Cursul 12

Valori şi vectori proprii ai matricilor simetrice. Descompunerea SVD a unei matrici. Aplicaţii

12.1 Valori şi vectori proprii ale matricilor simetrice. Descompunerea unei matrici simetrice

O matrice pătratică $M \in \mathbb{R}^{n\times n}$ cu proprietatea că $M^T = M$ se numeşte matrice simetrică. În machine learning matricile simetrice studiate sunt cel mai adesea matrici obtinute dintr-o matrice de date $A \in \mathbb{R}^{m\times n}$, care stochează pe coloane datele pentru n entitaţi. Fiecare entitate are m caracteristici, numite atribute:

$$A = [X_1|X_2|\dots|X_n]$$

De exemplu, în diagnosticarea inteligentă sau in studiul eficacităţii unor medicamente în tratamentul unei boli, entităţile sunt n persoane. Pentru fiecare individ se înregistreaza valorile pentru un set de m analize medicale (atribute ale indivizilor).

Dacă dupa constituirea matricii de date se calculeaza versorul fiecărei coloane şi se notează cu B matricea:

$$B = [X_1^0|X_2^0|\dots|X_n^0],$$

atunci matricea simetrica $M = B^T B$, are ca element generic $M_{ij} =< X_i^0, X_j^0 >= \cos(X_i, X_j)$. Cu alte cuvinte un element M_{ij} indică similaritatea dintre individul i şi j.

Pe de alta parte elementele N_{ij} ale matricii $N = BB^T$ indica similaritatea dintre atribute.

Informaţia importantă codificată de matricea de date A se extrage din valorile proprii şi vectorii proprii corespunzători ai matricii simetrice M respectiv N.

Să studiem particularităţile matricilor simetrice, comparativ cu matricile pătratice generale, nesimetrice.

În cele ce urmează interpretăm produsul Av, dintre o matrice pătratică, $A \in \mathbb{R}^{n\times n}$, şi un vector, $v \in \mathbb{R}^n$, ca fiind vectorul w ce reprezintă efectul unui operator liniar $L : \mathbb{R}^n \to \mathbb{R}^n$ de matrice A în baza canonică, asupra vectorului v. Deci în loc de $L(v)$, scriem Av.

Propoziţia 12.1.1 *Fie $A \in \mathbb{R}^{n\times n}$ o matrice pătratică şi $\mathbb{R}^n$ înzestrat cu produsul scalar standard. Atunci avem următoarea relaţie:*

$$< Av, w >=< v, A^T w >, \forall\ v, w \in \mathbb{R}^n \tag{12.1}$$

Demonstraţie: Exprimăm produsul scalar $< x, y >= x^T y$. Astfel membrul stâng al egalităţii ce dorim s-o demonstrăm este:

$$< Av, w >= (Av)^T w = v^T A^T w$$

iar membrul drept:

$$< v, A^T w >= v^T (A^T w) = v^T A^T w$$

şi deci:

$$< Av, w >=< v, A^T w >$$

□

Observaţia 12.1.1 *Dacă A este o matrice simetrică atunci din $A^T = A$ şi relaţia (12.1) rezultă că:*

$$< Av, w >=< v, Aw > \tag{12.2}$$

Să enunţam (fără demontraţie) particularităţile valorilor şi vectorilor proprii ai unei matrici simetrice:

Propoziţia 12.1.2 *Polinomul caracteristic al unei matrici simetrice, $A \in \mathbb{R}^{n\times n}$, are toate n rădăcinile reale, adică o matrice simetrică are n valori proprii.*

Dimensiunea fiecărui subspaţiu propriu al unei matrici simetrice coincide cu ordinul de multiplicitate al valorii proprii corespunzătoare.

Propoziţia 12.1.3 *La valori proprii distincte ale unei matrici simetrice corespund vectori proprii ortogonali.*

Demonstraţie: Fie $\lambda_1 \neq \lambda_2$ două valori proprii distincte ale matricii simetrice, A, şi $v_1 \in S_{\lambda_1}$, $v_2 \in S_{\lambda_2}$, vectori proprii corespunzători, adică $Av_1 = \lambda_1 v_1$, $Av_2 = \lambda_2 v_2$. Din proprietatea (12.2) a matricilor simetrice, avem că:

$$< Av_1, v_2 >=< v_1, Av_2 >$$

ceea ce este echivalent cu:

$$< \lambda_1 v_1, v_2 >=< v_1, \lambda_2 v_2 >$$

sau

$$\lambda_1 < v_1, v_2 >= \lambda_2 < v_1, v_2 > \quad \Leftrightarrow \quad (\lambda_1 - \lambda_2) < v_1, v_2 >= 0$$

Cum $\lambda_1 \neq \lambda_2$, rezultă că $\lambda_1 - \lambda_2 \neq 0$ şi deci $< v_1, v_2 >= 0$, adică $v_1 \perp v_2$

□

Să analizăm consecinţele acestor particularităţi ale matricilor simetrice:

• Fie $A \in \mathbb{R}^{n\times n}$ o matrice simetrică. Polinomul caracteristic al lui A, $P_n(\lambda) = \det(A - \lambda I_n)$, având n valori proprii (simple sau multiple) admite descompunerea:

$$P_n(\lambda) = (-1)^n (\lambda - \lambda_1)^{k_1} (\lambda - \lambda_2)^{k_2} \cdots (\lambda - \lambda_s)^{k_s}, \quad k_1 + k_2 + \cdots + k_s = n$$

unde λ_i este rădăcină multiplă de ordin k_i, $i = \overline{1, s}$.

• Pentru fiecare rădăcină λ_i, se determină subspaţiul propriu $S_{\lambda_i} = \{v \in \mathbb{R}^n \mid Av = \lambda_i v \Leftrightarrow (A - \lambda_i I_n)v = \theta\}$.

• Deoarece dimensiunea subspaţiului propriu S_{λ_i} este egală cu ordinul de multipliciate, k_i, al lui λ_i, determinăm o bază arbitrară $\mathcal{B}_i$, în S_{λ_i} (care conţine k_i vectori) şi apoi o ortonormăm folosind procedeul Gramm–Schmidt şi obţinem baza ortonormată $\mathcal{B}'_i$, $i = \overline{1, s}$.

• Dacă $\mathcal{B}'_1 = (u_1, u_2, \dots, u_{k_1})$, $\mathcal{B}'_2 = \underbrace{(u_{k_1+1}, \dots, u_{k_1+k_2})}_{k_2 \text{ vectori}}$, şi în final $\mathcal{B}'_s = \underbrace{(u_{n-k_s+1}, \dots, u_n)}_{k_s \text{ vectori}}$, sunt baze ortonormate din subspaţiile proprii $S_{\lambda_1}, S_{\lambda_2}, \dots, S_{\lambda_s}$, atunci concatenând cele s baze ortonormate formate din vectori proprii ai matricii A, obţinem o bază ortonormată în $\mathbb{R}^n$:

$$\mathcal{B}' = (u_1, u_2, \dots, u_{k_1}, u_{k_1+1}, \dots, u_{k_1+k_2}, \dots, u_n)$$

deoarece vectorii din bazele $\mathcal{B}'_i$ sunt ortonormate şi pentru că la valori proprii distincte corespund vectori proprii ortogonali, rezultă că orice vector dintr-o bază $\mathcal{B}'_i$ este ortogonal pe orice vector dintr-o bază $\mathcal{B}'_j$, $i =\neq j$.

• Notăm cu $T_{\mathcal{B}\mathcal{B}'}$ matricea de trecere de la baza canonică din $\mathbb{R}^n$ la baza ortonormată $\mathcal{B}'$ formată din vectori proprii ai matricii A. Aceasta este o matrice ortogonală, fiind matricea de trecere dintre două baze ortonormate.

• Notând cu D matricea diagonală a valorilor proprii,

$$D = \text{diag}(\underbrace{\lambda_1, \dots, \lambda_1}_{k_1}, \underbrace{\lambda_2, \dots, \lambda_2}_{k_2}, \dots, \underbrace{\lambda_s, \dots, \lambda_s}_{k_s})$$

, rezultă că matricea simetrică A este similară cu această matrice diagonală şi în plus matricea T din relaţia de similaritate este matricea ortogonală $T_{\mathcal{B}\mathcal{B}'}$:

$$A = T_{\mathcal{B}\mathcal{B}'} D T^{-1}_{\mathcal{B}\mathcal{B}'} = T_{\mathcal{B}\mathcal{B}'} D T^{T}_{\mathcal{B}\mathcal{B}'}$$

Astfel suntem conduşi la unul din cele mai importante rezultate aplicative din algebra liniară şi anume:

Propoziţia 12.1.4 *Dacă $A \in \mathbb{R}^{n\times n}$ este o matrice simetrică, ce are valorile proprii λ_i cu ordinele de multipliciate k_i, $i = \overline{1, s}$, $k_1 + k_2 + \cdots + k_s = n$, atunci există o bază ortonormată $\mathcal{B}'$, în $\mathbb{R}^n$, formată din vectori proprii ai lui A şi notând cu Q matricea de trecere de la baza canonică la baza $\mathcal{B}'$, matricea A este similară cu matricea*

$$D = diag(\underbrace{\lambda_1, \dots \lambda_1}_{k_1\ ori}, \dots, \underbrace{\lambda_s, \dots \lambda_s}_{k_s\ ori}) \tag{12.3}$$

şi relaţia de similaritate este:

$$A = QDQ^T$$

Definiţia 12.1.1 *Descompunerea unei matrici simetrice în forma*

$$A = QDQ^T$$

unde D este matricea diagonală a valorilor proprii şi $Q = T_{\mathcal{B}\mathcal{B}'}$ este o matrice ortogonală se numeşte descompunere ortogonală.

Să ilustrăm această proprietate printr-un exemplu:

Exemplul 1. Să se determine valorile proprii şi subspaţiile proprii corespunzătoare, pentru matricea simetrică:

$$A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix}$$

Să se determine apoi câte o bază ortonormată în fiecare subspaţiu propriu al lui A şi o bază ortonormată în $\mathbb{R}^3$, formată din vectori proprii ai lui A.

Să se scrie descompunerea simetrică ortogonală a matricii A.

- Valorile proprii ale lui A sunt $\lambda_{1,2} = 1, \lambda_3 = 7$;
- Subspaţiile proprii corespunzătoare, $S_{\lambda=1}$:

$$S_{\lambda=1} = \{v = (x, y, z) \mid Av = 1v \Leftrightarrow (A - 1I_3)v = 0\}$$

$$\begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$S_{\lambda=1} = \{v = \alpha(-1, 1, 0)^T + \beta(-1, 0, 1)^T, \alpha, \beta \in \mathbb{R}\}$$

Baza în acest subspaţiu este $\mathcal{B}_1 = (v_1, v_2)$. Evident că baza $\mathcal{B}_1$ nu este ortonormată. Aplicând procedeul Gramm-Schmidt obţinem baza

$$\mathcal{B}'_1 = (q_1 = \frac{1}{\sqrt{2}}(-1, 1, 0)^T, q_2 = \frac{1}{\sqrt{6}}(-1, -1, 2)^T)$$

Pentru a determina subspaţiul propriu $S_{\lambda=7}$ determinăm soluţiile sistemului $(A - 7I_3)v = 0$:

$$\begin{bmatrix} -4 & 2 & 2 \\ -2 & -4 & 2 \\ 2 & 2 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Alegem drept determinant principal determinantul constituit din elementele de intersecţie ale liniilor 1,2 cu coloanele 1,2. Astfel $z := \gamma$, este necunoscută secundară şi obţinem:

$$S_{\lambda=7} = \{v = \gamma(1, 1, 1)^T, \gamma \in \mathbb{R}\}$$

Baza ortonormată în $S_{\lambda=7}$ este:

$$\mathcal{B}'_2 = (q_3 = \frac{1}{\sqrt{3}}(1, 1, 1)^T)$$

iar $\mathcal{B}' = \mathcal{B}'_1 \cup \mathcal{B}'_2 = (q_1, q_2, q_3)$ este o bază ortonormată în $\mathbb{R}^3$ formată din vectori proprii ai matricii A. Notând: $Q = [q_1|q_2|q_3]$ avem descompunerea:

$$A = [q_1|q_2|q_3] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{bmatrix} [q_1|q_2|q_3]^T$$

12.2 Forme pătratice

În Inteligenţa artificială deseori într-o etapă a unui algoritm trebuie determinate punctele în care o funcţie, $f : D \subset \mathbb{R}^n \to \mathbb{R}$, de clasă C^2, ia valoarea minimă sau maximă. Problema aflării punctelor de minim sau maxim se numeşte problemă de optimizare şi se notează astfel:

$$\text{argmin} f(x), x \in D, \quad \text{argmax} f(x), x \in D$$

şi se citeşte să se determine argumentul $x \in D$ care minimizează funcţia f sau să se determine argumentul $x \in D$ care maximizează funcţia f.

Exemple de probleme de optimizare: minimizarea erorii în clasificare, sau să se determine drumul de lungime minimă pe care trebuie sa-l parcurgă un agent inteligent pentru a deservi n noduri/puncte de lucru, etc.

Decizia dacă un punct din D este punct de minim sau maxim pentru o funcţie f se ia analizând o formă pătratică asociată funcţiei f.

Considerăm spaţiul spaţiul vectorial $\mathbb{R}^n$ înzestrat cu produsul scalar standard şi notat $< v, w >= v^T w$.

Definiţia 12.2.1 *Fie $\mathbb{R}^n$ raportat la o bază ortonormată $\mathcal{B}$ (de obicei baza canonică) şi $A = (a_{ij})$, $i, j = \overline{1, n}$ o matrice simetrică. Aplicaţia $q : \mathbb{R}^n \to \mathbb{R}$, definită prin $q(v_\mathcal{B}) =< v_\mathcal{B}, Av_\mathcal{B} >= v_\mathcal{B}^T Av_\mathcal{B}$, se numeşte formă pătratică.*

Dacă un vector arbitrar $v \in \mathbb{R}^n$ are relativ la baza $\mathcal{B}$, coordonatele $v_\mathcal{B} = (x_1, x_2, \dots, x_n)^T$, atunci expresia analitică a formei pătratice în această bază este: ,

$$q(v_\mathcal{B}) = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{12.4}$$

Efectuând produsele, obţinem

$$q(v_\mathcal{B}) = a_{11}x_1^2 + a_{22}x_2^2 + \cdots + a_{nn}x_n^2 + 2a_{12}x_1x_2 + \cdots + 2a_{1n}x_1x_n + \cdots + 2a_{n-1\,n}x_{n-1}x_n$$

Această expresie ilustrează de ce funcţia se numeşte pătratică: expresia ei este o sumă de termeni de grad 2 în $x_1, x_2, \dots x_n$, adică ceea ce se numeşte **polinom omogen de grad 2**.

Exemplul 2. Formă pătratică definită pe $\mathbb{R}^2$:

$$q(v) = Q(x_1, x_2) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} -2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = -2x_1^2 + 2x_1x_2 + 3x_2^2$$

Dacă cunoaştem expresia analitică a unei forme pătratice $q : \mathbb{R}^n \to \mathbb{R}$, adică un polinom omogen de grad 2 în $x_1, x_2, \dots, x_n$, matricea simetrică ce o defineşte conform relaţiei $q(v) = v^T Av$ se determină astfel:

• coeficienţii pătratelor $x_1^2, x_2^2, \dots, x_n^2$ sunt respectiv elementele $a_{11}, a_{22}, \dots, a_{nn}$ din matricea simetrică A;

• coeficienţii produselor x_ix_j împărţiti la 2 sunt elementele a_{ij} şi a_{ji} din matricea A, $i, j = \overline{1, n}$.

Exemplul 3. Se dă forma pătratică $q : \mathbb{R}^3 \to \mathbb{R}$ definită prin $q(x_1, x_2, x_3) = 2x_1^2 + 3x_1x_2 + 6x_1x_3 - 5x_2^2 - 8x_2x_3 + x_3^2$. Matricea simetrică asociată este:

$$A = \begin{bmatrix} 2 & \frac{3}{2} & 3 \\ \frac{3}{2} & -5 & -4 \\ 3 & -4 & 1 \end{bmatrix}$$

O formă pătratică ia valori reale care pot fi pozitive, negative sau zero. $q(\theta) =< \theta, A\theta >= 0$. Deci o formă pătratică aplică pe $(0, 0, \dots, 0)$ în 0.

• Forma pătratică $q : \mathbb{R}^n \to \mathbb{R}$ care ia valori strict pozitive, $q(v) > 0$, oricare ar fi vectorul $v \in \mathbb{R}^n \setminus \{\theta\}$, se numeşte formă pătratică pozitiv definită.

• Dacă $q(v) \geq 0$, pentru orice $v \in \mathbb{R}^n$, atunci q se numeşte formă semipozitiv definită (mai precis în acest caz q ia valoarea zero şi pentru vectori nenuli).

• Dacă $q(v) < 0$, oricare ar fi $v \in \mathbb{R}^n \setminus \{\theta\}$ atunci q se numeşte formă negativ definită, iar dacă $q(v) \leq 0, \forall\ v \in \mathbb{R}^n$, forma q se numeşte seminegativ definită.

• Dacă pe anumiţi vectori q ia valori pozitive, iar pe alţii negative, atunci q se numeşte formă pătratică nedefinită.

Analizând expresia analitică a formei pătratice din Exemplul 3 este greu să ne pronunţam dacă ea este pozitiv definită, negativ definită sau nedefinită. Este însă foarte simplu să indicăm tipul formei pătratice dacă ea conţine doar termeni în x_i^2, $i = \overline{1, n}$.

Exemplul 4. Forma pătratică $Q(x_1, x_2, x_3) = -3x_1^2 - x_2^2 - 4x_3^2$ este evident negativ definită, forma $Q(x_1, x_2, x_3) = 2x_1^2 - x_2^2 + 6x_3^2$ este nedefinită, deoarece $Q(1, 0, 1) = 2 + 6 = 8 > 0$, iar $Q(0, 1, 0) = -1 < 0$.

Observăm că putem deduce rapid tipul unei forme pătratice dacă ea este definită de o matrice diagonală, care evident este simetrică:

$$\begin{aligned} q(x_1, x_2, \dots, x_n) &= \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= d_1x_1^2 + d_2x_2^2 + \cdots + d_nx_n^2, \quad d_i \in \mathbb{R} \end{aligned}$$

O formă pătratică a cărei matrice de definiţie este diagonală se zice că este în *forma canonică*.

Să exploatăm faptul că orice matrice simetrică, A, este similară cu o matrice diagonală, adică există $\lambda_1, \lambda_2, \dots \lambda_n \in \mathbb{R}$, ce sunt valorile proprii ale lui A, şi matricea inversabilă $T := T_{\mathcal{B}\mathcal{B}'}$, ce este matricea de trecere de la baza canonică $\mathcal{B}$ la baza ortonormată, $\mathcal{B}'$, formată din vectori proprii ai lui A, astfel încât $A = TDT^{-1} = T_{\mathcal{B}\mathcal{B}'}DT^T_{\mathcal{B}\mathcal{B}'}$, cu

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

Echivalent, putem scrie că

$$D = T^T_{\mathcal{B}\mathcal{B}'}AT_{\mathcal{B}\mathcal{B}'} \tag{12.5}$$

Relaţia dintre coordonatele unui vector $v \in \mathbb{R}^n$ relativ la cele două baze este: $v_{\mathcal{B}} = T_{\mathcal{B}\mathcal{B}'}v'_{\mathcal{B}}$ Să deducem expresia analitică a formei pătratice de matrice A, relativ la baza $\mathcal{B}'$ formată din vectori proprii ai lui A. Pentru aceasta notăm cu $X_1, X_2, \dots, X_n$, coordonatele vectorului arbitrar v relativ la baza $\mathcal{B}'$:

$$\begin{aligned} q(v_{\mathcal{B}}) &= < v_{\mathcal{B}}, Av_{\mathcal{B}} > = < T_{\mathcal{B}\mathcal{B}'}v'_{\mathcal{B}}, AT_{\mathcal{B}\mathcal{B}'}v'_{\mathcal{B}} > \stackrel{(12.1)}{=} < v_{\mathcal{B}'}, T^T_{\mathcal{B}\mathcal{B}'}AT_{\mathcal{B}\mathcal{B}'}v_{\mathcal{B}'} > \stackrel{(12.5)}{=} < v_{\mathcal{B}'}, Dv_{\mathcal{B}'} > \\ &= \begin{bmatrix} X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \lambda_1 X_1^2 + \lambda_2 X_2^2 + \dots + \lambda_n X_n^2 \end{aligned}$$

Astfel am arătat ca o formă patratică q definită de matricea simetrică A asociază unui vector exprimat în baza $\mathcal{B}$, aceeaşi valoare ca valoarea asociată de forma pătratică Q (definită de matricea diagonala a valorilor proprii ale lui A) aceluiaşi vector, dar exprimat în baza $\mathcal{B}'$.

În concluzie pentru a decide tipul formei pătratice care relativ la baza ortonormată iniţială, $\mathcal{B}$ din $\mathbb{R}^n$, are matricea simetrică, A, se determină valorile proprii ale matricii A.

• Dacă toate valorile proprii sunt strict pozitive, forma pătratică este pozitiv definită;

Dacă toate valorile proprii sunt mai mari sau egale cu 0, forma pătratică este semipozitiv definită;

• Dacă $\lambda_i < 0, \forall\ i = \overline{1,n}$, forma este negativ definită, respectiv seminegativ definită dacă $\lambda_i \leq 0, \forall\ i = \overline{1,n}$.

• Dacă o parte dintre valorile proprii sunt pozitive şi restul negative, forma pătratică este nedefinită.

Exemplul 5. Se dă forma pătratică $q : \mathbb{R}^2 \to \mathbb{R}$, $q(x_1, x_2) = 4x_1^2 - 4x_1x_2 + x_2^2$. Să se determine matricea formei pătratice, valorile ei proprii şi să se precizeze tipul formei: pozitiv, negativ definită sau nedefinită.

Matricea formei pătratice este

$$A = \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}$$

Valorile proprii ale lui A sunt $\lambda_1 = 0, \lambda_2 = 5$. Astfel dacă

$$\mathcal{B}' = (u_1, u_2)$$

este o bază ortonormată în $\mathbb{R}^2$ formată din vectori proprii ai matricii A şi a $v = X_1 u_1 + X_2 u_2$ este un vector din $\mathbb{R}^2$ exprimat în baza $\mathcal{B}'$ atunci forma pătratică are relativ la baza $\mathcal{B}'$ expresia: $q(v) = \lambda_1 X_1^2 + \lambda_2 X_2^2 = 0X_1^2 + 5X_2^2$, deci este semipozitiv definită.

În analiza matematică unei funcţii $f : D \subset \mathbb{R}^n \to \mathbb{R}$, de clasă C^2, i se asociază matricea simetrică $\mathrm{Hess}(f)(x_0)$ a derivatelor de ordin 2 într-un punct (x_0), numită Hessiana funcţiei în acest punct. Elementele a_{ij} ale acestei matrici sunt:

$$a_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0), i, j = \overline{1, n}$$

Dacă x_0 este un punct critic al funcţiei f, adică

$$\frac{\partial f}{\partial x_i}(x_0) = 0, \forall\ i = \overline{1, n},$$

atunci tipul formei pătratice având ca matrice, matricea Hessiană în x_0 indică dacă punctul x_0 este punct de maxim, minim, sau punct şa.

În Fig.12.1 este ilustrat graficul unei forme pătratice $q : \mathbb{R}^2 \to \mathbb{R}$, $q(x_1, x_2) = \lambda_1 x_1^2 + \lambda_2 x_2^2$, pentru cazul q – pozitiv definită ($\lambda_1, \lambda_2 > 0$), negativ definită, $\lambda_1, \lambda_2 < 0$ şi respectiv nedefinită, $\lambda_1 > 0, \lambda_2 < 0$. Observăm că în primul caz $(0, 0)$ este punct de minim, pentru că $q(x_1, x_2) > 0$, $\forall\ v = (x_1, x_2)^T \neq 0$, în al doilea este punct de maxim şi în al treielea este punct şa. La analiză aflaţi că o funcţie $f : D \subset \mathbb{R}^2 \to \mathbb{R}$, de clasă C^r pe D, $r \geq 2$, este aproximată în vecinătatea unui punct critic (x_{01}, x_{02}) de o astfel de formă pătratică şi deci tipul extremal al punctului critic depinde de tipul punctului $(0, 0)$ pentru forma pătratică asociată.

12.3 Descompunerea singulară a unei matrici

Necesitatea de a minimiza volumul de informaţie digitală ce trebuie să fie stocată sau transmisă printr-un canal de comunicaţie a condus la dezvoltarea a numeroase metode de reducere a dimensiunii acestora (comprimarea datelor). Una din metodele de comprimare a datelor, oferite de algebra liniară se bazează pe descompunerea singulară a unei matrici $A \in \mathbb{R}^{m \times n}$.

12.3.1 Noţiuni şi rezultate preliminare

Definiţia 12.3.1 *O matrice simetrică $A \in \mathbb{R}^{n \times n}$ pentru care forma pătratică asociată, q este pozitiv definită, adică:*

$$q(v) =< v, Av > > 0, \forall v \in \mathbb{R}^n \setminus \{\theta\} \tag{12.6}$$

se numeşte matrice pozitiv definită, iar dacă verifică:

$$q(v) =< v, Av > \geq 0, \forall\ v \in \mathbb{R}^n \tag{12.7}$$

se numeşte matrice simetrică semi–pozitiv definită.

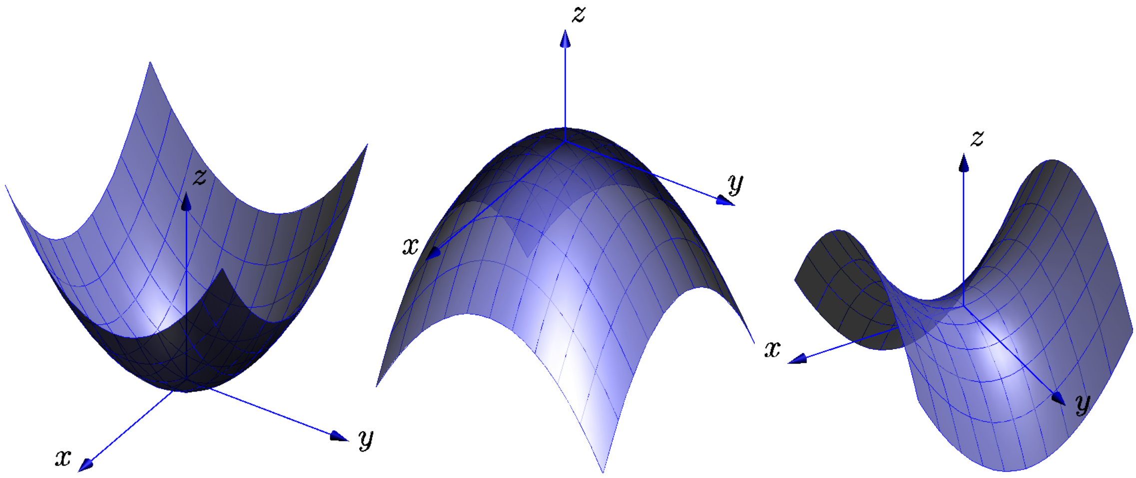

Fig.12.1: Graficele a 3 forme pătratice aduse la forma canonică, $q(x_1, x_2) = \lambda_1 x_1^2 + \lambda_2 x_2^2$: stânga, forma este pozitiv definită, centru, negativ definită şi cea din dreapta, nedefinită.

.

Propoziţia 12.3.1 *O matrice simetrică este semi-pozitiv definită dacă şi numai dacă are toate valorile proprii mai mari sau egale cu zero.*

Demonstraţie: Fie $A \in \mathbb{R}^{n\times n}$ o matrice simetrică şi semi-pozitiv definită. Fiind simetrică are n valori proprii distincte sau nu. Fie $\lambda \in \mathbb{R}$ o valoare proprie şi v un vector propriu corespunzător, $Av = \lambda v$. Condiţia de semipozitiv definită este echivalentă cu

$$< v, Av > \geq 0 \quad \Leftrightarrow \quad < v, \lambda v > \geq 0 \quad \lambda < v, v > \geq 0$$

v fiind vector propriu este nenul şi deci $< v, v > > 0$. Astfel $\lambda < v, v > \geq 0$ dacă şi numai dacă $\lambda \geq 0$. $\square$

Considerăm o matrice arbitrară de m linii şi n coloane, $A \in \mathbb{R}^{m\times n}$. Matricile asociate $A^T A$, AA^T sunt matrici pătratice de tip $n \times n$, respectiv $m \times m$ şi simetrice deoarece coincid cu transpusele lor. De exemplu, $(A^T A)^T = A^T (A^T)^T = A^T A$.

Propoziţia 12.3.2 *Dacă $A \in \mathbb{R}^{m\times n}$ este o matrice de tip $m \times n$, atunci matricile simetrice asociate, $A^T A$, AA^T, sunt semipozitiv definite.*

Demonstraţie: Deoarece produsul scalar al unui vector cu el însuşi este mai mare sau egal cu zero, avem că pentru orice vector $v \in \mathbb{R}^n$: $< Av, Av > \geq 0$. Dar

$$< Av, Av > \stackrel{\text{cf } (12.1)}{=} < v, A^T Av >,$$

ceea ce implică $< v, A^T Av > \geq 0$, $\forall\ v \in \mathbb{R}^n$, adică matricea $A^T A$ este semipozitiv definită.

Analog, $< A^T v, A^T v > \geq 0$ şi din

$$< A^T v, A^T v > \stackrel{\text{cf (12.1)}}{=} < v, AA^T v >, \forall\, v \in \mathbb{R}^m$$

rezultă că şi matricea AA^T este semipozitiv definită. □

Propoziţia 12.3.3 *Rangul matricii $A^T A$ coincide cu rangul matricii $A \in \mathbb{R}^{m\times n}$.*

Demonstraţie: Vezi Cursul 7, partea relativ la soluţia celor mai mici pătrate a unui sistem $Ax = b$.

□

12.4 Calculul descompunerii singulare

Propoziţia 12.4.1 (**Descompunerea singulară a unei matrici**) *Pentru orice matrice $A \in \mathbb{R}^{m\times n}$ de rang $r \leq \min(m, n)$ există două matrici ortogonale $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ şi numerele reale pozitive $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ astfel încât A se descompune în produsul $A = U\Sigma V^T$, adică:*

$$\underbrace{A}_{m\times n} = \underbrace{U}_{m\times m} \underbrace{\begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \sigma_r & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix}}_{m\times n} \underbrace{V^T}_{n\times n} \tag{12.8}$$

Definiţia 12.4.1 *Descompunerea $A = U\Sigma V^T$ se numeşte descompunerea singulară a matricii A (**singular value decomposition, SVD**). Valorile pozitive $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ din matricea Σ se numesc **valorile singulare ale matricii A**, vectorii u_i, **vectori singulari stângi**, iar vectorii v_i, **vectori singulari drepţi** , $i = \overline{1, r}$.*

Pentru a interpreta descompunerea SVD şi pentru a prezenta aplicaţii ale ei, definim câteva noţiuni şi rezultate de calcul matricial:

• Produsul exterior a doi vectori $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$ este o matrice de tip $m \times n$, obţinută ca produsul uv^T.

A nu se confunda produsul exterior (*outer product* în l. engleză) uv^T, cu produsul scalar $< u, v >= u^T v$ (*inner* sau *dot product* în engleză).

Dacă $u = (x_1, x_2, \dots, x_m)^T$ şi $v = (y_1, y_2, \dots, y_n)^T$ atunci produsul lor exterior este:

$$uv^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & \dots & x_1y_n \\ x_2y_1 & x_2y_2 & \dots & x_2y_n \\ \vdots & \vdots & \dots & \vdots \\ x_my_1 & x_my_2 & \dots & x_my_n \end{bmatrix}$$

Prin urmare matricea produs exterior, $B = uv^T$, are elementele $b_{ij} = x_i y_j$, $i = \overline{1,m}$, $j = \overline{1,n}$.

Dacă vectorii u, v sunt nenuli, atunci coloanele matricii uv^T sunt proporţionale şi prin urmare rangul matricii B este 1.

Propoziţia 12.4.2 *Produsul $U\Sigma V^T$ din descompunerea SVD a unei matrici $A \in \mathbb{R}^{m\times n}$, de rang r, este egal cu:*

$$U\Sigma V^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_1^T + \cdots + \sigma_r u_1 v_r^T$$

Demonstraţie: • Calculăm mai întâi produsul dintre matricea $U = [u_1|u_2|\ldots|u_m] \in \mathbb{R}^{m\times m}$ şi matricea pseudodiagonală $\Sigma \in \mathbb{R}^{m\times n}$, $\Sigma = [\Sigma_1|\Sigma_2|\ldots|\Sigma_r|\ldots|\Sigma_n]$, care are 0 în toate poziţiile de indici diferiţi este:

$$U\Sigma \stackrel{Curs\,1}{=} [U\Sigma_1|U\Sigma_2|\ldots|U\Sigma_r|\ldots|U\Sigma_n]$$

Dar coloana j, $j = \overline{1,r}$, a produsului $U\Sigma$ este:

$$U\begin{bmatrix} 0\\ 0\\ \vdots\\ \Sigma_{jj}\\ \vdots\\ 0\end{bmatrix} = U\begin{bmatrix} 0\\ 0\\ \vdots\\ \sigma_j\\ \vdots\\ 0\end{bmatrix} = \sigma_j U\begin{bmatrix} 0\\ 0\\ \vdots\\ 1\\ \vdots\\ 0\end{bmatrix} = \sigma_j U e_j = \sigma_j u_j$$

Ultima egalitate exprimă faptul că coloana j a unei matrici este produsul dintre acea matrice si vectorul bazei canonice e_j. Astfel rezultă că produsul $U\Sigma = [\sigma_1 u_1|\sigma_2 u_2|\ldots|\sigma_r u_r|0u_{r+1}|\ldots 0u_m]$.

• Notând cu $P = U\Sigma$ să evaluăm produsul $A = PV^T$, unde $P = [p_1|p_2|\ldots|p_n] \in \mathbb{R}^{m\times n}$ şi $V = [v_1|v_2|\ldots|v_n] \in \mathbb{R}^{n\times n}$:

Din primul curs (şi notebook-ul relativ la programarea functională în Python) avem că produsul PV^T este suma produselor exterioare dintre coloanele matricii P şi liniile de acelaşi indice ale matricii V^T. Dar o linie j în matricea $V^T = [v_1|v_2|\ldots|v_n]^T$ este reprezentată de vectorul v_j^T. Atfel rezultă că

$$PV^T = U\Sigma V^T = p_1 v_1^T + p_2 v_2^T + \cdots + p_n v_n^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

□

Prin urmare ideea de bază a descompunerii SVD a unei matrici $A \in \mathbb{R}m \times n$, de rang r este că matricea A se poate descompune ca o combinaţie liniară cu coeficienţi pozitivi, descrescatori a r matrici de rangul 1, $M_j = u_j v_j^T$, $j = \overline{1,r}$:

$$A = U\Sigma V^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

Aproximarea de rang k a matricii A: Observăm că exprimarea matricii A ca o combinaţie liniară de r matrici are coeficienţii σ_i descrescători, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Dacă ultimele valori singulare sunt mici (apropiate de zero), atunci renunţând la termenii ce le conţin obţinem o aproximare a matricii A:

$$A_k = \sigma_1(u_1 v_1^T) + \cdots + \sigma_k(u_k v_k^T)$$

Matricea aproximare A_k se factorizează astfel:

$$A_k = \underbrace{[u_1|u_2|\dots|u_k]}_{m\times k} \underbrace{\begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \sigma_k \end{bmatrix}}_{k\times k} \underbrace{\begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_k^T \end{bmatrix}}_{k\times n} = U_k \Sigma_k V_k^T$$

şi are rangul k (pentru a arăta că are rangul k se demonstrează că subspaţiul Null al produsului din membrul drept are dimensiunea egală cu dimensiunea subspaţiului Null al matricii Σ_k, adică $n-k$).

Ori de câte ori se aproximează un element al unei mulţimi înzestrate cu o distanţă (metrică), cu alt element al aceleaşi mulţimi ne interesează ”cât de bună este acea aproximare”, evaluând distanţa dintre element şi aproximantul său. În multimea matricilor din $\mathbb{R}^{m\times n}$ se definesc diferite norme şi atunci dist(A,B)=$\|A-B\|$.

O normă este cea definită de produsul scalar a două matrici:

$$< A, B >= \text{trace}(A^T B)$$

şi anume

$$\|A\| = \sqrt{< A, A >} = \sqrt{\text{trace}(A^T A)}$$

Această normă se numeşte **norma Frobenius a unei matrici** şi pentru a o distinge de alte norme se notează $\|A\|_F$.

Teorema 12.4.1 *(Teorema Eckart) Fie matricea $A \in \mathbb{R}^{m\times n}$ şi $A_k \in \mathbb{R}^{m\times n}$ aproximarea sa de rang k. Dintre toate matricile $B \in \mathbb{R}^{m\times n}$ de rang k, distanţa de la A la B este minimă pentru $B = A_k$, adică:*

$$\min_{B\,|\,\mathit{rang}(B)=k} \{\|A-B\|_F\} = \|A - A_k\|_F = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2}$$

Cu alte cuvinte A_k este cea mai bună aproximare de rang k a matricii A Această proprietate se exploatează în numeroase domenii din Computer Science, printre care: comprimarea datelor în general şi a imaginilor în particular, information retrieval, machine learning.

Dacă A este o matrice imagine ai cărei pixeli au diverse nivele de gri, între negru şi alb, un element a_{ij} al matricii fiind codul nivelului de gri $c \in \{0, 1, 2, \dots, 255\}$ sau normalizat $c \in [0, 1]$ (depinde de tipul de imagine şi limbajul de programare care o citeşte; de exemplu în Python/numpy imaginile în nivele de gri din mulţimea $\{0, 1, 2, \dots, 255\}$ sunt convertite la citire în imagini cu nivelul de gri in $[0, 1]$, adică dacă codul pt gri este $c = 135$, el este convertit la $135.0/255 \in [0, 1]$), atunci determinând descompunerea SVD a matricii $A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$, şi renunţând la termenii ce au coeficienţii $\sigma_{k+1}, \ldots, \sigma_r$ suficient de mici în comparaţie cu $\sigma_1, \sigma_2, \ldots, \sigma_k$ obţinem o aproximare A_k a imaginii a imaginii A. Aceasta este o modalitate de comprimare a imaginii în scopul stocării sau transmiterii ei pe un canal de comunicaţie. Adică în locul transmiterii vectorilor $u_1, u_2, \ldots, u_r$, $v_1, v_2, \ldots, v_r$ şi respectiv a valorilor singulare $\sigma_1, \sigma_2, \ldots, \sigma_r$ se transmit doar vectorii $u_1, \ldots, u_k, v_1, \ldots, v_k$ şi valorile singulare $\sigma_1, \sigma_2, \ldots, \sigma_k$, $k << r$.

Descompunerea trunchiată, A_k, a unei imagini filtrează o parte din zgomotul conţinut în imagine fără a pierde o informaţie semnificativă din A.

Ca exemplu aveţi Notebook-ul in care se aplică descompunerea SVD imaginii lui Cristian Avramescu. Un alt Notebook cu explicaţii detailate este inclus in arhivă cu acest curs.

12.5 Contrucţia matricilor U, V, Σ din descompunerea SVD

- **Construcţia matricii V**:

Matricea $A^T A$ fiind o matrice de tip $n \times n$, simetrică şi semipozitiv definită, are n valori proprii $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. Construind câte o bază ortonormată în fiecare subspaţiu propriu şi reunind aceste baze obţinem baza ortonormată din $\mathbb{R}^n$ formată din vectori proprii ai matricii $A^T A$, $\mathcal{B}'_n = (v_1, v_2, \ldots, v_n)$. Notăm cu $V = [v_1|v_2|\ldots|v_n]$ matricea de trecere de la baza canonică la baza $\mathcal{B}'_n$, care evident este matrice ortogonală. Astfel matricea $A^T A$ este similară cu matricea diagonală a valorilor proprii:

$$A^T A = V \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) V^T$$

Dar cum două matrici similare au aceelaşi rang, rezulta că rang(diag$(\lambda_1, \ldots, \lambda_n)$)=rang($A^T A$). Însă rang($A^T A$)=rang(A)=r. Prin urmare matricea diagonală diag$(\lambda_1, \lambda_2, \ldots, \lambda_n)$ are rangul r şi deci doar primele r valori proprii în ordinea descrescătoare sunt nenule:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0, \text{şi } \lambda_{r+1} = \cdots = \lambda_n = 0$$

- **Construcţia matricii U**:

Vectorii proprii ortonormaţi, $v_1, v_2, \ldots, v_r$, corespund la valori proprii nenule: $A^T A(v_i) = \lambda_i v_i$, $i = \overline{1, n}$. Considerăm vectorii din $\mathbb{R}^m$, $w_i = A v_i$, $i = \overline{1, r}$. Să arătăm că sistemul $w_1, w_2, \ldots, w_r$ este un sistem ortogonal de vectori:

$$< w_i, w_j >=< Av_i, Av_j >=< v_i, (A^T A) v_j >=< v_i, \lambda_j v_j >= \lambda_j < v_i, v_j >= \lambda_j \delta_{ij}$$

Astfel pentru $i \neq j$, $< w_i, w_j >= \lambda_j \cdot 0 = 0$ (deci vectorii sunt ortogonali), iar pentru $i = j$ avem: $< w_i, w_i >= \lambda_i$ ceea ce este echivalent cu $\|w_i\|^2 = \lambda_i$ sau echivalent:

$$\|w_i\| = \sqrt{\lambda_i} \overset{\text{notaţie}}{=} \sigma_i, i = \overline{1, r}$$

Normând vectorii $(w_1, w_2, \ldots, w_r)$ obţinem sistemul ortonormat de r vectori din $\mathbb{R}^m$, $(u_1, u_2, \ldots, u_r)$ unde

$$u_i = \frac{w_i}{\|w_i\|} = \frac{w_i}{\sigma_i} = \frac{A v_i}{\sigma_i}, i = \overline{1, r}$$

Prin urmare avem următoarea relaţie între vectorii ortonormaţi $v_1, v_2, \dots, v_r$ din $\mathbb{R}^n$ şi vectorii ortonormaţi $u_1, u_2, \dots, u_r$ din $\mathbb{R}^m$:

$$Av_i = \sigma_i u_i, \quad i = \overline{1,r}$$

Vectorii ortonormaţi $(u_1, u_2, \dots u_r)$ îi completăm la o bază ortonormată $(u_1, u_2, \dots, u_r, u_{r+1}, \dots, u_m)$ în $\mathbb{R}^m$.

Şi anume determinăm o bază arbitrară în subspaţiul $Null(A^T)$. Deoarece A^T are rangul lui A, adică r, rezultă că dimensiunea lui este $m-r$. Baza arbitrară $t_1, t_2, \dots t_{m-r}$ se ortonormează cu metoda Gramm-Schmidt şi obţinem din ea baza ortonormată $u_{r+1}, u_{r+2}, \dots u_m$ în $Null(A^T)$.

Bază ortonormată $u_{r+1}, \dots, u_m$, din subspaţiul Null(A^T) completează sistemul ortonormat $(u_1, u_2, \dots, u_r)$ la o bază ortonormată în $\mathbb{R}^m$,$(u_1, \dots, u_r, u_{r+1}, \dots, u_m)$.

Notăm cu U matricea de trecere dela baza canonică din $\mathbb{R}^m$ la baza ortonormată $(u_1, u_2, \dots, u_m)$. Matricea $U = [u_1|u_2|\dots|u_m]$ este matricea ortogonală din descompunerea SVD a matricii A.

Opţional: De ce baza ortonormată $u_{r+1}, \dots, u_m$ concatenată la sistemul ortogonal de vectori $(u_1, \dots, u_r)$ conduce la o bază ortonormată în $\mathbb{R}^m$?

Având construit sistemul ortonormat de vectori $u_1, u_2, \dots, u_r$, din $\mathbb{R}^m$ determinăm o bază arbitrară în complementul ortogonal, $S^\perp$, al subspaţiului S=span$(u_1, u_2, \dots, u_r)$, pe care apoi o ortonormăm cu ajutorul procedeului lui Gramm-Schmidt.

Complementul ortogonal $S^\perp$ este format din mulţimea vectorilor di $\mathbb{R}^m$ ce sunt simultan ortogonali pe $u_1, u_2, \dots, u_r$:

$$S^\perp = \{w = (y_1, y_2, \dots, y_m)^T \in \mathbb{R}^m \,|\, < w, u_i >= 0, i = \overline{1,r}\}$$

Dar $< u_i, w >= 0$ este echivalent cu $u_i^T w = 0, \forall i = \overline{1,r}$, ceea ce înseamnă că coordonatele vectorilor w sunt soluţii ale sistemului omogen:

$$[u_1|u_2|\dots|u_r]^T w = 0$$

sau echivalent:

$$\left[\frac{1}{\sigma_1}Av_1|\frac{1}{\sigma_2}Av_2|\dots|\frac{1}{\sigma_r}Av_r\right]^T w = 0 \;\Leftrightarrow\; \left(A\left[\frac{1}{\sigma_1}v_1|\frac{1}{\sigma_2}v_2|\dots|\frac{1}{\sigma_r}v_r\right]\right)^T w = 0$$

Aplicând relaţia $(PQ)^T = Q^T P^T$ avem:

$$\left[\frac{1}{\sigma_1}v_1|\frac{1}{\sigma_2}v_2|\dots|\frac{1}{\sigma_r}v_r\right]^T A^T w = 0$$

Notăm cu C matricea $\left[\frac{1}{\sigma_1}v_1|\frac{1}{\sigma_2}v_2|\dots|\frac{1}{\sigma_r}v_r\right]^T$ Înmultind la stanga ultima relaţie cu C^T obţinem:

$$(C^T C)A^T w = 0$$

Matricea C are rangul r şi conform Propoziţiei 12.3.3 rezultă că şi C^TC are rangul r, prin urmare matricea C^TC de tip $r \times r$ este nesingulară şi deci sistemul omogen de matrice C^TC are doar soluţia banală, adică din $(C^TC)A^Tw = 0$ rezultă că $A^Tw = 0$.

Prin urmare,

$$\left[\frac{1}{\sigma_1}v_1|\frac{1}{\sigma_2}v_2|\dots|\frac{1}{\sigma_r}v_r\right]^T A^T w = 0 \;\Leftrightarrow\; A^T w = 0$$

adică ecuaţiile complementului ortogonal $S^{\perp}$, unde S=span$(u_1, u_2, \dots, u_r)$ sunt:

$$A^T w = 0$$

şi deci o bază ortonormată $u_{r+1}, \dots, u_m$, în subspaţiul Null(A^T) completează sistemul $(u_1, u_2, \dots, u_r)$ la o bază $(u_1, \dots, u_r, u_{r+1}, \dots, u_m)$ ortonormată în $\mathbb{R}^m$.

Notăm cu U matricea de trecere dela baza canonică din $\mathbb{R}^m$ la baza ortonormată $(u_1, u_2, \dots, u_m)$. $U = [u_1|u_2|\dots|u_m]$ este matrice ortogonală.

End Opţional

Etapizarea calculelor pentru determinarea descompuneii singulare a unei matrici $A \in \mathbb{R}^{m\times n}$

• se calculează produsul $M = A^TA$;

• se determină polinomul caracteristic al matricii simetrice semipozitiv definite $M = A^TA$, $P_n(\lambda) = \det(M - \lambda I_n)$ şi i se determină rădăcinile $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$;

• se determină câte o bază în fiecare subspaţiu propriu al matricii $M = A^TA$, care apoi se ortonormează folosind procedeul Gramm-Schmidt, şi reuniunea acestor baze conduce la o baza ortonormată în $\mathbb{R}^n$, formată din vectori proprii $(v_1, v_2, \dots, v_n)$;

• se constituie matricea $V = [v_1|v_2|\dots|v_n]$;

• se separă vectorii $v_1, v_2, \dots, v_r$ ce corespund respectiv valorilor proprii nenule $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$, se calculează valorile singulare $\sigma_i = \sqrt{\lambda_i}$, $i = \overline{1, r}$.

• se determină vectorii ortonormaţi $(u_1, u_2, \dots, u_r)$ din $\mathbb{R}^m$, prin $u_i = \dfrac{1}{\sigma_i}Av_i$, $i = \overline{1, r}$.

• se determină o bază arbitrară în subspaţiul Null al matricii A^T, rezolvând sistemul liniar şi omogen $A^Tx = 0$. Aplicând procedeul Gramm-Schmidt se ortonormează baza determinată obţinând astfel m-r vectori ortonormaţi, $u_{r+1}, \dots, u_m$ (deoarece dimensiunea lui Null(A^T) este egală $m - r$ adică cu numărul m de coloane ale matricii minus rang(A^T)=rang(A)=r);

• se constituie matricea $U = [u_1|u_2|\dots|u_m]$;

• Se scrie descompunerea SVD: $A = U\Sigma V^T$, ţinând seama că matricea Σ are aceleaşi dimensiuni ca şi A, adică este de tip $m \times n$.

Exemplul 6. Să se determine descompunearea singulară a matricii:

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

• Calculăm

$$M = A^T A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

• $P_2(\lambda) = (2-\lambda)^2 - 1$ şi $\lambda_1 = 3, \lambda_2 = 1$ (le-am ordonat descrescător).

• Determinăm câte o bază în fiecare subspaţiu propriu:

$$S_{\lambda=3} = \{v = \alpha(1,1)^T, \alpha \in \mathbb{R}\}, \mathcal{B}_1 = ((1,1)^T)$$

$$S_{\lambda=1} = \{v = \beta(1,-1)^T, \beta \in \mathbb{R}\}, \mathcal{B}_2 = ((1,-1)^T)$$

Vectorii celor două baze sunt ortogonali pentru că corespund la valori proprii distincte ale unei matrici simetrice ($A^T A$). Îi normăm şi obţinem baza ortonormată în $\mathbb{R}^2$:

$$\mathcal{B}' = (v_1 = (\sqrt{2}/2, \sqrt{2}/2)^T, v_2 = (\sqrt{2}/2, -\sqrt{2}/2)^T))$$

şi matricea

$$V = [v_1|v_2] = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}$$

• Calculăm valorile singulare: $\sigma_1 = \sqrt{3}, \sigma_2 = \sqrt{1} = 1$.

• Determinăm coordonatele vectorilor $u_i = \dfrac{1}{\sigma_i} A v_i$, $i = 1, 2$, unde 2 este rangul matricii $A^T A$:

$$u_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} \sqrt{6}/3 \\ \sqrt{6}/6 \\ \sqrt{6}/6 \end{bmatrix}$$

$$u_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 \\ -\sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} 0 \\ -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}$$

• Completăm sistemul ortonormat (u_1, u_2) la o bază ortonormată în $\mathbb{R}^3$. Teoretic ar trebui să determinăm o bază în subspaţiul soluţiilor sistemului $A^T x = 0$, adică:

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Notând $x_2 = \gamma$ avem:

$$\text{Null}(A^T) = \{v = \gamma(-1,1,1)^T, \gamma \in \mathbb{R}\}$$

şi deci $u_3 = (-1,1,1)^T/\sqrt{3}$.

În acest caz special am fi putut determina pe $u_3 = u_1 \times u_2$.

- matricea U este:

$$U = [u_1|u_2|u_3] = \begin{bmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \end{bmatrix}$$

- Descompunerea SVD a matricii A este:

$$A = \begin{bmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}$$