LA DECOMPOSIZIONE AI VALORI SINGOLARI (SVD)

La decomposizione in valori singolari: SVD.

Si dimostra che per ogni matrice $A \in \mathcal{L}(R^n, R^m)$ esiste una matrice ortogonale $U \in \mathcal{L}(R^m, R^m)$, una matrice ortogonale $V \in \mathcal{L}(R^n, R^n)$ ed una matrice diagonale $\Sigma \in \mathcal{L}(R^n, R^m)$ con i valori singolari $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_n$ sulla diagonale, tale che $A = U\Sigma V^T$:

$$A = U \begin{Vmatrix} \sigma_1 & & & & \\ & \sigma_2 & & 0 & \\ & & .. & & \\ & & & .. & \\ & & & & \sigma_n \\ & 0 & & & \end{Vmatrix} V^T$$

Tale decomposizione di A, detta **decomposizione in valori singolari** o più brevemente **SVD** (Singular Value Decomposition), trova molte applicazioni tra le quali il calcolo della soluzione di norma minima per il problema dei minimi quadrati.

Poichè U è ortogonale, si ha:

$$\|Ax-b\|_2^2 = \|U\Sigma V^T x-b\|_2^2 = \|U^T(U\Sigma V^T x-b)\|_2^2 = \|\Sigma V^T x-U^T b\|_2^2.$$

Con il cambio di variabile $V^T x = z$ e $U^T b = d$, si ha infine:

$$\|Ax-b\|_2^2 = \|\Sigma z-d\|_2^2$$

dove

$$\|\Sigma z-d\|_2^2 = (\sigma_1 z_1-d_1)^2 + \ldots + (\sigma_n z_n-d_n)^2 + d_{n+1}^2 + \ldots + d_m^2.$$

E' evidente che la soluzione z che minimizza $||\Sigma z-d||_2^2$ è data, nel caso che i valori singolari siano tutti non nulli, da $z_1=\dfrac{d_1}{\sigma_1}$,..., $z_n=\dfrac{d_n}{\sigma_n}$; in tal caso il residuo è d_{n+1}^2 $+...+d_m^2$.. Se viceversa qualche valore singolare è nullo , diciamo σ_i =0, allora z_i è arbitrario ed il residuo è $d_i^2+d_{n+1}^2+...+d_m^2$ per ogni valore di z_i. Le altre componenti di z sono indipendenti dalla scelta di z_i e quindi il vettore z di norma minima si ottiene per z_i = 0. Per ogni z, il corrispondente vettore x=Vz conserva la norma di z e quindi x è la soluzione di norma minima per il problema originale Ax=b.

Si osservi che l'utilizzo della SVD per il calcolo della soluzione ai minimi quadrati richiede la risoluzione di un sistema lineare la cui matrice è la parte triangolare di Σ ed il cui indice di condizionamento, σ_1/σ_n , è uguale a quello della matrice R'. Se σ_n è molto piccolo l'indice di condizionamento può risultare ancora troppo alto e fornire una soluzione inaccettabile. Può risultare più stabile porre σ_n=0 e calcolare la soluzione di norma minima. Infatti ciò comporta la risoluzione di un sistema di dimensione n-1 ottenuto dal precedente togliendo l'ultima riga e l'ultima colonna. Il suo indice di condizionamento risulta allora σ_1/σ_{n-1}. Il miglioramento dell'indice di condizionamento può compensare la perturbazione introdotta dalla soppressione dell'ultimo valore singolare.

Esempio:

Consideriamo i dati esposti all'inizio del capitolo sulla popolazione degli Stati Uniti negli anni dal 1900 al 1970 ed interpoliamoli con un polinomio di secondo grado in forma canonica:

$$p(t)=c_1+c_2t+c_3t^2.$$

La matrice A del sistema sopradimensionato (8×3) ha i seguenti valori singolari:

$$\sigma_1 = 0.106\ 10^8 \qquad \sigma_2 = 0.648\ 10^2 \qquad \sigma_3 = 0.346\ 10^{-3}$$

e quindi la matrice A^TA del sistema di equazioni normali ha un indice di condizionamento molto elevato :

$$K(A^TA)=0.1\ 10^{22}.$$

Escludiamo dunque l'uso del sistema normale e applichiamo la tecnica della SVD per il quale l'indice di condizionamento è $\sigma_1/\sigma_n=0.306\ 10^{11}$.

Essa fornisce i seguenti valori dei coefficienti della parabola:

in semplice precisione (8 cifre significative):

$$c_1 = -\,0.372\ 10^5 \quad c_2 = 0.368\ 10^2 \quad c_3 = -\,0.905\ 10^{-2}$$

con i quali si ha il valore estrapolato: p(1980)=145.21 milioni,
ed in doppia precisione (16 cifre significative):

$$c_1 = 0.375\ 10^5 \quad c_2 = -\,0.402\ 10^2 \quad c_3 = 0.108\ 10^{-1}$$

per i quali si ottiene p(1980)= 227.78 milioni.

I due valori differiscono tra loro in modo consistente e quindi i risultati ottenuti in semplice precisione sono inaccettabili (vedremo che quello ottenuto in doppia precisione è corretto). L'indice di condizionamento σ_1/σ_n è ancora troppo grande per la semplice precisione. Proviamo allora a porre $\sigma_3=0$ e calcoliamo la soluzione di norma minima. Si ottiene:

in semplice precisione:

$$c_1 = -\,0.166\ 10^{-2} \quad c_2 = -\,0.162\ 10^1 \quad c_3 = -\,0.869\ 10^{-3}$$

con i quali si ha il valore estrapolato: p(1980)=214.96 milioni,

ed in doppia precisione:

$$c_1 = -\,0.167\ 10^{-2} \quad c_2 = -\,0.162\ 10^1 \quad c_3 = -\,0.871\ 10^{-3}$$

con i quali si ha il valore estrapolato: p(1980)=212.91 milioni.

Anche la semplice precisione fornisce questa volta un valore "accettabile", di gran lunga migliore di quello ottenuto tenendo conto di tutti i valori singolari.

A conclusione di questo esempio, si valutino i valori singolari ed i corrispondenti indici di condizionamento dei sistemi risultanti dall'uso delle seguenti rappresentazioni alternative del polinomio approssimante:

$$p(t)=c_1+c_2(t-1900)+c_3(t-1900)^2.$$

$$p(t)=c_1+c_2\left(\frac{t-1935}{10}\right)+c_3\left(\frac{t-1935}{10}\right)^2.$$

In entrambi i casi si troverà la stima p(1980)= 227.78 milioni che risulta quindi accettabile.

Un'altra applicazione interessante della SVD è la compressione dei dati. Per n=m=500, la matrice A è costituita da 250.000 coefficienti. Supponiamo che essi abbiano valori compresi tra 0 ed 1 e rappresentino, in modo crescente, le varie tonalità di grigio comprese tra il bianco (=0) ed il nero (=1) in un fotogramma quadrato costituito, appunto, da 250.000 punti. Supponiamo che un satellite esegua una sequenza, a distanza molto ravvicinata, di tali fotogrammi e che li debba inviare a terra. A volte è comodo, a costo di una perdita di precisione dell'immagine, poter ridurre la massa di dati da trasmettere. Ciò può essere fatto attraverso la SVD nel seguente modo.

Si osservi che indicando con u_i e con v_i le colonne di U e V rispettivamente, si ha:

$$A=U\Sigma V^T=\sum_{i=1}^{n} \sigma_i u_i v_i^T .$$

cioè la matrice A è data dalla somma di n matrici di rango 1. Se $\sigma_i \approx 0$ per i>r, allora si può troncare la precedente somma ed approssimare A con

$$A \approx \sum_{i=1}^{r} \sigma_i u_i v_i^T.$$

Nel nostro esempio, se potessimo accontentarci dei primi 20 termini, sarebbe sufficiente inviare a terra 20 colonne di U e di V e 20 valori singolari: in tutto 20020 dati. (vedi esempio numerico)