Contents

Chapter 6

Valors singulars d'una matriu

Sabem que qualsevol matriu real simètrica A es pot descompondre en la forma $A = PDP^T$ on D és una matriu diagonal i P és una matriu ortogonal. Ara veurem que una matriu real qualsevol A, no necessàriament quadrada es pot descompondre en la forma $A = U\Sigma V^T$ on Σ és diagonal i U, V ortogonals (no precissament inversa una de l'altra).

Es considera l'espai vectorial $\mathbb{R}^k$ amb un producte escalar. Es recorda que si $x, y \in \mathbb{R}^k$, aleshores $x.y = X^T Y$, on X, Y són les matrius (columna) de les coordenades de x i y en una base ortonormal.

Definició 6.1
Direm que una matriu simètrica A és positiva si verifica la condició $Ax.x \geq 0$ per tot $x \in E$.

Notau que Ax indica el vector imatge de x mitjançant l'endomorfisme associat a la matriu A en una base. Si la base fixada és ortonormal aleshores la condició de positivitat es pot escriure així: $Ax.x = (AX)^T X = X^T AX \geq 0$.

Proposició 6.1
Sigui A una matriu real $m \times n$, les matrius $A^T A$ i AA^T són simètriques, positives i tenen el mateix rang que A.

Prova. Vegem-ho només per $A^T A \in M_{n\times n}(\mathbb{R})$. Fixam bases ortonormals a $\mathbb{R}^n$ i $\mathbb{R}^m$. La simetria és òbvia: $(A^T A)^T = A^T (A^T)^T = A^T A$.

Vegem que és positiva: $X^T(A^T A)X = (AX)^T(AX) = (Ax).(Ax) \geq 0$. Finalment, vegem que $\operatorname{rang} A^T A = \operatorname{rang} A$. En efecte, si φ és l'endomorfisme (autoadjunt) de $\mathbb{R}^n$ associat a la matriu $A^T A$, aleshores $\operatorname{rang} A^T A = n - \dim \operatorname{Nuc} \varphi$ i $\operatorname{rang} A = n - \dim \operatorname{Nuc} f$, on f és l'aplicació lineal associada a A. Basta veure doncs que $\operatorname{Nuc} \varphi = \operatorname{Nuc} f$. Si $f(x) = 0$ aleshores $AX = 0$ que implica $A^T AX = 0$ i per tant $\varphi(x) = 0$. Recíprocament, si $\varphi(x) = 0$ tenim $A^T A = 0$ per tant $X^T A^T AX = 0$ que podem escriure $(AX)^T(AX) = 0$ o sigui $(Ax).(Ax) = 0$ i finalment $Ax = 0$ o sigui $f(x) = 0$. Per $AA^T \in M_{m\times m}(\mathbb{R})$ la demostració és anàloga. ∎

Proposició 6.2
Els valors propis de $A^T A$ (i de AA^T) són no negatius.

Prova. Sigui λ un valor propi de $A^T A$ (de φ). Sigui $x \neq 0$ tal que $A^T Ax = \lambda x$. Sabem que $A^T A$ és positiva: $(A^T Ax).x = (\lambda x).x = \lambda(x.x) \geq 0$. Per tant $\lambda \geq 0$ ja que $x.x > 0$.

Anàlogament per AA^T. ∎

PROPOSICIÓ 6.3
Els valors propis no nuls de A^TA i AA^T són els mateixos.

Prova. Sigui $\lambda > 0$ un valor propi de A^TA. Existeix $x \neq 0$ tal que $A^TAx = \lambda x$ d'on $AA^TAx = A(\lambda x) = \lambda Ax$, per tant λ és valor propi de AA^T amb vector propi Ax (és $Ax \neq 0$).

Recíprocament, sigui $\lambda > 0$ un valor propi de AA^T. Existeix $x \neq 0$ tal que $AA^Tx = \lambda x$ d'on $A^TAA^Tx = (\lambda x) = \lambda A^Tx$, per tant λ és valor propi de A^TA amb vector propi A^Tx (és $A^Tx \neq 0$). ■

PROPOSICIÓ 6.4
Sigui $f : \mathbb{R}^n \to \mathbb{R}^m$ lineal amb $\operatorname{rang} f = r$. Aleshores existeix una base ortonormal $\{e_1, \dots, e_n\}$ de $\mathbb{R}^n$, una base ortonormal $\{v_1, \dots, v_m\}$ de $\mathbb{R}^m$ i escalars $\sigma_1 \geq \dots \geq \sigma_r > 0$ tales que $f(e_i) = \sigma_i v_i, i = 1, \dots, r$ i $f(e_i) = 0$ si $i = r + 1 \dots, n$.

Prova. Fixam bases ortonormals a $\mathbb{R}^n$ i $\mathbb{R}^m$. Sigui A la matriu de f, i sigui φ l'endomorfisme de $\mathbb{R}^n$ associat a la matriu A^TA. Com sabem φ és autoadjunt ja que A^TA és simètrica, i té rang r. Per tant existeix una base ortonormal de E formada per vectors propis: $\{v_1, \dots, v_n\}$, tal que $\varphi(v_i) = \lambda_i v_i, i = 1, \dots, n$. La base es pot escollir de manera que $\lambda_1 \geq \dots \geq \lambda_r > 0$ i $\lambda_i = 0, i = r + 1, \dots, n$. Definim $u_i = \frac{f(v_1)}{\sigma_i}$ on $\sigma_i = \sqrt{\lambda_i}, i = 1, \dots, r$.

Es pot veure que $\{u_1, \dots, u_r\}$ és un conjunt ortonormal de $\mathbb{R}^m$ (i, per tant, linealment independent):

$$u_i.u_j = \frac{f(v_i)}{\sigma_i}\frac{f(v_j)}{\sigma_j} = \frac{1}{\sigma_i\sigma_j}(Av_i)^T(Av_j) = \frac{1}{\sigma_i\sigma_j}v_i^TA^TAv_j = \frac{1}{\sigma_i\sigma_j}v_i\lambda_jv_j = 0,$$

ja que $\{v_1, \dots, v_n\}$ és un conjunt ortogonal.

Si consideram ara $S = \langle u_1, \dots, u_r\rangle$, podem escollir una base ortonormal $\{u_{r+1}, \dots, u_m\}$ de l'ortogonal de S i així tenim que $\{u_1, \dots, u_r, u_{r+1}, \dots, u_m\}$ és una base ortonormal de $\mathbb{R}^m$ amb les propietats desitjades. ■

DEFINICIÓ 6.2
Els nombres reals $\sigma_1 \geq \cdots \sigma_r > 0$ i $\sigma_{r+1} = \cdots = \sigma_n = 0$ són els valors singulars de l'aplicació lineal f (o de A).

COROL·LARI 6.1 (DESCOMPOSICIÓ D'UNA MATRIU EN VALORS SINGULARS)
Una matriu real qualsevol A de tipus $m \times n$ i rang r es pot descompondre en la forma $A = U\Sigma V^T$ on Σ és diagonal de rang r, i U, V són ortogonals.

Prova. Fixades bases ortonormals a $\mathbb{R}^n$ i $\mathbb{R}^m$, sigui $f : \mathbb{R}^n \to \mathbb{R}^m$ lineal amb matriu A. Pel resultat anterior existeixen bases ortonormals de $\mathbb{R}^n$ i $\mathbb{R}^m$ tals que la matriu de f en aquestes bases és $\Sigma = diag(\sigma_1, \dots, \sigma_r, 0, \dots, 0)$. Podem escriure idò $\Sigma = U^TAV$ o també $A = U\Sigma V^T$, on U i V són ortogonals (no precisament inverses una de l'altra). Òbviament, $\operatorname{rang} \Sigma = r$. ■

COROL·LARI 6.2 (DESCOMPOSICIÓ POLAR D'UNA MATRIU)
Sigui $A \in M_{n\times n}(K)$. Existeix una matriu hermítica $P \in M_{n\times n}(K)$ amb valors propis no negatius, i una matriu ortogonal $Q \in M_{n\times n}(K)$ tals que $A = PQ$.

Prova. Si A és una matriu quadrada, llavors en la seva descomposició en valors singulars les matrius $U, \Sigma, V \in M_{n\times n}(K)$. En aquest cas tenim que

$$A = U\Sigma V^T = U\Sigma(U^TU)V^T = (U\Sigma U^T)(UV^T) = PQ.$$

Vegem que $P = U\Sigma U^T$ és hermítica:

$$P^T = (U\Sigma U^T)^T = U\Sigma^T U^T = U\Sigma U^T = P,$$

ja que Σ és simètrica. A més a més, com que U, V són unitàries llavors

$$Q^{-1} = (UV^T)^{-1} = (V^T)^{-1}U^{-1} = VU^T = Q^T.$$

■

COROL·LARI 6.3 (DESCOMPOSICIÓ POLAR D'UN ENDOMORFISME)
Sigui $f : \mathbb{R}^n \to \mathbb{R}^n$ un automorfisme. Existeixen $g, h : \mathbb{R}^n \to \mathbb{R}^n$ lineals amb g autoadjunt i positiu i h ortogonal (conserva el producte escalar) tals que $f = gh$.

Prova. Existeixen bases ortonormals $\{e_1, \ldots, e_n\}$ i $\{v_1, \ldots, v_n\}$ de $\mathbb{R}^n$ i escalars $\sigma_1 \geq \cdots \geq \sigma_n > 0$ tals que $f(e_i) = \sigma_i v_i, i = 1, \ldots, n$. Definim h de la següent manera: $h(e_i) = v_i, i = 1, \ldots, n$ i també $g : g(v_i) = \sigma_i v_i, i = 1, \ldots, n$. Les aplicacions g i h satisfan les condicions de l'enunciat. ■

PROPOSICIÓ 6.5
Si r és el nombre de valors singular diferents de zero de A llavors $\operatorname{rang}(A) = r$.

Prova. Aquest resultat és conseqüència de que el rang d'una matriu no varia si la multiplicam per matrius invertibles, com son U, V. ■

PROPOSICIÓ 6.6
Si $A = U\Sigma V^T$ és un SVD de $A, \operatorname{rang}(A) = r, U = [u_1\ u_2\ \cdots\ u_m], V = [v_1\ v_2\ \cdots\ v_n]$ llavors $\operatorname{Im} A = \langle u_1, \ldots, u_r \rangle$ *i* $\operatorname{Nuc} A = \langle v_{r+1}, \ldots, v_n \rangle$.

Prova. Com que U, V són matrius invertibles, llavors

$$\operatorname{Im}(AV) = \operatorname{Im}(A), \quad \operatorname{Nuc}(U^T A) = \operatorname{Nuc}(A).$$

Però com que $\operatorname{Im}(AV) = \operatorname{Im}(U\Sigma) = \langle \sigma_1 u_1, \sigma_r u_r \rangle = \langle u_1, \ldots, u_r \rangle$.

Per altra part, com que $\{v_1, \ldots, v_n\}$ és una base ortonormal de K^n, si $x \in K^n$ llavors $x = c_1 v_1 + \cdots + c_n v_n = Vc$ amb $c = (c_1, \ldots, c_n)$. Així

$$\begin{aligned} & x \in \operatorname{Nuc}(A) \iff Ax = 0 \iff AVc = 0 \iff U^T AVc = 0 \iff \Sigma c = 0 \\ & \iff \sigma_i c_i = 0,\ 1 \leq i \leq r \iff x = c_{r+1} v_{r+1} + \cdots + c_n v_n. \end{aligned}$$

Aleshores, $\operatorname{Nuc} A = \langle v_{r+1}, \ldots, v_n \rangle$. ■

Una altra propietat immediata de la SVD és la següent.

PROPOSICIÓ 6.7
Si $A = U\Sigma V^T$ és la SVD de la matriu $A \in M_{m\times n}(K)$ tal que $\operatorname{rang} A = r, U = [u_1\ u_2\ \cdots\ u_m], V = [v_1\ v_2\ \cdots\ v_n]$, *llavors*

$$A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T.$$

6.1 Càlcul de la Descomposició en Valors Singulars

Sigui $A \in M_{p\times n}(K)$ amb rang $A = r$. Com hem vist a la secció anterior, el càlcul de la descomposició en valors singulars es basa en la diagonalització de la matriu A^TA i la podem fer en les següents etapes:

a) Els vectors $v_1, \dots, v_r$ s'obtenen calculant les bases ortonormals dels subespais propis associats als valors propis no nuls de A^TA, ordenats de major a menor.

b) Denotems $V = (v_1|\cdots|v_n)$ i $U = (u_1|\cdots|u_p)$. Com que $A = U\Sigma V^T$ llavors $AV = U\Sigma$ i per tant $Av_i = \sigma_i u_i$, per tot $i = 1, \dots, r$. En conseqüència, les r primeres files de U s'obtenen directament de les de v fent

$$u_i = \frac{1}{\sigma_i}Av_i, \quad i = 1, \dots, r.$$

c) Ja tenim la descomposició que serà de la forma

$$A = (u_1|\cdots|u_r)\begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & \sigma_r \end{pmatrix}\begin{pmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_r^T \end{pmatrix} = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T.$$

Exemples 1

(1) Cercam la SVD de la matriu $A_2 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Calculam la matriu simètrica A^TA, els seus valors propis i els seus valors singular:

$$A^TA = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},$$

que té com a valors propis $\lambda_1 = 3, \lambda_2 = 1$. Per tant, els valors singulars de A són $\sigma_1 = \sqrt{3}, \sigma_2 = 1$. Els vectors propis normalitzats de A^TA són $v_1 = \frac{1}{\sqrt{2}}(1,1), v_2 = \frac{1}{\sqrt{2}}(-1,1)$ i així tenim

$$V = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{pmatrix}.$$

Per calcular la matriu U feim els següents càlculs:

$$u_1 = \frac{Av_1}{\sigma_1} = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix},$$

$$u_2 = \frac{Av_2}{\sigma_2} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}.$$

(2) Calculam una descomposició en valors singulars de la matriu

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

Calculam la matriu A^TA i els seus valors propis:

$$A^TA = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 9 \end{pmatrix}, \quad \lambda_1 = 9, \lambda_2 = 4, \lambda_3 = 2.$$

Per tant, els seus valors singulars són $\sigma_1 = 3, \sigma_2 = 2, \sigma_3 = \sqrt{2}$. *Així la matriu* Σ *és*

$$\Sigma = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \sqrt{2} \\ 0 & 0 & 0 \end{pmatrix}.$$

Calculam ara el vectors propis de $A^T A$ *i els normalitzam:*

$$\begin{aligned} V_9 &= \mathrm{Nuc}\,(A^T A - 9I) = \langle(0,0,1)\rangle, \quad v_1 = (0,0,1) \\ V_4 &= \mathrm{Nuc}\,(A^T A - 4I) = \langle(1,-1,0)\rangle, \quad v_2 = (1/\sqrt{2}, -1/\sqrt{2}, 0) \\ V_2 &= \mathrm{Nuc}\,(A^T A - 2I) = \langle(1,1,0)\rangle, \quad v_3 = (1/\sqrt{2}, 1/\sqrt{2}, 0). \end{aligned}$$

D'aquesta forma tenim ja la matriu V :

$$V = \begin{pmatrix} 0 & 1\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \end{pmatrix}.$$

Ens falta calcular els vectors de la matriu U :

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{3} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 0 \\ 2/3 \\ 2/3 \end{pmatrix},$$

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix},$$

$$u_3 = \frac{1}{\sigma_3} A v_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

La descomposió en valors singulars de A *és*

$$A = 3u_1 v_1^T + 2u_2 v_2^T + \sqrt{2} u_3 v_3^T.$$

(3) Sigui $A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. *Cercam una descomposició en valors singulars de* A.

Calculam $A^T A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}$, *que té valors propis* $\lambda_1 = 2$ *i* $\lambda_2 = \lambda_3 = 0$. *Una base ortonormal de vectors propis de* $A^T A$ *és* $v_1 = (0,0,1), v_2 = (1,0,0), v_3 = (0,1,0)$. *Tenim un únic valor singular no nul:* $\sigma_1 = \sqrt{2}$. *Una base ortonormal* $\{u_1, u_2\}$ *de* $\mathbb{R}^2$ *tal que* $f(v_1) = u_1, f(v_2) = f(v_3) = 0$ *és* $\{\frac{1}{\sqrt{2}}(1,1), \frac{1}{\sqrt{2}}(-1,1)\}$. *Llavors* A *descomposa com* $A = U\Sigma V^T$ *on*

$$\Sigma = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \quad V = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

6.2 Interpretació geomètrica dels Valors Singulars

En aquesta secció volem donar una interpretació geomètrica de què són els valors singulars d'una aplicació lineal, o equivalentment, d'una matriu. Per resoldre aquesta qüestió hem de contestar primer la pregunta de quina és la imatge de la circumferència unitat per una aplicació lineal.

Abans de constestar aquesta darrera qüestió pensem en com es transforma un segment recte per una aplicació lineal. Només hi ha dues possibilitats: o un altre segment recte o un punt; i en aquest darrer cas, podem pensar que és un segment degenerat que ha col·lapsat a un punt. Per tant, si això és així, la imatge per una aplicació lineal d'un polígon de n costats només té tres possibilitats: un altre polígon de n costats, un polígon col·lapsat, que seria un segment, o bé, el cas més degenerat, un punt. En conseqüència, la imatge d'un polígon regular és un altre polígon regular.

Podem contestar ara la pregunta inicial relacionada amb la circumferència. Podem pensar la circumfereència com un pas al límit d'un polígon de n costats i per tant, la imatge de la circumferència unitat per una aplicació lineal és una el·lipse centrada a l'origen de coordenades. Aquesta el·lipse pot ser degenerada y derivar en un segment o en un punt, però no hi ha més possibilitats.

Evidentment, aquest raonament que hem fet a $\mathbb{R}^2$ ho podem generalitzar a $\mathbb{R}^n$ i tenim un resultat simil·lar: la imatge de l'esfera unitat a un espai de dimensió n és una hiperel·lipse que pot ser degenerada, és a dir, el nombre de semieixos pot ser menor que n, i habitar, per tant, en un espai de dimensió inferior.

Podem donar ara ja una nova definició de valor singular d'una matriu, que ens donarà la seva interpretació geomètrica.

Definició 6.3
Els valors singulars d'una matriu A són les longituts dels semieixos de la hiperel·lipse en què es converteix l'esfera unitat per A.

Si tornam a pensar en el pla, observem que una conseqüència de tot el que s'ha dit anteriorment és que les matrius que deixen invariant la circumferència unitat són les matrius ortogonals. Això és així ja que si els vectors columna de la matriu A són ortogonals, els semieixos de l'el·lipse transformada també seran perpendiculars i tendran longitut 1, per tant, l'el·lipse serà una cirumferència.

Recordem el Teorema de Descomposició en Valors Singulars. Aquest teorema diu que qualsevol matriu A se pot descomposar de la forma $A = U\Sigma V^T$. Relacionant-ho amb la interpretació geomètrica que hem donat anteriorment, resulta que la transformació ortogonal V només produeix la rotació de la circumferència (de la esfera de dimensió n si estam a dimensió superior), Σ la deforma a una el·lipse (o hiperel·lipse en dimensió n) i finalment la matriu U rota l'el·lipse (hiperel·lipse en dimensió n).

6.3 Aplicacions de la SVD

6.3.1 Càlcul del rang aproximat d'una matriu

El problema de calcular el rang d'una matriu es un problema delicat desde el punt de vista numèric, ja que el condicionament de dit problema es muy dolent. Aquest mal condicionament condueix a replantejar la questió del rang en termes dels valors singulars de la matriuz donada. Aquest canvi d'enfocament se basa en què, si tenim la SVD per a una matriu $m \times n$ de ranr r llavors podem escriure:

$$A = U\Sigma V^T = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T,$$

és a dir, la matriu de rang r pot expressar-se com a suma de r matrius de rang 1.

El següent resultat mostra la distància que hi ha entre la matriu A i les matrius de rang $p \leq r$.

Proposició 6.8
Si $p \leq r$, definim $A_p = \sigma_1 U - 1 v_1^T + \cdots + \sigma_p u_p v_p^T$. Llavors A_p és la matriu de rang menor o igual que p més propera a A, és a dir $\|A - A_p\| \leq \|A - B\|$ per qualsevol B tal que $\operatorname{rang}(B) \leq p$.

A més a més, $\|A - A_p\| = \sigma_{p+1}$.

Figure 6.1: Gravat Melancolia del pintor Durero.

6.3.2 Compresió d'imatges digitals

Una de les aplicacions de la descomposició en valors singulars és la seva utilització en la compressió d'imatges digitals de manera que puguin ser transmeses de manera eficient per mitjans electrònics (internet, satèl·lit,etc). El problema que es vol considerar és el de saber quina és la quantitat mínima d'informació que es necessita transmetre per aconseguir imatges ntidas, sense que es perdin les parts essencials, i d'altra banda s'estalviï emmagatzematge.

Suposem que una matriu $A \in M_{m\times n}(K)$ representa els tons de gris d'una imatge amb una grandària de $m \times n$ pixels. La imatge està en un rectangle i es considera una graella de $m \times n$. Cada element de la graella (píxel) tendrà un nombre associat indicant el valor del to de gris, entre 0 i 255, de la imatge. As una matriu $A \in M_{m\times n}(K)$ té en cada lloc (i, j) el valor del gris corresponent a l'element de la graella o píxel de la fila i i columna j de la graella.

Suposem que coneixem la descomposició en valors singulars de la matriu A, és a dir $A = \sigma_1 u_1 v_1^T + \cdots \sigma_r u_r v_r^T$, essent els valors singulars no nuls $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Els valors singulars més petits provenen de les parts de la imatge amb menor interès. Això es pot observar si es representa parcialment $A_k = \sigma_1 u_1 v_1^T + \cdots \sigma_k u_k v_k^T$ amb $k \leq r$. Llavors, com hem vist a la Proposició 6.8, A_k és una aproximació de A que correspon únicament als k primers valors singulars i els corresponents vectors singulars. Es pot veure que es necessita considerar un $k << r$ per rescatar la imatge nítida.

Anem a veure un exemple concret. Considerem el gravat anomenat Melancolia, pintat pel famós pintor aleman Durero. El podeu veure a la Figura 6.1. Com sabem aquesta imatge digital no és més que una matriu de pixels. Suposem per exemple que al nostre ordinador ocupa un grandària de 736×566 pixels. Si volem enviar aquesta informació no podem enviar-la sense comprimir, ja que els clients de correu electrònic de deixen enviar fitxer tan grossos. Per tant, volem comprimir la imatge sense perdre molta resolució.

Diguem A a la matriu de grandària 736×566 que conté les dades necessàries per obtenir la imatge. És a dir, se necessiten 416576 nombres en punt flotant (diguem, dades) per aconseguir la imatge. D'acord amb el Teorema de Descomposició en Valors Singulars, existeixen matrius U, Σ i V tals que $A = U\Sigma V^T$, essen U de grandària 736×736, V de grandària 566×566 i

$$\Sigma = \begin{pmatrix} \Sigma_{566} \\ 0 \end{pmatrix}, \text{ amb } \Sigma_{566} = diag(\sigma_1, \ldots, \sigma_{566}),$$

amb $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{566}$ els valors singular de A.

El nombre de files especificades com a zero a Σ és, al manco, 736 - 566 = 170, així que les darreres 170 columnes de U no juguen cap paper. Denotam com U_1 la submatriu de U formada per les seves primeres 566 columnes. Podem escriure $A = U_1 \Sigma V^T$.

Si u_i representa la $i-$éssima columna de U_1 i v_i^T la $i-$éssima fila de V^T llavors sabem que

$$A = \sum_{i=1}^{566} \sigma_i u_i v_i^T.$$

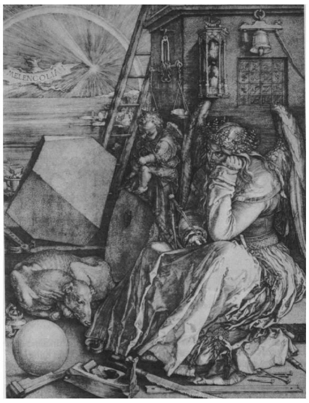

Figure 6.2: Gravat Melancolia del pintor Durero amb 150 modes.

Cada una de les matrius $\sigma_i u_i v_i^T$ és de grandària 736×566, pero només es necessiten $736 + 566 + 1 = 1303$ dades per formar-la: els 736 elements de u_i , els 566 de v_i^T i 1 de σ_i. A la matriu $\sigma_i u_i v_i^T$ se li diu $i-$éssim mode de A. Hem de notar que aquest modes no estan determinats de forma única per A. El que és més important és que alguns d'aquests modes són més dominants que altres per formar la matriu A. Les matrius $u_i v_i^T$ són matrius de rang 1 de forma que determinen subespais de dimensió 1, que podem identificar amb direccions en $\mathbb{R}^{566}$. A més a més, com que els vectors u_i i v_i són ortogonals, aquestes direccions són mutuament ortogonals. El valor σ_i ens proporciona la coordenada de A en la direcció $u_i v_i^T$. Si hhi ha valors singulars molt més grossos que altres, les direccions d'aquests valors singulars són més determinants en la formació de A que els dels valors singulars més petits.

Aproximant A per la suma dels seus primers modes aconseguim matrius que prenen en consideració les direccions fonamentals de A. Les imatges d'aquestes matrius han d'assemblar-se a les de A a mesura que el nombre de modes aumenta. Malgrat això, amb poc modes se poden conseguir resultats bastant bons. Observem que, per exemple, 150 modes requereixen (recordem que cada mode precisa de 1303 dades) $150 \times 1303 = 195450$ dades; menys de la meitat de les dades necessàries per formar la matriu A completa.

Amb 100 modes la resolució és bastant bona, però amb 150 modes (menys de la meitat de les dades de la matriu original) no es distingeix fàcilment l'original de la còpia.

Abans d'acabar, cal insistir de nou que hi ha algorismes més econòmics per al tractament d'imatges i que no utilitzen els valors singulars.

6.3.3 El problema de mínims quadrats i la inversa de Moore-Penrose

El problema de l'ajust de dades, és a dir, descobrir una funció matemàtica que pugui explicar de la millor forma possible el comportament d'algun mecanisme o grup d'éssers o objectes que pot ser mesurat, i de com coneixem algunes dades (amb els seus possibles errors de medició), és un problema clssic i ha suposat un repte per a la comunitat matemtica des del seu plantejament per Gauss i Legendre cap a 1800.

En llenguatge d'àlgebra lineal consisteix a trobar la solució on d'un sistema lineal $Ax = b$ essent $A \in K^{m \times n}$ amb $m \geq n$. Sabem que el sistema té solució si i només si $b \in \operatorname{Im} A$ i això normalment no es cumpleix si m és molt més gran que n. Si $m > n$ direm que el sistema està sobredeterminat. Si no existeix cap $x \in K^{n \times 1}$ tal que $Ax = b$ podem intentar cercar un x de forma que el vector $r = Ax - b \in K^m$ sigui el més petit possible. El problema lineal de mínims quadrats consisteix en trobar x perquè el vector r tengui la norma dos més petita possible.

Aquest problema pot no tenir solució única i llavors d'entre tots el vectors x tals que $\|Ax - b\|_2$ té norma mínima ens quedarem amb aquell tal que també la seva norma dos sigui mínima. Anam a definir-ho formalment.

Definició 6.4
$x \in \mathbb{R}^n$ és una solució per mínims quadrats si

$$\|Ax - b\|_2 = inf\{\|Az - b\|, z \in \mathbb{R}^n\}.$$

$x \in \mathbb{R}^n$ *és una solució òptima del problema de mínims quadrats si és una solució per mínims quadrats i a més a més*

$$\|x\|_2 = inf\{\|z\|, z \text{ és una solució per mínims quadrats}\}.$$

Comentari: Per a la definir anterior se podrien utilitzar altres normes que durien a diferents nocions de solucions generalitzades.

Anam a fer ara un petit parèntesi en el problema dels mínims quadrats. Com veurem més envant, en realitat el concepte que introduïm a la següent definició està molt relacionat amb el problema que ens ocupa.

DEFINICIÓ 6.5
Sigui $A \in M_{m\times n}(K)$ y $A = U\Sigma V^T$ la seva SVD, amb $\sigma_1 \geq \cdots \geq \sigma_r$ els seus valors singulars no nuls. Definima la inversa generalitzada de Moore-Penrose de A com la matriu

$$A^+ = V\Sigma^+U^T = V\begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix}U^T = V\begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 & 0 \\ \vdots & & \ddots & & \\ 0 & 0 & \cdots & \frac{1}{\sigma_r} & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}U^T.$$

De la definició anterior no sembla que se pugui assegurar que existeix una única inversa de Moore-Penrose per a cada A. En efect, la definició depen de l'elecció de les matrius U, V en la descomposició de A en valors singulars i aquestes no són, en general, úniques. Anem a demostrar que, malgrat aquesta arbitrarietat, la inversa de Moore-Penrose és única.

PROPOSICIÓ 6.9
Per a cada $A \in M_{m\times n}(k)$ existeix una únca inversa generatlizada de Moore-Penrose.

Prova. Sigui $A = U\Sigma V^T$ la SVD de A amb $\sigma_1 \geq \cdots \geq \sigma_r > 0$ els seus valors singulars. Per la Proposició 6.6, sabem que $\operatorname{Im} A = \langle u_1 \ \ldots \ u_r\rangle, \operatorname{Nuc} A = \operatorname{Im} A^T = \langle v_1 \ \ldots \ v_r\rangle$, essent bases ortonormals d'aquests subespais. Si escrivim ara $V = [V_1 \ V_2], V_1 \in M_{n\times r}(K), U = [U_1 \ U_2], U_1 \in M_{m\times r}(K)$, i $\Sigma_r = diag(\sigma_1, \ldots, \sigma_r)$ llavors $\Sigma_r^{-1} = diag(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_r})$ i

$$A = U_1\Sigma_r V_1^T, \ A^+ = V_1\Sigma_r^{-1}U_1^T$$

Si hi hagués una altra SVD, com que els valors singulars si són únics, existirien matrius $\tilde{U} \in M_{m\times m}(K), \tilde{V} \in M_{n\times n}(k)$ tals que $A = \tilde{U}\Sigma\tilde{V}^T$. Dividint $\tilde{U}, \tilde{V}$ com hem fet amb U, V tendriem $A = \tilde{U}_1\Sigma_r\tilde{V}_1$ amb $\tilde{U}_1, \tilde{V}_1$ matrius tals que les seves columnes són bases ortonormals de $\operatorname{Im} A, \operatorname{Im} A^T$, respectivament. Per a aquesta descomposició, la inversa de Moore-Penrose serà $\tilde{A}^+ = \tilde{V}_1\Sigma\tilde{U}_1^T$. Hem de veure que $A^+ = \tilde{A}^+$.

Pel que hem comentat al paràgraf anterior, existeixen matrius unitàries $P, Q \in M_{r\times r}(k)$ tals que

$$\tilde{U}_1 = U_1P, \quad \tilde{V}_1 = V_1Q.$$

Per altra part, com que $\tilde{U}_1\Sigma_r\tilde{V}_1^T = U_1\Sigma_r V_1^T$ llavors $U_1P\Sigma_rQ^TV_1^T = U_1\Sigma_rV_1^T$. Multiplicant per U_1^T a l'esquerra i per V_1 a la dreta obtenim $P\Sigma_rQ^T = \Sigma_r$. I com que Σ_r és invertible (és diagonal amb tots els seus elements de la diagonal diferents de zero) llavors $Q\Sigma_r^{-1}P^T = \Sigma_r^{-1}$, i també $V_1Q\Sigma_r^{-1}P^TU_1^T = V_1\Sigma_r^{-1}U_1^T$. Però recordem que $V_1Q = \tilde{V}_1$ i $P^TU_1^T = (U_1P)^T = \tilde{U}_1^T$, pel que $\tilde{A}^+ = A^+$. ∎

Tornam ara a intentar trobar la solució del problema dels mínims quadrats, és a dir, donat el sistema (no quadrat generalment) $Ax = b$ intentar trobar la solució x que fa mínima la norma dos del vector $Ax - b$, essent també mínima la norma de x.

PROPOSICIÓ 6.10
Sigui $A \in M_{m\times n}(K)$ i A^+ la seva inversa generalitzada de Moore-Penrose. Llavors, per cada $b \in K^m$ la solució òptima del problema de mínims quadrats és $x = A^+b$.

Prova. Observem que volem minimitzar $\|b - Ax\|_2$. Tenim la següent cadena d'igualtats:

$$\|b - Ax\|_2^2 = \|U^T(b - Ax)\|_2^2 = \|U^t b - U^T Ax\|_2^2 = \|U^T b - \Sigma V^T x\|_2^2 = \|c - \Sigma y\|_2^2,$$

essent $U^T b = c, V^T x = y$. Continuam la cadena de igualtats:

$$\|b - Ax\|_2^2 = \sum_{i=1}^{r} |c_i - \sigma_i y_i|^2 + \sum_{i=r+1}^{m} |c_i|^2.$$

El mínim de la suma anterior s'assoleix quan $y_i = \frac{c_i}{\sigma_i}$ i en aquest cas el mínim val $\sum_{i=r+1}^{m} |c_i|^2$.

Quan $r < n, y_{r+1}, \ldots, y_n$ no apareixen explícitament a l'expressió anterior i de totes les solucions obtingudes, la que té norma mínima és aquella tal que $y_{r+1} = \cdots = y_n = 0$. Com que $x = Vy$ i V és ortogonal, $\|x\|_2 = \|y\|_2$ i la norma de x serà mínima si i només si la de y ho és.

Llavors, la solució del problema de mínims quadrats és $x = Vy$. Si denotam $c = U^T b = \begin{pmatrix} \hat{c} \\ d \end{pmatrix}, y = \begin{pmatrix} \hat{y} \\ 0 \end{pmatrix}$, amb $\hat{c}, \hat{y} \in K^r$, el vector x ve donat per

$$x = Vy = \begin{pmatrix} \hat{y} \\ 0 \end{pmatrix} = B \begin{pmatrix} \Sigma_r^{-1}\hat{c} \\ 0 \end{pmatrix} = V \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{c} \\ d \end{pmatrix} = V \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} c = V \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^T b,$$

d'on deduïm el nostre resultat. ∎