



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA TRIENNALE IN
INGEGNERIA DELL'INFORMAZIONE

Decomposizione in valori singolari e sue applicazioni nel campo dell'Ingegneria

Relatore:

PROF. FRANCESCO BOTTACIN

Laureando:

FRANCESCO FORTUNATO

1217632

Anno Accademico 2021/2022

Data di laurea: 23/09/2022

Abstract

La decomposizione in valori singolari estende il concetto di diagonalizzazione di matrici quadrate alle matrici rettangolari.

Prima di spiegare nel dettaglio come questo sia possibile nel campo dei numeri reali e in quello dei numeri complessi, verranno richiamati concetti di Algebra Lineare come autovalori e autovettori. Inoltre, per la piena comprensione dell'argomento, daremo una panoramica delle proprietà che possiede lo spazio vettoriale complesso.

Successivamente verranno presentate applicazioni della decomposizione in valori singolari, come la matrice pseudoinversa e l'approssimazione di rango basso, utili in diversi settori ingegneristici, come quello industriale e biomedico.

Indice

Introduzione	1
1 Richiami di Algebra Lineare	5
1.1 Autovalori e autovettori	5
1.2 Matrici simmetriche	9
1.2.1 Introduzione e prime proprietà	9
1.2.2 Forme bilineari e matrici	9
1.2.3 Teorema spettrale	11
2 Decomposizione in valori singolari (SVD)	13
2.1 Decomposizione in valori singolari per matrici reali	13
2.2 Lo spazio vettoriale \mathbb{C}^n e le matrici complesse	18
2.2.1 Forme hermitiane e matrici associate	18
2.2.2 Cambiamenti di base	21
2.3 Decomposizione in valori singolari per matrici complesse	23
3 Applicazioni della SVD	27
3.1 La matrice pseudoinversa	27
3.2 Approssimazione di rango basso	32
3.3 Riconoscimento di movimenti e gesti umani tramite SVD	34

Ringraziamenti

Prima di cominciare la trattazione, vorrei ringraziare alcune persone che mi sono state vicine in questo percorso.

Un sentito e particolare grazie va al mio relatore, il professor Bottacin, per la sua immensa disponibilità e pazienza nei miei confronti e per avermi fornito tutto il materiale utile per scrivere questa tesi che senza di lui non esisterebbe; una persona veramente cordiale che, in questi mesi, mi ha dato preziosi consigli e conoscenze.

Ringrazio infinitamente i miei genitori, mia sorella e mia nonna, che mi hanno sempre sostenuto e hanno creduto in me, anche nei momenti più difficili.

Non riesco a nominare tutti, ma ringrazio i miei amici, colleghi e professori incontrati per aver reso l'esperienza meno pesante e meno noiosa.

Infine, vorrei ringraziare me stesso, per l'impegno e la dedizione che ho messo in questi tre anni per giungere a questo piccolo traguardo.

Grazie a tutti voi per aver condiviso con me sia gli attimi più belli che quelli un po' meno belli. :)

Introduzione

Quando si studia Algebra Lineare e ci si addentra nel vasto mondo delle matrici, uno dei problemi più importanti che si affronta è quello della diagonalizzazione degli endomorfismi, ovvero di funzioni lineari in cui dominio e codominio sono lo stesso spazio vettoriale V . Tipicamente, data una matrice quadrata A , la quale rappresenta una funzione lineare f in una base arbitraria di V (generalmente quella canonica), la domanda che ci si chiede è l'esistenza o meno di un'altra base di vettori di V tale per cui la matrice di f rispetto a tale base sia diagonale. Una matrice diagonale è infatti molto più semplice e comoda da usare, poiché le sue proprietà sono immediatamente visibili e manipolarla non richiede un lavoro computazionale elevato.

In questa tesi ci poniamo l'interessante quesito di “diagonalizzare” una matrice rettangolare, che in letteratura risponde al nome di **decomposizione in valori singolari (SVD)**. Introduciamo svariati concetti utili alla comprensione esaustiva dell'argomento e il tutto sarà corredato da alcuni esempi.

Innanzitutto verranno riviste delle nozioni “di base” di Algebra Lineare per gli endomorfismi che sfrutteremo successivamente per spiegare i principi della decomposizione in valori singolari, sia nel campo dei numeri reali \mathbb{R} , che in quello dei numeri complessi \mathbb{C} . Partendo dagli spazi vettoriali definiti all'interno del campo dei numeri reali, ricordiamo che la matrice quadrata A risulta diagonalizzabile solo se vengono soddisfatte alcune condizioni. Incontreremo dunque il problema degli autovalori e degli autovettori. Gli autovettori sono essenzialmente dei vettori che, se mappati dalla funzione lineare f , mantengono la direzione che avevano in precedenza e vengono “stirati” o “ridotti” in modulo in base alla grandezza degli autovalori corrispondenti (in formule un autovettore $v \in V$ è tale che $Av = \lambda v$, dove $\lambda \in \mathbb{C}$ è l'autovalore corrispondente). Se si trovano tanti autovettori linearmente indipendenti quanto è la dimensione dello spazio vettoriale V , allora la funzione f sarà diagonalizzabile tramite un opportuno cambio di base; la matrice di f rispetto a tale base sarà perciò diagonale. In realtà, come vedremo meglio nel Capitolo 1, il problema della diagonalizzabilità impiega ulteriori concetti, come il polinomio caratteristico, che permette tra le altre cose di trovare gli autovalori di una data matrice e le relative molteplicità algebriche. La molteplicità geometrica invece fa riferimento alla dimensione dello spazio vettoriale generato dagli autovettori corrispondenti a un dato autovalore. Noteremo infatti che, per ogni autovalore, se la molteplicità algebrica equivale a quella geometrica, si ha diagonalizzabilità.

Vi è inoltre una classe di matrici quadrate reali che è particolare perché possiede delle proprietà che la rendono molto interessante e queste saranno molto utili quando parleremo della SVD e in generale quando estenderemo l'analisi al caso delle matrici complesse: la classe delle matrici simmetriche. Relativamente a questa categoria di matrici, pre-

senteremo e dimostreremo il noto Teorema Spettrale, il quale afferma che ogni matrice simmetrica a coefficienti reali è sempre diagonalizzabile.

Tuttavia, nella maggior parte dei casi, quando si rilevano dei dati o quando si modella un qualche fenomeno fisico, emergono matrici rettangolari: a questo punto entra in gioco la decomposizione in valori singolari. Essa permette di trovare due basi tali che la matrice di una data funzione g rispetto a tali basi è diagonale ed eventualmente contiene delle righe o colonne di soli zeri, il rango infatti non può che essere minore o uguale della minima dimensione della matrice. Nel Capitolo 2 verrà presentato nel dettaglio come, date per l'appunto una funzione lineare g qualunque che ha come dominio e codominio spazi vettoriali diversi e la sua matrice M rispetto alle basi canoniche, sia sempre possibile portare la matrice in una forma diagonale con i cosiddetti valori singolari sulla diagonale. Questi valori singolari hanno una particolare relazione con gli autovalori della matrice tMM , ovvero con una speciale funzione composta. Il risultato che ne deriva è che esistono delle matrici del cambiamento di base che permettono di analizzare la matrice M attraverso la matrice equivalente decomposta nei suoi valori singolari, mantenendone le dimensioni e il rango.

Seppur meno usata nell'Ingegneria, verrà data una esposizione della SVD anche nel campo complesso per completezza, il che però richiede di spiegare come differisce lo spazio vettoriale \mathbb{C}^n , per qualche n , e le matrici complesse dagli analoghi reali. Forniremo quindi, nella seconda parte del Capitolo 2 una rapida carrellata di proprietà e definizioni per queste matrici, che comunque estendono il caso reale in maniera "quasi" naturale, eccetto per alcune (nuove) nozioni. Infine, grazie a questi concetti, verrà spiegato come calcolare la SVD anche per una matrice complessa.

Per concludere, nel Capitolo 3, presenteremo alcuni risultati che la SVD ha portato nel campo dell'Ingegneria, utilizzando articoli scientifici che trattano di processi di produzione e provengono dal settore elettronico e biomedico.

Daremo un'esposizione teorica della cosiddetta matrice pseudoinversa, una matrice molto utile nella pratica perché sostituisce il concetto di matrice inversa per le matrici rettangolari. La matrice inversa è infatti definita solo per matrici quadrate. Si vedrà che il modo più veloce e intuitivo per determinare la pseudoinversa utilizza la SVD, in particolare calcolando il reciproco dei valori singolari non nulli.

Applicheremo questa teoria ad un problema di valutazione delle condizioni operative adatte per rispondere alle richieste di qualità della produzione e del design di prodotti ([1]). Vedremo che, in questo contesto, è necessario invertire una matrice rettangolare, il che porta all'impiego della pseudoinversa come possibile soluzione del problema.

Vedremo successivamente l'approssimazione di rango basso, la quale permette per l'appunto di approssimare, tramite una funzione di costo opportuna, una matrice M con un'altra matrice di rango inferiore, sfruttando la decomposizione in valori singolari della matrice M .

Analizzeremo brevemente come viene usata per ridurre l'influenza del rumore in un sistema elettronico che rileva la composizione chimica dei gas tramite dei sensori ([4]). Grazie alla SVD e all'analisi dei valori singolari, sarà possibile ricavare le informazioni più importanti sul riconoscimento di gas e odori nonostante la presenza del rumore costruendo una matrice di rango più basso e contenente i valori singolari con peso maggiore.

Per concludere, indagheremo brevemente su come la SVD permetta il riconoscimento di semplici gesti della mano e l'identificazione del livello di disabilità di una persona nel camminare ([5]). Tramite la misurazione di punti fissati in istanti diversi con dei sensori verranno costruite tre matrici sulla base delle rilevazioni eseguite. La SVD di queste tre matrici fornisce delle informazioni più chiare sui movimenti effettuati ed è quindi un metodo efficace da applicare per un sistema di riconoscimento automatico.

Per riassumere, il documento sarà così organizzato:

- Nel Capitolo 1, ripasseremo concetti di Algebra Lineare come autovalori, autovettori, diagonalizzazione di matrici quadrate e tutti i loro aspetti nel campo \mathbb{R} ; successivamente rivedremo le proprietà delle matrici simmetriche e, a loro proposito, enunceremo e dimostreremo il Teorema Spettrale, che esprime uno dei risultati più importanti dell'intera teoria delle matrici.
- In seguito, nel Capitolo 2, analizzeremo nel dettaglio la decomposizione in valori singolari per matrici reali, fornendo anche un'interpretazione geometrica. Esamineremo poi lo spazio vettoriale \mathbb{C}^n e le sue proprietà e, grazie a tali concetti, comprenderemo come funziona la SVD anche per matrici complesse.
- Infine, nel Capitolo 3, apprezzeremo la potenza della SVD osservando alcune delle applicazioni più importanti che ha portato al mondo dell'Ingegneria, quali ad esempio la matrice pseudoinversa e l'approssimazione di rango basso. Ne daremo una trattazione teorica e vedremo come vengono utilizzate aiutandoci con articoli scientifici di svariati settori ingegneristici.

Capitolo 1

Richiami di Algebra Lineare

In questo capitolo verranno rivisti alcuni concetti fondamentali di Algebra Lineare, quali autovalori, autovettori e autospazi e il problema della diagonalizzazione di matrici quadrate. Enunceremo successivamente un teorema utile che assicura la diagonalizzazione di una classe di particolari matrici: le matrici simmetriche. Verranno anche mostrati degli esempi per chiarire tali nozioni.

1.1 Autovalori e autovettori

Sia $f: V \rightarrow V$ una funzione lineare, dove V è uno spazio vettoriale di dimensione n definito sul campo K . La diagonalizzazione di un endomorfismo (trasformazione lineare in cui dominio e codominio coincidono) equivale a domandarsi se esiste una base di V per cui la corrispondente matrice di f sia in forma diagonale, ovvero ci siano elementi diversi da 0 al più sulla diagonale, che è la struttura più semplice e conveniente per una matrice.

Definizione 1.1.1. Un endomorfismo f di V è *diagonalizzabile* se esiste una base di V per cui la matrice di f rispetto a tale base sia diagonale.

Inoltre, tale matrice deve rappresentare la stessa funzione f , quindi la matrice iniziale A è *diagonalizzabile* se esiste una matrice diagonale D e una matrice invertibile S tale che $A = SDS^{-1}$.

Ricordiamo che le due matrici A e D possiedono questa precisa relazione tramite la matrice S , detta *matrice del cambiamento di base*, che a sua volta rappresenta la funzione identità rispetto alle due basi diverse di V .

Dalla definizione vista sopra, richiedere l'esistenza di una tale base significa che devono esistere n vettori linearmente indipendenti v_1, \dots, v_n tali che $f(v_i) = \lambda_i v_i$ per $i = 1, \dots, n$. La matrice di f rispetto a questa base assume quindi la seguente forma:

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

Definizione 1.1.2. Un *autovalore* di f è un valore $\lambda \in K$ per cui esiste almeno un vettore $v \in V$, $v \neq \mathbf{0}$ tale che $f(v) = \lambda v$. In aggiunta, v viene detto *autovettore* di f relativo all'autovalore λ .

Se prendiamo l'equazione $f(v) = \lambda v$ e vediamo λv come un'applicazione lineare del tipo λid_V (funzione identica moltiplicata per lo scalare λ), possiamo riscrivere la relazione in questo modo: $(f - \lambda)(v) = 0$. Di conseguenza, l'insieme degli autovettori relativi all'autovalore λ è il nucleo della nuova funzione lineare che prende un vettore v e lo trasforma in $f(v) - \lambda v$, ovvero poniamo $V_\lambda = \text{Ker}(f - \lambda)$. L'insieme V_λ è quindi un sottospazio vettoriale di V , essendo il nucleo di un'applicazione lineare e viene chiamato *autospatio* di f relativo all'autovalore λ .

Per calcolare autovalori e autovettori, vediamo subito che dalla definizione la condizione necessaria e sufficiente perché $\lambda \in K$ sia un autovalore è che $\text{Ker}(f - \lambda) \neq \{\mathbf{0}\}$, ovvero che la funzione $f - \lambda$ non sia iniettiva, cioè non sia un isomorfismo e quindi valga $\det(f - \lambda) = 0$. Per calcolare questo determinante, basta prendere una matrice A di f relativa a una base qualsiasi di V e ricordare che la matrice di λid_V è la matrice ottenuta prendendo la matrice identità I_n o (I) e moltiplicandola per λ ; a questo punto si ha $\det(A - \lambda I_n)$.

Definizione 1.1.3. Data una matrice A di ordine n a coefficienti in K e data x una variabile incognita, il *polinomio caratteristico* di A è

$$P_A(x) = \det(A - x \cdot I).$$

Osservazione 1.1.4. Il polinomio caratteristico è quindi di grado n a coefficienti in K e si dimostra essere invariante rispetto a basi diverse (e quindi matrici diverse) della stessa funzione lineare, ovvero si può parlare di *polinomio caratteristico* $P_f(x)$ di un endomorfismo f di V .

Osservazione 1.1.5. $\lambda \in K$ è un autovalore di f se e solo se λ è una radice di $P_f(x)$, cioè se $P_f(\lambda) = 0$. Quest'ultima equazione viene chiamata *equazione caratteristica*.

Vediamo un esempio chiarificatore.

Esempio 1.1.6. Consideriamo la matrice

$$A = \begin{pmatrix} -5 & 8 \\ -4 & 7 \end{pmatrix}$$

Il polinomio caratteristico di A è

$$\det(A - x \cdot I) = \det \begin{pmatrix} -5-x & 8 \\ -4 & 7-x \end{pmatrix} = x^2 - 2x - 3.$$

Ponendo $x^2 - 2x - 3 = 0$ troviamo le radici reali $x_1 = -1$ e $x_2 = 3$, ovvero i due autovalori di A .

Per trovare gli autovettori corrispondenti all'autovalore $\lambda_1 = -1$ cerchiamo le soluzioni del seguente sistema:

$$(A - (-1) \cdot I) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{0}$$

che equivale a

$$\begin{pmatrix} -4 & 8 \\ -4 & 8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

le cui soluzioni sono, per entrambe le equazioni, date da $x_1 = 2x_2$.
L'autospazio relativo a λ_1 è quindi

$$V_{-1} = \left\{ \begin{pmatrix} 2\alpha \\ \alpha \end{pmatrix} \mid \alpha \in \mathbb{R} \right\}$$

che ha dimensione 1 e una base del sottospazio è, ad esempio, formata dal solo vettore $v_1 = (2, 1)$.

Per quanto riguarda l'autovalore $\lambda_2 = 3$, gli autovettori si trovano dalle soluzioni del sistema

$$(A - 3 \cdot I) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{0}$$

cioè

$$\begin{pmatrix} -8 & 8 \\ -4 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

e le sue soluzioni provengono dalla sola equazione $x_1 = x_2$.
L'autospazio relativo a λ_2 è quindi

$$V_3 = \left\{ \begin{pmatrix} \beta \\ \beta \end{pmatrix} \mid \beta \in \mathbb{R} \right\}$$

il quale ha anch'esso dimensione 1 e una base del sottospazio è, ad esempio, formata dal solo vettore $v_2 = (1, 1)$.

I due vettori v_1 e v_2 sono linearmente indipendenti (non sono paralleli) e quindi formano una base di \mathbb{R}^2 , perciò esiste una base di \mathbb{R}^2 formata da autovettori di A e quindi A è diagonalizzabile. Si può infatti verificare che $D = S^{-1}AS$, dove $D = \begin{pmatrix} -1 & 0 \\ 0 & 3 \end{pmatrix}$ e $S = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$.

Per caratterizzare meglio autovalori e autovettori e capire quando una matrice arbitraria è diagonalizzabile o meno, è necessario introdurre ulteriori concetti.

Definizione 1.1.7. Sia f un endomorfismo di uno spazio vettoriale di dimensione finita V , sia $P_f(x)$ il suo polinomio caratteristico e $\lambda \in K$ un autovalore di f . La *molteplicità algebrica* di λ è il più grande intero m tale che $P_f(x)$ sia divisibile per $(x - \lambda)^m$, d'altra parte la *molteplicità geometrica* di λ è la dimensione dell'autospazio $V_\lambda = \text{Ker}(f - \lambda)$.

Proposizione 1.1.8. Dato un autovalore $\lambda \in K$ di molteplicità algebrica m di un endomorfismo f di V si ha che la molteplicità geometrica di λ è minore o uguale della sua molteplicità algebrica, ovvero

$$\dim V_\lambda \leq m.$$

Dimostrazione. Sia $r = \dim V_\lambda$ e sia v_1, \dots, v_r una base di V_λ . Completiamo tale base ad una base di V aggiungendo gli $n - r$ vettori arbitrari v_{r+1}, \dots, v_n ai precedenti r . Rispetto a questa base la matrice A di f assume la seguente forma a blocchi:

$$A = \begin{pmatrix} \lambda \cdot I_r & B \\ \mathbf{0} & C \end{pmatrix}$$

dove B è una matrice con r righe e $n - r$ colonne, mentre C è quadrata di ordine $n - r$. Il polinomio caratteristico di f è quindi

$$P_f(x) = \det(A - x \cdot I) = \det((\lambda - x) \cdot I_r) \det(C - x \cdot I) = (\lambda - x)^r \det(C - x \cdot I).$$

Dall'arbitrarietà delle matrici B e C vediamo quindi che λ è una radice del polinomio caratteristico con molteplicità *almeno* uguale a r perciò si ha $m \geq r = \dim V_\lambda$, che dimostra la tesi. \square

Teorema 1.1.9. *Sia $f: V \rightarrow V$ un endomorfismo di uno spazio vettoriale V di dimensione n sul campo K . Siano inoltre $\lambda_1, \dots, \lambda_r$ gli autovalori di f e m_1, \dots, m_r le rispettive molteplicità algebriche. Allora f è diagonalizzabile se e solo se $m_1 + \dots + m_r = n$ e, per ogni autovalore, la sua molteplicità algebrica coincide con la molteplicità geometrica.*

Dimostrazione. Dobbiamo dimostrare entrambe le implicazioni.

Se f è diagonalizzabile esiste una base di V rispetto alla quale la matrice di f è diagonale e a blocchi del tipo

$$\begin{pmatrix} \lambda_1 \cdot I_{m_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \lambda_2 \cdot I_{m_2} & \dots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \lambda_r \cdot I_{m_r} \end{pmatrix}$$

Il polinomio caratteristico, usando questa matrice, risulta quindi

$$P_f(x) = (\lambda_1 - x)^{m_1} (\lambda_2 - x)^{m_2} \dots (\lambda_r - x)^{m_r},$$

perciò è vero che $m_1 + \dots + m_r = \deg P_f(x) = n$. Inoltre esiste una base di V di autovettori di f , da cui segue che

$$\dim V_{\lambda_1} + \dim V_{\lambda_2} + \dots + \dim V_{\lambda_r} = n = \dim V$$

e vale di conseguenza che $\dim V_{\lambda_i} = m_i \forall i = 1, \dots, r$ dal fatto che $m_i \geq \dim V_{\lambda_i}$ (Proposizione 1.1.8).

Viceversa, supponiamo che $m_1 + \dots + m_r = n$ e che, per ogni autovalore λ_i , la sua molteplicità geometrica coincida con la molteplicità algebrica; vogliamo dimostrare che f è diagonalizzabile.

Per ipotesi, $\dim V_{\lambda_i} = m_i$, quindi esiste una base v_{i1}, \dots, v_{im_i} dell'autospazio $V_{\lambda_i} \forall i = 1, \dots, r$. Per ogni autospazio ci sono m_i autovettori e, sempre per ipotesi, $m_1 + \dots + m_r = n$. Di conseguenza, l'insieme di tutti gli autovettori di f contiene n vettori linearmente indipendenti (si dimostra che autovettori relativi ad autovalori distinti sono linearmente indipendenti) e quindi esiste una base di V formata da autovettori e pertanto f è diagonalizzabile. \square

1.2 Matrici simmetriche

1.2.1 Introduzione e prime proprietà

Definizione 1.2.1. Sia $A \in M_n(K)$ una matrice quadrata di ordine n a coefficienti in un campo K . A è *simmetrica* se ${}^tA = A$, è *antisimmetrica* se ${}^tA = -A$.

Lemma 1.2.2. Dato $v = {}^t(a_1, \dots, a_n) \in \mathbb{C}^n$, sia $\bar{v} = {}^t(\bar{a}_1, \dots, \bar{a}_n)$ il vettore complesso coniugato di v . Allora, $\forall v \in \mathbb{C}^n$, si ha ${}^tv\bar{v} \geq 0$ e ${}^tv\bar{v} = 0$ se e solo se $v = \mathbf{0}$.

Dimostrazione.

$$\begin{aligned} {}^tv\bar{v} &= (a_1, \dots, a_n) \begin{pmatrix} \bar{a}_1 \\ \vdots \\ \bar{a}_n \end{pmatrix} \\ &= a_1\bar{a}_1 + \dots + a_n\bar{a}_n \\ &= |a_1|^2 + \dots + |a_n|^2, \end{aligned}$$

dove abbiamo usato il fatto che $\forall z \in \mathbb{C}$ si ha $z\bar{z} = |z|^2$. Pertanto ${}^tv\bar{v}$ è una somma di quadrati reali e notiamo che vale 0 solo quando $a_i = 0 \forall i = 1, \dots, n$, il che dimostra il lemma. Si veda la sezione 2.2 per chiarimenti sulle operazioni tra numeri complessi. \square

Teorema 1.2.3. Sia $A \in M_n(\mathbb{R})$ una matrice simmetrica, di ordine n , a coefficienti reali. Allora A possiede n autovalori reali (non necessariamente distinti).

Dimostrazione. Il polinomio caratteristico di A $P_A(x) = \det(A - x \cdot I)$ è un polinomio di grado n a coefficienti reali. Per il Teorema Fondamentale dell'Algebra, l'equazione $P_A(x) = 0$ ha n soluzioni nel campo complesso, quindi A possiede n autovalori $\lambda_1, \dots, \lambda_n \in \mathbb{C}$.

Si prenda un autovalore $\lambda \in \mathbb{C}$ di A e un corrispondente autovettore $v \in \mathbb{C}^n$, vale quindi $Av = \lambda v$. Ricordando che $\bar{A} = A$ perché A è una matrice a coefficienti reali, si ha che $A\bar{v} = \bar{\lambda}\bar{v}$, che si ottiene eseguendo l'operazione di complesso coniugato all'uguaglianza precedente. Se invece si applica l'operazione di trasposizione a $Av = \lambda v$ si trova ${}^tvA = \lambda{}^tv$, che equivale a ${}^tvA = \lambda{}^tv$, dato che ${}^tA = A$ per ipotesi. Moltiplicando entrambi i membri di questa uguaglianza a destra per \bar{v} si ha:

$${}^tvA\bar{v} = \lambda{}^tv\bar{v}.$$

Visto che $A\bar{v} = \bar{\lambda}\bar{v}$, si ottiene $\bar{\lambda}{}^tv\bar{v} = \lambda{}^tv\bar{v}$, ovvero $(\bar{\lambda} - \lambda){}^tv\bar{v} = 0$. Ovviamente $v \neq \mathbf{0}$, inoltre per il lemma 1.2.2 si ha ${}^tv\bar{v} > 0$, perciò deve essere $\bar{\lambda} - \lambda = 0$, ovvero $\bar{\lambda} = \lambda$, che significa $\lambda \in \mathbb{R}$. \square

1.2.2 Forme bilineari e matrici

Prima di enunciare altri risultati per la diagonalizzazione di matrici simmetriche, introduciamo un concetto utile anche in seguito, che è quello delle forme bilineari.

Definizione 1.2.4. Una *forma bilineare* su V è una funzione

$$g: V \times V \rightarrow K, \quad (v, w) \mapsto g(v, w),$$

lineare rispetto a ciascuno dei due argomenti, cioè tale che

$$(i) \quad g(\lambda_1 v_1 + \lambda_2 v_2, w) = \lambda_1 g(v_1, w) + \lambda_2 g(v_2, w),$$

$$(ii) \quad g(v, \mu_1 w_1 + \mu_2 w_2) = \mu_1 g(v, w_1) + \mu_2 g(v, w_2),$$

per ogni $\lambda_1, \lambda_2, \mu_1, \mu_2 \in K$ e ogni $v, v_1, v_2, w, w_1, w_2 \in V$.

Definizione 1.2.5. Una forma bilineare $g: V \times V \rightarrow K$ è detta *simmetrica* se $g(v, w) = g(w, v)$, per ogni $v, w \in V$.

Definizione 1.2.6. Sia V uno spazio vettoriale dotato di una forma bilineare simmetrica non degenere (ovvero con $\text{Ker}(g) = \mathbf{0}$). Due vettori v, w si dicono *ortogonali* se $g(v, w) = 0$. Due sottospazi vettoriali $U_1, U_2 \subseteq V$ si dicono *ortogonali* se $g(u_1, u_2) = 0$, per ogni $u_1 \in U_1$ e ogni $u_2 \in U_2$. Inoltre, dato un sottoinsieme S di V definiamo *l'ortogonale* di S :

$$S^\perp = \{v \in V \mid g(v, w) = 0, \forall w \in S\}.$$

Sia ora $K = \mathbb{R}$. Una forma bilineare simmetrica viene detta *definita positiva* se $g(v, v) > 0, \forall v \in V, v \neq \mathbf{0}$, *semidefinita positiva* con \geq nella disuguaglianza invece di $>$ e analogamente con $<$ e \leq per la forma *definita (e semidefinita) negativa* ed è *indefinita* negli altri casi.

Infine, una base formata da vettori di V a due a due ortogonali è detta una *base ortogonale* di V mentre una *base ortonormale* è una base ortogonale con la condizione aggiuntiva che $g(v, v) = 1$ per ogni vettore v della base.

Nel nostro caso, ci interessano le forme bilineari simmetriche definite positive.

Definizione 1.2.7. Dato V uno spazio vettoriale dotato di una forma bilineare simmetrica definita positiva, la *norma* di un vettore $v \in V$ è così definita:

$$\|v\| = \sqrt{g(v, v)}.$$

Segue infatti che $\|v\| \geq 0$ e $\|\lambda v\| = |\lambda| \|v\| \forall v \in V$ e $\forall \lambda \in \mathbb{R}$, e vale $\|v\| = 0$ se e solo se $v = \mathbf{0}$.

Osservazione 1.2.8. Il prodotto scalare usuale in \mathbb{R}^n è una forma bilineare simmetrica non degenere definita positiva:

$$v \cdot w = a_1 b_1 + \cdots + a_n b_n = \sum_{i=1}^n a_i b_i,$$

dove $v = (a_1, \dots, a_n)$ e $w = (b_1, \dots, b_n)$. Uno spazio vettoriale dotato di una forma bilineare simmetrica definita positiva viene detto *spazio vettoriale euclideo*.

La matrice di una forma bilineare simmetrica rispetto a una base \mathbf{v} di V è la matrice G con elementi $g_{ij} = g(v_i, v_j)$.

Quando si effettua un cambio di base, anche la matrice cambia e si trova una relazione particolare; denominate G la matrice scritta rispetto alla base \mathbf{v} e G' rispetto alla base \mathbf{v}' che rappresentano la stessa forma bilineare simmetrica e P la matrice del cambiamento di base con le componenti dei vettori della base \mathbf{v}' scritte rispetto alla base \mathbf{v} , si ottiene la seguente formula:

$$G' = {}^t P G P,$$

in tal caso G e G' sono dette *congruenti*. In particolare se le basi v e v' sono ortonormali si trova che ${}^t P P = I$. Una matrice P tale che ${}^t P P = I$ viene detta *matrice ortogonale*.

Si dimostra inoltre che ogni spazio vettoriale V di dimensione finita sul campo \mathbb{R} , dotato di una forma bilineare simmetrica definita positiva g , possiede una base ortonormale, perciò, operando un cambio di base, data una matrice G simmetrica definita positiva (che rappresenta la forma bilineare definita positiva), esiste sempre una matrice P invertibile tale che ${}^t P G P = I$.

1.2.3 Teorema spettrale

Definizione 1.2.9. Una funzione lineare $f: V \rightarrow V$ è detta *simmetrica* se

$$f(v) \cdot w = v \cdot f(w)$$

per ogni $v, w \in V$. È vero in realtà che una funzione lineare è simmetrica se e solo se la relazione sopra vale per i soli vettori di una base arbitraria di V . Inoltre una funzione lineare è simmetrica se e solo se, presa una base ortonormale di V , la matrice rispetto a tale base è simmetrica.

Teorema 1.2.10. (TEOREMA SPETTRALE). *Siano V uno spazio vettoriale euclideo e $f: V \rightarrow V$ una funzione lineare. Allora f è ortogonalmente diagonalizzabile, cioè esiste una base ortonormale di autovettori di f , se e solo se f è simmetrica. Di conseguenza, data una matrice simmetrica definita positiva a coefficienti reali, essa può sempre essere trasformata in una matrice diagonale tramite una matrice di cambiamento di base ortogonale.*

Dimostrazione. Supponiamo che f sia ortogonalmente diagonalizzabile, cioè che esista una base ortonormale di V fatta di autovettori di f . La matrice rispetto a tale base è una matrice diagonale e quindi simmetrica. Ciò implica che f è simmetrica.

Viceversa, dimostriamo per induzione che ogni funzione simmetrica è ortogonalmente diagonalizzabile. Se $\dim V = 1$, sia v_1 una base ortonormale di V , si ha dunque $f(v_1) = \lambda_1 v_1$, ovvero v_1 è un autovettore di f .

Supponiamo che V abbia dimensione $n > 1$ e che l'enunciato sia vero per spazi vettoriali di dimensione $n - 1$. Presa una base ortonormale $\{u_1, \dots, u_n\}$ di V , la matrice A di f rispetto a tale base è simmetrica per ipotesi, quindi ha n autovalori reali (si veda il Teorema 1.2.3).

Sia λ_1 un autovalore di f e v_1 un autovettore corrispondente, si prenda $W = \langle v_1 \rangle^\perp$. Dalla simmetria di f , vale infatti che $f(v_1) \cdot w = \lambda_1 v_1 \cdot w = 0 = v_1 \cdot f(w)$, dunque $f(w) \in \langle v_1 \rangle^\perp = W$

per ogni $w \in W$. La funzione $f|_W: W \rightarrow W$ ristretta a W è anch'essa una funzione lineare simmetrica con $\dim W = n - 1$.

Per ipotesi induttiva essa è ortogonalmente diagonalizzabile tramite una base ortonormale di W $\{w_2, \dots, w_n\}$ di autovettori di $f|_W$. Basta quindi porre $w_1 = v_1 / \|v_1\|$ ed osservare che i vettori w_1, \dots, w_n formano una base ortonormale di V costituita da autovettori di f . \square

Capitolo 2

Decomposizione in valori singolari (SVD)

Abbiamo visto nel Capitolo 1 come calcolare autovalori e autovettori e, come stabilito dal Teorema Spettrale (Teorema 1.2.10), le matrici reali simmetriche siano sempre diagonalizzabili. Ora invece andremo a presentare un'estensione di tale teoria a matrici generiche (quindi anche rettangolari) e come esse siano decomponibili in valori singolari. Vedremo poi di estendere i risultati al caso complesso.

2.1 Decomposizione in valori singolari per matrici reali

Si considerino due spazi vettoriali V e W dotati ciascuno di forme bilineari simmetriche definite positive g e h , ciò significa che in V e W esistono basi ortonormali; essi sono quindi *isometrici* rispettivamente a \mathbb{R}^n e \mathbb{R}^m , per qualche n, m . Possiamo quindi supporre che essi siano \mathbb{R}^n e \mathbb{R}^m ; sia $f: V \rightarrow W$ una funzione lineare, è possibile considerare $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ senza perdita di generalità.

Sia $A = M_{\mathbf{e}}^{\mathbf{e}}(f)$ la matrice di f rispetto alle basi canoniche, quindi A ha m righe e n colonne.

Definizione 2.1.1. La *trasposta* di f è una funzione lineare $f^*: \mathbb{R}^m \rightarrow \mathbb{R}^n$ tale che soddisfi la seguente uguaglianza:

$$v \cdot f^*(w) = f(v) \cdot w,$$

per ogni $v \in \mathbb{R}^n$ e ogni $w \in \mathbb{R}^m$.

Osservazione 2.1.2. Notiamo che tale definizione è molto simile a quella per la funzione simmetrica (Definizione 1.2.9) che valeva per endomorfismi, la trasposta allarga il concetto a funzioni lineari qualunque.

La funzione f^* è detta la *trasposta* di f perché la sua matrice rispetto alle basi canoniche $\mathbf{e}^{(n)}(e_1, \dots, e_n)$ e $\mathbf{e}^{(m)}(e_1, \dots, e_m)$ è la trasposta della matrice $A = (a_{ij})$ di f : $M_{\mathbf{e}}^{\mathbf{e}}(f^*) = {}^tA$. Indichiamo con $B = (b_{ij})$ la matrice di f^* rispetto alle basi canoniche, vogliamo dimostrare che $B = {}^tA$.

Si ha $f^*(e_j) = \sum_{l=1}^n b_{lj}e_l$, per ogni $j = 1, \dots, m$. Vale quindi $e_i \cdot f^*(e_j) = b_{ij}$, ricordando che $e_i \cdot e_j = 0 \forall i \neq j$ mentre $e_i \cdot e_i = 1$.

Analogamente, $f(e_i) = \sum_{l=1}^m a_{li}e_l$, per ogni $i = 1, \dots, n$ e si ottiene $f(e_i) \cdot e_j = a_{ji}$.

Dalla definizione di trasposta, risulta infine che $e_i \cdot f^*(e_j) = f(e_i) \cdot e_j$, cioè $b_{ij} = a_{ji}$, ovvero $B = {}^tA$.

Si può dimostrare, similmente a quanto fatto sopra per la matrice di f^* , che nucleo e immagine delle due funzioni f e f^* sono legati dalle seguenti uguaglianze:

$$\text{Ker}(f^*) = (\text{Im}f)^\perp, \quad \text{Im}(f^*) = (\text{Ker}f)^\perp.$$

Consideriamo ora le funzioni composte

$$f^* \circ f: \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad f \circ f^*: \mathbb{R}^m \rightarrow \mathbb{R}^m,$$

le cui matrici, rispetto alle basi canoniche, sono tAA e $A{}^tA$, rispettivamente. tAA è una matrice quadrata di ordine n , simmetrica poiché ${}^t({}^tAA) = {}^tAA$. Analogamente $A{}^tA$ è una matrice quadrata simmetrica di ordine m .

Proposizione 2.1.3. *La matrice tAA è semidefinita positiva, inoltre si ha $\text{Ker}({}^tAA) = \text{Ker}(A)$.*

La matrice $A{}^tA$ è anch'essa semidefinita positiva, inoltre si ha $\text{Ker}(A{}^tA) = \text{Ker}{}^tA$.

Dimostrazione. Verifichiamo l'enunciato solo per tAA , la dimostrazione per $A{}^tA$ è del tutto analoga.

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

$${}^tv({}^tAA)v = ({}^tv{}^tA)(Av) = {}^t(Av)(Av) = {}^tw w = w \cdot w = \|w\|^2 \geq 0,$$

dove si è posto $w = Av$. In questo modo tAA è semidefinita positiva.

Inoltre, osserviamo dal calcolo fatto sopra che ${}^tv({}^tAA)v = 0$ se e solo se $w = \mathbf{0}$, ovvero se e solo se $v \in \text{Ker}A$. Quindi, se $v \in \text{Ker}({}^tAA)$ si ha sicuramente ${}^tv({}^tAA)v = 0$, segue che $v \in \text{Ker}A$. Ciò dimostra che $\text{Ker}({}^tAA) \subseteq \text{Ker}A$. Banalmente $\text{Ker}A \subseteq \text{Ker}({}^tAA)$ dato che, se $Av = \mathbf{0}$, allora anche $({}^tAA)v = {}^tA(Av) = \mathbf{0}$; ciò dimostra che $\text{Ker}({}^tAA) = \text{Ker}A$. \square

La matrice tAA è simmetrica, quindi è ortogonalmente diagonalizzabile, ovvero esiste una base ortonormale di \mathbb{R}^n formata da autovettori di tAA . Essendo tale matrice semidefinita positiva, i suoi autovalori λ_i sono ≥ 0 . Indichiamo ora con r il rango di A , dunque $\dim(\text{Ker}A) = n - r$. Ricordando che $\text{Ker}({}^tAA) = \text{Ker}A$, si deduce che la matrice tAA ha un autovalore $\lambda = 0$ di molteplicità $n - r$ e il corrispondente autospazio è esattamente il nucleo di A .

Sia quindi \mathbf{v} una base ortonormale di \mathbb{R}^n formata da autovettori di tAA , dove $\mathbf{v} = \{v_1, \dots, v_r, u_1, \dots, u_{n-r}\}$. Poniamo dunque $({}^tAA)v_i = \lambda_i v_i$, dove $\lambda_1, \dots, \lambda_r$ sono gli autovalori non nulli di tAA , e $({}^tAA)u_j = \mathbf{0}$ per $j = 1, \dots, n-r$. I vettori u_1, \dots, u_{n-r} formano una base ortonormale di $\text{Ker}f$, mentre v_1, \dots, v_r costituiscono una base ortonormale di $V' = (\text{Ker}f)^\perp$. Si ha quindi che V è decomposto nella somma diretta $V = V' \oplus \text{Ker}f = (\text{Ker}f)^\perp \oplus \text{Ker}f$. Applicando la funzione f si ottiene:

$$f(v_1) = w_1, \dots, f(v_r) = w_r, f(u_1) = \mathbf{0}, \dots, f(u_{n-r}) = \mathbf{0},$$

dove $Imf = \langle w_1, \dots, w_r \rangle$. Dimostriamo ora che i vettori w_1, \dots, w_r sono a due a due ortogonali. Infatti se $i \neq j$ si ha:

$$w_i \cdot w_j = f(v_i) \cdot f(v_j) = (Av_i) \cdot (Av_j) = {}^t(Av_i)(Av_j) = {}^t v_i {}^t A A v_j = {}^t v_i \lambda_j v_j = \lambda_j v_i \cdot v_j = 0$$

poiché v_j è un autovettore di ${}^t A A$ relativo all'autovalore λ_j e i vettori v_i e v_j sono ortogonali per ipotesi.

Possiamo quindi normalizzare i vettori w_i , ponendo $\hat{w}_i = w_i / \|w_i\|$, così da ottenere una base ortonormale $\{\hat{w}_1, \dots, \hat{w}_r\}$ di Imf . Consideriamo poi una base ortonormale $\{z_1, \dots, z_{m-r}\}$ di $(Imf)^\perp$ e otteniamo perciò una base ortonormale di \mathbb{R}^m , $\mathbf{w} = \{\hat{w}_1, \dots, \hat{w}_r, z_1, \dots, z_{m-r}\}$.

Proposizione 2.1.4. *La matrice $M_{\mathbf{v}}^{\mathbf{w}}(f)$ di f , rispetto alle basi \mathbf{v} di \mathbb{R}^n e \mathbf{w} di \mathbb{R}^m , è una matrice a blocchi del tipo*

$$D = M_{\mathbf{v}}^{\mathbf{w}}(f) = \left(\begin{array}{cccc|cccc} \alpha_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & \alpha_r & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{array} \right)$$

dove gli α_i sono le radici quadrate (positive) degli autovalori non nulli λ_i di ${}^t A A$: $\alpha_i = \sqrt{\lambda_i}$.

Dimostrazione. Osserviamo che $f(v_1) = w_1 = \alpha_1 \hat{w}_1, \dots, f(v_r) = w_r = \alpha_r \hat{w}_r$, e poi $f(u_1) = \mathbf{0}, \dots, f(u_{n-r}) = \mathbf{0}$, dove $\alpha_i = \|w_i\| > 0$, per $i = 1, \dots, r$. Ciò dimostra la forma della matrice rispetto a tali basi.

Dimostriamo ora che $\alpha_i^2 = \lambda_i$. Ricordando che v_i è un autovettore di ${}^t A A$ associato all'autovalore λ_i , si ha:

$$v_i \cdot ({}^t A A)v_i = v_i \cdot \lambda_i v_i = \lambda_i \|v_i\|^2 = \lambda_i.$$

Usando la stessa espressione, si ha anche:

$$v_i \cdot ({}^t A A)v_i = {}^t v_i {}^t A A v_i = {}^t (A v_i)(A v_i) = {}^t w_i w_i = \|w_i\|^2 = \alpha_i^2$$

perciò vale $\alpha_i^2 = \lambda_i$, come volevasi dimostrare. \square

Definizione 2.1.5. Gli elementi sulla diagonale della matrice D sono detti i *valori singolari* della funzione f , o della matrice A , solitamente si intende i valori non nulli $\alpha_1, \dots, \alpha_r$. Questi valori singolari vengono normalmente disposti in ordine decrescente $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_r \geq 0$. Ovviamente il numero di valori singolari non nulli è il rango di f .

Le matrici $A = M_{\mathbf{e}}^{\mathbf{e}}(f)$ e $D = M_{\mathbf{v}}^{\mathbf{w}}(f)$ rappresentano quindi entrambe la stessa funzione lineare f rispetto a basi diverse del dominio e del codominio. Le matrici di cambiamento di

base sono: in \mathbb{R}^n la matrice $P = M_{\mathbf{v}}^e(id)$, le cui colonne sono le coordinate dei vettori della base ortonormale \mathbf{v} scritte rispetto alla base canonica di \mathbb{R}^n ; in \mathbb{R}^m la matrice $Q = M_{\mathbf{w}}^e(id)$, le cui colonne sono le coordinate dei vettori della base ortonormale \mathbf{w} scritte rispetto alla base canonica di \mathbb{R}^m . P e Q sono dunque matrici ortogonali e si ha $P^{-1} = {}^tP$ e $Q^{-1} = {}^tQ$. Dalla formula del cambiamento di base

$$M_e^e(f)M_{\mathbf{v}}^e(id) = M_{\mathbf{w}}^e(id)M_{\mathbf{v}}^w(f)$$

segue che $AP = QD$ e quindi

$$A = QDP^{-1} = QD{}^tP, \quad D = Q^{-1}AP = {}^tQAP.$$

Queste formule rappresentano la *decomposizione in valori singolari* della matrice A . Le colonne di Q , cioè i vettori della base ortonormale \mathbf{w} , sono detti i *vettori singolari sinistri* di A , mentre le colonne di P , cioè i vettori della base ortonormale \mathbf{v} , sono detti i *vettori singolari destri*.

Osservazione 2.1.6. Le considerazioni fatte fino ad ora hanno riguardato solo la matrice tAA . Per quanto concerne la matrice $A{}^tA$, di ordine m , essa è a sua volta simmetrica e, pertanto, ortogonalmente diagonalizzabile. Sfruttando la decomposizione in valori singolari di A , si ha:

$$A{}^tA = (QD{}^tP)({}^tQD{}^tP) = QD{}^tPP{}^tD{}^tQ = Q(D{}^tD){}^tQ.$$

La matrice $D{}^tD$ è una matrice diagonale i cui elementi non nulli sulla diagonale principale sono $\alpha_1^2 = \lambda_1, \dots, \alpha_r^2 = \lambda_r$. Perciò gli autovalori non nulli $\lambda_1, \dots, \lambda_r$ di tAA sono anche gli autovalori non nulli di $A{}^tA$. Inoltre le colonne della matrice Q sono formate dalle coordinate degli autovettori di $A{}^tA$, che sono quindi i vettori $\hat{w}_1, \dots, \hat{w}_r, z_1, \dots, z_{m-r}$.

È importante notare che, in generale, i valori singolari di una matrice quadrata sono *diversi* dai suoi autovalori: un esempio è dato da $A = \begin{pmatrix} 1 & 0 \\ 1 & -2 \end{pmatrix}$ i cui autovalori sono 1 e -2 come è facile vedere (la matrice è triangolare inferiore), mentre i valori singolari sono $\alpha_1 = \sqrt{3 + \sqrt{5}}$ e $\alpha_2 = \sqrt{3 - \sqrt{5}}$, i quali sono stati calcolati sfruttando il procedimento usato nell'Esempio 2.1.7 che verrà analizzato in seguito.

Siano V e W due spazi vettoriali euclidei (dotati quindi di forme bilineari simmetriche definite positive), possiamo riassumere il contenuto geometrico della decomposizione in valori singolari di una funzione lineare $f: V \rightarrow W$ come segue: lo spazio vettoriale V si decompone nella somma diretta $V = V' \oplus K$, dove $K = \text{Ker}f$ e $V' = (\text{Ker}f)^\perp$ e, analogamente, W si decompone nella somma diretta $W = W' \oplus Z$, dove $W' = \text{Im}f$ e $Z = (\text{Im}f)^\perp$. Esiste una base ortonormale $\mathbf{v} = \{v_1, \dots, v_r, u_1, \dots, u_{n-r}\}$ di V , dove $V' = \langle v_1, \dots, v_r \rangle$ e $K = \langle u_1, \dots, u_{n-r} \rangle$ e una base ortonormale $\mathbf{w} = \{\hat{w}_1, \dots, \hat{w}_r, z_1, \dots, z_{m-r}\}$, dove $W' = \langle \hat{w}_1, \dots, \hat{w}_r \rangle$ e $Z = \langle z_1, \dots, z_{m-r} \rangle$. La funzione f induce un isomorfismo \tilde{f} tra il sottospazio V' di V e il sottospazio W' di W perché le immagini dei vettori della base di V' formano essi stessi una base di W' . La matrice di questo isomorfismo, rispetto alle basi ortonormali $\{v_1, \dots, v_r\}$ di V' e $\{\hat{w}_1, \dots, \hat{w}_r\}$ di W' è

$$\begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_r \end{pmatrix}.$$

Infine, la funzione $f: V \rightarrow W$ si ottiene come composizione della proiezione ortogonale π di V sul sottospazio V' , seguita dall'isomorfismo $\tilde{f}: V' \rightarrow W'$, seguita dall'inclusione di W' in W , nel caso W' abbia dimensione minore di W :

$$f: V \xrightarrow{\pi} V' \xrightarrow{\tilde{f}} W' \hookrightarrow W.$$

Esempio 2.1.7. Consideriamo la funzione $f: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ la cui matrice, rispetto alle basi canoniche, è

$$A = \begin{pmatrix} 2 & 1 & 1 \\ -1 & 2 & 2 \end{pmatrix}$$

Vogliamo determinare la decomposizione in valori singolari di A .

Calcoliamo quindi la matrice tAA e troviamo i suoi autovalori e autovettori.

$${}^tAA = \begin{pmatrix} 2 & -1 \\ 1 & 2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 & 1 \\ -1 & 2 & 2 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 5 \\ 0 & 5 & 5 \end{pmatrix}$$

Il polinomio caratteristico vale $P_{AA}(x) = \det(sI - A) = (s - 5)s(s - 10)$. Gli autovalori di questa matrice sono: $\lambda_1 = 10$, $\lambda_2 = 5$ e $\lambda_3 = 0$. Gli autovettori risultano invece essere $(0, 1, 1)$ per l'autovalore $\lambda_1 = 10$, $(1, 0, 0)$ per l'autovalore $\lambda_2 = 5$ e $(0, -1, 1)$ per l'autovalore $\lambda_3 = 0$, ovvero $(0, -1, 1)$ è una base del nucleo di f .

I tre autovettori sono ortogonali tra loro, quindi per avere una base ortonormale di \mathbb{R}^3 basta normalizzarli, ottenendo $\mathbf{v} = \{v_1, v_2, u_1\}$, dove $v_1 = (0, 1/\sqrt{2}, 1/\sqrt{2})$ e $v_2 = (1, 0, 0)$ sono gli autovettori relativi agli autovalori (non nulli) $\lambda_1 = 10$ e $\lambda_2 = 5$ e formano una base ortonormale del sottospazio $V' = (\text{Ker } f)^\perp$, e $u_1 = (0, -1/\sqrt{2}, 1/\sqrt{2})$ è una base ortonormale di $\text{Ker } f$.

Applicando f ai vettori v_1 e v_2 troviamo i vettori $w_1 = f(v_1) = Av_1 = (2/\sqrt{2}, 4/\sqrt{2})$ e $w_2 = f(v_2) = Av_2 = (2, -1)$.

I valori singolari di f sono dunque $\alpha_1 = \|w_1\| = \sqrt{10}$ e $\alpha_2 = \|w_2\| = \sqrt{5}$. Per ottenere una base ortonormale di $W' = \text{Im } f$ bisogna normalizzare i vettori w_1 e w_2 , perciò si trovano i vettori $\hat{w}_1 = (1/\sqrt{5}, 2/\sqrt{5})$ e $\hat{w}_2 = (2/\sqrt{5}, -1/\sqrt{5})$. Possiamo quindi scrivere le matrici ortogonali P e Q :

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix} \quad Q = \begin{pmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{pmatrix}$$

e la matrice

$$D = \begin{pmatrix} \sqrt{10} & 0 & 0 \\ 0 & \sqrt{5} & 0 \end{pmatrix}$$

Vale infatti la relazione ottenuta nella Definizione 2.1.5 tra la matrice A e la matrice D data da $A = QD{}^tP$.

Seguendo il ragionamento fatto in precedenza, possiamo descrivere l'applicazione della funzione f a un vettore $v \in \mathbb{R}^3$ come composizione di varie funzioni: dapprima la proiezione ortogonale di v sul piano π generato dai vettori v_1 e v_2 (il piano π ha equazione $y = z$, infatti una combinazione lineare di v_1 e v_2 è del tipo $(\alpha, \beta, \beta) \forall \alpha, \beta \in \mathbb{R}$). A questo

punto il piano π si identifica al piano \mathbb{R}^2 tramite la funzione f . Nella trasformazione la direzione di $v_1 \in \pi$ diventa quella indicata da $w_1 \in \mathbb{R}^2$ però, ricordando che i vettori v_1 e v_2 sono autovettori di f , i vettori paralleli a v_1 vengono riscalati tramite il fattore $\alpha_1 = \sqrt{10}$ e similmente la direzione di $v_2 \in \pi$ diventa quella indicata da $w_2 \in \mathbb{R}^2$ e i vettori paralleli a v_2 vengono riscalati tramite il fattore $\alpha_2 = \sqrt{5}$.

L'effetto di questa funzione lineare è pertanto quello di “schiacciare” la sfera di raggio 1 centrata nell'origine di \mathbb{R}^3 sul piano π ottenendo un disco di raggio 1 e infine ruotare e dilatare tale disco trasformando i vettori v_1 e v_2 nei vettori w_1 e w_2 , ottenendo un'ellisse in \mathbb{R}^2 .

In questo esempio, da un punto di vista più astratto, ci sono due vettori (più precisamente sono gli autovettori della funzione composta $f^* \circ f$ relativi agli autovalori non nulli) normalizzati v_1 e v_2 che vengono mappati nell'immagine di f nei vettori non nulli w_1 e w_2 , i quali generano il sottospazio $\text{Im } f$, mentre il vettore ortonormale u_1 genera il sottospazio $\text{Ker } f$ (è un autovettore di $f^* \circ f$ relativo all'autovalore $\lambda = 0$). A questo punto, normalizzando i vettori w_1 e w_2 e operando un cambio di basi (dalle basi canoniche che descrivono la matrice A alle basi determinate dai vettori $\{v_1, v_2, u_1\}$ in \mathbb{R}^3 e dai vettori $\{w_1, w_2\}$ in \mathbb{R}^2 , si ottiene la matrice D , che contiene i valori singolari sulla diagonale principale.

2.2 Lo spazio vettoriale \mathbb{C}^n e le matrici complesse

Prima di descrivere la decomposizione in valori singolari nel campo \mathbb{C} dei numeri complessi, è opportuno mostrare alcuni risultati utili in tale campo per le matrici complesse. Fino ad ora infatti, abbiamo usato il campo \mathbb{R} e, tra le altre cose, i concetti delle forme bilineari; vedremo adesso che anche in \mathbb{C}^n vi è un prodotto scalare definito tramite le forme hermitiane e altre nozioni che estendono naturalmente i concetti visti fino ad ora per le matrici reali.

2.2.1 Forme hermitiane e matrici associate

Ricordiamo innanzitutto che un numero complesso ha una forma del tipo $z = a + ib \in \mathbb{C}$, dove $a = \Re(z)$ è la parte reale, mentre $b = \Im(z)$ è la parte immaginaria. Inoltre, il coniugato \bar{z} è così definito: $\bar{z} = a - ib$. Infine, il modulo di z possiede la seguente relazione: $|z|^2 = z\bar{z} = a^2 + b^2$.

Il campo \mathbb{C} può essere identificato con lo spazio vettoriale \mathbb{R}^2 associando a $z = a + ib$ il vettore $(a, b) \in \mathbb{R}^2$. Il quadrato della norma di tale vettore è uguale al quadrato del modulo di z , ovvero a $a^2 + b^2$.

Analogamente, possiamo quindi identificare lo spazio vettoriale \mathbb{C}^n con \mathbb{R}^{2n} associando al vettore $(z_1, \dots, z_n) \in \mathbb{C}^n$ il vettore $(a_1, b_1, \dots, a_n, b_n) \in \mathbb{R}^{2n}$, dove $z_j = a_j + ib_j$ per

$j = 1, \dots, n$. Il quadrato della norma di tale vettore è ora¹ :

$$\|(z_1, \dots, z_n)\|^2 = \sum_{j=1}^n z_j \bar{z}_j = \sum_{j=1}^n |z_j|^2 = \sum_{j=1}^n (a_j^2 + b_j^2).$$

Introduciamo quindi un *prodotto scalare* in \mathbb{C}^n che generalizza il prodotto scalare euclideo di \mathbb{R}^n .

Definizione 2.2.1. Siano $v = (z_1, \dots, z_n)$ e $w = (\zeta_1, \dots, \zeta_n)$ elementi di \mathbb{C}^n , il loro *prodotto scalare* (o *prodotto scalare hermitiano*) è definito ponendo

$$v \cdot w = z_1 \bar{\zeta}_1 + z_2 \bar{\zeta}_2 + \dots + z_n \bar{\zeta}_n = \sum_{j=1}^n z_j \bar{\zeta}_j.$$

Grazie a questa definizione, ritroviamo la notevole proprietà che valeva anche per il prodotto scalare in \mathbb{R}^n : $v \cdot v = \|v\|^2 \forall v \in \mathbb{C}^n$. Infatti, per come abbiamo definito la norma di un vettore complesso, risulta

$$v \cdot v = \sum_{j=1}^n z_j \bar{z}_j = \sum_{j=1}^n |z_j|^2 = \|v\|^2 \quad \text{e quindi } \|v\| = \sqrt{v \cdot v},$$

per ogni $v \in \mathbb{C}^n$.

Il prodotto scalare hermitiano soddisfa altre proprietà:

$$(i) \quad (\lambda_1 v_1 + \lambda_2 v_2) \cdot w = \lambda_1 v_1 \cdot w + \lambda_2 v_2 \cdot w,$$

$$(ii) \quad v \cdot (\mu_1 w_1 + \mu_2 w_2) = \bar{\mu}_1 v \cdot w_1 + \bar{\mu}_2 v \cdot w_2,$$

$$(iii) \quad w \cdot v = \overline{v \cdot w},$$

per ogni $\lambda_1, \lambda_2, \mu_1, \mu_2 \in \mathbb{C}$ e ogni $v, v_1, v_2, w, w_1, w_2 \in \mathbb{C}^n$.

La proprietà (i) afferma che il prodotto scalare è lineare nel primo argomento, la (ii) afferma che esso è coniugato-lineare (o antilineare) nel secondo, mentre la (iii) ci mostra che non è simmetrico a differenza del prodotto scalare in \mathbb{R}^n . Il prodotto scalare in \mathbb{C}^n è in realtà una forma sesquilineare, il che ci permette di dare la definizione:

Definizione 2.2.2. Sia V uno spazio vettoriale definito sul campo dei numeri complessi \mathbb{C} . Una *forma sesquilineare* su V è una funzione

$$g: V \times V \rightarrow \mathbb{C}, \quad (v, w) \mapsto g(v, w),$$

lineare rispetto al primo argomento e antilineare rispetto al secondo, che soddisfa quindi le seguenti proprietà (che derivano da quelle viste sopra):

$$(i) \quad g(\lambda_1 v_1 + \lambda_2 v_2, w) = \lambda_1 g(v_1, w) + \lambda_2 g(v_2, w),$$

¹Notiamo che se usassimo il prodotto scalare euclideo di \mathbb{R}^n anche per lo spazio \mathbb{C}^n , ovvero per un vettore complesso $\vec{z} = (z_1, \dots, z_n) \in \mathbb{C}^n$ la sua norma diverrebbe $\|z\| = \sum_{j=1}^n z_j^2$ si incontrerebbe subito un problema: tale forma bilineare non sarebbe più definita positiva in quanto, ad esempio, un vettore con componenti immaginarie avrebbe norma negativa!

$$(ii) \quad g(v, \mu_1 w_1 + \mu_2 w_2) = \bar{\mu}_1 g(v, w_1) + \bar{\mu}_2 g(v, w_2),$$

per ogni $\lambda_1, \lambda_2, \mu_1, \mu_2 \in \mathbb{C}$ e ogni $v, v_1, v_2, w, w_1, w_2 \in V$.
Una forma sesquilineare g è detta *hermitiana* se

$$g(w, v) = \overline{g(v, w)},$$

per ogni $v, w \in V$.

Osservazione 2.2.3. Se $g: V \times V \rightarrow \mathbb{C}$ è una forma hermitiana, per ogni vettore $v \in V$ si ha $g(v, v) = \overline{g(v, v)}$, ovvero $g(v, v)$ è un numero reale: nel nostro caso questa proprietà è essenziale in quanto possiamo ora estendere le definizioni di forma hermitiana definita positiva, definita negativa, ecc . . . , proprio come nel caso delle forme bilineari. Notiamo infatti che la definizione di forma sesquilineare ricalca quasi del tutto l'analoga definizione della forma bilineare, differendo solo per l'antilinearità rispetto al secondo argomento e il fatto che sia definita solamente sul campo \mathbb{C} , mentre una forma hermitiana differisce da una forma bilineare simmetrica anche per la presenza del coniugato rendendola appunto non simmetrica.

Diremo quindi che la forma hermitiana g è *definita positiva* se $g(v, v) > 0, \forall v \in V, v \neq \mathbf{0}$, *semidefinita positiva* con \geq nella disuguaglianza invece di $>$ e analogamente con $<$ e \leq per la forma *definita (e semidefinita) negativa* ed è *indefinita* negli altri casi. Sfruttiamo inoltre altri risultati che provengono dalle forme bilineari simmetriche, ad esempio due vettori $v, w \in V$ si dicono ortogonali se $g(v, w) = 0$, un vettore v è normalizzato se $g(v, v) = 1$ e tutte le altre definizioni che abbiamo visto nella Sezione 1.2.2 a cui rimandiamo.

Similmente a quanto fatto per le forme bilineari, fissata una base $\mathbf{v} = \{v_1, \dots, v_n\}$ di V e posto $g_{ij} = g(v_i, v_j)$, per ogni $i, j = 1, \dots, n$, abbiamo una matrice quadrata di ordine n a coefficienti complessi.

Definizione 2.2.4. La matrice $G = (g_{ij})$ è detta la *matrice della forma sesquilineare* g , rispetto alla base \mathbf{v} di V .

Grazie alla sesquilinearità di g si ha:

$$g(v, w) = g\left(\sum_{i=1}^n \lambda_i v_i, \sum_{j=1}^n \mu_j v_j\right) = \sum_{i,j=1}^n \lambda_i \bar{\mu}_j g_{ij},$$

dato che v e w sono combinazioni lineari dei vettori della base di V : $v = \sum_{i=1}^n \lambda_i v_i$ e $w = \sum_{j=1}^n \mu_j v_j$, essi sono rappresentabili anche come vettori del tipo $(\lambda_1, \dots, \lambda_n)^{\mathbf{v}}$ e $(\mu_1, \dots, \mu_n)^{\mathbf{v}}$. Possiamo quindi esprimere $g(v, w)$ in forma matriciale:

$$g(v, w) = (\lambda_1, \dots, \lambda_n)^{\mathbf{v}} G \begin{pmatrix} \bar{\mu}_1 \\ \vdots \\ \bar{\mu}_n \end{pmatrix}^{\mathbf{v}}.$$

La forma sesquilineare g è hermitiana se e solo se $g_{ij} = \bar{g}_{ji}$, per ogni $i, j = 1, \dots, n$, cioè se e solo se la matrice G è tale che $G = {}^t \bar{G}$.

Definizione 2.2.5. Data una matrice $A \in M_n(\mathbb{C})$, la matrice $A^* = {}^t\bar{A}$ è detta la *trasposta coniugata* di A . Inoltre A viene detta matrice *hermitiana* se $A^* = A$, A è detta *anti-hermitiana* se si ha $A^* = -A$.

Osservazione 2.2.6. Se $A = (a_{ij}) \in M_n(\mathbb{C})$ è una matrice hermitiana ciò significa che gli elementi sulla diagonale a_{ii} rappresentano la forma hermitiana $g(v_i, v_i)$ per qualche $v_i \in V$ e possiedono quindi la proprietà di essere numeri reali, come visto in precedenza. La diagonale principale di A contiene quindi numeri reali e in particolare la traccia di A $tr(A)$ è reale².

L'operazione di trasposizione coniugata soddisfa le seguenti proprietà, che sono del tutto identiche a quelle della trasposizione di matrici reali:

$$(A^*)^* = A, \quad (A + B)^* = A^* + B^*, \quad (AB)^* = B^* A^*,$$

osserviamo infatti ancora una volta che la trasposizione coniugata estende la trasposizione (reale) al campo dei numeri complessi.

2.2.2 Cambiamenti di base

Come precedentemente avevamo visto con le forme bilineari, ci chiediamo ora, data una forma sesquilineare g definita su uno spazio vettoriale V , complesso di dimensione finita n , come cambi la matrice di g , se cambiamo la base di V .

Siano $\mathbf{v} = \{v_1, \dots, v_n\}$ e $\mathbf{v}' = \{v'_1, \dots, v'_n\}$ due basi di V , siano inoltre $G = (g_{ij}) = M_{\mathbf{v}\mathbf{v}}(g)$ e $G' = (g'_{ij}) = M_{\mathbf{v}'\mathbf{v}'}(g)$ rispettivamente le matrici di g rispetto alle basi \mathbf{v} e \mathbf{v}' . Ciò significa che $g_{ij} = g(v_i, v_j)$ e che $g'_{ij} = g(v'_i, v'_j)$ per ogni $i, j = 1, \dots, n$.

Definiamo ora la matrice $P = M_{\mathbf{v}}^{\mathbf{v}'}(id_V)$, che rappresenta un isomorfismo di \mathbb{C}^n in sé, le cui colonne sono le componenti dei vettori v'_1, \dots, v'_n della base \mathbf{v}' calcolate rispetto alla base \mathbf{v} : possiamo quindi scrivere che per ogni vettore v'_j della base \mathbf{v}' si ha

$$v'_j = \sum_{i=1}^n p_{ij} v_i,$$

dove i vettori v_i fanno parte della base \mathbf{v} e i coefficienti p_{ij} formano la j -esima colonna di P per ogni $i = 1, \dots, n$ fissato un certo j .

La relazione tra le matrici G e G' è quindi la seguente:

$$\begin{aligned} g'_{ij} &= g(v'_i, v'_j) \\ &= g\left(\sum_{h=1}^n p_{hi} v_h, \sum_{k=1}^n p_{kj} v_k\right) \\ &= \sum_{h,k=1}^n p_{hi} \bar{p}_{kj} g(v_h, v_k) \\ &= \sum_{h,k=1}^n p_{hi} g_{hk} \bar{p}_{kj}. \end{aligned}$$

²Ricordiamo che la traccia di una matrice quadrata A di ordine n è la somma degli elementi sulla diagonale principale: $tr(A) = \sum_{i=1}^n a_{ii}$.

In termini di matrici, è possibile riscrivere quest'uguaglianza come segue:

$$G' = {}^tPG\bar{P} = \bar{P}^*G\bar{P}.$$

Definizione 2.2.7. Due matrici quadrate G e G' di ordine n a coefficienti in \mathbb{C} si dicono *congruenti* se esiste una matrice invertibile P , sempre di ordine n a coefficienti complessi tale che valga la relazione scritta in precedenza, ovvero $G' = \bar{P}^*G\bar{P}$.

Osservazione 2.2.8. Due matrici $G, G' \in M_n(\mathbb{C})$ rappresentano la stessa forma sesquilineare g definita su uno spazio vettoriale V di dimensione n su \mathbb{C} , rispetto a basi diverse, se e solo se esse sono congruenti.

Osserviamo inoltre che se $\mathbf{v} = \{v_1, \dots, v_n\}$ e $\mathbf{v}' = \{v'_1, \dots, v'_n\}$ sono entrambe basi ortonormali di V , si ha $G = G' = I_n$, perciò la formula $G' = {}^tPG\bar{P}$ si riduce a ${}^tP\bar{P} = I_n$, che equivale a ${}^t\bar{P}P = I$, cioè $P^* = {}^t\bar{P} = P^{-1}$. Questo fatto ci permette di dare la seguente definizione:

Definizione 2.2.9. Una matrice $P \in M_n(\mathbb{C})$ è detta *unitaria* se $P^*P = I$. Quindi la matrice di cambiamento di base da una base ortonormale di V ad un'altra base ortonormale di V è una matrice unitaria.

Concludiamo questa sezione illustrando alcuni risultati utili anche per la diagonalizzazione di endomorfismi nel campo complesso, in particolare per le matrici hermitiane.

Teorema 2.2.10. Sia $A \in M_n(\mathbb{C})$ una matrice hermitiana di ordine n , a coefficienti complessi. Allora gli autovalori di A sono tutti reali.

Dimostrazione. Questo teorema è molto simile al Teorema 1.2.3 e anche la dimostrazione è del tutto identica.

Sia $\lambda \in \mathbb{C}$ un autovalore di A e $v \in \mathbb{C}^n$ un autovettore corrispondente a λ . Vale dunque $Av = \lambda v$. Applichiamo ora l'operazione di trasposizione a quest'ultima uguaglianza: ${}^t(Av) = {}^t(\lambda v)$, si ottiene ${}^t v {}^t A = \lambda {}^t v$. Applichiamo l'operazione di coniugio e ricaviamo:

$${}^t \bar{v} {}^t \bar{A} = \bar{\lambda} {}^t \bar{v}.$$

Dato che A è hermitiana e quindi $A = {}^t \bar{A}$, possiamo riscrivere l'uguaglianza nel seguente modo, moltiplicando inoltre entrambi i membri a destra per v :

$${}^t \bar{v} Av = \bar{\lambda} {}^t \bar{v} v,$$

ovvero

$${}^t \bar{v} \lambda v = \bar{\lambda} {}^t \bar{v} v,$$

da cui otteniamo infine che $(\bar{\lambda} - \lambda) {}^t \bar{v} v = 0$. Ricordando che ${}^t \bar{v} v = \|v\|^2 > 0$, si ha che $\bar{\lambda} = \lambda$, cioè $\lambda \in \mathbb{R}$. \square

Definizione 2.2.11. Sia V uno spazio vettoriale definito sul campo \mathbb{C} . Una funzione lineare $f: V \rightarrow V$ è detta *hermitiana* se

$$f(v) \cdot w = v \cdot f(w)$$

per ogni $v, w \in V$. La definizione scritta sopra è del tutto simile a quella per le funzioni simmetriche (si veda Definizione 1.2.9). Inoltre un endomorfismo è hermitiano se e solo se la matrice di f rispetto ad una base ortonormale di V è hermitiana.

Teorema 2.2.12. (TEOREMA SPETTRALE PER MATRICI COMPLESSE). *Siano V uno spazio vettoriale complesso e $f: V \rightarrow V$ una funzione lineare. Allora f possiede solo autovalori reali ed è unitariamente diagonalizzabile, cioè esiste una base ortonormale formata da autovettori di f , se e solo se f è hermitiana. Di conseguenza, data una matrice hermitiana a coefficienti complessi, essa può sempre essere trasformata in una matrice diagonale tramite una matrice di cambiamento di base unitaria.*

Dimostrazione. Supponiamo che f sia unitariamente diagonalizzabile, cioè che esista una base ortonormale di V fatta di autovettori di f . La matrice rispetto a tale base è una matrice diagonale (con autovalori reali) e quindi hermitiana per il Teorema 2.2.10. Ciò implica che f è hermitiana.

Viceversa, dimostriamo per induzione che ogni funzione hermitiana è unitariamente diagonalizzabile. Se $\dim V = 1$, sia v_1 una base ortonormale di V , si ha dunque $f(v_1) = \lambda_1 v_1$, ovvero v_1 è un autovettore di f e $\lambda_1 \in \mathbb{R}$ essendo la funzione hermitiana.

Supponiamo che V abbia dimensione $n > 1$ e che l'enunciato sia vero per spazi vettoriali di dimensione $n - 1$. Presa una base ortonormale $\{u_1, \dots, u_n\}$ di V , la matrice A di f rispetto a tale base è hermitiana per ipotesi, perciò possiede autovalori reali (sempre dal Teorema 2.2.10).

Sia λ_1 uno degli autovalori di f e v_1 un autovettore corrispondente, si prenda $W = \langle v_1 \rangle^\perp$. Dato che f è hermitiana, vale infatti che $f(v_1) \cdot w = \lambda_1 v_1 \cdot w = 0 = v_1 \cdot f(w)$, dunque $f(w) \in \langle v_1 \rangle^\perp = W$ per ogni $w \in W$. La funzione $f|_W: W \rightarrow W$ ristretta a W è anch'essa una funzione lineare hermitiana con $\dim W = n - 1$.

Per ipotesi induttiva essa è unitariamente diagonalizzabile tramite una base ortonormale di W $\{w_2, \dots, w_n\}$ di autovettori di $f|_W$. Basta quindi porre $w_1 = v_1 / \|v_1\|$ ed osservare che i vettori w_1, \dots, w_n formano una base ortonormale di V costituita da autovettori di f . \square

2.3 Decomposizione in valori singolari per matrici complesse

Nella trattazione della decomposizione in valori singolari nel caso complesso faremo uso delle nozioni date dalla Sezione 2.2, le quali costituiscono le differenze principali con l'analogo del caso reale.

Siano dunque V e W due spazi vettoriali definiti sul campo \mathbb{C} , dotati di forme hermitiane definite positive g e h , rispettivamente, e sia $f: V \rightarrow W$ una funzione lineare. In ogni spazio vettoriale complesso dotato di forma hermitiana definita positiva esistono basi ortonormali (proprio come negli spazi vettoriali reali) perciò, essendo V e W isometrici a \mathbb{C}^n e \mathbb{C}^m

rispettivamente, possiamo considerare la funzione $f: \mathbb{C}^n \rightarrow \mathbb{C}^m$, per qualche n, m , senza perdita di generalità. Possiamo dunque utilizzare il prodotto scalare usuale (hermitiano) definito per gli spazi vettoriali complessi.

Indichiamo con $A = M_{\mathbf{e}}^{\mathbf{e}}(f)$ la matrice di f rispetto alle basi canoniche, A ha quindi m righe e n colonne.

Definiamo la *trasposta* di f , $f^*: \mathbb{C}^m \rightarrow \mathbb{C}^n$, la quale, come nel caso reale, soddisfa la seguente uguaglianza:

$$v \cdot f^*(w) = f(v) \cdot w,$$

per ogni $v \in \mathbb{C}^n$ e ogni $w \in \mathbb{C}^m$. A differenza però del caso reale, la matrice della funzione f^* rispetto alle basi canoniche è la *trasposta coniugata* della matrice A di f : $M_{\mathbf{e}}^{\mathbf{e}}(f^*) = {}^t \bar{A} = A^*$.

Indichiamo con $B = (b_{ij})$ la matrice di f^* rispetto alle basi canoniche (ortonormali), che invece indichiamo con $\mathbf{e}^{(n)}(e_1, \dots, e_n)$ e $\mathbf{e}^{(m)}(e_1, \dots, e_m)$ rispettivamente, vogliamo dimostrare che $B = A^*$.

Si ha $f^*(e_j) = \sum_{l=1}^n b_{lj} e_l$, per ogni $j = 1, \dots, m$. Vale quindi $e_i \cdot f^*(e_j) = \bar{b}_{ij}$, ricordando che $e_i \cdot e_j = 0 \forall i \neq j$, $e_i \cdot e_i = 1$ e infine vale l'antilinearità del prodotto scalare, ovvero $e_i \cdot \lambda e_i = \bar{\lambda}$.

Analogamente, $f(e_i) = \sum_{l=1}^m a_{li} e_l$, per ogni $i = 1, \dots, n$ e si ottiene $f(e_i) \cdot e_j = a_{ji}$.

Dalla definizione di funzione trasposta, risulta infine che $e_i \cdot f^*(e_j) = f(e_i) \cdot e_j$, cioè $\bar{b}_{ij} = a_{ji}$, ovvero $b_{ij} = \bar{a}_{ji}$, che in forma matriciale equivale a $B = A^*$.

Consideriamo ora le funzioni composte

$$f^* \circ f: \mathbb{C}^n \rightarrow \mathbb{C}^n, \quad f \circ f^*: \mathbb{C}^m \rightarrow \mathbb{C}^m,$$

le cui matrici, rispetto alle basi canoniche, sono A^*A e AA^* , rispettivamente. La matrice A^*A è una matrice quadrata di ordine n , hermitiana poiché si ha $(A^*A)^* = A^*A$. Analogamente, la matrice AA^* è quadrata, di ordine m ed è hermitiana. Entrambe le matrici sono semidefinite positive, come nel caso reale.

Dato che la matrice A^*A è hermitiana, essa è unitariamente diagonalizzabile, cioè esiste una base ortonormale di \mathbb{C}^n formata da autovettori di A^*A . Siccome A^*A è semidefinita positiva, i suoi autovalori (tutti reali) sono ≥ 0 .

Come nel caso reale vale che $\text{Ker} A = \text{Ker}(A^*A)$, indichiamo inoltre con r il rango di A , perciò $\dim(\text{Ker} A) = n - r$, si deduce che la matrice A^*A ha un autovalore $\lambda = 0$ di molteplicità $n - r$ e il corrispondente autospazio è quindi il nucleo di A (o di f).

Sia quindi \mathbf{v} una base ortonormale di \mathbb{C}^n formata da autovettori di A^*A , dove $\mathbf{v} = \{v_1, \dots, v_r, u_1, \dots, u_{n-r}\}$, tali che $(A^*A)v_i = \lambda_i v_i$, dove $\lambda_1, \dots, \lambda_r$ sono gli autovalori non nulli di A^*A , e $(A^*A)u_j = \mathbf{0}$ per $j = 1, \dots, n-r$. I vettori u_1, \dots, u_{n-r} formano una base ortonormale di $\text{Ker} f$, mentre i vettori v_1, \dots, v_r costituiscono una base ortonormale di $V' = (\text{Ker} f)^\perp$. Lo spazio vettoriale V si decompone nella somma diretta $V = V' \oplus \text{Ker} f = (\text{Ker} f)^\perp \oplus \text{Ker} f$. Applicando la funzione f ai vettori della base \mathbf{v} si ottiene:

$$f(v_1) = w_1, \dots, f(v_r) = w_r, f(u_1) = \mathbf{0}, \dots, f(u_{n-r}) = \mathbf{0},$$

dove $\text{Im} f = \langle w_1, \dots, w_r \rangle$. Dimostriamo ora che i vettori w_1, \dots, w_r sono a due a due ortogonali (nel prodotto scalare hermitiano), come nel caso reale.

$$\begin{aligned} w_i \cdot w_j &= f(v_i) \cdot f(v_j) = (Av_i) \cdot (Av_j) = {}^t(Av_i)(\bar{A}v_j) = {}^t v_i {}^t A \bar{A} v_j \\ &= {}^t v_i {}^t \overline{AA} v_j = {}^t v_i \overline{\lambda_j} v_j = \bar{\lambda}_j v_i \cdot v_j = 0 \end{aligned}$$

2.3. DECOMPOSIZIONE IN VALORI SINGOLARI PER MATRICI COMPLESSE 25

poiché v_j è un autovettore di A^*A relativo all'autovalore λ_j , i vettori v_i e v_j sono ortogonali per ipotesi e abbiamo utilizzato il fatto che ${}^t A \bar{A} v_j = \overline{{}^t A A} v_j$ sfruttando l'operazione di coniugio.

Possiamo quindi normalizzare i vettori w_i , ponendo $\hat{w}_i = w_i / \|w_i\|$, così da ottenere una base ortonormale $\{\hat{w}_1, \dots, \hat{w}_r\}$ di Imf . Consideriamo poi una base ortonormale $\{z_1, \dots, z_{m-r}\}$ di $(Imf)^\perp$ e otteniamo perciò una base ortonormale di \mathbb{C}^m , $\mathbf{w} = \{\hat{w}_1, \dots, \hat{w}_r, z_1, \dots, z_{m-r}\}$. Si ottiene quindi il seguente risultato di cui non verrà data la dimostrazione, essendo del tutto analoga quella fatta nel caso reale (si veda la Proposizione 2.1.4):

Proposizione 2.3.1. *La matrice $M_{\mathbf{v}}^{\mathbf{w}}(f)$ di f , rispetto alle basi \mathbf{v} di \mathbb{C}^n e \mathbf{w} di \mathbb{C}^m , è una matrice a blocchi del tipo*

$$D = M_{\mathbf{v}}^{\mathbf{w}}(f) = \left(\begin{array}{cccc|ccc} \alpha_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & \alpha_r & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{array} \right)$$

dove gli α_i sono le radici quadrate (positive) degli autovalori non nulli λ_i di A^*A : $\alpha_i = \sqrt{\lambda_i}$.

Gli elementi sulla diagonale della matrice D sono i valori singolari della funzione f , o della matrice A .

Le matrici $A = M_{\mathbf{e}}^{\mathbf{e}}(f)$ e $D = M_{\mathbf{v}}^{\mathbf{w}}(f)$ rappresentano quindi entrambe la stessa funzione lineare f rispetto a basi diverse del dominio e del codominio. Le matrici di cambiamento di base sono: in \mathbb{C}^n la matrice $P = M_{\mathbf{v}}^{\mathbf{e}}(id)$, le cui colonne sono le coordinate dei vettori della base ortonormale \mathbf{v} scritte rispetto alla base canonica di \mathbb{C}^n ; in \mathbb{C}^m la matrice $Q = M_{\mathbf{w}}^{\mathbf{e}}(id)$, le cui colonne sono le coordinate dei vettori della base ortonormale \mathbf{w} scritte rispetto alla base canonica di \mathbb{C}^m . P e Q sono dunque matrici unitarie e si ha $P^{-1} = P^*$ e $Q^{-1} = Q^*$. Dalla formula del cambiamento di base

$$M_{\mathbf{e}}^{\mathbf{e}}(f) M_{\mathbf{v}}^{\mathbf{e}}(id) = M_{\mathbf{w}}^{\mathbf{e}}(id) M_{\mathbf{v}}^{\mathbf{w}}(f)$$

segue che $AP = QD$ e quindi

$$A = QDP^{-1} = QDP^*, \quad D = Q^{-1}AP = Q^*AP.$$

Queste formule rappresentano la decomposizione in valori singolari della matrice A .

Capitolo 3

Applicazioni della SVD

In questo capitolo analizzeremo alcuni dei risultati più importanti che la SVD ha portato al mondo dell'Ingegneria, quali la generalizzazione della matrice inversa, anche chiamata pseudoinversa, e la cosiddetta approssimazione di rango basso (low-rank approximation). Vedremo poi delle applicazioni reali di questi due potenti strumenti.

3.1 La matrice pseudoinversa

La matrice pseudoinversa viene anche detta matrice inversa di *Moore-Penrose* grazie agli omonimi matematici che l'hanno per primi descritta nel XX secolo, indipendentemente l'uno dall'altro. Essa esiste ed è unica per qualsiasi matrice, che sia essa definita nel campo \mathbb{R} dei numeri reali o nel campo \mathbb{C} dei numeri complessi. Riveste grande importanza in quanto, tra le altre cose, permette di trovare la soluzione "più adatta" in un sistema di equazioni lineari che non ha una soluzione (minimi quadrati).

Definizione 3.1.1. Sia A una matrice di m righe e n colonne definita su un campo K , la matrice pseudoinversa A^+ è una matrice definita sullo stesso campo K , ha n righe e m colonne. Inoltre soddisfa le seguenti 4 condizioni, anche dette condizioni di *Moore-Penrose*:

- (1) AA^+ non è necessariamente la matrice identica ma l'applicazione di AA^+ al vettore Av ($v \in K^m$ vettore arbitrario) è uguale all'applicazione della sola matrice A sul vettore v (un altro modo per vederla è dire che l'applicazione di AA^+ dopo di A mappa le colonne di A su loro stesse):

$$AA^+A = A.$$

- (2) A^+ è una inversa debole, il che significa:

$$A^+AA^+ = A^+.$$

- (3) AA^+ è hermitiana, quindi $(AA^+)^* = AA^+$.

- (4) A^+A è anch'essa hermitiana, quindi $(A^+A)^* = A^+A$.

La matrice A^+ esiste sempre ma, nel caso in cui la matrice A abbia rango pieno (il che equivale a dire che il suo rango è $\min\{m, n\}$), allora c'è una semplice formula.

Se A ha le colonne linearmente indipendenti (quindi la matrice A^*A è invertibile) si ha $A^+ = (A^*A)^{-1}A^*$, in questo caso è facile vedere che $A^+A = I$ e A^+ viene chiamata infatti “inversa sinistra”.

Se A invece ha le righe linearmente indipendenti (quindi la matrice AA^* è invertibile) si ha $A^+ = A^*(AA^*)^{-1}$, anche in questo caso è facile vedere che vale $AA^+ = I$ e A^+ viene chiamata “inversa destra”.

Ciò che però è più interessante è vedere come la pseudoinversa sia facilmente calcolabile tramite la decomposizione in valori singolari della matrice A , il cui procedimento non dipende dal rango (pieno) della matrice stessa, ma è valido per qualsiasi matrice. Prima di descrivere tale metodo di calcolo, diamo qualche altra proprietà della pseudoinversa, anche chiamata inversa generalizzata.

- Se A è invertibile, la pseudoinversa equivale all'inversa, cioè $A^+ = A^{-1}$.
- La pseudoinversa della pseudoinversa corrisponde alla matrice originale, ovvero $(A^+)^+ = A$.
- La pseudoinversa inoltre commuta con la trasposizione, con la coniugazione complessa e con l'operazione di trasposta coniugata:

$$({}^tA)^+ = {}^t(A^+) \quad (\overline{A})^+ = \overline{A^+} \quad (A^*)^+ = (A^+)^*.$$

- Se A è una matrice quadrata ($n = m$) diagonale del tipo $A = \text{diag}(v)$, con $v = (a_1, \dots, a_n)$, allora A^+ è ancora una matrice diagonale i cui elementi sono i reciproci delle componenti non nulle di v , altrimenti valgono 0, ovvero $A^+ = \text{diag}(w)$, dove w è un vettore le cui componenti sono i reciproci di quelle di v esclusi i valori nulli, i quali sono nulli anche in w .
- Se A è una matrice rettangolare di dimensioni $m \times n$ diagonale (dato un arbitrario vettore $v \in K^{\min\{m, n\}}$, la matrice A è del tipo $\text{diag}(v)$ dove i valori di v occupano nella matrice gli elementi a_{ii} per $i = 1, \dots, \min\{m, n\}$), la matrice A^+ è ancora diagonale di dimensioni $n \times m$ e si ottiene, similmente al caso precedente, facendo i reciproci degli elementi non nulli in a_{ii} (e quindi in v) della matrice A e operando una trasposizione della risultante matrice.
- La pseudoinversa della matrice nulla è la sua trasposta.
- La seguente proprietà riguarda invece la pseudoinversa di una matrice moltiplicata per uno scalare $\alpha \neq 0$:

$$(\alpha A)^+ = \alpha^{-1}A^+$$
- L'uguaglianza $(AB)^+ = B^+A^+$ non è vera in generale, ma vi sono condizioni che la rendono valida ([7]).

Vediamo adesso qualche caso speciale di pseudoinversa:

Scalari Si può definire la pseudoinversa di uno scalare x , che come è facile immaginare è il reciproco se lo scalare non è zero, vale zero altrimenti:

$$x^+ = \begin{cases} 0 & \text{se } x = 0 \\ x^{-1} & \text{altrimenti} \end{cases}$$

Vettori Anche la pseudoinversa di un vettore x è definita ed equivale al vettore a cui viene applicata l'operazione di trasposta coniugata diviso per la norma al quadrato:

$$x^+ = \begin{cases} {}^t\mathbf{0} & \text{se } x = \mathbf{0} \\ \frac{x^*}{x^*x} & \text{altrimenti} \end{cases}$$

Colonne o righe ortonormali Se la matrice A possiede rango pieno e ha le righe ortonormali ($AA^* = I$) o le colonne ortonormali ($A^*A = I$) allora vale $A^+ = A^*$.

Mostriamo adesso la formula per il calcolo della pseudoinversa tramite la SVD, essa è molto utile in quanto semplice e indipendente dalla struttura della matrice originale.

Senza perdita di generalità chiamiamo $r \leq \min\{m, n\}$ il rango della matrice A . Sappiamo inoltre che la decomposizione in valori singolari di una generica matrice A definita nel campo dei numeri complessi è la seguente: $A = QDP^*$, dove Q e P sono le solite matrici del cambiamento di base, mentre D è una matrice con r valori diversi da 0 nella diagonale principale (si veda la Sezione 2.3). Si può quindi dimostrare che $A^+ = PD^+Q^*$ verifica tutte le proprietà della pseudoinversa, perciò la definizione è valida.

Vediamo ora un paio di esempi di calcolo di pseudoinversa, sfruttando la SVD.

Esempio 3.1.2. Consideriamo la matrice A dell'esempio 2.1.7, di cui avevamo già calcolato la decomposizione in valori singolari D e le relative matrici di cambiamento di base P e Q . Cerchiamo ora di trovare la sua inversa generalizzata. Ricordiamo che

$$A = \begin{pmatrix} 2 & 1 & 1 \\ -1 & 2 & 2 \end{pmatrix} \quad D = \begin{pmatrix} \sqrt{10} & 0 & 0 \\ 0 & \sqrt{5} & 0 \end{pmatrix}$$

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix} \quad Q = \begin{pmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{pmatrix}.$$

Dalle proprietà viste prima sappiamo che $D^+ = \begin{pmatrix} 1/\sqrt{10} & 0 \\ 0 & 1/\sqrt{5} \\ 0 & 0 \end{pmatrix}$. Applicando ora la formula definita in precedenza, ovvero $A^+ = PD^+Q^*$, troviamo:

$$A^+ = \begin{pmatrix} 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{10} & 0 \\ 0 & 1/\sqrt{5} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{pmatrix} = \begin{pmatrix} 2/5 & -1/5 \\ 1/10 & 1/5 \\ 1/10 & 1/5 \end{pmatrix},$$

dove $Q^* = {}^tQ$, dato che ci troviamo in un caso reale. Si possono effettivamente verificare tutte le proprietà della pseudoinversa, qui ne viene mostrata una, le altre vengono lasciate come esercizio al lettore. Dimostriamo che $AA^+A = A$, infatti

$$\begin{pmatrix} 2 & 1 & 1 \\ -1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 2/5 & -1/5 \\ 1/10 & 1/5 \\ 1/10 & 1/5 \end{pmatrix} \begin{pmatrix} 2 & 1 & 1 \\ -1 & 2 & 2 \end{pmatrix}$$

$$= I \begin{pmatrix} 2 & 1 & 1 \\ -1 & 2 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 \\ -1 & 2 & 2 \end{pmatrix} = A.$$

Esempio 3.1.3. Nell'esempio precedente la matrice aveva rango pieno, in quanto uguale a 2, cioè a una delle dimensioni della matrice; prendiamo adesso una matrice non dotata di rango pieno, ad esempio $A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ e cerchiamo di calcolarne la pseudoinversa.

Prima di tutto ricaviamo la SVD, cercando gli autovalori della matrice ${}^tAA = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$, i quali sono $\lambda_1 = 2$ e $\lambda_2 = 0$. I relativi autovettori sono: $v_1 = (1, 0)$ per λ_1 e $v_2 = (0, 1)$ per λ_2 , essi sono già normalizzati quindi formano una base ortonormale di \mathbb{R}^2 e la matrice P è quindi la matrice identica $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Per calcolare la matrice Q serve normalizzare $Av_1 = (1, 1)$ ottenendo il vettore $(1/\sqrt{2}, 1/\sqrt{2})$ mentre $Av_2 = (0, 0)$ (il rango di A vale 1 e l'immagine dell'applicazione lineare è infatti un sottospazio vettoriale di dimensione 1). Serve quindi completare la base di \mathbb{R}^2 aggiungendo un vettore normalizzato e ortogonale al vettore $(1/\sqrt{2}, 1/\sqrt{2})$: una possibile scelta è il vettore $(-1/\sqrt{2}, 1/\sqrt{2})$. La matrice Q è quindi uguale a $\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$, poiché la matrice del cambio di base rappresenta comunque una mappa lineare il cui dominio e codominio sono entrambi \mathbb{R}^2 ma in questo caso la funzione data dalla matrice A possiede un nucleo non triviale e quindi il solo vettore $(1/\sqrt{2}, 1/\sqrt{2})$ forma una base ortonormale dell'immagine della funzione lineare. Sfruttando la formula della SVD, troviamo che $D = {}^tQAP = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{pmatrix}$. Anche in questo caso D^+ è facile da calcolare, infatti vale $\begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & 0 \end{pmatrix}$, perciò la pseudoinversa della matrice A si calcola con la stessa formula dell'esempio precedente (anche in questo caso $Q^* = {}^tQ$):

$$A^+ = PD^+Q^* = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ 0 & 0 \end{pmatrix}.$$

Anche in questo esempio è possibile verificare correttamente le 4 proprietà della pseudoinversa.

Diamo ora, similmente a quanto fatto con la decomposizione in valori singolari, una interpretazione geometrica della pseudoinversa. Usando gli stessi termini della SVD (si veda il Capitolo 2), la matrice A rappresenta una funzione lineare $f: V \rightarrow W$, dove V, W sono spazi vettoriali euclidei (se definiti nel campo dei reali, dotati di forme hermitiane definite positive se definiti nel campo dei complessi). La pseudoinversa è una funzione lineare $f^+: W \rightarrow V$ che esegue il processo inverso della SVD. Infatti essa è data dalla composizione di alcune funzioni: prima la proiezione ortogonale di W nel sottospazio W' , a cui segue la composizione con la funzione inversa di \tilde{f} , ovvero $\tilde{f}^{-1}: W' \rightarrow V'$, infine si compone con l'inclusione di V' in V :

$$f^+: W \rightarrow W' \xrightarrow{\tilde{f}^{-1}} V' \hookrightarrow V.$$

La funzione \tilde{f}^{-1} è esattamente l'inversa di \tilde{f} poiché avevamo visto che quest'ultima rappresenta un isomorfismo ed è quindi invertibile. Chiamati α_i i valori singolari per $i = 1, \dots, r$ con r il rango della matrice A , la matrice di \tilde{f}^{-1} rispetto alle basi che avevamo scelto nel caso della decomposizione in valori singolari è la seguente:

$$\begin{pmatrix} 1/\alpha_1 & 0 & \cdots & 0 \\ 0 & 1/\alpha_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\alpha_r \end{pmatrix}.$$

Descriviamo ora come viene utilizzata la costruzione della matrice pseudo-inversa in un recente lavoro di E. Arnese-Feffin, P. Facco, F. Bezzo e M. Barolo [1].

In questo articolo viene trattato il problema dell'individuazione delle condizioni operative e dell'utilizzo delle materie prime (gli input) in modo da raggiungere delle specifiche di qualità predeterminate all'interno di un processo di produzione. Daremo qui una breve descrizione del metodo proposto dagli autori per risolvere tale problema, il quale sfrutta la matrice inversa generalizzata.

Sia $X \in \mathbb{R}^{N \times V_X}$ la matrice di input composta da N righe e V_X colonne, la quale raccoglie N misure di V_X variabili, che in questo caso vengono associate alle condizioni iniziali del processo di produzione controllabili a priori.

Sia invece $Y \in \mathbb{R}^{N \times V_Y}$ la matrice di output composta da N righe e V_Y colonne, che invece contiene le N misure di V_Y variabili corrispondenti alle specifiche di qualità del prodotto. A questo punto gli autori propongono di sfruttare la PLS (Partial Least-Squares Regression), che in sintesi permette di stabilire una relazione lineare tra le matrici X e Y costruendo delle basi ortonormali di opportuni sottospazi di $\mathbb{R}^N \times \mathbb{R}^{V_X}$ e $\mathbb{R}^N \times \mathbb{R}^{V_Y}$ e proiettando le matrici X e Y su questi sottospazi; ciò che si ottiene sono delle matrici T e U , che sono legate a X e Y dalle seguenti relazioni:

$$X = T^t P + E, \quad Y = U^t Q + F,$$

dove le matrici P , Q , E , F provengono anch'esse dalla PLS.

Usando questo metodo il modello matematico che si ottiene stabilisce una relazione lineare diretta tra le variabili in input e quelle in output. Difatti, si costruisce una matrice W^* tale per cui, data una osservazione in input $x_{new} \in \mathbb{R}^{V_X}$, si può ottenere il valore del vettore di qualità in output $\hat{y}_{new} \in \mathbb{R}^{V_Y}$ tramite la formula

$${}^t \hat{y}_{new} = {}^t \hat{x}_{new} (W^* {}^t Q),$$

dove si impone $B = W^* {}^t Q \in \mathbb{R}^{V_X} \times \mathbb{R}^{V_Y}$, la quale viene chiamata matrice dei coefficienti di regressione.

Il problema che quindi gli autori si sono posti è quello di risolvere il sistema

$${}^t \hat{x}_{new} B = {}^t \hat{y}_{new},$$

nello specifico di trovare le condizioni iniziali adatte \hat{x}_{new} in modo che la qualità del prodotto \hat{y}_{new} sia quella desiderata. Ciò significa invertire la matrice B , che però, in generale,

è una matrice rettangolare e quindi non invertibile (inoltre la matrice B potrebbe non avere rango pieno).

Tra i possibili metodi per risolvere tale sistema (si rimanda all'articolo [1] per ulteriori approfondimenti) si distinguono diversi casi a seconda che la matrice Y abbia rango massimo o meno, ovvero che le misurazioni delle variabili di qualità siano linearmente indipendenti o meno. Nel metodo algebrico, serve invertire la matrice tQQ , che è invertibile quando le variabili di qualità non sono correlate tra loro.

Nel caso le variabili di qualità fossero correlate fra loro, una possibile soluzione è quella di trascurare alcune variabili e quindi alcune colonne della matrice Y , ma così facendo non è garantito che il vettore \hat{y}_{new} abbia i valori desiderati alla fine del processo. Gli autori propongono un approccio diverso che include tutte le variabili di output e usa la SVD per decomporre la matrice $S = {}^tQQ$. Ricordiamo infatti che vale la seguente formula:

$$S = LD{}^tM = \begin{pmatrix} L_1 & L_2 \end{pmatrix} \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix} {}^t \begin{pmatrix} M_1 & M_2 \end{pmatrix},$$

in cui le matrici L e M sono le matrici di cambiamento di base, D_1 contiene i valori singolari diversi da 0, mentre D_2 è la matrice nulla.

Uguualmente a quello che succede nel caso della pseudoinversa di *Moore-Penrose*, gli autori propongono di invertire solo D_1 eseguendo il reciproco dei soli valori singolari diversi da 0, ottenendo una inversa generalizzata:

$$S_{reg}^{-1} = M_1 D_1^{-1} {}^t L_1.$$

L'inversa $S^{-1} = MD^{-1}{}^tL$ infatti non esiste e quindi quest'ultima formula non è applicabile.

Nella stragrande maggioranza dei casi, accade che le variabili di qualità siano quasi correlate, ciò significa che le colonne di Y sono "quasi" dipendenti, perciò la matrice $S = {}^tQQ$ risulta invertibile ma alcuni suoi valori singolari hanno modulo molto vicino a zero. Quando si calcola S^{-1} bisogna invertire i valori singolari e si possono quindi ottenere valori molto grandi, la qual cosa può generare errori nella ricerca della soluzione del sistema. La tecnica che risolve questo problema consiste nel trascurare tutti i valori singolari $< \varepsilon$ (dove ε è un parametro che viene fissato a priori) e invertire i rimanenti valori singolari. Naturalmente in questo modo viene persa una quantità di informazione che può essere stimata dalla somma dei valori singolari trascurati.

3.2 Approssimazione di rango basso

L'approssimazione di rango basso è un problema di minimizzazione, dove lo scopo è approssimare nel modo migliore possibile una matrice data con una matrice di rango inferiore. È usato principalmente nella modellazione matematica e nella compressione dei dati.

Questo problema ha una soluzione analitica che si basa sulla SVD e costituisce in realtà un risultato importante: il **teorema di Eckart-Young-Mirsky**.

In termini matematici, data una matrice M , si tratta di minimizzare la seguente funzione:

$$\|M - \hat{M}\|_F,$$

dove \hat{M} è una matrice con il requisito aggiuntivo che $\text{rank}(\hat{M}) \leq r$ per un certo r , mentre il pedice ‘ F ’ si riferisce alla norma di Frobenius¹.

Dunque la decomposizione in valori singolari di M è $M = UD^tV \in \mathbb{R}^{m \times n}$, con $m \geq n$ senza perdita di generalità e indichiamo le matrici U , $D = \text{diag}(\sigma_1, \dots, \sigma_n)$ e V come segue:

$$U = (U_1 \ U_2), \quad D = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}, \quad V = (V_1 \ V_2),$$

dove $D_1 = \text{diag}(\sigma_1, \dots, \sigma_r)$ è una matrice $r \times r$ che contiene gli r valori singolari di D con valore assoluto maggiore, mentre U_1 è una matrice $m \times r$ e V_1 è una matrice $n \times r$.

Si calcola infine:

$$\hat{M}^* = U_1 D_1^t V_1,$$

la quale è una matrice di rango r . Il teorema di Eckart-Young-Mirsky afferma che tale matrice \hat{M}^* minimizza la funzione di costo

$$\|M - \hat{M}^*\|_F = \min_{\text{rank}(\hat{M}) \leq r} \|M - \hat{M}\|_F = \sqrt{\sigma_{r+1}^2 + \dots + \sigma_n^2}.$$

Per maggiori dettagli si veda [6].

Descriviamo ora come viene utilizzata l’approssimazione di rango basso nell’articolo [4]. In questo lavoro gli autori affrontano il problema di ridurre gli effetti del rumore presente nei dati rilevati da un array di sensori di un “naso” elettronico, che consiste in un sistema di riconoscimento di gas e odori anche complessi; il procedimento che viene presentato sfrutta l’approssimazione di rango basso, come in breve vedremo.

Si consideri un array di N sensori che analizza M campioni di gas: i dati vengono raccolti in una matrice X di M righe e N colonne. Sia inoltre r il rango di tale matrice, dove solitamente si ha $M \geq N$. Se analizzassimo lo stesso gas M volte ci aspetteremmo, in condizioni ideali, una matrice con le righe tutte uguali e quindi rango $r = 1$; sfortunatamente questo non succede a causa del rumore elettronico, delle piccole variazioni nella composizione chimica del gas ed eventuali miscugli con altri gas. Il rango r , seppur ≥ 1 , non dovrebbe essere molto grande, perciò il problema che si sono posti gli autori è quello di capire se c’è un modo per ottenere le informazioni essenziali (e quindi riconoscere uno o più gas correttamente) nonostante l’influenza del rumore all’interno della matrice X .

Gli autori propongono di usare la SVD per decomporre la matrice X :

$$X = U\Sigma^tV$$

dove U e V sono le matrici del cambiamento di base e $\Sigma \in \mathbb{R}^{M \times N}$ è una matrice diagonale i cui elementi sono gli N valori singolari di X disposti in ordine decrescente

¹Essendo lo spazio delle matrici anch’esso uno spazio vettoriale, si può definire una norma anche per una matrice A di dimensioni $m \times n$, chiamata norma di Frobenius. Essa corrisponde alla radice quadrata della somma dei moduli al quadrato di tutti gli elementi appartenenti ad essa:

$$\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2(A)},$$

dove σ_i sono i valori singolari di A .

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N.$$

Indichiamo con u_i le colonne di U e con v_i le colonne di V . L'uguaglianza $X = U\Sigma^tV$ equivale quindi alla seguente:

$$X = \sum_{i=1}^N \sigma_i u_i^t v_i.$$

Il prodotto $u_i^t v_i$ è una matrice di rango 1 e quindi X viene espressa come una somma di matrici di rango 1 con “pesi” dati dai valori singolari σ_i . L'idea proposta dagli autori è quella di troncatura la sommatoria ad un certo $K < N$ e ottenere una matrice di rango K che approssima la matrice X considerando solo i primi K valori singolari: $X_K = \sum_{i=1}^K \sigma_i u_i^t v_i$. Questa idea è, nella pratica, l'uso dell'approssimazione di rango basso per portare la matrice originale ad un rango K più basso.

Si fissa di solito un valore di soglia ε e si trascurano i valori singolari $< \varepsilon$. Le informazioni più significative presenti nella matrice X dovrebbero infatti essere presenti nei valori singolari più grandi (quelli $> \varepsilon$) mentre il rumore, che viene solitamente modellato come una variabile casuale, si distribuisce in tutte le direzioni (ha valore medio nullo nella maggior parte dei casi e una varianza fissata). Così facendo, l'informazione contenuta nella rilevazione rimane all'incirca la stessa, mentre il rumore ha un effetto molto inferiore, perciò il cosiddetto rapporto segnale-rumore, indice della bontà di un sistema di elaborazione dei dati e che mette in relazione la potenza del segnale utile rispetto a quella del rumore, aumenta.

3.3 Riconoscimento di movimenti e gesti umani tramite SVD

Infine, esponiamo sinteticamente in questa sezione un'altra applicazione della SVD, presentata nell'articolo [5], che si basa sull'analisi e sul riconoscimento di movimenti del corpo umano.

Gli autori di questo lavoro presentano un metodo per identificare gesti e il livello di disabilità motoria costruendo delle matrici partendo dai dati rilevati e applicando la decomposizione in valori singolari.

Vengono misurate le posizioni in tutte e tre gli assi x, y, z di m punti del corpo p_i , $i = 1, \dots, m$, mentre un soggetto esegue un determinato movimento. Le misure sono ripetute in n istanti successivi t_j , $j = 1, \dots, n$, così da ottenere una sequenza temporale di dati per ogni punto. Si costruiscono pertanto tre matrici

$$M_X = (x_{i,j}) \quad M_Y = (y_{i,j}) \quad M_Z = (z_{i,j}),$$

di dimensioni $m \times n$, che contengono tutte le informazioni sul movimento del soggetto.

A questo punto si esegue la SVD della matrice M_X attraverso la formula: $M_X = U_X \Sigma_X^t V_X$, dove Σ_X è una matrice $m \times n$ che contiene i valori singolari $\sigma_1, \dots, \sigma_{\min\{m,n\}}$ in ordine decrescente di valore assoluto ($\sigma_1 \geq \dots \geq \sigma_{\min\{m,n\}}$) sulla diagonale principale e 0 altrove, mentre U_X è una matrice $m \times m$ le cui colonne corrispondono ai vettori singolari sinistri. Il procedimento viene ripetuto anche per le matrici M_Y e M_Z .

L'idea che gli autori hanno proposto è la seguente: i vettori singolari sinistri delle tre

matrici M_X, M_Y, M_Z esprimono le caratteristiche di tutto il movimento in misura direttamente proporzionale al corrispondente valore singolare e permettono di identificare cenni semplici della mano e il livello di disabilità di una persona nel camminare (si consulti l'articolo [5] per maggiori informazioni sugli esperimenti).

In breve, se il valore singolare aumenta, ciò influenza la caratteristica del movimento espressa dal vettore singolare sinistro che diviene più preponderante. Questo approccio viene usato nei due esperimenti per il riconoscimento automatico dei gesti della mano e nell'individuazione del livello di difficoltà motoria nel camminare discussi nell'articolo e i risultati ottenuti suggeriscono che la SVD rappresenti un metodo efficace per estrarre le informazioni essenziali relative al movimento acquisite dalla serie temporale dei dati.

Bibliografia

- [1] ARNESE-FEFFIN E., FACCO P., BEZZO F., BAROLO M.: Digital design of new products: accounting for output correlation via a novel algebraic formulation of the latent-variable model inversion problem, *Chemometrics and Intelligent Laboratory Systems*, 227 (2022), 104610. <https://doi.org/10.1016/j.chemolab.2022.104610>
- [2] BEN-ISRAEL, A., GREVILLE, T.: *Generalized Inverses, Theory and Application*, 2nd Ed., Springer-Verlag, New York (2003).
- [3] BOTTACIN, F.: *Algebra Lineare e Geometria*, III Ed., Esculapio, Bologna (2021).
- [4] JHA S.K., YADAVA R.D.S.: Denoising by Singular Value Decomposition and its application to electronic nose data processing, *IEEE Sensors Journal*, Vol. 11, No. 1 (2011), 35–44. DOI 10.1109/JSEN.2010.2049351
- [5] JIANG Y., HAYASHI I., WANG S.: Knowledge acquisition method based on Singular Value Decomposition for human motion analysis, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 26, No. 12 (2014), 3038–3050. DOI 10.1109/TKDE.2014.2316521
- [6] Low-rank approximation: https://en.wikipedia.org/wiki/Low-rank_approximation
- [7] Moore-Penrose Inverse: https://en.wikipedia.org/wiki/Moore-Penrose_inverse
- [8] STRANG, G.: *Linear Algebra and Its Applications*, 4th Ed., Brooks/Cole Pub. (2005).