langle x, y\rangle = \left\langle \sum_{k=1}^{K} \alpha_k v^{(i_k)}, \sum_{l=1}^{L} \beta_l v^{(j_l)} \right\rangle = \sum_{k=1}^{K} \sum_{l=1}^{L} \alpha_k \beta_l \underbrace{\langle v^{(i_k)}, v^{(j_l)}\rangle}_{=0 \text{ essendo } i_k \neq j_l} = 0.$$

4 Forme quadratiche

Sia $A \in \mathbb{R}^{n\times n}$. La funzione $g_A : \mathbb{R}^n \to \mathbb{R}$ data da

$$g_A(x) = \langle x, Ax\rangle = x^T Ax,\ x \in \mathbb{R}^n,$$

è detta la *forma quadratica associata* alla matrice A.

Notare che, per $x \in \mathbb{R}^n$, si ha

$$\begin{aligned} g_A(x) &= \langle x, Ax \rangle = \sum_{i=1}^{n} x_i (Ax)_i = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} a_{ij} x_j \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j . \end{aligned}$$

Esaminiamo la questione:

- a matrici di $\mathbb{R}^{n\times n}$ diverse corrispondono forme quadratiche associate diverse?

La risposta è no se si considerano matrici generali, mentre è si se si considerano matrici simmetriche.

Per la risposta negativa nel caso di matrici generali si consideri il seguente esercizio.

Esercizio. Trovare due matrici distinte $A, B \in \mathbb{R}^{n\times n}$, $n > 2$, tali che $g_A = g_B$. Suggerimento: considerare $B = 0$ e un precedente esercizio.

La risposta positiva nel caso di matrici simmetriche viene dal seguente teorema.

Teorema 5 *Sia $A \in \mathbb{R}^{n\times n}$. Esiste una e una sola matrice simmetrica $E \in \mathbb{R}^{n\times n}$ tale che $g_A = g_E$, cioè*

$$\forall x \in \mathbb{R}^n : \langle x, Ax \rangle = \langle x, Ex \rangle .$$

Essa è data da

$$E = \frac{1}{2}\left(A + A^T\right).$$

Sia $\mathcal{G}$ l'applicazione che associa ad ogni $A \in \mathbb{R}^{n\times n}$ la forma quadratica g_A.

Il teorema precedente dice che la restrizione di $\mathcal{G}$ all'insieme delle matrici simmetriche è una corrispondenza biunivoca (applicazione biettiva) tra l'insieme delle matrici simmetriche di $\mathbb{R}^{n\times n}$ e l'insieme $\mathcal{G}(\mathbb{R}^{n\times n})$ delle forme quadratiche.

Dimostrazione. Proviamo l'esistenza. Consideriamo la matrice

$$E = \frac{1}{2}\left(A + A^T\right).$$

Esercizio. Provare che tale matrice è simmetrica.

Per ogni $x \in \mathbb{R}^n$, si ha

$$\begin{aligned}
\langle x, Ex \rangle &= \left\langle x, \frac{1}{2}\left(A + A^T\right) x \right\rangle = \frac{1}{2}\left\langle x, \left(A + A^T\right) x \right\rangle \\
&= \frac{1}{2}\left\langle x, Ax + A^T x \right\rangle = \frac{1}{2}\langle x, Ax \rangle + \frac{1}{2} \underbrace{\left\langle x, A^T x \right\rangle}_{=\langle Ax, x \rangle = \langle x, Ax \rangle} \\
&= \frac{1}{2}\langle x, Ax \rangle + \frac{1}{2}\langle x, Ax \rangle = \langle x, Ax \rangle .
\end{aligned}$$

Proviamo infine l'unicità. Sia $E_1 \in \mathbb{R}^{n \times n}$ matrice simmetrica tale che

$$\forall x \in \mathbb{R}^n : \langle x, Ax \rangle = \langle x, Ex \rangle = \langle x, E_1 x \rangle .$$

Allora

$$\forall x \in \mathbb{R}^n : 0 = \langle x, Ex \rangle - \langle x, E_1 x \rangle = \langle x, Ex - E_1 x \rangle = \langle x, (E - E_1)\, x \rangle .$$

Essendo E e E_1 simmetriche, si ha che anche $E - E_1$ è simmetrica. Quindi si conclude $E - E_1 = 0$ ricordando un teorema precedente, da cui $E = E_1$. ■

La matrice E nel precedente teorema è detta la *matrice simmetrica associata* alla forma quadratica g_A.

Ad esempio, la forma quadratica associata alla matrice non simmetrica

$$A = \begin{bmatrix} a & -2 \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{2 \times 2}$$

è

$$\begin{aligned}
g_A(x) &= \sum_{i=1}^{2} \sum_{j=1}^{2} a_{ij} x_i x_j = a x_1^2 - 2 x_1 x_2 + 0 x_2 x_1 + 1 x_2^2 \\
&= a x_1^2 - 2 x_1 x_2 + x_2^2, \; x \in \mathbb{R}^2.
\end{aligned}$$

La matrice simmetrica associata alla forma quadratica è

$$E = \frac{1}{2}\left(A + A^T\right) = \frac{1}{2}\left(\begin{bmatrix} a & -2 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} a & 0 \\ -2 & 1 \end{bmatrix}\right) = \begin{bmatrix} a & -1 \\ -1 & 1 \end{bmatrix}.$$

Esercizio. Trovare le matrici simmetriche associate alle forme quadratiche $g : \mathbb{R}^2 \to \mathbb{R}$ data da

$$g(x) = 2x_1^2 - 3x_1 x_2 + x_2^2, \; x \in \mathbb{R}^2$$

e $h : \mathbb{R}^3 \to \mathbb{R}$ data da

$$h(x) = 2x_1^2 - x_2^2 - 4x_3^2 + x_1 x_2 - x_1 x_3 + 2 x_2 x_3, \; x \in \mathbb{R}^3.$$

5 Matrici simmetriche semidefinite e definite positive

Definizione 6 *Sia $A \in \mathbb{R}^{n\times n}$ simmetrica. Si dice che:*

- *A è* semidefinita positiva *se* $\forall x \in \mathbb{R}^n : g_A(x) = \langle x, Ax\rangle \geq 0$;
- *A è* definita positiva *se* $\forall x \in \mathbb{R}^n \setminus \{\underline{0}\} : g_A(x) > 0$; *notare che, per $x = \underline{0}$, si ha $g_A(x) = \langle x, Ax\rangle = 0$;*
- *A è* semidefinita negativa *se* $\forall x \in \mathbb{R}^n : g_A(x) \leq 0$
- *A è* definita negativa *se* $\forall x \in \mathbb{R}^n \setminus \{\underline{0}\} : g_A(x) < 0$.

Le definizioni di semidefinita positiva, definita positiva, ecc. potrebbero essere applicate a forme quadratiche invece che a matrici simmetriche, dal momento che le forme quadratiche sono in corrispondenza uno a uno con le matrici simmetriche.

Consideriamo, ad esempio la matrice simmetrica

$$E = \begin{bmatrix} a & -1 \\ -1 & 1 \end{bmatrix}$$

vista in precedenza con forma quadratica associata

$$g_E(x) = ax_1^2 - 2x_1x_2 + x_2^2,\ x \in \mathbb{R}^2.$$

Per $a \geq 1$, tale matrice è semidefinita positiva : per $x \in \mathbb{R}^2$, si ha

$$\begin{aligned} g_E(x) &= ax_1^2 - 2x_1x_2 + x_2^2 = (a-1)x_1^2 + x_1^2 - 2x_1x_2 + x_2^2 \\ &= (a-1)x_1^2 + (x_1 - x_2)^2 \geq 0. \end{aligned}$$

Per $a > 1$, la matrice è definita positiva: per $x \in \mathbb{R}^2$ si ha

$$\begin{aligned} & g_E(x) = (a-1)x_1^2 + (x_1 - x_2)^2 = 0 \\ \Leftrightarrow\ & (a-1)x_1^2 = 0 \ \text{ e } \ (x_1 - x_2)^2 = 0 \\ \Leftrightarrow\ & x_1 = 0 \ \text{ e } \ x_1 - x_2 = 0 \\ \Leftrightarrow\ & x = \underline{0}. \end{aligned}$$

Per $a = 1$, la matrice non è definita positiva: per $x \in \mathbb{R}^2$, si ha

$$g_E(x) = (x_1 - x_2)^2 = 0 \quad \Leftrightarrow \quad x_1 - x_2 = 0$$

e quindi $g_E(x) = 0$ per $x = (t, t)$, $t \in \mathbb{R}$.

Esercizio. Provare che per $a < 1$ la matrice non è semidefinita positiva e nemmeno semidefinita negativa.

Sia $A \in \mathbb{R}^{n \times n}$ simmetrica. Si ha che A è semidefinita (definita) negativa se e solo se $-A$, che è simmetrica, è semidefinita (definita) positiva.

Esercizio. Provare questo fatto.

Per cui, ci si può limitare a studiare solo le matrici semidefinite e definite positive.

Sia $B \in \mathbb{R}^{m \times n}$. Si è visto in precedenza che la matrice $B^T B \in \mathbb{R}^{n \times n}$ è simmetrica. La matrice $B^T B$ è semidefinita positiva ed è definita positiva se e solo se $\mathrm{rank}(B) = n$.

Infatti, per $x \in \mathbb{R}^n$, si ha

$$\langle x, B^T B x \rangle = \langle Bx, Bx \rangle = \|Bx\|_2^2 \geq 0$$

e

$$\|Bx\|_2^2 = 0 \quad \Leftrightarrow \quad Bx = \underline{0} \quad \Leftrightarrow \quad x \in \ker(B)$$

Quindi $B^T B$ è semidefinita positiva ed è definita positiva se e solo se $\ker(B) = \{\underline{0}\}$, cioè $\dim(\ker(B)) = n - \mathrm{rank}(B) = 0$.

Esercizio. In precedenza si è visto che anche la matrice $BB^T \in \mathbb{R}^{m \times m}$ è simmetrica. Provare che BB^T è semidefinita positiva ed è definita positiva se e solo se $\mathrm{rank}\big(B^T\big) = m$. Dal corso di Geometria è noto che $\mathrm{rank}(B^T) = \mathrm{rank}(B)$.

5.1 Proprietà delle matrici simmetriche semidefinite (definite) positive

Vediamo come si comportano le matrici simmetriche semidefinite (definite) positive rispetto ad alcune operazioni matriciali.

1) $\forall A, B \in \mathbb{R}^{n \times n}$ simmetriche semidefinite positive: $A+B$, che è simmetrica, è semidefinita positiva; inoltre, se almeno una tra A e B è definita positiva, allora $A + B$ è definita positiva.

2) $\forall \alpha \geq 0$ $\forall A \in \mathbb{R}^{n \times n}$ simmetrica semidefinita positiva: αA, che è simmetrica, è semidefinita positiva; inoltre, se $\alpha > 0$ e A è definita positiva, allora αA è definita positiva.

3) $\forall A \in \mathbb{R}^{n \times n}$ simmetrica semidefinita positiva:

$$A \text{ è non singolare} \Leftrightarrow A \text{ è definita positiva}$$

e

$$A \text{ è non singolare} \Rightarrow A^{-1}\text{, che è simmetrica, è definita positiva.}$$

Dimostrazione di 1). Siano $A, B \in \mathbb{R}^{n \times n}$ simmetriche semidefinite positive. Per $x \in \mathbb{R}^n$ si ha

$$\langle x, (A + B)\, x \rangle = \langle x, Ax + Bx \rangle = \underbrace{\langle x, Ax \rangle}_{\geq 0} + \underbrace{\langle x, Bx \rangle}_{\geq 0} \geq 0.$$

Se $x \neq \underline{0}$ e almeno una tra A e B è definita positiva, allora $\langle x, Ax \rangle > 0$ o $\langle x, Bx \rangle > 0$ e quindi $\langle x, (A+B)x \rangle > 0$.

Dimostrazione di 2). Siano $\alpha \geq 0$ e $A \in \mathbb{R}^{n \times n}$ simmetrica semidefinita positiva. Per $x \in \mathbb{R}^n$ si ha

$$\langle x, \alpha Ax \rangle = \alpha \underbrace{\langle x, Ax \rangle}_{\geq 0} \geq 0.$$

Se $x \neq \underline{0}$, $\alpha > 0$ e A è definita positiva, allora $\langle x, Ax \rangle > 0$ e quindi $\langle x, \alpha Ax \rangle = \alpha \langle x, Ax \rangle > 0$.

Esercizio. Siano $\alpha < 0$ e $A \in \mathbb{R}^{n \times n}$ simmetrica semidefinita (definita) positiva. Quale proprietà ha la matrice simmetrica αA?

La dimostrazione di 3) è data come conseguenza del prossimo importante teorema.

Ricordiamo che una matrice simmetrica A è diagonalizzabile, cioè simile ad una matrice diagonale D, con trasformazione ortogonale. Gli autovalori di A sono gli elementi diagonali della matrice D.

La semidefinita e definita positività di A possono essere caratterizzate in termini del segno degli autovalori di A.

Teorema 7 *Sia $A \in \mathbb{R}^{n \times n}$ simmetrica. Si ha che A è semidefinita (definita) positiva se e solo se A ha tutti gli autovalori non negativi (positivi).*

Dimostrazione. Parte "solo se". Assumiamo che A abbia un autovalore $\lambda < 0$ ($\lambda \leq 0$) e proviamo che A non è semidefinita (definita) positiva.

Sia $x \in \mathbb{R}^n \setminus \{\underline{0}\}$ un autovettore relativo all'autovalore λ. Si ha

$$\langle x, Ax \rangle = \langle x, \lambda x \rangle = \lambda \langle x, x \rangle = \lambda \|x\|_2^2 < 0 \ (\leq 0),$$

essendo $\|x\|_2^2 > 0$. Per cui A non è semidefinita (definita) positiva.

Parte "se". Siano $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ gli autovalori della matrice simmetrica A. Assumiamo $\lambda_1, \ldots, \lambda_n \geq 0$ (> 0) e proviamo che A è semidefinita (definita) positiva.

Poichè A è simmetrica, esiste $Q \in \mathbb{R}^{n \times n}$ ortogonale tale che $D = Q^T A Q = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Si ha

$$A = QDQ^T.$$

Quindi, per $x \in \mathbb{R}^n$,

$$\langle x, Ax \rangle = \langle x, QDQ^T x \rangle = \langle Q^T x, DQ^T x \rangle = \langle y, Dy \rangle,$$

dove $y = Q^T x$.

Pertanto,

$$\begin{aligned}
\langle x, Ax\rangle &= \langle y, Dy\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}y_iy_j = \sum_{i=1}^{n} \lambda_i y_i^2 \\
&\geq \sum_{i=1}^{n}\left(\min_{k=1,\dots,n}\lambda_k\right) y_i^2 = \left(\min_{k=1,\dots,n}\lambda_k\right)\sum_{i=1}^{n} y_i^2 \\
&= \left(\min_{k=1,\dots,n}\lambda_k\right)\|y\|_2^2 = \left(\min_{k=1,\dots,n}\lambda_k\right)\|x\|_2^2\,,
\end{aligned}$$

dove l'ultima uguaglianza segue dal fatto che Q^T è ortogonale e quindi $\|y\|_2^2 = \left\|Q^Tx\right\|_2^2 = \|x\|_2^2$.

Essendo $\lambda_1, \dots, \lambda_n \geq 0$ (> 0) risulta

$$\min_{k=1,\dots,n}\lambda_k \geq 0 \ \ (> 0)\,,$$

e quindi

$$\langle x, Ax\rangle \geq \left(\min_{k=1,\dots,n}\lambda_k\right)\|x\|_2^2 \geq 0 \ \ (> 0 \text{ per } x \neq \underline{0})\,.$$

■

Esercizio. Caratterizzare la semidefinita (definita) negatività di una matrice simmetrica in termini di autovalori della matrice.

Come conseguenza del teorema proviamo la precedente proprietà 3) della matrici simmetriche semidefinite positive.

Sia $A \in \mathbb{R}^{n\times n}$ simmetrica semidefinita positiva e siano $\lambda_1, \dots, \lambda_n \geq 0$ i suoi autovalori. Poichè A è non singolare se e solo se $\lambda_1, \dots, \lambda_n > 0$ si ha che A è non singolare se e solo se A è definita positiva.

Inoltre, se A è non singolare, cioè $\lambda_1, \dots, \lambda_n > 0$, la matrice simmetrica A^{-1} ha autovalori $\lambda_1^{-1}, \dots, \lambda_n^{-1} > 0$ e quindi è definita positiva.

Determiniamo le matrici simmetriche semidefinite e definite positive 2×2.

Sia

$$A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$$

simmetrica. L'equazione caratteristica di A, le cui due radici sono gli autovalori di A, è

$$\lambda^2 - \mathrm{tr}\,(A)\,\lambda + \det(A) = 0.$$

Poiche la somma e il prodotto delle radici sono $\mathrm{tr}\,(A)$ e $\det(A)$, rispettivamente, ne viene che

- A è semidefinita positiva, cioè le radici sono entrambe non negative, se e solo se
$$\mathrm{tr}\,(A) = a + d \geq 0 \ \text{ e } \ \det(A) = ad - b^2 \geq 0;$$

- A è definita positiva, cioè le radici sono entrambe positive, se e solo se

$$\mathrm{tr}\,(A) = a + d > 0 \quad \text{e} \quad \det(A) = ad - b^2 > 0.$$

Consideriamo ad esempio la matrice simmetrica vista in precedenza

$$E = \begin{bmatrix} a & -1 \\ -1 & 1 \end{bmatrix}.$$

Avendosi

$$\mathrm{tr}\,(E) = a + 1 \text{ e } \det(E) = a - 1,$$

si ottiene che E è semidefinita positiva se e solo se

$$a + 1 \geq 0 \text{ e } \ a - 1 \geq 0,$$

vale a dire $a \geq 1$. Inoltre, E è definita positiva se e solo se

$$a + 1 > 0 \text{ e } a - 1 > 0,$$

vale a dire $a > 1$. Tutto questo si accorda con quanto precedentemente trovato.

Esercizio. Sia $A \in \mathbb{R}^{n \times n}$. Provare che se A è simmetrica, allora A^k, dove k è un intero positivo, è simmetrica. Provare inoltre che se A è simmetrica semidefinita (definita) positiva, allora A^k è simmetrica semidefinita (definita) positiva. Suggerimento: guardare agli autovalori di A^k.

Esercizio. Dire se la forma quadratica

$$g(x) = 2x_1^2 - 3x_1x_2 + x_2^2, \ x \in \mathbb{R}^2,$$

vista in un esercizio precedente è semidefinita positiva, definita positiva, semidefinita negativa o definita negativa.

Esercizio. Si consideri una matrice simmetrica 2×2. Quale condizione sul determinante della matrice caratterizza il fatto che la matrice sia definita positiva o negativa? Quale condizione caratterizza che sia semidefinita positiva o negativa?

6 Le matrici $A^T A$ e AA^T

Sia $A \in \mathbb{R}^{m \times n}$. Si è visto che la matrice

$$A^T A \in \mathbb{R}^{n \times n}$$

è simmetrica semidefinita positiva.

Per cui $A^T A \in \mathbb{R}^{n \times n}$ è diagonalizzabile con trasformazione ortogonale e quindi esiste $U \in \mathbb{R}^{n \times n}$ ortogonale tale che

$$U^T A^T A U = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_r, \lambda_{r+1} = 0, \ldots, \lambda_n = 0) \in \mathbb{R}^{n \times n},$$

dove $\lambda_1, \dots, \lambda_n$ sono gli autovalori di A^TA e $\lambda_1, \dots, \lambda_r > 0$, $r \in \{0, 1, \dots, n\}$, sono quelli non nulli.

Qui senza perdita di generalità si sta assumendo $\lambda_1 \geq \cdots \geq \lambda_n$. Se così non fosse basta riordinare gli autovalori in ordine decrescente e, di conseguenza, anche le colonne di U in modo che, per ogni $j \in \{1, \dots, n\}$, la colonna $j-$esima di U sia un autovettore corrispondente all'autovalore $j-$esimo.

Le colonne $u^{(1)}, \dots, u^{(n)}$ di U costituiscono una base ortonormale di $\mathbb{R}^n$ di autovettori di A^TA relativi agli autovalori $\lambda_1, \dots, \lambda_n$, rispettivamente.

La lista $\lambda_1, \dots, \lambda_n$ contiene gli autovalori distinti di A^TA ciascuno ripetuto un numero di volte pari alla sua molteplicità. Quindi, la somma delle molteplicità degli autovalori non nulli è r e la molteplicità dell'autovalore 0 è $n - r$.

Anche la matrice

$$AA^T \in \mathbb{R}^{m \times m}$$

è simmetrica semidefinita positiva.

Per cui anche $AA^T \in \mathbb{R}^{m \times m}$ è diagonalizzabile con trasformazione ortogonale e quindi esiste $V \in \mathbb{R}^{m \times m}$ ortogonale tale che

$$V^TAA^TV = \mathrm{diag}\,(\mu_1, \dots, \mu_s, \mu_{s+1} = 0, \dots, \mu_m = 0) \in \mathbb{R}^{m \times m},$$

dove $\mu_1, \dots, \mu_n$ sono gli autovalori di AA^T e $\mu_1, \dots, \mu_s > 0$, $s \in \{0, 1, \dots, m\}$ sono quelli non nulli.

Di nuovo, senza perdita di generalità si sta assumendo $\mu_1 \geq \cdots \geq \mu_m$.

Le colonne $v^{(1)}, \dots, v^{(m)}$ di V costituiscono una base ortonormale di $\mathbb{R}^m$ di autovettori di AA^T relativi agli autovalori $\mu_1, \dots, \mu_m$, rispettivamente.

La lista $\mu_1, \dots, \mu_m$ contiene gli autovalori distinti di AA^T ciascuno ripetuto un numero di volte pari alla sua molteplicità. Quindi, la somma delle molteplicità degli autovalori non nulli è s e la molteplicità dell'autovalore 0 è $m - s$.

Si ha che A^TA e AA^T hanno gli stessi autovalori distinti non nulli con le stesse molteplicità. Questo segue dal seguente teorema.

Teorema 8 *Sia $B \in \mathbb{R}^{p \times q}$.*

Se $\lambda \in \mathbb{R}$ un autovalore non nullo di B^TB con autovettore relativo $x \in \mathbb{R}^q \setminus \{\underline{0}\}$, allora λ un autovalore di BB^T con autovettore relativo Bx.

Inoltre, se $x^{(1)}, \dots, x^{(l)}$ sono autovettori linearmente indipendenti di B^TB relativi all'autovalore λ, allora anche gli autovettori $Bx^{(1)}, \dots, Bx^{(l)}$ di BB^T sono linearmente indipendenti e, quindi, la molteplicità geometrica di λ come autovalore di B^TB è minore o uguale della molteplicità geometrica come autovalore di BB^T.

Dimostrazione. Sia λ un autovalore non nullo di $B^TB \in \mathbb{R}^{q \times q}$ e sia $x \in \mathbb{R}^q \setminus \{\underline{0}\}$ un autovettore relativo.

Posto

$$y = Bx,$$

si ha

$$BB^Ty = B\underbrace{B^TBx}_{=\lambda x} = B\lambda x = \lambda Bx = \lambda y.$$

Per concludere che λ è un autovalore di BB^T occorre mostrare che $y \neq \underline{0}$. Se fosse $y = \underline{0}$, allora

$$\underline{0} = B^Ty = B^TBx = \lambda x$$

e quindi si avrebbe $x = \underline{0}$ essendo $\lambda \neq 0$. Pertanto $y \neq \underline{0}$.

Siano ora $x^{(1)}, \ldots, x^{(l)}$ autovettori linearmente indipendenti di B^TB relativi all'autovalore λ. Proviamo che gli autovettori $Bx^{(1)}, \ldots, Bx^{(l)}$ di BB^T sono linearmente indipendenti, mostrando che una combinazione lineare nulla

$$\alpha_1 Bx^{(1)} + \cdots + \alpha_l Bx^{(l)} = \underline{0}$$

di $Bx^{(1)}, \ldots, Bx^{(l)}$ deve avere coefficienti nulli. Si ha

$$\begin{aligned}
&B^T\left(\alpha_1 Bx^{(1)} + \cdots + \alpha_l Bx^{(l)}\right) = B^T\alpha_1 Bx^{(1)} + \cdots + B^T\alpha_l Bx^{(l)}\\
&= \alpha_1 B^TBx^{(1)} + \cdots + \alpha_l B^TBx^{(l)} = \alpha_1\lambda x^{(1)} + \cdots + \alpha_l\lambda x^{(l)}\\
&= B^T\underline{0} = \underline{0}.
\end{aligned}$$

Essendo $x^{(1)}, \ldots, x^{(l)}$ linearmente indipendenti, si ha

$$\alpha_1\lambda = \cdots = \alpha_l\lambda = 0,$$

da cui $\alpha_1 = \cdots = \alpha_l = 0$, avendosi $\lambda \neq 0$. ■

Applicando il teorema con $B = A$, si vede che gli autovalori non nulli di A^TA sono autovalori di AA^T e le molteplicità geometriche non diminuiscono nel passare da A^TA a AA^T.

Applicando il teorema con $B = A^T$ e quindi $B^T = A$, si vede che gli autovalori non nulli di AA^T sono autovalori di A^TA e le molteplicità geometriche non diminuiscono nel passare da AA^T a A^TA.

Per cui, A^TA e AA^T hanno gli stessi autovalori distinti non nulli con le stesse molteplicità geometriche.

Esercizio. Verificare quanto appena asserito per la matrice

$$A = \begin{bmatrix} 0 & 2 & 1 \\ -1 & 0 & -2 \end{bmatrix} \in \mathbb{R}^{2\times 3}$$

calcolando gli autovalori di $A^TA \in \mathbb{R}^{3\times 3}$ e $AA^T \in \mathbb{R}^{2\times 2}$.

Esercizio. Sia $x \in \mathbb{R}^n \setminus \{\underline{0}\}$. Provare che la matrice $xx^T \in \mathbb{R}^{n\times n}$ ha come autovalori distinti $\|x\|_2^2$ e 0. Trovare le molteplicità di tali autovalori e determinare il polinomio caratteristico della matrice. Suggerimento: usare il fatto che A^TA e AA^T hanno gli stessi autovalori non nulli.

Esercizio. Verificare quanto asserito nel precedente esercizio per $x = (1, 2, -1) \in \mathbb{R}^3$, andando a calcolare gli autovalori di xx^T.

La lista $\lambda_1, \ldots, \lambda_r$ contiene gli autovalori non nulli distinti di $A^T A$ in ordine decrescente, ciascuno ripetuto un numero di volte pari alla sua molteplicità. La lista $\mu_1, \ldots, \mu_s$ contiene gli autovalori non nulli distinti di AA^T in ordine decrescente, ciascuno ripetuto un numero di volte pari alla sua molteplicità.

Il teorema precedente ci dice che $A^T A$ e AA^T hanno gli stessi autovalori non nulli con le stesse molteplicità. Si conclude che

$$r = s \text{ e } \lambda_1 = \mu_1, \ldots, \lambda_r = \mu_r.$$

6.1 Costruire V da U

Partendo da una matrice ortonormale $U \in \mathbb{R}^{n \times n}$ tale che

$$U^T A^T A U = \operatorname{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0) \in \mathbb{R}^{n \times n},$$

le cui colonne $u^{(1)}, \ldots, u^{(n)}$ formano una base ortonormale di $\mathbb{R}^n$ fatta di autovettori di $A^T A$, mostriamo come si può ottenere una particolare matrice ortonormale $V \in \mathbb{R}^{n \times n}$ tale che

$$V^T A A^T V = \operatorname{diag}(\mu_1 = \lambda_1, \ldots, \mu_r = \lambda_r, \mu_{r+1} = 0, \ldots, \mu_m = 0) \in \mathbb{R}^{m \times m},$$

le cui colonne $v^{(1)}, \ldots, v^{(m)}$ formano una base ortonormale di $\mathbb{R}^m$ fatta di autovettori di AA^T.

Definiamo in maniera separata le colonne $v^{(1)}, \ldots, v^{(r)}$ dalle colonne $v^{(r+1)}, \ldots, v^{(m)}$.

Definiamo $v^{(1)}, \ldots, v^{(r)}$.

Dal precedente teorema si ha che, per ogni $j \in \{1, \ldots, r\}$,

$$c^{(j)} = Au^{(j)}$$

è un autovettore di AA^T relativo all'autovalore λ_j, essendo $u^{(j)}$ autovettore di $A^T A$ relativo all'autovalore non nullo λ_j.

I vettori $c^{(1)}, \ldots, c^{(r)}$ sono ortogonali: per $i, j \in \{1, \ldots, r\}$ si ha

$$\begin{aligned} \langle c^{(i)}, c^{(j)} \rangle &= \langle Au^{(i)}, Au^{(j)} \rangle = \langle u^{(i)}, A^T A u^{(j)} \rangle = \langle u^{(i)}, \lambda_j u^{(j)} \rangle \\ &= \lambda_j \langle u^{(i)}, u^{(j)} \rangle = \begin{cases} \lambda_j \text{ se } i = j \\ 0 \text{ se } i \neq j. \end{cases} \end{aligned}$$

Poichè

$$\|c^{(j)}\|_2 = \sqrt{\langle c^{(j)}, c^{(j)} \rangle} = \sqrt{\lambda_j},\ j \in \{1, \ldots, r\},$$

i vettori

$$v^{(j)} = \frac{1}{\|c^{(j)}\|_2} c^{(j)} = \frac{1}{\sqrt{\lambda_j}} c^{(j)} = \frac{1}{\sqrt{\lambda_j}} A u^{(j)},\ j \in \{1, \ldots, r\},$$

risultano essere autovettori ortonormali di AA^T, con $v^{(j)}$, $j \in \{1, \ldots, r\}$, autovettore relativo all'autovalore λ_j.

Definiamo ora $v^{(r+1)}, \ldots, v^{(m)}$.

I vettori $v^{(r+1)}, \ldots, v^{(m)}$ sono una qualunque base ortonormale dell'autospazio relativo all'autovalore 0 di AA^T (la molteplicità di tale autovalore è $m - r$). Tale base si può ottenere da una base qualsiasi dell'autospazio con il processo di ortonormalizzazione di Gram-Schimdt.

Osserviamo poi che ogni autovettore $v^{(k)}$, $k \in \{r+1, \ldots, m\}$, è ortogonale a ogni autovettore $v^{(j)}$, $j \in \{1, \ldots, r\}$. Questo segue dal fatto che autovettori della matrice simmetrica AA^T relativi a autovalori distinti sono ortogonali: $v^{(k)}$ è autovettore relativo all'autovalore 0 e $v^{(j)}$ è autovettore relativo all'autovalore $\lambda_j \neq 0$.

Allora, aggiungendo ai vettori ortonormali $v^{(1)}, \ldots, v^{(r)}$ i vettori ortonormali $v^{(r+1)}, \ldots, v^{(m)}$, si ottiene una base ortonormale

$$v^{(1)}, \ldots, v^{(r)}, v^{(r+1)}, \ldots, v^{(m)}$$

di $\mathbb{R}^m$ di autovettori di AA^T relativi agli autovalori

$$\mu_1 = \lambda_1, \ldots, \mu_r = \lambda_r, \mu_{r+1} = 0, \ldots, \mu_m = 0,$$

rispettivamente. I vettori $v^{(1)}, \ldots, v^{(m)}$ sono le colonne di V.

Esercizio. Si consideri la matrice

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ -1 & -1 \\ 0 & 2 \end{bmatrix} \in \mathbb{R}^{4 \times 2}$$

Si determini la matrice ortogonale U che diagonalizza $A^TA \in \mathbb{R}^{2\times 2}$. A partire da questa, con la costruzione appena indicata si determini la matrice ortogonale V che diagonalizza $AA^T \in \mathbb{R}^{4\times 4}$.

7 La SVD di una matrice

La *decomposizione ai valori singolari* di una matrice rettangolare, nota con l'acronimo *SVD (Singular Value Decomposition)*, permette di ottenere in modo agevole informazioni importanti riguardo la matrice, come ad esempio il rango, i sottospazi nucleo e immagine e la norma 2.

Teorema 9 (Teorema della SVD). *Per ogni $A \in \mathbb{R}^{m\times n}$ esistono $V \in \mathbb{R}^{m\times m}$ ortogonale, $U \in \mathbb{R}^{n\times n}$ ortogonale e $\Sigma \in \mathbb{R}^{m\times n}$ del tipo*

$$\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix},$$

dove $\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r\times r}$ *con* $\sigma_1 \geq \cdots \geq \sigma_r > 0$ *per qualche* $r \in \{0, 1, \ldots, \min\{m, n\}\}$, *tali che*

$$A = V\Sigma U^T.$$

Dimostrazione. Sia $A \in \mathbb{R}^{m\times n}$. Andiamo prima a definire le matrici U, Σ e V, poi a stabilire alcune relazioni e infine a a verificare che

$$V^T AU = \Sigma,$$

da cui seguirà $A = V\Sigma U^T$.

Matrice U. La matrice $U \in \mathbb{R}^{n\times n}$ è una matrice ortogonale tale che

$$U^T A^T AU = \operatorname{diag}(\lambda_1, \ldots, \lambda_r, \lambda_{r+1} = 0, \ldots, \lambda_n = 0) \in \mathbb{R}^{n\times n},$$

dove $\lambda_1 \geq \cdots \geq \lambda_r > 0$ sono gli autovalori non nulli di $A^T A$. Le colonne $u^{(1)}, \ldots, u^{(n)}$ di U sono una base ortonormale di $\mathbb{R}^n$ di autovettori di $A^T A$. In particolare, si ha che $u^{(1)}, \ldots, u^{(r)}$ sono autovettori relativi agli autovalori non nulli $\lambda_1, \ldots, \lambda_r$, rispettivamente, e che $u^{(r+1)}, \ldots, u^{(n)}$ sono autovettori relativi all'autovalore 0.

Matrice Σ. Poniamo

$$\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r\times r},$$

dove

$$\sigma_j = \sqrt{\lambda_j},\ j \in \{1, \ldots, r\},$$

e poi

$$\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{m\times n}.$$

Matrice V. La matrice $V \in \mathbb{R}^{m\times m}$ è la matrice ortogonale tale che

$$V^T AA^T V = \operatorname{diag}(\mu_1 = \lambda_1, \ldots, \mu_r = \lambda_r, \mu_{r+1} = 0, \ldots, \mu_m = 0).$$

costruita nella precedente sezione a partire dalla matrice U. Le colonne $v^{(1)}, \ldots, v^{(m)}$ sono una base ortonormale di $\mathbb{R}^m$ di autovettori di AA^T.

In particolare, si ha che, per $j \in \{1, \ldots, r\}$,

$$v^{(j)} = \frac{1}{\sqrt{\lambda_j}} Au^{(j)} = \frac{1}{\sigma_j} Au^{(j)}$$

è un autovettore relativo all'autovalore non nullo λ_j e che $v^{(r+1)}, \ldots, v^{(m)}$ sono autovettori ortonormali relativi all'autovalore 0 costruiti a partire da una qualunque base dell'autospazio relativo all'autovalore 0.

Relazioni. Scriviamo
$$U = [\ U_r \quad W\ ],$$
dove
$$U_r = [u^{(1)} \dots u^{(r)}] \in \mathbb{R}^{n\times r} \ \text{ e } \ W = [u^{(r+1)} \dots u^{(n)}] \in \mathbb{R}^{n\times(n-r)},$$
e
$$V = [\ V_r \quad Z\ ],$$
dove
$$V_r = [v^{(1)} \dots v^{(r)}] \in \mathbb{R}^{m\times r} \ \text{ e } \ Z = [v^{(r+1)} \dots v^{(m)}] \in \mathbb{R}^{m\times(m-r)}.$$

Si ha
$$AU_r = V_r\Sigma_r$$
Infatti, ricordando $v^{(j)} = \frac{1}{\sigma_j}Au^{(j)}$, $j \in \{1,\dots,r\}$, risulta
$$AU_r = A\left[u^{(1)} \dots u^{(r)}\right] = \left[Au^{(1)} \dots Au^{(r)}\right] = \left[\sigma_1 v^{(1)} \dots \sigma_r v^{(r)}\right] = V_r\Sigma_r.$$

Si ha
$$AW = 0.$$
Infatti,
$$U^TA^TAU = \operatorname{diag}(\lambda_1,\dots,\lambda_n) = \begin{bmatrix} \Sigma_r^2 & 0 \\ 0 & 0 \end{bmatrix}$$
e
$$\begin{aligned} U^TA^TAU &= \begin{bmatrix} U_r^T \\ W^T \end{bmatrix} A^TA \begin{bmatrix} U_r & W \end{bmatrix} \\ &= \begin{bmatrix} U_r^TA^TA \\ W^TA^TA \end{bmatrix} \begin{bmatrix} U_r & W \end{bmatrix} = \begin{bmatrix} U_r^TA^TAU_r & U_r^TA^TAW \\ W^TA^TAU_r & W^TA^TAW \end{bmatrix}. \end{aligned}$$
Corrispondendo le dimensioni dei blocchi nelle due precedenti forme 2×2 a blocchi di U^TA^TAU, si ha
$$0 = W^TA^TAW = (AW)^T AW$$
e quindi $AW = 0$ come segue dall'esercizio sottostante.

Esercizio. Sia $B \in \mathbb{R}^{m\times n}$. Provare che se B^TB ha diagonale nulla, allora $B = 0$. Suggerimento: per $i \in \{1,\dots,n\}$, esprimere l'elemento $B^TB(i,i)$ come prodotto scalare della $i-$ esima riga di B^T con la $i-$esima colonna di B.

Verifica. Proviamo $V^TAU = \Sigma$. Risulta
$$\begin{aligned} V^TAU &= \begin{bmatrix} V_r^T \\ Z^T \end{bmatrix} A \begin{bmatrix} U_r & W \end{bmatrix} = \begin{bmatrix} V_r^TA \\ Z^TA \end{bmatrix} \begin{bmatrix} U_r & W \end{bmatrix} \\ &= \begin{bmatrix} V_r^TAU_r & V_r^TAW \\ Z^TAU_r & Z^TAW \end{bmatrix}. \end{aligned}$$

Ora, si conclude che

$$V^T AU = \left[\begin{array}{cc} V_r^T AU_r & V_r^T AW \\ Z^T AU_r & Z^T AW \end{array} \right] = \left[\begin{array}{cc} \Sigma_r & 0 \\ 0 & 0 \end{array} \right] = \Sigma.$$

Infatti, si ha:

- $V_r^T AU_r = V_r^T V_r \Sigma_r = \Sigma_r$, essendo $V_r^T V_r = I_r$ in quanto le colonne $v^{(1)}, \ldots, v^{(r)}$ di V_r sono ortonormali;
- $Z^T AU_r = Z^T V_r \Sigma_r = 0$, essendo $Z^T V_r = 0$ in quanto le colonne $v^{(r+1)}, \ldots, v^{(m)}$ di Z sono ortogonali alle colonne di $v^{(1)}, \ldots, v^{(r)}$ di V_r (si ha $Z^T V_r\,(i,j) = \left\langle v^{(r+i)}, v^{(j)} \right\rangle = 0$, $(i,j) \in \{1, \ldots, m-r\} \times \{1, \ldots, r\}$);
- $V_r^T AW = 0$ e $Z^T AW = 0$, essendo $AW = 0$.

■

Data $A \in \mathbb{R}^{m \times n}$, una *decomposizione ai valori singolari*, o una *SVD* (*Singular Value Decomposition*), di A è una fattorizzazione

$$A = V \Sigma U^T,$$

dove V, Σ e U sono come nell'enunciato del precedente teorema.

Il precedente teorema afferma che esiste una SVD di A. Nella dimostrazione si è costruita una particolare SVD di A in cui

- gli elementi positivi decrescenti $\sigma_1, \ldots, \sigma_r$ in Σ_r sono le radici quadrate degli autovalori non nulli di $A^T A$ e AA^T;
- le colonne ortonormali $u^{(1)}, \ldots, u^{(n)}$ di U sono autovettori di $A^T A$ relativi agli autovalori $\lambda_1 = \sigma_1^2, \ldots, \lambda_r = \sigma_r^2, \lambda_{r+1} = 0, \ldots, \lambda_n = 0$;
- le colonne ortonormali $v^{(1)}, \ldots, v^{(m)}$ di V sono autovettori di AA^T relativi agli autovalori $\mu_1 = \sigma_1^2, \ldots, \mu_r = \sigma_r^2, \mu_{r+1} = 0, \ldots, \mu_m = 0$ e si ha

$$v^{(j)} = \frac{1}{\sigma_j} Au^{(j)},\ j \in \{1, \ldots, r\}.$$

In ogni SVD di A accade questo.

Infatti, sia $A = V \Sigma U^T$ una SVD di A. Si ha

$$\begin{aligned} A^T A &= \left(V \Sigma U^T\right)^T V \Sigma U^T = U \Sigma^T V^T V \Sigma U^T = U \Sigma^T \Sigma U^T \\ AA^T &= V \Sigma U^T \left(V \Sigma U^T\right)^T = V \Sigma U^T U \Sigma^T V^T = V \Sigma \Sigma^T V^T \end{aligned}$$

e quindi $A^T A$ è simile alla matrice diagonale

$$\begin{aligned}
\Sigma^T \Sigma &= \begin{bmatrix} \Sigma_r & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{bmatrix} \begin{bmatrix} \Sigma_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \\
&= \begin{bmatrix} \Sigma_r^2 & 0_{r\times(n-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{bmatrix} \in \mathbb{R}^{n\times n}
\end{aligned}$$

con matrice di trasformazione U e AA^T è simile alla matrice diagonale

$$\begin{aligned}
\Sigma \Sigma^T &= \begin{bmatrix} \Sigma_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \begin{bmatrix} \Sigma_r & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{bmatrix} \\
&= \begin{bmatrix} \Sigma_r^2 & 0_{r\times(m-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(m-r)} \end{bmatrix} \in \mathbb{R}^{m\times m}
\end{aligned}$$

con matrice di trasformazione V.

Pertanto,

- gli elementi diagonali $\sigma_1^2, \ldots, \sigma_r^2$ di Σ_r^2 sono gli autovalori non nulli di $A^T A$ e AA^T;
- le colonne di U sono autovettori di $A^T A$ relativi agli autovalori $\lambda_1 = \sigma_1^2, \ldots, \lambda_r = \sigma_r^2, \lambda_{r+1} = 0, \ldots, \lambda_n = 0$;
- le colonne di V sono autovettori di AA^T relativi agli autovalori $\mu_1 = \sigma_1^2, \ldots, \mu_r = \sigma_r^2, \mu_{r+1} = 0, \ldots, \mu_n = 0$.

Infine si ha

$$\begin{aligned}
&AU = [Au^{(1)} \ \ldots \ Au^{(r)} \ Au^{(r+1)} \ \ldots \ Au^{(n)}] && (3) \\
&= V\Sigma U^T U = V\Sigma = [\sigma_1 v^{(1)} \ \ldots \ \sigma_r v^{(r)} \ \underline{0} \ \ldots \ \underline{0}] && (4)
\end{aligned}$$

da cui

$$v^{(j)} = \frac{1}{\sigma_j} Au^{(j)}, \ j \in \{1, \ldots, r\}.$$

I $p = \min\{m, n\}$ elementi diagonali di Σ sono detti i *valori singolari* di A. Gli elementi positivi decrescenti $\sigma_1, \ldots, \sigma_r$ in Σ_r sono i valori singolari non nulli. I rimanenti $p - r$ elementi diagonali nulli $\sigma_{r+1} = \cdots = \sigma_p = 0$ di Σ sono i valori singolari nulli.

Le colonne $u^{(1)}, \ldots, u^{(p)}$ di U sono dette *vettori singolari destri* di A e le colonne $v^{(1)}, \ldots, v^{(p)}$ di V sono dette *vettori singolari sinistri* di A.

In particolare, per $j \in \{1, \ldots, p\}$, $u^{(j)}$ e $v^{(j)}$ sono detti vettore singolare destro e sinistro, rispettivamente, relativi al valore singolare σ_j.

Si ha

$$Au^{(j)} = \sigma_j v^{(j)}, \ j \in \{1, \ldots, p\}.$$

Per $j \in \{1, \ldots, r\}$, questa è la relazione vista più volte in precedenza. Esercizio. Mostrare che essa vale anche per $j \in \{r + 1, \ldots, p\}$. Suggerimento: usare (3) e (4) della pagina precedente.

Data $A \in \mathbb{R}^{m\times n}$ e una sua SVD $A = V\Sigma U^T$, determiniamo una SVD di A^T. Si ha

$$A^T = \left(V\Sigma U^T\right)^T = U\Sigma^T V^T$$

con

$$\Sigma^T = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{n\times m}.$$

Quindi $A^T = U\Sigma^T V^T$ è una SVD di A^T: nel passare da A a A^T, i valori singolari restano gli stessi e le matrici U e V si scambiano tra loro e quindi i vettori singolari destri e sinistri si scambiano tra loro.

Risulta allora

$$A^T v^{(j)} = \sigma_j u^{(j)},\ j \in \{1,\ldots,p\}.$$

Esercizio. A partire da una SVD $A = V\Sigma U^T$ di $A \in \mathbb{R}^{m\times n}$, determinare:

- Una SVD di αA, dove $\alpha \geq 0$. Suggerimento $\alpha A = \alpha V\Sigma U^T = V(\alpha\Sigma)U^T$.
- Una SVD di $-A$ e una SVD di αA dove $\alpha < 0$. Suggerimento $A = -V\Sigma U^T = (-V)\Sigma U^T$.
- Una SVD di $A^T A$ e di AA^T.

Determinare la SVD significa determinare le matrici U, Σ e V.

Esercizio. Determinare, seguendo la costruzione indicata nella dimostrazione del precedente teorema, delle SVD per le matrici A viste negli esercizi delle sezioni precedenti "Le matrici $A^T A$ e AA^T" e "Costruire V da U" e per le matrici

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix},\ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},\ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \text{ e } \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Determinare una SVD anche per le trasposte di queste matrici.

Esercizio. Sia $x \in \mathbb{R}^n$. Descrivere una SVD di $x \in \mathbb{R}^{n\times 1}$ e di $x^T \in \mathbb{R}^{1\times n}$. Determinare tali SVD per $x = (1, 2, -1, -1)$.

Esercizio. Sia $A \in \mathbb{R}^{n\times n}$ e siano $\sigma_1, \sigma_2, \ldots, \sigma_n$ i suoi valori singolari. Mostrare che

$$|\det(A)| = \sigma_1\sigma_2\cdots\sigma_n.$$

Esercizio. Determinare una SVD di una matrice ortogonale.

7.1 Nucleo, immagine e rango

Il nucleo, l'immagine e il rango di una matrice si ottengono immediatamente dalla sua SVD.

Teorema 10 *Sia $A \in \mathbb{R}^{m\times n}$ e sia $A = V\Sigma U^T$ una sua SVD con*

$$\begin{aligned} U &= [u^{(1)} \ \cdots \ u^{(r)} \ u^{(r+1)} \ \cdots \ u^{(n)}] \\ V &= [v^{(1)} \ \cdots \ v^{(r)} \ v^{(r+1)} \ \cdots \ v^{(m)}], \end{aligned}$$

dove r è la dimensione di Σ_r in Σ. Si ha

$$\ker(A) = \operatorname{span}\left(u^{(r+1)}, \ldots, u^{(n)}\right)$$

e

$$\operatorname{Im}(A) = \operatorname{span}\left(v^{(1)}, \ldots, v^{(r)}\right).$$

Dimostrazione. Si ha

$$\begin{aligned} &\ker(A) = \{x \in \mathbb{R}^n : Ax = \underline{0}\} = \left\{x \in \mathbb{R}^n : V\Sigma U^T x = \underline{0}\right\} \\ &= \left\{x \in \mathbb{R}^n : \Sigma U^T x = \underline{0}\right\} = \{Uy : y \in \mathbb{R}^n \text{ e } \Sigma y = \underline{0}\} \\ &\ \text{ponendo } y = U^T x \text{ e quindi } x = Uy \\ &= \{Uy : y \in \mathbb{R}^n \text{ e } (\sigma_1 y_1, \ldots, \sigma_r y_r, 0, \ldots, 0) = \underline{0}\} \\ &= \{Uy : y \in \mathbb{R}^n \text{ e } y_1 = \cdots = y_r = 0\} \\ &= \operatorname{span}\left(u^{(r+1)}, \ldots, u^{(n)}\right) \end{aligned}$$

e

$$\begin{aligned} &\operatorname{Im}(A) = \{Ax : x \in \mathbb{R}^n\} = \left\{V\Sigma U^T x : x \in \mathbb{R}^n\right\} \\ &= \{V\Sigma y : y \in \mathbb{R}^n\} \text{ ponendo } y = U^T x \\ &= \{Vz : z \in \mathbb{R}^m \text{ e } z_{r+1} = \ldots = z_m = 0\} \\ &\text{ponendo } z = \Sigma y = (\sigma_1 y_1, \ldots, \sigma_r y_r, 0, \ldots, 0) \\ &= \operatorname{span}\left(v^{(1)}, \ldots, v^{(r)}\right). \end{aligned}$$

■

Come conseguenza del precedente teorema si ha

$$\operatorname{rank}(A) = r,$$

essendo $v^{(1)}, \ldots, v^{(r)}$ una base per $\operatorname{Im}(A)$ (sono vettori linearmente indipendenti) e $\operatorname{rank}(A) = \dim(\operatorname{Im}(A))$.

Essendo poi $A^T = U\Sigma^T V^T$ una SVD di A^T, risulta anche

$$\ker(A^T) = \operatorname{span}\left(v^{(r+1)}, \ldots, v^{(m)}\right)$$

e

$$\operatorname{Im}(A^T) = \operatorname{span}\left(u^{(1)}, \ldots, u^{(r)}\right).$$

Pertanto

$$\operatorname{rank}(A^T) = \dim(\operatorname{Im}(A^T)) = r = \operatorname{rank}(A).$$

Esercizio. Tramite la SVD, determinare nucleo e immagine delle matrici A e delle matrici

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \text{ e } \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

considerate in un esercizio della precedente sezione, e delle loro trasposte.

7.2 Norma 2 e norma di Frobenius

Una norma di matrice

$$\|\cdot\| : \bigcup_{\substack{m,n \text{ interi} \\ \text{positivi}}} \mathbb{R}^{m\times n} \to \mathbb{R}$$

si dice *ortogonalmente invariante* se

$$\begin{aligned}&\forall m, n \text{ interi positivi } \forall A \in \mathbb{R}^{m\times n}\ \forall Q \in \mathbb{R}^{m\times m} \text{ ortogonale} \\ &\forall Z \in \mathbb{R}^{n\times n} \text{ ortogonale: } \|QAZ\| = \|A\|.\end{aligned}$$

La norma 2 e la norma di Frobenius sono ortogonalmente invarianti, come ora mostriamo.

Siano $A \in \mathbb{R}^{m\times n}$, $Q \in \mathbb{R}^{m\times m}$ ortogonale e $Z \in \mathbb{R}^{n\times n}$ ortogonale.

Per quanto riguarda la norma 2, si ha

$$\begin{aligned}\|QAZ\|_2 &= \max_{x\in\mathbb{R}^n\setminus\{\underline{0}\}} \frac{\|QAZx\|_2}{\|x\|_2} \\ &= \max_{x\in\mathbb{R}^n\setminus\{\underline{0}\}} \frac{\|QAZx\|_2}{\|Zx\|_2} \quad \text{essendo } Z \text{ ortogonale} \\ &= \max_{y\in\mathbb{R}^n\setminus\{\underline{0}\}} \frac{\|QAy\|_2}{\|y\|_2} \quad \text{ponendo } y = Zx \\ &= \max_{y\in\mathbb{R}^n\setminus\{\underline{0}\}} \frac{\|Ay\|_2}{\|y\|_2} \quad \text{essendo } Q \text{ ortogonale} \\ &= \|A\|_2 .\end{aligned}$$

Per provare l'invarianza ortogonale della norma di Frobenius osserviamo che, per $B \in \mathbb{R}^{m\times n}$, risulta

$$\operatorname{tr}\left(B^T B\right) = \operatorname{tr}\left(BB^T\right)$$

e

$$\|B\|_F = \sqrt{\operatorname{tr}\left(B^T B\right)} = \sqrt{\operatorname{tr}\left(BB^T\right)}.$$

Infatti,

$$\operatorname{tr}\left(B^T B\right) = \sum_{i=1}^{n} \left(B^T B\right)(i,i) = \sum_{i=1}^{n}\sum_{j=1}^{m} \underbrace{B^T(i,j)}_{=B(j,i)} B(j,i) = \sum_{i=1}^{n}\sum_{j=1}^{m} b_{ji}^2 = \|B\|_F^2$$

e

$$\operatorname{tr}\left(BB^T\right) = \sum_{i=1}^{m}\left(BB^T\right)(i,i) = \sum_{i=1}^{m}\sum_{j=1}^{n} B(i,j)\underbrace{B^T(j,i)}_{=B(i,j)} = \sum_{i=1}^{m}\sum_{j=1}^{n} b_{ij}^2 = \|B\|_F^2 .$$

Per cui,

$$\begin{aligned}
&\|QAZ\|_F^2 = \operatorname{tr}\left((QAZ)^T QAZ\right) = \operatorname{tr}\left(Z^T A^T Q^T QAZ\right) = \operatorname{tr}\left(Z^T A^T AZ\right) \\
&= \operatorname{tr}\left((AZ)^T AZ\right) = \operatorname{tr}\left(AZ\,(AZ)^T\right) = \operatorname{tr}\left(AZZ^T A^T\right) = \operatorname{tr}\left(AA^T\right) = \|A\|_F^2 .
\end{aligned}$$

Esercizio. Siano $B \in \mathbb{R}^{m\times n}$ e $C \in \mathbb{R}^{n\times m}$. Provare che

$$\operatorname{tr}(BC) = \operatorname{tr}(CB).$$

Il seguente teorema fornisce la norma 2 e la norma di Frobenius di una matrice in termini di valori singolari.

Teorema 11 *Sia $A \in \mathbb{R}^{m\times n}$ e siano $\sigma_1, \ldots, \sigma_p$ i valori singolari di A. Si ha*

$$\|A\|_2 = \sigma_1 \quad e \quad \|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_p^2}.$$

Il motivo per cui nell'enunciato compaiono tutti i valori singolari $\sigma_1, \ldots, \sigma_p$, invece che solo i valori singolari non nulli $\sigma_1, \ldots, \sigma_r$, è per includere anche il caso $r = \operatorname{rank}(A) = 0$, cioè $A = 0$.

Dimostrazione. Sia $A = V\Sigma U^T$ una SVD di A. Dal momento che la norma 2 e la norma di Frobenius sono ortogonalmente invarianti e U e V^T sono ortogonali, si ha

$$\|A\|_2 = \|\Sigma\|_2 \text{ e } \|A\|_F = \|\Sigma\|_F .$$

Risulta

$$\|\Sigma\|_2 = \max_{x\in\mathbb{R}^n\setminus\{\underline{0}\}} \frac{\|\Sigma x\|_2}{\|x\|_2} = \sigma_1,$$

Infatti, per $x \in \mathbb{R}^n \setminus \{\underline{0}\}$, si ha

$$\begin{aligned}
&\|\Sigma x\|_2 = \|(\sigma_1 x_1, \ldots, \sigma_p x_p, 0, \ldots, 0)\|_2 \\
&= \sqrt{(\sigma_1 x_1)^2 + \cdots + (\sigma_p x_p)^2} = \sqrt{\sigma_1^2 x_1^2 + \cdots + \sigma_p^2 x_p^2} \\
&\leq \sqrt{\sigma_1^2 x_1^2 + \cdots + \sigma_1^2 x_p^2} \text{ essendo } \sigma_1 \geq \cdots \geq \sigma_p \\
&= \sqrt{\sigma_1^2\left(x_1^2 + \cdots + x_p^2\right)} = \sigma_1\sqrt{x_1^2 + \cdots + x_p^2} \\
&\leq \sigma_1\sqrt{x_1^2 + \cdots + x_p^2 + x_{p+1}^2 + \cdots + x_n^2} = \sigma_1 \|x\|_2
\end{aligned}$$

da cui

$$\frac{\|\Sigma x\|_2}{\|x\|_2} \leq \sigma_1.$$

Poi, per $x = e^{(1)}$, si ha $\|x\|_2 = 1$ e

$$\begin{aligned}
\|\Sigma x\|_2 &= \|\Sigma(:,1)\|_2 \\
&= \|(\sigma_1, 0, \ldots, 0, \ldots, 0)\|_2 = \sigma_1,
\end{aligned}$$

da cui

$$\frac{\|\Sigma x\|_2}{\|x\|_2} = \sigma_1.$$

Risulta poi per la norma di Frobenius

$$\|\Sigma\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} \Sigma\,(i,j)^2} = \sqrt{\sigma_1^2 + \cdots + \sigma_p^2}.$$

■

Per cui, ora abbiamo anche una formula per la norma 2 di una matrice: per $A \in \mathbb{R}^{m \times n}$, si ha

$$\begin{aligned}
\|A\|_2 &= \sigma_1 = \sqrt{\max\{\lambda : \lambda \text{ è autovalore di } A^T A\}} \\
&= \sqrt{\max\{\lambda : \lambda \text{ è autovalore di } AA^T\}}.
\end{aligned}$$

Si osservi che

$$\|A\|_2 = \max_{x \in \mathbb{R}^n \setminus \{\underline{0}\}} \frac{\|Ax\|_2}{\|x\|_2} = \frac{\|Au^{(1)}\|_2}{\|u^{(1)}\|_2}$$

dove $u^{(1)}$ è il vettore singolare destro relativo a σ_1.

Infatti, $\|u^{(1)}\|_2 = 1$ e

$$\|Au^{(1)}\|_2 = \|\sigma_1 v^{(1)}\|_2 = \sigma_1 \underbrace{\|v^{(1)}\|_2}_{=1} = \sigma_1 = \|A\|_2\,,$$

dove $v^{(1)}$ è il vettore singolare sinistro relativo a σ_1.

Esercizio. Determinare la norma 2 delle matrici

$$\begin{bmatrix} a & 1 \\ a & 1 \\ a & 1 \end{bmatrix}, \begin{bmatrix} a & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} a & 1 \\ 0 & 0 \end{bmatrix} \text{ e } \begin{bmatrix} 0 & -a \\ 1 & 0 \end{bmatrix}, \text{ dove } a \in \mathbb{R}.$$

Esercizio. Quale relazione sussiste tra la norma 2 di $A \in \mathbb{R}^{m \times n}$ e la norma 2 di A^T?

Esercizio. Dato $x \in \mathbb{R}^n$, determinare la norma 2 della matrice $x \in \mathbb{R}^{n\times 1}$ utilizzando la formula sopra per la norma 2 di matrice.

Esercizio. Determinare la norma 2 di una matrice $Q \in \mathbb{R}^{n\times n}$ ortogonale utilizzando la definizione di norma 2 di matrice come norma di operatore. Fare lo stesso utilizzando la formula per la norma 2.

Esercizio. Si consideri la matrice

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Si determini un vettore $x \in \mathbb{R}^2$ tale che

$$\frac{\|Ax\|_2}{\|x\|_2} = \|A\|_2\,.$$

Fare lo stesso per la matrice

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

Si osservi che

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{\operatorname{rank}(A)}\ \|A\|_2.$$

Infatti

$$\|A\|_2 = \sigma_1 \leq \|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_p^2}$$

e, essendo $r = \operatorname{rank}(A)$ il numero di valori singolari non nulli,

$$\begin{aligned} \|A\|_F &= \sqrt{\sigma_1^2 + \cdots + \sigma_p^2} = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} \\ &\leq \sqrt{\sigma_1^2 + \cdots + \sigma_1^2} = \sqrt{r\sigma_1^2} = \sqrt{r}\sigma_1 = \sqrt{r}\|A\|_2. \end{aligned}$$

Esercizio. Sia $A \in \mathbb{R}^{m\times n}$. Provare che $\|A\|_2 < \|A\|_F$ se e solo se $\operatorname{rank}(A) \geq 2$. Descrivere poi per quali matrici si ha $\|A\|_F = \sqrt{\operatorname{rank}(A)}\ \|A\|_2$.

7.3 La SVD dell'inversa

Vediamo ora come è legata la SVD di A^{-1} alla SVD di A, dove $A \in \mathbb{R}^{n\times n}$ è non singolare. Si osservi che $r = \operatorname{rank}(A) = n$, essendo A non singolare.

Sia $A = V\Sigma U^T$ una SVD di A. Qui $U, V \in \mathbb{R}^{n\times n}$ sono ortogonali e $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n) \in \mathbb{R}^{n\times n}$ con $\sigma_1 \geq \cdots \geq \sigma_n > 0$ essendo $r = n$. Si ha

$$A^{-1} = \left(V\Sigma U^T\right)^{-1} = U\Sigma^{-1}V^T,$$

ma questa non è una SVD di A^{-1} dal momento che in

$$\Sigma^{-1} = \operatorname{diag}\left(\sigma_1^{-1}, \ldots, \sigma_n^{-1}\right)$$

gli elementi diagonali sono in ordine crescente.

Dette $u^{(1)}, \ldots, u^{(n)}$ le colonne di U e $v^{(1)}, \ldots, v^{(n)}$ le colonne di V si ha però

$$
\begin{aligned}
A^{-1} &= U\Sigma^{-1}V^T \\
&= \left[u^{(1)} \ldots u^{(n)}\right] \operatorname{diag}\left(\sigma_1^{-1}, \ldots, \sigma_n^{-1}\right) \begin{bmatrix} \left(v^{(1)}\right)^T \\ \vdots \\ \left(v^{(n)}\right)^T \end{bmatrix} \\
&= \sigma_1^{-1} u^{(1)} (v^{(1)})^T + \cdots + \sigma_n^{-1} u^{(n)} (v^{(n)})^T \\
&= \sigma_n^{-1} u^{(n)} (v^{(n)})^T + \cdots + \sigma_1^{-1} u^{(1)} (v^{(1)})^T \\
&= \left[u^{(n)} \ldots u^{(1)}\right] \operatorname{diag}\left(\sigma_n^{-1}, \ldots, \sigma_1^{-1}\right) \begin{bmatrix} \left(v^{(n)}\right)^T \\ \vdots \\ \left(v^{(1)}\right)^T \end{bmatrix} \\
&= \left[u^{(n)} \ldots u^{(1)}\right] \operatorname{diag}\left(\sigma_n^{-1}, \ldots, \sigma_1^{-1}\right) \left[v^{(n)} \ldots v^{(1)}\right]^T.
\end{aligned}
$$

Per cui

$$A^{-1} = \left[u^{(n)} \ldots u^{(1)}\right] \operatorname{diag}\left(\sigma_n^{-1}, \ldots, \sigma_1^{-1}\right) \left[v^{(n)} \ldots v^{(1)}\right]^T$$

e questa è una SVD di A^{-1}, dal momento che $\left[u^{(n)} \ldots u^{(1)}\right]$ è ortogonale essendo $u^{(n)}, \ldots, u^{(1)}$ ortonormali, $\left[v^{(n)} \ldots v^{(1)}\right]$ è ortogonale essendo $v^{(n)}, \ldots, v^{(1)}$ ortonormali e in $\operatorname{diag}\left(\sigma_n^{-1}, \ldots, \sigma_1^{-1}\right)$ gli elementi diagonali sono in ordine decrescente.

Esercizio. Determinare la SVD dell'inversa delle matrici

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \text{ e } \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

precedentemente considerate.

Si ha

$$\|A^{-1}\|_2 = \text{massimo valore singolare di } A^{-1} = \sigma_n^{-1}.$$

Esercizio. Determinare la norma 2 dell'inversa delle matrici

$$\begin{bmatrix} a & 1 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & -a \\ 1 & 0 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} -1 & -a \\ a & -1 \end{bmatrix},$$

dove $a \neq 0$ nelle prime due, senza calcolare l'inversa.

7.4 SVD di matrici simmetriche

Una SVD di una matrice simmetrica si ottiene immediatamente da una diagonalizzazione con trasformazione ortogonale della matrice.

Sia $A \in \mathbb{R}^{n\times n}$ simmetrica e sia $U^T AU = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ una diagonalizzazione di A con matrice di trasformazione

$$U = \left[u^{(1)} \ldots u^{(n)}\right] \in \mathbb{R}^{n\times n}$$

ortogonale, dove $u^{(1)}, \ldots, u^{(n)}$ sono le colonne di U.

Senza perdita di generalità si può supporre $|\lambda_1| \geq \cdots \geq |\lambda_n|$ con $|\lambda_1| \geq \cdots \geq |\lambda_r| > 0$.

Si ha $A = U\operatorname{diag}(\lambda_1, \ldots, \lambda_n)\, U^T$, ma questa non è una SVD di A in quanto gli elementi diagonali in $\lambda_1, \ldots, \lambda_n$ potrebbero essere negativi.

Si ha però

$$\begin{aligned}
&A = U\operatorname{diag}(\lambda_1, \ldots, \lambda_n)\, U^T \\
&= \left[u^{(1)} \ldots u^{(r)}\ u^{(r+1)} \ldots u^{(n)}\right] \operatorname{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)\, U^T \\
&= \left[u^{(1)} \ldots u^{(r)}\ u^{(r+1)} \ldots u^{(n)}\right] \operatorname{diag}(\operatorname{sign}(\lambda_1), \ldots, \operatorname{sign}(\lambda_r), 1, \ldots, 1) \\
&\quad \cdot \operatorname{diag}(|\lambda_1|, \ldots, |\lambda_r|, 0, \ldots, 0)\, U^T \\
&= \left[\operatorname{sign}(\lambda_1)\, u^{(1)} \ldots \operatorname{sign}(\lambda_r)\, u^{(r)}\ u^{(r+1)} \ldots u^{(n)}\right] \\
&\quad \cdot \operatorname{diag}(|\lambda_1|, \ldots, |\lambda_r|, 0, \ldots, 0)\, U^T.
\end{aligned}$$

Esercizio. Mostrare che

$$\operatorname{sign}(\lambda_1)\, u^{(1)}, \ldots, \operatorname{sign}(\lambda_r)\, u^{(r)}, u^{(r+1)}, \ldots, u^{(n)}$$

sono vettori ortonormali.

Quindi

$$V = \left[\operatorname{sign}(\lambda_1)\, u^{(1)} \ldots \operatorname{sign}(\lambda_r)\, u^{(r)}\ u^{(r+1)} \ldots u^{(n)}\right] \in \mathbb{R}^{n\times n}$$

è ortogonale e pertanto

$$A = V\operatorname{diag}(|\lambda_1|, \ldots, |\lambda_r|, 0, \ldots, 0)\, U^T$$

è una SVD di A, dal momento che U e V sono ortogonali e $|\lambda_1| \geq \cdots \geq |\lambda_r| > 0$.

Pertanto, i valori singolari di A sono i moduli degli autovalori di A:

$$\sigma_i = |\lambda_i|,\ i \in \{1, \ldots, n\}.$$

Esercizio. Determinare una SVD della matrice simmetrica

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

Esercizio. Come si può determinare il rango di una matrice simmetrica a partire dai suoi autovalori?

Per una matrice simmetrica $A \in \mathbb{R}^{n\times n}$, si ha quindi

$$\|A\|_2 = |\lambda_1| \quad \text{e} \quad \|A\|_F = \sqrt{|\lambda_1|^2 + \cdots + |\lambda_n|^2} = \sqrt{\lambda_1^2 + \cdots + \lambda_n^2}.$$

Risulta allora

$$\|A\|_2 = \rho(A) := \max\{|\lambda| : \lambda \text{ è un autovalore di } A\}.$$

Il numero $\rho(A)$ è detto il *raggio spettrale* di A.

Esercizio. Determinare la norma 2 delle matrici

$$\begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} \text{ e } \begin{bmatrix} 0 & a \\ a & 0 \end{bmatrix},$$

dove $\theta \in [0, 2\pi)$ e $a \in \mathbb{R}$.

Esercizio. Determinare la norma 2 della matrice diagonale $\operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn})$.

Esercizio. A cosa è uguale la norma dell'inversa di una matrice simmetrica non singolare, in termini di autovalori della matrice?

Esercizio. Descrivere la SVD di una matrice simmetrica semidefinita positiva.

Esercizio. Utilizzando il procedimento precedentemente visto per trovare la SVD di una qualunque matrice, ritrovare la SVD di una matrice simmetrica $A \in \mathbb{R}^{n\times n}$ descritta in questa sezione. Suggerimento: mostrare che se

$$U^T A U = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$$

è una diagonalizzazione con trasformazione ortogonale di A, dove

$$|\lambda_1| \geq \cdots \geq |\lambda_r| > 0 = |\lambda_{r+1}| = \cdots = |\lambda_n|,$$

allora

$$U^T A^T A U = U^T A^2 U = \operatorname{diag}\left(\lambda_1^2, \ldots, \lambda_n^2\right)$$

e poi da qui costruire la SVD di A.

7.5 SVD leggera e decomposizione come somma di matrici di rango 1

Sia $A = V\Sigma U^T$ una SVD di $A \in \mathbb{R}^{m\times n}$, siano $\sigma_1, \ldots, \sigma_r$ i valori singolari non nulli di A e siano $u^{(1)}, \ldots, u^{(r)}$ e $v^{(1)}, \ldots, v^{(r)}$ i corrispondenti vettori singolari destri e sinistri, rispettivamente.

Teorema 12 *Si ha*

$$A = V_r \Sigma_r U_r^T = \sum_{k=1}^{r} \sigma_k E_k,$$

dove $V_r = \begin{bmatrix} v^{(1)} & \ldots & v^{(r)} \end{bmatrix} \in \mathbb{R}^{m\times r}$, $\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r\times r}$, $U_r = \begin{bmatrix} u^{(1)} & \ldots & u^{(r)} \end{bmatrix} \in \mathbb{R}^{n\times r}$ *e*

$$E_k = v^{(k)}(u^{(k)})^T \in \mathbb{R}^{m\times n}, \ k \in \{1, \ldots, r\}.$$

Osservare che, per $k \in \{1, \ldots, r\}$, E_k ha elementi

$$E_k(i,j) = v_i^{(k)} u_j^{(k)}, \ (i,j) \in \{1, \ldots, m\} \times \{1, \ldots, n\}.$$

Dimostrazione. Poichè

$$V = \begin{bmatrix} V_r & Z \end{bmatrix} \ \text{ e } \ U = \begin{bmatrix} U_r & W \end{bmatrix},$$

dove $Z = \left[v^{(r+1)} \ \ldots \ v^{(m)}\right]$ e $W = \left[u^{(r+1)} \ \ldots \ u^{(n)}\right]$, si ha

$$\begin{aligned} A &= V\Sigma U^T = \begin{bmatrix} V_r & Z \end{bmatrix} \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_r^T \\ W^T \end{bmatrix} = \begin{bmatrix} V_r\Sigma_r & 0 \end{bmatrix} \begin{bmatrix} U_r^T \\ W^T \end{bmatrix} \\ &= V_r \Sigma_r U_r^T. \end{aligned}$$

Si ha poi

$$\begin{aligned} V_r\Sigma_r U_r^T &= \left[v^{(1)} \ \ldots \ v^{(r)}\right] \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \begin{bmatrix} \left(u^{(1)}\right)^T \\ \vdots \\ \left(u^{(r)}\right)^T \end{bmatrix} \\ &= \left[\sigma_1 v^{(1)} \ \ldots \ \sigma_r v^{(r)}\right] \begin{bmatrix} \left(u^{(1)}\right)^T \\ \vdots \\ \left(u^{(r)}\right)^T \end{bmatrix} = \sum_{k=1}^{r} \left(\sigma_k v^{(k)}\right) \left(u^{(k)}\right)^T = \sum_{k=1}^{r} \sigma_k \cdot \underbrace{v^{(k)} \left(u^{(k)}\right)^T}_{=E_k}. \end{aligned}$$

■

La

$$A = V_r \Sigma_r U_r^T$$

è detta una *SVD leggera* di A, in quanto contiene solo le colonne delle matrici V e U e gli elementi diagonali di Σ che sono indispensabili per la ricostruzione della matrice A.

La SVD leggera scritta come

$$A = \sum_{k=1}^{r} \sigma_k E_k$$

è una decomposizione di A come somma di matrici di rango 1: le matrici

$$\sigma_k E_k = (\sigma_k v^{(k)})(u^{(k)})^T, k \in \{1, \ldots, r\},$$

come pure le matrici $E_K = v^{(k)}(u^{(k)})^T$, hanno rango 1.

Infatti, come ora proviamo, per $v \in \mathbb{R}^m \setminus \{\underline{0}\}$ e $u \in \mathbb{R}^n \setminus \{\underline{0}\}$ risulta $\operatorname{rank}(vu^T) = 1$, dove $vu^T \in \mathbb{R}^{m \times n}$.

Poiché $u \neq \underline{0}$, esiste $j^* \in \{1,\ldots,n\}$ tale che $u_{j^*} \neq 0$. Per $j \in \{1,\ldots,n\}$, si ha

$$\begin{aligned} j-\text{esima colonna di } vu^T &= \begin{bmatrix} v_1 u_j \\ \vdots \\ v_m u_j \end{bmatrix} = \frac{u_j}{u_{j^*}} \underbrace{\begin{bmatrix} v_1 u_{j^*} \\ \vdots \\ v_m u_{j^*} \end{bmatrix}}_{\neq \underline{0} \text{ essendo } v \neq \underline{0}} \\ &= \frac{u_j}{u_{j^*}} \cdot \underbrace{j^* - \text{esima colonna di } vu^T}_{\neq \underline{0}}. \end{aligned}$$

Si conclude che $\operatorname{rank}(vu^T) = 1$, avendo vu^T una colonna non nulla, la j^*-esima, e tutte le altre multiple di questa.

Esercizio. Determinare la SVD leggera e le matrici E_k di rango 1 per le matrici A e

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \text{ e } \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

viste in precedenti esercizi.

Esercizio. Dare la decomposizione di A^T come somma di matrici di rango 1 in termini dei valori singolari non nulli σ_k, $k \in \{1,\ldots,r\}$, e dei corrispondenti vettori singolari $u^{(k)}, v^{(k)}$, $k \in \{1,\ldots,r\}$, di A.

8 Approssimazione con matrici di rango inferiore

Sia $A = V\Sigma U^T$ una SVD di $A \in \mathbb{R}^{m\times n}$, siano $\sigma_1,\ldots,\sigma_r$ i valori singolari non nulli di A e siano $u^{(1)},\ldots,u^{(r)}$ e $v^{(1)},\ldots,v^{(r)}$ i corrispondenti vettori singolari destri e sinistri, rispettivamente.

Per $l \in \{0,1,\ldots,r\}$, sia

$$A_l = V_l \Sigma_l U_l^T \in \mathbb{R}^{m\times n},$$

dove $V_l = \begin{bmatrix} v^{(1)} & \ldots & v^{(l)} \end{bmatrix} \in \mathbb{R}^{m\times l}$, $\Sigma_l = \operatorname{diag}(\sigma_1,\ldots,\sigma_l) \in \mathbb{R}^{l\times l}$ e $U_l = \begin{bmatrix} u^{(1)} & \ldots & u^{(l)} \end{bmatrix} \in \mathbb{R}^{n\times l}$. Per convenzione si pone $A_0 = 0$.

Si osservi che i casi estremi sono $A_0 = 0$ e $A_r = V_r\Sigma_r U_r^T = A$, essendo $A = V_r\Sigma_r U_r^T$ la SVD leggera di A.

Per $l \in \{0,1,\ldots,r\}$, procedendo come nel caso della SVD leggera, si prova che

$$A_l = V_l\Sigma_l U_l^T = \sum_{k=1}^{l} \sigma_k E_k,$$

dove $E_k = v^{(k)}(u^{(k)})^T \in \mathbb{R}^{m\times n},\ k \in \{1, \ldots, r\}$.

Le $r+1$ matrici $A_0, A_1, \ldots, A_r$ di $\mathbb{R}^{m\times n}$ risultano essere approssimazioni sempre migliori della matrice A.

Teorema 13 *Sia* $l \in \{0, 1, \ldots, r\}$. *Si ha* $\operatorname{rank}(A_l) = l$ *e*

$$\min_{\substack{B\in\mathbb{R}^{m\times n}\\ \operatorname{rank}(B)=l}} \|B - A\|_2 = \|A_l - A\|_2 = \begin{cases} \sigma_{l+1} \ se\ l < r \\ 0\ se\ l = r \end{cases}.$$

Per cui, per $l \in \{0, 1, \ldots, r\}$, A_l (che ha rango l) è la miglior approssimazione della matrice $A \in \mathbb{R}^{m\times n}$ tra tutte le matrici B di $\mathbb{R}^{m\times n}$ di rango l, quando si misura l'errore di approssimazione con la metrica su $\mathbb{R}^{m\times n}$ dedotta dalla norma 2 di matrice.

L' errore di questa miglior approssimazione è σ_{l+1} se $l < r$. L'errore è zero se $l = r$ in quanto $A_r = A$.

La dimostrazione di questo teorema non sarà data.

Esercizio. Verificare la correttezza del teorema precedente nei due casi estremi $l = 0$ e $l = r$.

Esercizio. Determinare la miglior approssimazione della matrice

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

tra tutte le matrici di $\mathbb{R}^{2\times 2}$ di rango 1, quando si misura l'errore di approssimazione nella metrica su $\mathbb{R}^{2\times 2}$ dedotta dalla norma 2 di matrice.

9 SVD e compressione dei dati

Una matrice $A \in \mathbb{R}^{m\times n}$ viene usualmente rappresentata tramite i suoi mn elementi. Il "rappresentare" la matrice può anche voler dire memorizzarla in qualche dispositivo come RAM, hard disk, ...

La SVD (leggera) fornisce un modo di rappresentare la matrice A che è alternativo a quello di rappresentarla tramite i suoi elementi: rappresentando A con i suoi valori singolari non nulli $\sigma_1, \ldots, \sigma_r$ e i corrispondenti vettori singolari destri $u^{(1)}, \ldots, u^{(r)} \in \mathbb{R}^n$ e sinistri $v^{(1)}, \ldots, v^{(r)} \in \mathbb{R}^m$, A può essere ricostruita tramite la sua SVD leggera

$$A = V_r \Sigma_r U_r^T.$$

Tramite i suoi elementi la matrice viene rappresentata con mn scalari, mentre con la SVD si utilizzano

$$\underbrace{r}_{\text{valori singolari}} + \underbrace{rn}_{\text{vettori singolari destri}} + \underbrace{rm}_{\text{vettori singolari sinistri}} = r\,(1 + n + m)$$

scalari.

La rappresentazione con la SVD risulta migliore della rappresentazione tramite gli elementi se il *coefficiente di riduzione*

$$C = \frac{mn}{r\,(1+n+m)}$$

è maggiore di 1.

Dette $q = \min\{m,n\}$ e $kq = \max\{m,n\}$ le dimensioni di A, dove

$$k = \frac{\max\{m,n\}}{\min\{m,n\}} \geq 1$$

è il *coefficiente di rettangolarità* di A, si ha

$$\begin{aligned} C &= \frac{mn}{r\,(1+n+m)} = \frac{kq^2}{r\,(q+kq+1)} \\ &= \frac{q}{r} \cdot \frac{1}{1+\frac{1}{k}+\frac{1}{kq}}. \end{aligned}$$

Per matrici di grandi dimensioni, si ha $kq >> 1$ e quindi

$$C \approx \frac{q}{r} \cdot \frac{1}{1+\frac{1}{k}}.$$

Per tali matrici, la rappresentazione con la SVD è conveniente se e solo se

$$r = \operatorname{rank}(A) < \frac{q}{1+\frac{1}{k}}.$$

Esercizio. Come deve essere r rispetto a q in modo che la rappresentazione con la SVD sia conveniente per qualunque coefficiente di rettangolarità?

Invece di rappresentare esattamente la matrice A tramite la sua SVD (leggera)

$$A = V_r \Sigma_r U_r^T,$$

si può anche pensare di rappresentarla in maniera approssimata tramite

$$A \approx A_l = V_l \Sigma_l U_l^T,$$

per qualche $l \in \{0,1,\ldots,r\}$, utilizzando la sua miglior approssimazione A_l di rango l.

Nel rappresentare l'approssimazione A_l si utilizzano solo i primi l valori singolari e i primi l vettori singolari destri e sinistri.

Quando si utilizza A_l il coefficiente di riduzione diventa

$$C \approx \frac{q}{l} \cdot \frac{1}{1+\frac{1}{k}},$$

ottenuto rimpiazzando r con l nel precedente coefficiente di riduzione. La rappresentazione è conveniente se e solo se

$$l < \frac{q}{1+\frac{1}{k}}.$$

La differenza rispetto a prima, dove si aveva r al posto di l, è che ora l può essere scelto, mentre r è fissato.

Esercizio. Come deve essere l rispetto a q in modo che la rappresentazione sia conveniente per qualunque coefficiente di rettangolarità?

Utilizzando la norma 2 di matrice per misurare la grandezza delle matrici, l'errore relativo dell'approssimazione A_l di A è

$$\frac{\|A_l - A\|_2}{\|A\|_2} = \frac{\sigma_{l+1}}{\sigma_1}.$$

Nella pratica, spesso si fissa una tolleranza TOL per l'errore relativo e si cerca il minimo l tale che

$$\frac{\sigma_{l+1}}{\sigma_1} \leq \text{TOL} .$$

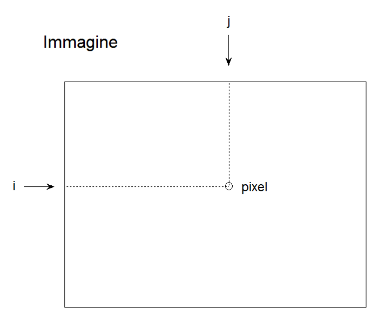
Per una matrice A la cui sequenza (decrescente) dei valori singolari $\sigma_1, \ldots, \sigma_r$ decade rapidamente, si ottengono buone approssimazioni di A per l non grande e con un conseguente buon coefficiente di riduzione.

Le considerazioni svolte hanno applicazione nella compressione di dati, pensando questi come organizzati in una matrice A.

Un esempio importante è la compressione di immagini:

- un'immagine in bianco e nero è una matrice A^{BW} di grandi dimensioni in cui ogni coppia (i, j) di indici di riga e colonna rappresenta un pixel e l'elemento a_{ij} è il livello (numero in $[0, 1]$) di grigio di quel pixel;

- un'immagine a colori RGB è invece una terna (A^R, A^G, A^B) di matrici delle stesse grandi dimensioni, i cui elementi sono le percentuali dei tre colori Red, Green e Blue, rispettivamente, nei pixel.



Molto spesso, per l non grande, l'immagine A_l^{BW} approssima l'immagine A^{BW}, e l'immagine (A_l^R, A_l^G, A_l^B) approssima l'immagine (A^R, A^G, A^B), con una qualità soddisfacente.

Un aspetto da tenere presente nella compressione di immagini con la SVD è che si ha

$$A_1 = \sigma_1 v^{(1)} (u^{(1)})^T$$

e

$$A_{k+1} = A_k + \sigma_{k+1} v^{(k+1)} (u^{(k+1)})^T, \; k \in \{1, \ldots, r-1\},$$

e, quindi, se si tratta di spedire a un ricevente un'immagine A in una linea di trasmissione lenta, conviene trasmettere prima $\sigma_1, u^{(1)}, v^{(1)}$, poi $\sigma_2, u^{(2)}, v^{(2)}$, poi $\sigma_3, u^{(3)}, v^{(3)}$ e così avanti in modo che l'acquisizione dell'immagine sia ottenuta per raffinamenti successivi.

Un altro aspetto importante è che vale

$$\begin{array}{l} A_{k+1}(i,j) = A_k(i,j) + \sigma_{k+1} v_i^{(k+1)} u_j^{(k+1)} \\ (i,j) \in \{1, \ldots, m\} \times \{1, \ldots, n\}, \end{array}$$

Quindi

$$A_{k+1}(I,J) = A_k(I,J) + \sigma_{k+1} v^{(k+1)}(I) (u^{(k+1)})^T (J),$$

dove $I = (i_1, \ldots, i_p)$ è una $p-$upla di indici di riga, $J = (j_1, \ldots, j_q)$ è una $q-$upla di indici di colonna e

$$A_{k+1}(I,J), A_k(I,J), v^{(k+1)}(I), (u^{(k+1)})^T(J)$$

sono le sottomatrici di $A_{k+1}, A_k, v^{(k+1)}, (u^{(k+1)})^T$, rispettivamente, relative agli indici di riga in I e agli indici di colonna in J.

Questo significa che il raffinamento dell'immagine può essere fatto anche solo in una parte (I,J) dell'immagine che è di interesse, trasmettendo solo $\sigma_{k+1}, u^{(k+1)}(J), v^{(k+1)}(I)$ invece dell'intero pacchetto $\sigma_{k+1}, u^{(k+1)}, v^{(k+1)}$.