

ALMA MATER STUDIORUM – UNIVERSITA' DI BOLOGNA

SCUOLA DI ECONOMIA, MANAGEMENT E STATISTICA

Corso di Laurea in Scienze Statistiche: curriculum Economia e Impresa

LA SINGULAR VALUE DECOMPOSITION E SUE GENERALIZZAZIONI

(Algebra Lineare)

Presentata da:

Raffaele Anselmo

0000764543

Relatore:

Prof.ssa Laura Guidotti

APPELLO PRIMO

ANNO ACCADEMICO 2017/2018

Indice

1	Introduzione	3
1.1	Cenni storici	3
1.2	La SVD di $A \in \mathbb{R}^{m \times n}$	5
2	GSVD, le generalizzazioni della SVD	7
2.1	GSVD - versione con pesi	7
2.2	GSVD simultanea di due matrici	9
2.2.1	Autovalori e autovettori generalizzati	9
2.2.2	Diagonalizzazione simultanea di due forme quadratiche	10
2.2.3	GSVD	13
2.2.4	Un'applicazione: La GSVD nella genomica comparati- va come strumento previsivo della speranza di vita di un paziente	16
	Bibliografia	17

Capitolo 1

Introduzione

Il crescente ammontare di dati raccolti ed utilizzati nei piú svariati campi, dalla psicologia alla finanza, ha reso necessaria l'implementazione di misure di sintesi, in grado di rendere i dati piú facilmente interpretabili e di mostrarne le strutture nascoste dal volume e dalla varietà di questi dati.

Nell'epoca dei Big Data un problema di grande attualità é quello relativo alla scomposizione (di solito fattorizzazione) di una matrice in matrici piú semplici e con proprietà che le rendono di facile interpretazione e piú efficienti dal punto di vista computazionale. La scomposizione in valori singolari (SVD) di una matrice A ad elementi reali (o complessi) di dimensione qualsiasi può considerarsi come punto di arrivo dell'algebra lineare numerica classica e come base di partenza per i tanti, ulteriori sviluppi richiesti dal mondo dei Big Data. Come per tutti i risultati fondamentali della matematica, anche per la SVD sono state proposte e si stanno ancora studiando delle generalizzazioni, da diversi punti di vista e con differenti campi di applicabilità. Argomento di questa relazione é la SVD di una matrice, le sue conseguenze algebriche e computazionali ed alcune sue generalizzazioni. Dopo aver illustrato il teorema sull'esistenza della SVD di $A \in \mathbb{R}^{m \times n}$, averne indicato le radici storiche, le proprietà e le applicazioni nella statistica multivariata, si espongono due versioni della SVD generalizzata (GSVD) ed una loro applicazione in ambito biostatistico.

1.1 Cenni storici

Sia $A \in \mathbb{R}^{m \times n}$, allora esistono tre matrici $U \in \mathbb{R}^{m \times m}$ ortogonale, $V \in \mathbb{R}^{n \times n}$ e Σ (pseudo)diagonale, tali che:

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

Questo risultato di teoria delle matrici affonda le sue radici in ricerche ottocentesche di geometria, quando ancora il concetto di matrice come oggetto algebrico era estraneo alla pratica matematica.

La Singular Value Decomposition é una delle scomposizioni matriciali piú usate e gode di solide fondamenta teoriche. Tale scomposizione é frutto del lavoro di cinque matematici: Eugenio Beltrami (1835-1899), Camille Jordan (1838-1921), James Joseph Sylvester (1814-1897), Erhard Schmidt (1876-1959) ed Hermann Weyl (1885-1955), che hanno lavorato decenni per fornirci le regole di un metodo esplorativo dell'analisi numerica in continua evoluzione.

Eugenio Beltrami (1835-1900) in un articolo apparso nel giornale di matematiche ad uso degli Studenti delle Università nel 1873, *Sulle funzioni bilineari*, pose (e risolse) il problema di determinare delle trasformazioni lineari che riducevano a forma canonica la funzione bilineare

$$f(x_1, \dots, x_n, y_1, \dots, y_n) = \sum_{i,j=1}^n c'_{ij} x_i y'_j$$

Quasi contemporaneamente al matematico italiano, Camille Jordan (1838-1922) affrontó e risolse il medesimo problema pubblicandolo nell'articolo *Mémoire sur les formes bilinéaire*. Entrambi possono essere considerati i progenitori della SVD poiché l'italiano fu il primo autore di una pubblicazione che trattasse tali temi, mentre al francese si deve una redazione piú completa ed elegante.

Successivamente, in una nota a margine del giornale "The Messenger of Mathematics" intitolata "A new proof that a general quadratic may be reduced to its canonical form by means of a real orthogonal substitution" (1889), James Joseph Sylvester raggiunse gli stessi risultati ottenuti precedentemente da Beltrami e Jordan, descrivendo un algoritmo iterativo per ridurre le forme bilineari in forme diagonali. Anch'esso lavoró in maniera indipendente dai predecessori. L'autonomia dei suoi lavori é confermata da una nota che lo stesso autore inviό a "Comptes Rendus", dove millantava l'importanza della propria scoperta, ignaro del fatto che Jordan avesse pubblicato i propri risultati una decade e mezzo prima, presso lo stesso giornale.

Erhard Schmidt fu il primo ad approcciarsi alla SVD fuori dal campo dell'algebra lineare. Nel suo studio sulle equazioni integrali con nucleo asimmetrico, introdusse l'analoga a dimensione infinita della SVD. Egli però non si limitó alla mera dimostrazione dell'esistenza della scomposizione, mostrando come questa possa essere usata per trovare approssimazioni ottimali di rango inferiore. Facendo ciò, Schmidt trasformó la SVD da una curiosità matematica ad un importante strumento teorico e computazionale.

Il termine "valori singolari" sembra derivi dalla letteratura sulle equazioni integrali. In una pubblicazione del 1910, Émile Picard riferendosi agli autovalori delle equazioni con nucleo simmetrico di Schmidt, li chiamó per la prima volta valori singolari.

Il contributo di Hermann Weyl alla teoria della SVD consiste nello sviluppo di una teoria sulle perturbazioni di una matrice, fornendo una prova elegante del teorema dell'approssimazione di Schmidt (1912).

Ulteriori estensioni della scomposizione in valori singolari sono dovute a Léon Autonne che, tramite lo studio delle matrici per decomposizione polare, estese la SVD alle matrici complesse (1913), ed a Carl Eckart e Gale Young, che nel 1936 estero il concetto alle matrici rettangolari e resero popolare la SVD presso gli statistici utilizzando tale scomposizione per l'approssimazione ai minimi quadrati di matrici di dati (soprattutto in ambito psicometrico) con matrici di rango inferiore.

1.2 La SVD di $A \in \mathbb{R}^{m \times n}$

SVD 1 Sia $A \in \mathbb{R}^{m \times n}$, $\text{rango} A = r$, allora esistono tre matrici

$U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m] \in \mathbb{R}^{m \times m}$ ortogonale,

$V = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n] \in \mathbb{R}^{n \times n}$ ortogonale,

Σ pseudo diagonale, $\sigma_{ij} = 0, i \neq j, \sigma_{ii} \geq 0$,

tali che

$$A = U \Sigma V^T \quad (1.1)$$

Le dimostrazioni di questo teorema sono molteplici; si può subito osservare che la SVD costituisce la generalizzazione ad una matrice di dimensione qualsiasi della scomposizione spettrale di una matrice simmetrica; se A é simmetrica e semidefinita positiva la scomposizione spettrale di A , $A = Q \Lambda Q^T$, é anche la SVD di A .

Nella costruzione della SVD intervengono le matrici $A^T A$ e $A A^T$ ¹ simmetriche e semidefinite positive, che hanno in comune gli autovalori non nulli $\lambda_i, i = 1, \dots, r$. Posto $\sqrt{\lambda_i} = \sigma_i > 0$, gli elementi delle matrici A, U, V, Σ sono legati dalle relazioni:

$$\begin{cases} A \mathbf{v}_i = \sigma_i \mathbf{u}_i & , i = 1, \dots, r \\ A^T \mathbf{u}_i = \sigma_i \mathbf{v}_i & , i = 1, \dots, r \\ A \mathbf{v}_i = \mathbf{0} & , i = r + 1, \dots, n \\ A^T \mathbf{u}_i = \mathbf{0} & , i = r + 1, \dots, m \end{cases}$$

¹ $A^T A$ e $A A^T$ sono le matrici di Gram associate alla matrice A

I numeri positivi σ_i sono i **valori singolari** di A , \mathbf{v}_i i vettori singolari a destra di A e \mathbf{u}_i i vettori singolari a sinistra di A .

Una presentazione equivalente della SVD di A é

$$A = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

che presenta A come combinazione lineare di matrici di rango 1.

La SVD di A fornisce un algoritmo stabile per stabilire il rango di una matrice (la SVD di A é *rank-revealing*) e un'espressione compatta della soluzione ai minimi quadrati del sistema lineare $A\mathbf{x} = \mathbf{b}$ poiché permette di formalizzare la nozione di inversa generalizzata di Moore-Penrose

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

dove $\Sigma^\dagger \in \mathbb{R}^{n \times m}$ con elementi diagonali $\frac{1}{\sigma_i}$. Allora la **soluzione ai minimi quadrati** é

$$\mathbf{x}_{LS} = A^\dagger \mathbf{b}$$

Per quanto riguarda gli ambiti di applicazione della SVD a problemi del mondo reale si può condividere l'affermazione

"The SVD constitutes one of science's superheroes in the fight against monstrous data, and it arise in seemingly every scientific discipline"²

Quest'affermazione risulta tutt'altro che esagerata. Oltre ad essere la base di molti metodi di analisi multivariata, come l'analisi delle componenti principali, l'analisi dei fattori e la cluster analysis, la scomposizione in valori singolari sta diventando parte integrante della vita di tutti i giorni.

La SVD rappresenta infatti uno strumento di base nelle più svariate applicazioni[1], fra le quali:

- indagare sulle ideologie dei Membri del Congresso americano;
- studiare il grado di cristallizzazione delle rocce ignee;
- analizzare le correlazioni nei mercati finanziari;
- sbloccare il cellulare con un'applicazione capace di riconoscere il proprio volto.

²[1] C.D.Martin, M.A.Manson

Capitolo 2

GSVD, le generalizzazioni della SVD

Vista l'importanza che la SVD di una matrice ha assunto nell'algebra lineare numerica é comprensibile che si siano moltiplicati gli studi su possibili estensioni e generalizzazioni di tale fattorizzazione.

In questa sede se ne considerano due, la GSVD di una matrice con la presenza di pesi e la GSVD simultanea di due matrici.

2.1 GSVD - versione con pesi

Questa tecnica consiste nella scomposizione in valori singolari di una matrice, imponendo dei vincoli sui vettori singolari.

GSVD - Weighted Version 1 *Siano $\Omega \in \mathbb{R}^{i \times i}$ e $\Phi \in \mathbb{R}^{j \times j}$ matrici definite positive e simmetriche, allora qualunque matrice reale $A \in \mathbb{R}^{i \times j}$ di rango k può essere espressa come segue:*

$$A = N\tilde{\Sigma}M^T \quad (2.1)$$

dove le colonne di N e M sono ortonormali rispetto a Ω e Φ , e contengono i vincoli sulle colonne e sulle righe di A rispettivamente:

$$N^T\Omega N = M^T\Phi M = I$$

La GSVD é facilmente dimostrata¹ partendo dalla SVD di \tilde{A} :

$$\tilde{A} = \Phi^{\frac{1}{2}}A\Omega^{\frac{1}{2}}$$

¹[2]Abdi H.

$$\begin{aligned}
\tilde{A} &= U\tilde{\Sigma}V^T & (U^TU = V^TV = I) \\
\Phi^{\frac{1}{2}}A\Omega^{\frac{1}{2}} &= U\tilde{\Sigma}V^T \\
A &= \Phi^{-\frac{1}{2}}U\tilde{\Sigma}\Omega^{-\frac{1}{2}} \\
A &= N\tilde{\Sigma}M^T & (N^T\Omega N = M^T\Phi M = I)
\end{aligned}$$

Le colonne di N e M possono essere chiamate vettori singolari generalizzati rispettivamente destro e sinistro. Sono ancora basi ortonormali per le righe e le colonne di A , ma la metrica utilizzata negli spazi i - e j -dimensionali non é piú quella Euclidea semplice, ma una metrica Euclidea generalizzata (pesata) definita rispettivamente da Ω e Φ .

Gli elementi diagonali non nulli della matrice $\tilde{\Sigma}$ sono chiamati valori singolari generalizzati.

É possibile adattare l'approssimazione low-rank di una matrice proposta inizialmente da Schmidt e formalizzata successivamente da Eckart e Young al contesto generalizzato, ovvero basato su uno spazio Euclideo generalizzato (o pesato).

Approssimazione low-rank 1 Sia la GSVD di A :

$$A = N\tilde{\Sigma}M^T = \sum_{i=1}^k \sigma_i n_i m_i^T$$

Se si tronca lo sviluppo della sommatoria negli ultimi $k - k^*$ termini, allora $A_{k^*} = \sum_{i=1}^{k^*} \sigma_i n_i m_i^T = N_{k^*} \tilde{\Sigma}_{k^*} M_{k^*}^T$ é l'approssimazione low-rank ai minimi quadrati generalizzata di A che minimizza la funzione:

$$\min_{\text{rank}(B) \leq k^*} \|A - B\|_2$$

fra tutte le matrici B di rango inferiore o uguale a k^* .

La funzione può anche essere scritta come segue:

$$\min_{\text{rank}(B) \leq k^*} \text{traccia}\{\Omega(A - B)\Phi(A - B)^T\}$$

Se la matrice Ω é una matrice diagonale D_w con elementi positivi (w_1, \dots, w_k) :

$$\min_{\text{rank}(B) \leq k^*} \text{traccia}\{D_w(A - B)\Phi(A - B)^T\} = \sum_{i=1}^k w_i (a_i - b_i)^T \Phi(a_i - x_i)$$

dove a_i e b_i sono rispettivamente le righe di A e B , scritte come vettori colonna. Questa funzione definisce un'analisi delle componenti principali (PCA). Le righe di A possono essere considerate come una nuvola di I -punti in uno spazio J -dimensionale, dove la metrica Euclidea è definita dalla matrice Φ (nel caso Φ sia diagonale si parla di "metrica diagonale". I valori (w_i, \dots, w_k) sono i "pesi" (da cui il nome della tecnica) assegnati ad ognuno dei punti riga, mentre le righe di B sono punti non conosciuti in un sottospazio k^* -dimensionale, e il minimo della funzione, raggiunto da A_{k^*} identifica il sottospazio più vicino alla nuvola di punti, in termini di una sommatoria pesata dei quadrati delle distanze. In questo caso i vettori (m_1, \dots, m_{k^*}) definiscono gli assi principali (ortonormali) del sottospazio, mentre le righe della matrice $N_{k^*} \tilde{\Sigma}_{k^*} = [\alpha_1 n_1, \dots, \alpha_{k^*} n_{k^*}]$ definiscono le coordinate delle proiezioni della nuvola dei punti nel sottospazio (rispetto agli assi definiti dai vettori di M).

In questo contesto l'ortonormalità degli assi e delle proiezioni è sempre definita nei termini della metrica Φ . Affinché la scomposizione sia unica è necessario che α_{k^*} sia strettamente maggiore di α_{k^*+1} , ciò significa che la dimensione k^* dell'approssimazione vada scelta dove è presente una differenza ben definita fra α_{k^*} e α_{k^*+1} .

2.2 GSVD simultanea di due matrici

La semplificazione simultanea di due o più matrici rispetto ad un determinato criterio è un argomento ormai classico dell'algebra lineare e dell'analisi numerica.

Due matrici quadrate diagonalizzabili sono riducibili a forma diagonale rispetto ad una stessa trasformazione per similitudine se e solo se le due matrici commutano fra loro:

$$P^{-1}AP = \Lambda_1, \quad P^{-1}BP = \Lambda_2 \Leftrightarrow AB = BA$$

2.2.1 Autovalori e autovettori generalizzati

Siano A e B due matrici quadrate $n \times n$ in generale ad elementi complessi; l'insieme

$$\lambda(A, B) = \{A - \lambda B; \lambda \in \mathbb{C}\}$$

è chiamato *fascio* di matrici. Gli elementi dell'insieme

$$\lambda(A, B) = \{z, z \in \mathbb{C}; \det(A - zB) = 0\}$$

sono gli *autovalori* del fascio $\lambda(A, B)$; se $\lambda \in \lambda(A, B)$ le soluzioni non banali del sistema lineare

$$(A - \lambda B)\mathbf{x} = \mathbf{0},$$

cioé

$$A\mathbf{x} = \lambda B\mathbf{x}; \quad \mathbf{x} \neq \mathbf{0}$$

sono chiamati *autovalori* del fascio di matrici. La ricerca degli autovalori e degli autovettori del fascio é indicato come *problema generalizzato degli autovalori e autovettori*.

Questo problema ha un' elegante soluzione nel caso particolare di due matrici hermitiane (reali simmetriche) di cui almeno una definita positiva.

2.2.2 Diagonalizzazione simultanea di due forme quadratiche

Teorema

Siano A, B due matrici simmetriche, B definita positiva e $\mathbf{x}^T A \mathbf{x}$ e $\mathbf{x}^T B \mathbf{x}$ le relative forme quadratiche.

Allora:

1. Esiste una matrice non singolare R tale che

$$R^T A R = \Lambda, \quad R^T B R = I,$$

2. il cambiamento di base in \mathbb{R}^n : $\mathbf{x} = R\mathbf{y}$ opera sulle forme quadratiche trasformandole cosí:

$$\mathbf{x}^T A \mathbf{x} = \mathbf{y}^T \Lambda \mathbf{y} = \lambda_1 |y_1|^2 + \lambda_2 |y_2|^2 + \cdots + \lambda_n |y_n|^2,$$

$$\mathbf{x}^T B \mathbf{x} = \mathbf{y}^T I \mathbf{y} = |y_1|^2 + |y_2|^2 + \cdots + |y_n|^2.$$

Dimostrazione

La matrice B é definita positiva, esiste quindi una matrice P unitaria tale che $P^T B P = K$ dove $K = D^2$ e $D = \text{diag}(k_1, k_2, \dots, k_n)$ con $k_i > 0$, $i = 1, \dots, n$. Da $P^T B P = D^2$ segue

$$D^{-1} P^T B P D^{-1} = (P D^{-1})^T B P D^{-1} = Q^T B Q = I,$$

dove $Q = P D^{-1}$.

La matrice B é stata resa congruente alla matrice unitá, come era prevedibile, mediante la congruenza indotta da Q .

La matrice $Q^T A Q$ é ancora simmetrica e pertanto esiste una matrice S non singolare tale che

$$S^T(Q^T A Q)S = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = (QS)^T A(QS) = R^T A R.$$

La matrice non singolare $R = QS$ diagonalizza rispetto alla relazione di congruenza contemporaneamente le matrici A e B ; realizza quindi la diagonalizzazione simultanea (per congruenza) di A e di B .

Eseguendo in ciascuna delle due forme quadratiche la sostituzione (cambiamento di coordinate) $\mathbf{x} = R\mathbf{y}$, si ottengono le equazioni sopra indicate in cui mancano i termini rettangolari $y_i y_j$ con $i \neq j$.

Gli elementi diagonali λ_i della matrice $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ e le colonne \mathbf{x}_i della matrice

$$R = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \dots \quad \mathbf{x}_n]$$

hanno un preciso significato nell' ambito del problema degli autosistemi generalizzati.

Infatti

$$\begin{aligned} (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda) &= \det(\Lambda - \lambda I) = \\ &= \det(R^T A R - \lambda R^T B R) = \det R^T \det(A - \lambda B) \det R. \end{aligned}$$

Visto che R é non singolare si conclude che

$$\det(A - \lambda B) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda),$$

gli elementi diagonali di Λ sono quindi gli autovalori generalizzati del fascio $\lambda(A, B)$.

Se λ_i é uno di questi autovalori generalizzati, il sistema lineare

$$(A - \lambda_i B)\mathbf{x} = \mathbf{0}$$

ha soluzioni non banali; sia \mathbf{x}_i una di queste soluzioni:

$$(A - \lambda_i B)\mathbf{x}_i = \mathbf{0} \iff A\mathbf{x}_i = \lambda_i B\mathbf{x}_i;$$

gli elementi della matrice R che realizza la simultanea diagonalizzazione di A e di B sono pertanto gli autovettori generalizzati del fascio $\lambda(A, B)$.

Le colonne della matrice R formano un insieme di n vettori indipendenti $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$; dall' uguaglianza $R^T B R = I$ segue:

$$\mathbf{x}_i^T B \mathbf{x}_i = 1, \text{ (per)} i = 1, \dots, n; \quad \mathbf{x}_i^T B \mathbf{x}_j = 0, \text{ (per)} i \neq j.$$

L'insieme $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ é un insieme *ortonormale* nello spazio \mathbb{R}^n unitario rispetto al prodotto scalare definito dalla matrice B :

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T B \mathbf{v}.$$

É banale osservare che se $B = I$ si ritrova la definizione di insieme ortonormale rispetto al prodotto scalare canonico.

Le precedenti considerazioni dimostrano il seguente

Teorema

Siano A, B due matrici simmetriche con B definita positiva e siano λ_i , $i = 1, \dots, n$ le soluzioni dell'equazione polinomiale $\det(A - \lambda B)$. Allora

1. esiste un insieme di n vettori indipendenti \mathbf{x}_i tali che

$$A\mathbf{x}_i = \lambda_i B\mathbf{x}_i; \quad i = 1, \dots, n.$$

I vettori $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ formano una base *B-ortonormale* di \mathbb{R}^n , cioè una base ortonormale rispetto al prodotto scalare definito da B .

2. Se $R = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_n]$ si ha:

$$R^T A R = \Lambda; \quad R^T B R = I.$$

In definitiva si é mostrato che, nelle ipotesi di due matrici hermitiane di cui una definita positiva, la diagonalizzazione simultanea equivale alla ricerca degli autovalori e degli autovettori generalizzati del fascio individuato dalle due matrici.

Il problema della diagonalizzazione simultanea di matrici simmetriche) é stato generalizzato in diverse direzioni; fra queste particolare interesse riveste il problema di trovare simultaneamente una scomposizione generalizzata in valori singolari di due o piú matrici mediante una stessa matrice non singolare X .

Prese $A \in \mathbb{R}^{m \times n}$ e $B \in \mathbb{R}^{p \times n}$ matrici con lo stesso numero di colonne, come si sa, i valori singolari di A sono gli autovalori non nulli della matrice simmetrica (semidefinita positiva) $A^T A$ e valori singolari di B sono gli autovalori non nulli della matrice simmetrica (semidefinita positiva) $B^T B$; se poi $\text{rango}(B)=n$, la matrice $B^T B$ é definita positiva.

In questo caso la contemporanea trasformazione di A e di B in matrice pseudodiagonali si puó ricondurre alla diagonalizzazione simultanea di $A^T A$ e $B^T B$ considerando il fascio $A^T A - \lambda B^T B$; i teoremi precedenti permettono di concludere che esiste $X \in \mathbb{R}^{n \times n}$ non singolare per cui $X^T(A^T A)X$ e

$X^T(B^T B)X$ sono entrambe diagonali.

Questo risultato, immediato dal punto di vista teorico, non é di grande utilit  dal punto di vista computazionale perch  la costruzione di $A^T A$ e $B^T B$ pu  rendere i calcoli molto pi  pesanti.

2.2.3 GSVD

Charles F. Van Loan introdusse nella sua tesi di dottorato nel 1976 una scomposizione simultanea di due matrici, chiamata B-SVD, in riferimento alla SVD di una matrice A , alla quale   associata una data matrice B con lo stesso numero di colonne

Van Loan mostr  come i valori $\sigma(A, B)$ possono essere ottenuti scomponendo A e B in tre matrici: una ortogonale, una pseudo-diagonale ed una non singolare.

Nel 1981 Paige e Saunders stilarono una formulazione pi  generale della B-SVD, a cui diedero il nome di Generalized Singular Value Decomposition.

GSVD - Forma triangolare 1 *Siano $A \in \mathbb{R}^{m \times n}$ e $B \in \mathbb{R}^{p \times n}$ con $\text{rango}(A^T, B^T) = k$, allora esisteranno due matrici ortogonali $U \in \mathbb{R}^{m \times m}$ e $V \in \mathbb{R}^{p \times p}$ e una matrice $Q \in \mathbb{R}^{n \times n}$ tali che:*

$$\begin{pmatrix} U^T & 0 \\ 0 & V^T \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} Q = \begin{pmatrix} \Sigma_1 \\ \Sigma_2 \end{pmatrix} \begin{pmatrix} R & 0 \end{pmatrix} \quad (2.2)$$

dove $R \in \mathbb{R}^{k \times k}$   una matrice non singolare triangolare superiore e le matrici $\Sigma_1 \in \mathbb{R}^{m \times k}$ e $\Sigma_2 \in \mathbb{R}^{p \times k}$ sono pseudo-diagonali

Dim. Operiamo la fattorizzazione ortogonale di $(A^T, B^T)^T$:

$$\begin{pmatrix} A \\ B \end{pmatrix} = U_1 \begin{pmatrix} R_1 & 0 \\ 0 & 0 \end{pmatrix} V_1^T$$

Dove U_1 e V_1 sono matrici ortogonali di dimensione rispettivamente $(m + p) \times (m + p)$ e $n \times n$ e R_1   non singolare triangolare superiore di ordine k . Usando la partizione di U_1 nella forma:

$$U_1 = \begin{pmatrix} \begin{matrix} k & n-k \\ U_{11} & U_{12} \\ U_{21} & u_{22} \end{matrix} \end{pmatrix} \begin{matrix} m \\ p \end{matrix}$$

La dimostrazione del teorema segue alla CSD (1) della matrice ortonormale $(U_{11}^T, U_{21}^T)^T$ riportata nel Teorema 1.

Possiamo riscrivere la scomposizione in funzione di A o B

$$\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \begin{pmatrix} \Sigma_1 \\ \Sigma_2 \end{pmatrix} \begin{pmatrix} R & 0 \end{pmatrix} Q^T$$

$$\begin{aligned} A &= U \Sigma_1(R_1, 0) Q^T \\ B &= U \Sigma_2(R_2, 0) Q^T \end{aligned} \quad (2.3)$$

Il vantaggio della fattorizzazione triangolare é che richiede solo trasformazioni ortogonali, rendendo i calcoli piú stabili. Se siamo in grado di effettuare trasformazioni non singolari, possiamo ricondurre la coppia di matrici (A, B) ad una forma diagonale, come spiegato dal seguente Teorema

GSVD - Forma Diagonale 1 *Siano $A \in \mathbb{R}^{m \times n}$ e $B \in \mathbb{R}^{p \times n}$ con $\text{rango}(A^T, B^T) = k$, allora esisteranno due matrici ortogonali $U \in \mathbb{R}^{m \times m}$ e $V \in \mathbb{R}^{p \times p}$ e una matrice ortogonale $X \in \mathbb{R}^{n \times n}$ tali che:*

$$\begin{pmatrix} U^T & 0 \\ 0 & V^T \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} X = \begin{pmatrix} \Sigma_1 \\ \Sigma_2 \end{pmatrix} \begin{pmatrix} I_k & 0 \end{pmatrix} \quad (2.4)$$

dove $\Sigma_1 \in \mathbb{R}^{m \times k}$ e $\Sigma_2 \in \mathbb{R}^{p \times k}$ sono pseudo-diagonali:

$$\Sigma_1 = \begin{matrix} & l & z & k-l-z \\ \begin{matrix} l \\ z \\ m-l-z \end{matrix} & \begin{pmatrix} I_1 & & \\ & D_1 & \\ & & O_1 \end{pmatrix} \end{matrix} \quad \Sigma_2 = \begin{matrix} & l & z & k-l-z \\ \begin{matrix} p-k+l \\ z \\ k-l-z \end{matrix} & \begin{pmatrix} O_2 & & \\ & D_2 & \\ & & I_2 \end{pmatrix} \end{matrix}$$

dove $I_1 \in \mathbb{R}^{l \times l}$ e $I_2 \in \mathbb{R}^{(k-l-z) \times (k-l-z)}$ sono matrici identità, $O_1 \in \mathbb{R}^{(m-l-z) \times (k-l-z)}$ e $O_2 \in \mathbb{R}^{(p-k+l) \times l}$ sono matrici nulle,

$$D_1 = \text{diag}(\alpha_{l+1}, \dots, \alpha_{l+z}), \quad D_2 = \text{diag}(\beta_{l+1}, \dots, \beta_{l+z}),$$

$$1 > \alpha_{l+1} \geq \dots \geq \alpha_{l+z} > 0, \quad 0 < \beta_{l+1} \leq \dots \leq \beta_{l+z} < 1, \quad \alpha_i^2 + \beta_i^2 = 1$$

Le coppie (α_i, β_i) definite dagli elementi diagonali di Σ_1 e Σ_2 sono chiamate coppie di valori singolari generalizzati (GSV pairs). Il quoziente $\lambda_i = \alpha_i / \beta_i$ é chiamato valore singolare generalizzato (GSV). Si noti che i λ_i sono le radici quadrate degli autovalori della matrice pencil $A^T A - \lambda B^T B$.

Dim.² Sia $C = (A^T B^T)^T$ di rango k e sia la sua SVD:

$$C = \begin{pmatrix} A \\ B \end{pmatrix} = U \begin{pmatrix} \Sigma_C & 0 \\ 0 & 0 \end{pmatrix} V^H$$

dove U e V sono matrici ortogonali e Σ_C é una matrice diagonale contenente i valori singolari non nulli di C come elementi diagonali. Ora operiamo la partizione di U come:

$$U = \begin{pmatrix} U_{A1} & U_{A2} \\ U_{B1} & U_{B2} \end{pmatrix}$$

Sia la SVD di U_{A1} :

$$U_{A1} = U_A \Sigma_A Q^T$$

e scriviamo

$$U_{B1} = U_{B1} Q Q^T = U_B L Q^T$$

dove $U_{B1} Q = U_B L$ con U_B ortogonale e L triangolare inferiore. Allora:

$$\begin{pmatrix} U_{A1} \\ U_{B1} \end{pmatrix} = \begin{pmatrix} U_A & 0 \\ 0 & U_B \end{pmatrix} \begin{pmatrix} \Sigma_A \\ L \end{pmatrix} Q^T$$

La quale é a colonne ortonormali dato che U é ortogonale. Allora:

$$\begin{pmatrix} U_{A1}^T & U_{B1}^T \end{pmatrix} \begin{pmatrix} U_{A1} \\ U_{B1} \end{pmatrix} = I_k$$

Combinando le ultime due equazioni si ricava:

$$Q \begin{pmatrix} \Sigma_A^T & L^T \end{pmatrix} \begin{pmatrix} U_A^T & 0 \\ 0 & U_B^T \end{pmatrix} \begin{pmatrix} U_A & 0 \\ 0 & U_B \end{pmatrix} \begin{pmatrix} \Sigma_A \\ L \end{pmatrix} Q^T = I_k$$

Dalla quale otteniamo $\Sigma_A^T \Sigma_A + L^T L = I_k$; $L^T L = I_k - \Sigma_A^T \Sigma_A = \Sigma_B^T \Sigma_B$ e, allora, $L = \Sigma_B$.

Possiamo riscrivere la GSVD di A e B come segue:

$$\begin{aligned} A &= U \Sigma_1(I_1, 0) X^T \\ B &= U \Sigma_2(I_2, 0) X^T \end{aligned} \tag{2.5}$$

Sebbene la formulazione della GSVD nei Teoremi 2 e 3 sia matematicamente

²[3]Qing Huang

equivalente, la computazione numerica delle due scomposizioni é significativamente diversa, per la possibilitá di trovare matrici R (nel caso del Teorema 2) e X (nel caso del Teorema 3) "ill-conditioned". Un algoritmo che mostra una computazione numerica stabile per uno dei due Teoremi, non necessariamente si rivela stabile anche per l'altro.

Si può osservare che:

- Se B é quadrata e non singolare, la GSVD di A e B é uguale alla SVD di AB^{-1} :

$$U^T AB^{-1}V = \Sigma_1 \Sigma_2^{-1}$$

- Se $B = I$ la GSVD di A e B si riconduce alla SVD di A , da qui il nome.
- Se $(A^T, B^T)^T$ é a colonne ortonormali, la GSVD di A e B corrisponde alla CSDecomposition.

2.2.4 Un'applicazione: La GSVD nella genomica comparativa come strumento previsivo della speranza di vita di un paziente

Il Glioblastoma Multiforme (GBM) é il tumore al cervello piú comune fra gli adulti e le prognosi sono finora state basate solo sull'età dei pazienti al momento della diagnosi. Può risultare utile usare la GSVD³ per confrontare le sequenze di genomi e le loro alterazioni (Copy Number Variations) fra un modello molecolare sano e quello malato. Patterns matematicamente significativi in entrambi i datasets rappresentano CNVs del genoma umano normale conservate anche in quello malato (come ad esempio l'amplificazione del cromosoma X nelle donne). Patterns significativi nel dataset normale ma non in quello del tumore rappresentano variazioni sperimentali che riguardano solamente i soggetti sani, mentre Patterns matematicamente significativi nel solo dataset di genomi del soggetto malato rappresentano variazioni sperimentali esclusive dei genomi malati. Questo pattern é correlato con la sopravvivenza e la risposta alla terapia del paziente, é indipendente dall'età e, combinato con essa, fornisce un previsore migliore rispetto alla sola età al momento della diagnosi.

Per formulare la GSVD, si estraggono diversi profili di espressione genica, sia sani (normal dataset) che malati (tumor dataset), dallo stesso numero di

³[4]Cheng H. Lee, Benjamin O. Alpert, Preethi Sankaranarayanan, Orly Alter. GSVD Comparison of Patient-Matched Normal and Tumor aCGH Probes Reveals Global Copy-Number Alterations Predicting Glioblastoma Multiforme Survival. PLOS ONE, January 2012.

pazienti (in modo da avere lo stesso numero di colonne nelle due matrici, condizione necessaria per la GSVD) e li si scompongono simultaneamente nelle matrici $U_{1,2}$, $\Sigma_{1,2}$ e X . I vettori x_n^T , comuni ad entrambe le scomposizioni, sono chiamati "probelet" ed indicano modelli di evoluzioni genomiche che si presentano fra i pazienti. I vettori $u_{1,n}^T$ sono le evoluzioni esclusive del dataset dei profili malati e sono chiamati "tumor arraylet", mentre con i vettori $u_{2,n}^T$ si intendono i "normal arraylet", ovvero i modelli di alterazioni esclusive dei profili sani.

I valori singolari possono essere intesi come dei pesi che misurano la significatività dei probelet x_n^T in termini di informazione totale catturata dal dataset. La significatività relativa dei probelet del tumor dataset relativa a quella del normal dataset è misurata in termini di distanza angolare θ_n , che è proporzionale al valore dei pesi (i valori singolari):

$$-\pi/4 \leq \theta_n = \arctan(\sigma_{1,n}/\sigma_{2,n}) - \pi/4 \leq \pi/4$$

Una distanza angolare di $\pm\pi/4$ indica un probelet esclusivo del tumor dataset o del normal dataset, mentre valori prossimi allo zero indicano probelet comuni ad entrambi i datasets.

Una volta effettuate queste analisi le si possono combinare con le altre informazioni disponibili per valutare lo stadio di avanzamento del tumore e la relativa speranza di vita. Dall'indagine di Lee, Alpert, Sankaranarayanan e Alter [4] si evince come ai pazienti che presentano bassi valori in corrispondenza del secondo probelet (il più significativo per la prognosi del tumore) viene conferita una speranza di vita più del doppio maggiore rispetto agli altri.

La GSVD non allungherà la vita alle persone affette da Glioblastoma Multiforme, ma riesce ad assolvere a due compiti molto importanti: fornire stime affidabili sulla speranza di vita e indicare l'avanzamento della malattia nel paziente, riuscendo così a supportare la scelta delle terapie più coerenti.

Bibliografia

- [1] C.D.Martin and M.A.Mason. The extraordinary SVD. *The American Mathematical Monthly*, 119(10), March 2011. <https://people.maths.ox.ac.uk./porterm/papers/s4.pdf>.
- [2] Abdi H. Singular Value Decomposition (SVD) and Generalized Singular Value Decomposition(GSVD). In: Salkind NJ, ed. *Encyclopedia of Measurement and Statistics*. Thousand Oaks: Sage Publications, 2007. <https://www.utdallas.edu/~herve/Abdi-SVD2007-pretty.pdf>.
- [3] Qing Huang. Some Topics Concerning the Singular Value Decomposition and Generalized Singular Value Decomposition. Technical report, A Dissertation Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy. ARIZONA STATE UNIVERSITY, August 2012.
- [4] Cheng H. Lee, Benjamin O. Alpert, Preethi Sankaranarayanan, and Orly Alter. GSVD Comparison of Patient-Matched Normal and Tumor aCGH Probes Reveals Global Copy-Number Alterations Predicting Glioblastoma Multiforme Survival. *PLOS ONE*, January 2012. <https://doi.org/10.1371/journal.pone.0030098>.
- [5] Orly Alter Patrick O. Brown and David Botstein. Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. *PNAS*, January 2003. www.pnas.org/cgi/doi/10.1073/pnas.0530258100.
- [6] G. W. Stewart. On the Early History of the Singular Value Decomposition. *SIAM Review*, 35(4):551–556, December 1993. <https://epubs.siam.org/doi/abs/10.1137/1035134>.
- [7] G.H Golub and Van Loan C.F. *Matrix Computations*. III edizione. The Johns Hopkins University Press, 1996.

- [8] Zhaojun Bai and James W. Demmel. Computing the Generalized Singular Value Decomposition. *SIAM J. Sci. Comput.*, 14(6):1464–1486, 1993. <https://people.maths.ox.ac.uk/porterm/papers/s4.pdf>.
- [9] Zhaojun Bai. The CSD, GSVD, their applications and computations. Technical report, IMA Preprint Series, April 1992. <https://conservancy.umn.edu/bitstream/handle/11299/1875/1/958.pdf>.
- [10] Shmuel Friedland. A New Approach to Generalized Singular Value Decomposition. Technical report, Department of Mathematics, Statistics and Computer Science. University of Illinois at Chicago, December 2004. <https://pdfs.semanticscholar.org/cfd2/f33c8e3c6c6867a855cf2dee0599f9e7bc49.pdf>.
- [11] Van Loan C.F. Generalizing the singular value decomposition. *SIAM J. Numer. Anal.*, 13(1), 1976.
- [12] Paige C.C. and Saunders M.A. Towards a generalized singular value decomposition. *SIAM J. Numer. Anal.*, 18(3), 1981.