

ROZDZIAŁ 4

Rozkłady macierzy

4.1 Metoda eliminacji Gaussa i rozkład LU

Rozpocznijmy od następujących prostych przykładów

Przykład 4.1.1 Dany jest układ równań $Ax = b$, gdzie $A := \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}$, $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ i $b := \begin{bmatrix} 3 \\ 5 \end{bmatrix}$ czyli

$$\begin{aligned}x_1 + 2x_2 &= 3 \\ 4x_1 + 5x_2 &= 5\end{aligned}$$

Mnożąc pierwsze równanie przez -4 i dodając je od drugiego równania otrzymamy

$$\begin{aligned}x_1 + 2x_2 &= 3 \\ -3x_2 &= -7\end{aligned}$$

Otrzymaliśmy układ równań z górną macierzą trójkątną

$$U = \begin{bmatrix} 1 & 2 \\ 0 & -3 \end{bmatrix},$$

który rozwiązujemy od "dołu do góry": $x_2 = \frac{7}{3}$, $x_1 = -\frac{5}{3}$. Powyższe operacje zapiszemy w postaci macierzowej. Pierwsze równanie pozostawiamy bez zmian, czyli możemy je zapisać w postaci

$$\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

Mnożenie pierwszego równania przez -4 o dodanie go do drugiego równania daje nam równanie, które w postaci macierzowej możemy zapisać jako

$$\begin{bmatrix} -4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -4 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix}.$$

Ponieważ

$$\begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & -3 \end{bmatrix} \text{ i } \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \\ -7 \end{bmatrix}$$

wiec obie powyższe równości możemy zapisać w postaci

$$\begin{bmatrix} 1 & 2 \\ 0 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \\ -7 \end{bmatrix}$$

Zauważmy, że macierzą odwrotną do $\begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix}$ jest $L := \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix}$. Zatem macierz $\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}$ możemy zapisać w postaci

$$\underbrace{\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}}_A = \underbrace{\begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix}}_L \underbrace{\begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix}}_{L^{-1}} \underbrace{\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}}_A = \underbrace{\begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix}}_L \underbrace{\begin{bmatrix} 1 & 2 \\ 0 & -3 \end{bmatrix}}_U$$

zaś układ $Ax = b$ w postaci

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix}}_L \underbrace{\begin{bmatrix} 1 & 2 \\ 0 & -3 \end{bmatrix}}_U \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_x = \underbrace{\begin{bmatrix} 3 \\ 5 \end{bmatrix}}_b$$

albo inaczej w postaci dwóch układów równań

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix}}_L \underbrace{\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}}_y = \underbrace{\begin{bmatrix} 3 \\ 5 \end{bmatrix}}_b \text{ i } \underbrace{\begin{bmatrix} 1 & 2 \\ 0 & -3 \end{bmatrix}}_U \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_x = \underbrace{\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}}_y.$$

Macierz pierwszego układu jest macierzą dolną trójkątną, zaś macierz drugiego – macierzą górną trójkątną. Pierwszy z tych układów rozwiązujemy "od góry do dołu", zaś drugi – "od dołu do góry".

W przypadku ogólnym, dla macierzy kwadratowej A stopnia n i dla układu równań $Ax = b$, jeśli uda nam się zapisać macierz A w postaci $A = LU$, gdzie L jest macierzą dolną trójkątną, zaś U – macierzą górną trójkątną, możemy ten układ zapisać w postaci dwóch układów

$$Ly = b \text{ i } Ux = y,$$

które można prosto rozwiązać. Opiszemy teraz, jak otrzymać taki rozkład macierzy nieosobliwej A . Proces ten, ściśle powiązany z metodą eliminacji Gaussa, opiera się na tzw. *transformacji Gaussa*. Przypuśćmy, że dla pewnego $k \in \{1, 2, \dots, n\}$, wektor $v = (v_1, v_2, \dots, v_n)$ ma k -tą współrzędną różną od zera. Niech $t = (\underbrace{0, 0, \dots, 0}_k, \frac{v_{k+1}}{v_k}, \frac{v_{k+2}}{v_k}, \dots, \frac{v_n}{v_k})$. Macierz transformacji Gaussa

M_k ma postać

$$\begin{aligned} M_k &= I_n - te_k^T = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \frac{v_{k+1}}{v_k} \\ \frac{v_{k+2}}{v_k} \\ \vdots \\ \frac{v_n}{v_k} \end{bmatrix} \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & -\frac{v_{k+1}}{v_k} & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & -\frac{v_{k+2}}{v_k} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -\frac{v_n}{v_k} & 0 & 0 & \dots & 1 \end{bmatrix}. \end{aligned}$$

Zauważmy, że dla dowolnego k macierz M_k jest dolną macierzą trójkątną,

$$M_k v = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & -\frac{v_{k+1}}{v_k} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -\frac{v_n}{v_k} & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_k \\ v_{k+1} \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} v_1 \\ \vdots \\ v_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

oraz, że $M_k^{-1} = I_n + te_k^T$. Zatem M_k^{-1} jest również dolną macierzą trójkątną. Przy pewnych założeniach, które przedstawimy później, dla każdorazowo odpowiednio dobranych do M_i wektorach v macierz $U := M_{n-1} \dots M_1 A$ jest górną macierzą trójkątną. Zatem $A = (M_{n-1} \dots M_1)^{-1} U = M_1^{-1} \dots M_{n-1}^{-1} U$. Przy tym macierz $L := M_1^{-1} \dots M_{n-1}^{-1}$ jest dolną macierzą trójkątną jako iloczyn dolnych macierzy trójkątnych. W ten sposób możemy otrzymać rozkład $A = LU$. Proces ten przedstawimy to na przykładzie.

Przykład 4.1.2 Dana jest macierz

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}.$$

Postępując podobnie jak w metodzie eliminacji zmiennych (mnożymy pierwszy wiersz przez -4 i dodajemy do wiersza drugiego oraz mnożymy pierwszy wiersz przez -7 i dodajemy do wiersza trzeciego) otrzymamy dla wektora v równego pierwszej kolumnie macierzy A

$$M_1 A = \begin{bmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ -7 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -11 \end{bmatrix}$$

Dla wynikowej macierzy znowu stosujemy metodę eliminacji Gaussa (mnożymy drugi wiersz przez -2 i dodajemy do wiersza trzeciego). Otrzymamy dla wektora v równego drugiej kolumnie wynikowej macierzy

$$U = M_2 M_1 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -11 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 1 \end{bmatrix}$$

czyli macierz górną trójkątną. Ponieważ

$$\begin{aligned} L &= M_1^{-1} M_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ -7 & 0 & 1 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 7 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 7 & 2 & 1 \end{bmatrix}, \end{aligned}$$

więc otrzymaliśmy rozkład LU macierzy A :

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 7 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 1 \end{bmatrix}.$$

Teraz układ równań $Ax = b$ można prosto rozwiązać, np. dla $b = (1, 2, 3)$ mamy do rozwiązania dwa układy równań

$$\begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 7 & 2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \text{i} \quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}.$$

Pierwszy układ rozwiązujemy "od góry do dołu" otrzymując $y = (1, -2, 0)$, po czym rozwiązujemy drugi układ "od dołu do góry" otrzymując rozwiązanie wyjściowego układu $x = (-\frac{1}{3}, \frac{2}{3}, 0)$. Proces ten mógł być przeprowadzony bez przeszkód, ponieważ na każdym etapie mogliśmy wyznaczyć wektor v , bowiem dla macierzy A mamy $a_{11} \neq 0$ i podobnie dla macierzy $M_1 A$ element znajdujący się w drugim wierszu i w drugiej kolumnie jest również różny od zera. Tak jednak nie musi być zawsze.

Przykład 4.1.3 Dana jest macierz $A = \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}$. Ponieważ jest ona nieosobliwa, więc dla dowolnego $b \in \mathbb{R}^2$ układ $Ax = b$ posiada dokładne rozwiązanie. Nie da się jednak wyznaczyć rozkładu LU dla tej macierzy w sposób podany w przykładzie 4.1.2, bowiem $a_{11} = 0$. W tej sytuacji możemy jednak zapisać tę macierz w postaci $\bar{A} = P^T A$, gdzie $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ jest macierzą permutacji. Oczywiście układ $Ax = b$ jest równoważny układowi $\bar{A}x = \bar{b}$, gdzie $\bar{b} = P^T b$ (w drugim układzie zamieniono kolejność równań). Zgodnie z ćwiczeniem 2.2.51,

$$\bar{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 2 \end{bmatrix}.$$

Dla tej macierzy możemy już zastosować procedurę opisaną w przykładzie 4.1.2. Mamy

$$M_1 = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \quad \text{i} \quad M_1^{-1} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$$

$$U = M_1 \bar{A} = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

oraz

$$L = M_1^{-1} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}.$$

Mamy więc-

$$P^T A = \bar{A} = \begin{bmatrix} 1 & 0 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

Jak zauważyliśmy wcześniej, rozwiązania obu układów $Ax = b$ i $\bar{A}x = \bar{b}$ są identyczne, a ten drugi układ rozwiązujemy metodą podaną wyżej korzystając z rozkładu LU macierzy $\bar{A}$. W przypadku większych macierzy postępujemy podobnie mnożąc macierz A przez odpowiednią macierz permutacji. Szczegóły pozostawiamy czytelnikowi.

4.2 Ortogonalizacja Grama–Schmidta i rozkład QR

Dany jest liniowo niezależny układ wektorów $a_1, a_2, \dots, a_m \in \mathbb{R}^n$ (oczywiście musi być $m \leq n$). Należy skonstruować układ wektorów ortogonalnych $u_1, u_2, \dots, u_m \in \mathbb{R}^n$ rozpinających tę samą podprzestrzeń liniową, co układ $a_1, a_2, \dots, a_m$. Układ taki można skonstruować stosując

tw. ortogonalizację Grama-Schmidta, którą poniżej opiszemy. Dla danego wektora $v \in \mathbb{R}^n$ zdefiniujemy macierz rzutu ortogonalnego na wektor v ,

$$P_v = \frac{1}{\|v\|^2} vv^T.$$

Nietrudno zauważyć, że dla dowolnego $u \in \mathbb{R}^n$ zachodzi $P_v u = \frac{v^T u}{\|v\|^2} v$, czyli $P_v u$ jest współliniowy z v oraz, że $u - P_v u$ jest ortogonalny do v , czyli $v^T(u - P_v u) = 0$. Uzasadnia to nazwę macierzy rzutu ortogonalnego. Pojedyncza k -ta iteracja w procesie ortogonalizacji Grama-Schmidta polega na wyznaczeniu wektora u_k będącego kombinacją liniową wektorów $a_1, a_2, \dots, a_k$ i wektora ortogonalnego do poprzednio wyznaczonych wektorów $u_1, u_2, \dots, u_{k-1}$. Przy takiej konstrukcji zachodzi $\text{span}\{u_1, u_2, \dots, u_k\} = \text{span}\{a_1, a_2, \dots, a_k\}$ dla dowolnego $k = 1, 2, \dots, m$. Po wyznaczeniu wektorów $u_1, u_2, \dots, u_m$ należy je znormalizować. Konstrukcję tę można wyrazić wzorami

$$y_k = a_k - \sum_{i=1}^{k-1} P_{y_i} a_k = a_k - \sum_{i=1}^{k-1} \frac{y_i^T a_k}{\|y_i\|^2} y_i \quad (4.1)$$

i

$$u_k = \frac{y_k}{\|y_k\|}, \quad (4.2)$$

$k = 1, 2, \dots, m$. Widzimy, że dla każdego $k = 1, 2, \dots, m$ wektor y_k jest kombinacją liniową wektorów $y_1, y_2, \dots, y_{k-1}, a_k$, skąd prosto wynika, że dla każdego $k = 1, 2, \dots, m$ wektor u_k jest kombinacją liniową wektorów $a_1, a_2, \dots, a_k$. Z dowodu poniższego twierdzenia wynika, że wektory $y_1, y_2, \dots, y_m$ są niezerowe, czyli procedura Grama-Schmidta jest dobrze określona.

Twierdzenie 4.2.1 Niech układ wektorów $\{a_1, a_2, \dots, a_m\} \subseteq \mathbb{R}^n$ będzie liniowo niezależny. Wówczas dla każdego $k = 1, 2, \dots, m$ układ wektorów $\{u_1, u_2, \dots, u_k\}$ otrzymany przez ortogonalizację Grama-Schmidta jest ortogonalny i $\text{span}\{u_1, u_2, \dots, u_k\} = \text{span}\{a_1, a_2, \dots, a_k\}$.

Dowód. Unormowanie wektorów $u_1, u_2, \dots, u_m$ wynika z konstrukcji. Jednocześnie widzimy, że $\text{span}\{u_1, u_2, \dots, u_k\} = \text{span}\{y_1, y_2, \dots, y_k\}$. Pokażemy przez indukcję względem k , że układ $\{y_1, y_2, \dots, y_k\}$ jest ortogonalny dla $k = 1, 2, \dots, m$.

¹⁰ Dla $k = 1$ mamy $y_1 = a_1$ i $u_1 = \frac{y_1}{\|y_1\|}$ (oczywiście $\|y_1\| \neq 0$, bo a_1 jako składnik układu liniowo niezależnego jest niezerowy), więc dla $k = 1$ teza jest oczywista.

²⁰ Przypuśćmy, że dla pewnego $k > 1$ układ wektorów $y_1, \dots, y_{k-1}$ jest ortogonalny i

$$\text{span}\{y_1, y_2, \dots, y_{k-1}\} = \text{span}\{a_1, a_2, \dots, a_{k-1}\}.$$

Oczywiście, wektory y_i , $i = 1, 2, \dots, k-1$, są niezerowe jako składniki układu ortogonalnego. Pokażemy, że układ $\{y_1, y_2, \dots, y_k\}$:

(a) jest ortogonalny i

(b) rozpina podprzestrzeń $\text{span}\{a_1, a_2, \dots, a_k\}$.

W celu pokazania (a) wystarczy pokazać, że y_k jest ortogonalny do wektorów y_j , $j = 1, 2, \dots, k-1$. Na mocy założenia mamy

$$y_j^T y_i = \delta_{ij} = \begin{cases} 1 & \text{dla } i = j \\ 0 & \text{dla } i \neq j \end{cases}, \quad i, j = 1, 2, \dots, k-1.$$

Zatem dla $j = 1, 2, \dots, k-1$ zachodzą równości

$$y_j^T (a_k - \sum_{i=1}^{k-1} P_{y_i} a_k) = y_j^T (a_k - \sum_{i=1}^{k-1} \frac{y_i y_i^T}{\|y_i\|^2} a_k) = y_j^T a_k - \sum_{i=1}^{k-1} \frac{1}{\|y_i\|^2} \underbrace{y_j^T y_i y_i^T}_{\delta_{ij}} a_k = 0,$$

a więc układ $\{y_1, y_2, \dots, y_k\}$ jest ortogonalny.

Pokażemy (b). Na mocy założenia $y_i \in \text{span}\{a_1, a_2, \dots, a_{k-1}\}$, $i = 1, 2, \dots, k-1$. Ponadto z konstrukcji wynika, że $y_k \in \text{span}\{a_k, y_1, \dots, y_{k-1}\}$. Nietrudno więc zauważyć, że $y_k \in \text{span}\{a_1, a_2, \dots, a_k\}$. Zatem $\text{span}\{y_1, y_2, \dots, y_k\} \subseteq \text{span}\{a_1, a_2, \dots, a_k\}$. Przeciwna inkluzja wynika z faktu, że układ ortogonalny jest liniowo niezależny i z założenia, że układ $\{a_1, a_2, \dots, a_m\}$ jest liniowo niezależny.

Teza twierdzenia wynika z 1^0 i 2^0 . ■

Niech $\{a_1, a_2, \dots, a_k\} \subseteq \mathbb{R}^n$ będzie układem liniowo niezależnym. Z procedury Grama–Schmidta wynika, że dla dowolnego k wektor u_k jest kombinacją liniową wektorów a_i , $i = 1, 2, \dots, k$, czyli

$$u_k = \sum_{i=1}^k \sigma_{ik} a_i,$$

przy czym $\sigma_{kk} \neq 0$, $k = 1, 2, \dots, m$. Przyjmując $\sigma_{ik} = 0$ dla $i > k$ mamy

$$u_k = \sum_{i=1}^m \sigma_{ik} a_i,$$

$k = 1, 2, \dots, m$. Niech $A := [a_1, a_2, \dots, a_m]$. Możemy więc wektor u_k przedstawić w postaci $u_k = A s_k$, gdzie $s_k = [\sigma_{1k}, \dots, \sigma_{mk}, 0, \dots, 0]^T$, $k = 1, 2, \dots, m$. Przyjmując $Q := [u_1, u_2, \dots, u_m]$ i $S := [s_1, s_2, \dots, s_m]$ otrzymamy

$$Q = AS,$$

przy czym, zgodnie z konstrukcją, S jest górną macierzą trójkątną o dodatnich elementach na głównej przekątnej, zgodnie ze wzorami (4.1)–(4.2) mamy bowiem $\sigma_{kk} = \frac{1}{\|y_k\|} > 0$. Zatem S jest macierzą nieosobliwą i przyjmując $R := S^{-1}$ otrzymamy

$$A = QR. \quad (4.3)$$

Macierz Q typu $n \times m$ ma ortogonalne kolumny, zaś macierz R typu $m \times m$ jest górną macierzą trójkątną jako macierz odwrotna do nieosobliwej górnej macierzy trójkątnej S oraz wszystkie elementy macierzy R leżące na przekątnej są dodatnie (patrz ćwiczenie 2.2.17). Przedstawienie macierzy A w postaci (4.3) nazywa się *rozkładem QR*. Okazuje się, że rozkład taki istnieje również dla macierzy o niepełnym rzędzie kolumnowym lub wierszowym. Szczegóły pomijamy. W podręcznikach do algebry liniowej można znaleźć algorytm służący wyznaczeniu rozkładu QR dla dowolnej macierzy A .

4.3 Rozkład Cholesky’ego

Definicja 4.3.1 Niech A będzie macierzą kwadratową. Jeśli istnieje macierz dolna trójkątna L taka, że

$$A = LL^T, \quad (4.4)$$

to, równość (4.4) nazywamy *rozkładem Cholesky’ego* macierzy A .

Twierdzenie 4.3.2 Niech A będzie macierzą typu $n \times n$ określoną dodatnio. Wówczas istnieje dokładnie jedna macierz dolna trójkątna L taka, że $A = LL^T$. Ponadto macierz L ma dodatnie elementy na głównej przekątnej.

Dowód. W dalszej części pokażemy niezależnie, że macierz A możemy przedstawić w postaci $A = UDU^T$, gdzie U jest macierzą ortogonalną zaś $D = \text{diag}(d_1, d_2, \dots, d_n)$ – macierzą diagonalną o dodatnich elementach na głównej przekątnej (patrz twierdzenie 4.6.11). Mamy więc $A = B^T B$, gdzie $B = UD^{\frac{1}{2}}U^T$ zaś $D^{\frac{1}{2}} = \text{diag}(\sqrt{d_1}, \sqrt{d_2}, \dots, \sqrt{d_n})$. Zauważmy, że B jest macierzą nieosobliwą, więc istnieje jej rozkład $B = QR$. Zdefiniujmy $L = R^T$. Mamy więc $A = B^T B = R^T Q^T Q R = LL^T$. ■

Jeśli znany jest rozkład Cholesky’ego macierzy określonej dodatnio A , to można go wykorzystać do prostego rozwiązywania układu równań $Ax = b$. Układ ten możemy bowiem zapisać w postaci dwóch układów $Ly = b$ i $L^T x = y$, z których pierwszy rozwiązujemy "od góry do dołu", a następnie drugi "od dołu do góry". Istnienie i jednoznaczność rozwiązań tych układów jest zagwarantowana tym, że elementy leżące na głównej przekątnej macierzy L są dodatnie. W podręcznikach do algebry liniowej można znaleźć algorytm służący wyznaczeniu rozkładu Cholesky’ego dowolnej macierzy A .

4.4 Rozkład ortogonalny przestrzeni $\mathbb{R}^n$

Definicja 4.4.1 Niech w $\mathbb{R}^n$ określony będzie iloczyn skalarny $\langle \cdot, \cdot \rangle$. Punkt $y \in V$ nazywamy *rzutem ortogonalnym* (lub *projekcją ortogonalną*) punktu $x \in \mathbb{R}^n$ na podprzestrzeń V jeśli

$$\forall_{v \in V} \quad \langle x - y, v \rangle = 0. \quad (4.5)$$

Zapisujemy $y = P_V x$.

Uwaga 4.4.2 Rzut ortogonalny na podprzestrzeń liniową zależy od przyjętego iloczynu skalarnego.

Twierdzenie 4.4.3 Rzut ortogonalny na podprzestrzeń liniową $V \subseteq \mathbb{R}^n$ istnieje i jest wyznaczony jednoznacznie.

Dowód. Niech $\langle \cdot, \cdot \rangle$ będzie iloczynem skalarnym w $\mathbb{R}^n$ i $\| \cdot \|$ będzie indukowaną przezeń normą. Jeśli $x = 0$, to $y = 0$ jest oczywiście jedynym punktem w V spełniającym (4.5). Niech więc $x \neq 0$ i niech $y \in V$ spełnia równość

$$\|x - y\| = \min_{v \in V} \|x - v\|. \quad (4.6)$$

Oczywiście

$$\|x - y\| = \min_{v \in V \cap B(x, \|x\|)} \|x - v\|,$$

ponieważ $0 \in V \cap B(x, \|x\|) \subseteq V$, a zatem minimum w (4.6) nie może leżeć poza tą kulą. Ponieważ $\|x - \cdot\|$ jest funkcją ciągłą i zbiór $V \cap B(x, \|x\|)$ jest zwarty (w $\mathbb{R}^n$ oznacza to, że jest on domknięty i ograniczony), więc z twierdzenia Weierstrassa wynika istnienie punktu y realizującego to minimum. Pokażemy, że $P_V x = y$. Gdyby bowiem było inaczej, to ponieważ $\langle x - P_V x, y - P_V x \rangle = 0$, więc z własności iloczynu skalarnego mielibyśmy

$$\begin{aligned} \|x - y\|^2 &= \|x - P_V x + P_V x - y\|^2 \\ &= \|x - P_V x\|^2 + 2\langle x - P_V x, P_V x - y \rangle + \|P_V x - y\|^2 \\ &> \|x - P_V x\|^2, \end{aligned}$$

co prowadziłoby do sprzeczności z (4.6).

Pokażemy jednoznaczność rzutu ortogonalnego. Gdyby istniał inny niż $y = P_V x$ punkt $y_1 \in V$ spełniający (4.5), to mielibyśmy po odjęciu stronami w (4.5) dla $y = y_1$ i dla $y = P_V x$

$$\forall_{v \in V} \quad \langle P_V x - y_1, v \rangle = 0.$$

W szczególności dla $v = P_V x - y_1 \in V$ oznaczałoby to, że

$$\|P_V x - y_1\|^2 = \langle P_V x - y_1, P_V x - y_1 \rangle = 0,$$

co stałoby w sprzeczności z założeniem $y_1 \neq P_V x$. ■

Poniższe twierdzenie wyraża w inny sposób to, co było przedstawione w twierdzeniu 4.4.3.

Twierdzenie 4.4.4 *Niech V będzie podprzestrzenią liniową przestrzeni $\mathbb{R}^n$ z określonym iloczynem skalarnym $\langle \cdot, \cdot \rangle$ i indukowaną przezeń normą $\|\cdot\|$, i niech $x \in \mathbb{R}^n$. Wówczas następujące warunki są równoważne:*

- (i) *Dla dowolnego $z \in V$ zachodzi $\|x - y\| \leq \|x - z\|$;*
- (ii) *$y = P_V(x)$, czyli dla dowolnego $z \in V$ zachodzi równość $\langle z, x - y \rangle = 0$.*

Dowód. (i) $\Rightarrow$ (ii). Niech $y = P_V(x)$ i niech $z \in V$. Jeśli $z = 0$, to $\langle z, x - y \rangle = 0$. Niech więc $z \neq 0$. Dalej, niech

$$z_\lambda = y + \lambda z$$

dla $\lambda \in \mathbb{R}$. Oczywiście $z_\lambda \in V$, ponieważ V jest podprzestrzenią liniową. Z (i) i z własności iloczynu skalarnego mamy więc

$$\begin{aligned} \|x - y\|^2 &\leq \|x - z_\lambda\|^2 = \|x - y - \lambda z\|^2 \\ &= \|x - y\|^2 - 2\lambda \langle x - y, z \rangle + \lambda^2 \|z\|^2, \end{aligned}$$

a więc dla dowolnego $\lambda \in \mathbb{R}$

$$\lambda^2 \|z\|^2 - 2\lambda \langle x - y, z \rangle \geq 0,$$

w szczególności dla $\lambda = \frac{\langle x - y, z \rangle}{\|z\|^2}$ mamy $-\frac{\langle x - y, z \rangle^2}{\|z\|^2} \geq 0$, co jest możliwe tylko wtedy gdy $\langle x - y, z \rangle = 0$.

(ii) $\Rightarrow$ (i). Z własności iloczynu skalarnego oraz z (ii) mamy dla dowolnego $z \in V$

$$\|z - x\|^2 = \|z - y\|^2 + \|y - x\|^2 + 2\langle z - y, y - x \rangle \geq \|y - x\|^2,$$

co daje (i). ■

Pojęcie rzutu ortogonalnego można uogólnić na zbiory inne niż podprzestrzeń liniowa.

Definicja 4.4.5 Niech w $\mathbb{R}^n$ określony będzie iloczyn skalarny $\langle \cdot, \cdot \rangle$ i indukowana przezeń norma $\|\cdot\|$. Niech $C \subseteq \mathbb{R}^n$ i niech $x \in \mathbb{R}^n$. Punkt $y \in C$ spełniający $\|x - y\| \leq \|x - z\|$ dla dowolnego $z \in C$ nazywa się *rzutem metrycznym* punktu $x \in \mathbb{R}^n$ na podzbiór C .

Uwaga 4.4.6 Jeśli $V \subseteq \mathbb{R}^n$ jest podprzestrzenią liniową, to zgodnie z twierdzeniem 4.4.4 rzut ortogonalny P_V jest rzutem metrycznym na V i odwrotnie. W ogólnym przypadku z samej definicji nie wynika istnienie i jednoznaczność rzutu metrycznego.

Wniosek 4.4.7 *Niech w $\mathbb{R}^n$ określony będzie iloczyn skalarny $\langle \cdot, \cdot \rangle$ i indukowana przezeń norma $\|\cdot\|$. Niech $V \subseteq \mathbb{R}^n$ będzie podprzestrzenią afiniczną i niech $x \in \mathbb{R}^n$ i $y \in V$. Wówczas następujące warunki są równoważne:*

- (i) Punkt y jest rzutem metrycznym punktu x na podprzestrzeń V , czyli dla dowolnego $z \in V$ zachodzi $\|x - y\| \leq \|x - z\|$;
- (ii) Dla dowolnego $z \in V$ zachodzi równość $\langle z - y, x - y \rangle = 0$.

Ponadto punkt y spełniający (i) bądź (ii) istnieje i jest wyznaczony jednoznacznie.

Ćwiczenie 4.4.8 Udowodnić wniosek 4.4.7.

Uwaga 4.4.9 Jeśli $C \subseteq \mathbb{R}^n$ i $x \in \mathbb{R}^n$, to z samej definicji nie wynika istnienie i jednoznaczność rzutu metrycznego punktu x na podzbiór C . Można natomiast pokazać, że jeśli zbiór C jest domknięty i wypukły, to dla dowolnego $x \in \mathbb{R}^n$ rzut metryczny punktu x na C istnieje i jest wyznaczony jednoznacznie. Oznaczamy go wówczas symbolem $P_C(x)$. W szczególności, jest tak dla zbioru $C \subseteq \mathbb{R}^n$ będącego podprzestrzenią afiniczną, co wynika również z wniosku 4.4.7.

Ćwiczenie 4.4.10 Korzystając z wniosku 4.4.7 sprawdzić słuszność wzorów na rzuty metryczne punktu $x \in \mathbb{R}$ dla podanych podanych zbiorów $V \subseteq \mathbb{R}^n$:

- (a) Jeśli V jest hiperpłaszczyzną, tzn. $V = H(a, \beta) := \{z \in \mathbb{R}^n : \langle a, z \rangle = \beta\}$, gdzie $a \in \mathbb{R}^n$, $a \neq 0$, $\beta \in \mathbb{R}$, to zachodzi wzór

$$P_V(x) = x - \frac{\langle a, x \rangle - \beta}{\|a\|^2} a.$$

- (b) Jeśli V jest półprzestrzenią, tzn. $V = H_-(a, \beta) := \{z \in \mathbb{R}^n : \langle a, z \rangle \leq \beta\}$, gdzie $a \in \mathbb{R}^n$, $a \neq 0$, $\beta \in \mathbb{R}$, to zachodzi wzór

$$P_V(x) = x - \frac{(\langle a, x \rangle - \beta)_+}{\|a\|^2} a.$$

- (c) Niech $V \subseteq \mathbb{R}^n$ będzie zbiorem rozwiązań układu równań liniowych, tzn. $V = \{y \in \mathbb{R}^n : Ay = b\}$, gdzie A jest macierzą typu $m \times n$ pełnego rzędu wierszowego i $b \in \mathbb{R}^m$ (zauważyć, że układ ten posiada rozwiązanie). Wówczas zachodzi wzór

$$P_V(x) = x - A^\top(AA^\top)^{-1}(Ax - b).$$

W szczególności, dla $x = 0$ otrzymamy $P_V(0) = A^\top(AA^\top)^{-1}b$, co oznacza, że $A^\top(AA^\top)^{-1}b$ jest rozwiązaniem o najmniejszej normie równania $Ax = b$

Definicja 4.4.11 Dla macierzy kwadratowej A definiujemy rekurencyjnie jej *potęgę*: $A^1 = A$, $A^{n+1} = A^n A$, $n \in \mathbb{N}$.

Definicja 4.4.12 Mówimy, że macierz kwadratowa A jest *idempotentna*, jeśli $A^2 = A$.

Ćwiczenie 4.4.13 Niech A będzie macierzą pełnego rzędu wierszowego. Pokazać, że macierze $A^\top(AA^\top)^{-1}A$ i $I - A^\top(AA^\top)^{-1}A$ są idempotentne.

Ćwiczenie 4.4.14 Niech A będzie macierzą pełnego rzędu kolumnowego. Pokazać, że macierze $A(A^\top A)^{-1}A^\top$ i $I - A(A^\top A)^{-1}A^\top$ są idempotentne.

Definicja 4.4.15 Niech w $\mathbb{R}^n$ będzie określony iloczyn skalarny $\langle \cdot, \cdot \rangle$ i niech $V \subseteq \mathbb{R}^n$ będzie podprzestrzenią liniową. Zbiór

$$V^\perp := \{y \in \mathbb{R}^n : \forall_{x \in V} \quad \langle y, x \rangle = 0\}$$

nazywamy *podprzestrzenią ortogonalną* do V (lub *uzupełnieniem ortogonalnym* podprzestrzeni V)

Ćwiczenie 4.4.16 Niech $V \subseteq \mathbb{R}^n$ będzie podprzestrzenią liniową. Pokazać, że $V^\perp$ jest podprzestrzenią liniową.

Twierdzenie 4.4.17 (rozkład ortogonalny przestrzeni) Niech w $\mathbb{R}^n$ będzie określony iloczyn skalarny $\langle \cdot, \cdot \rangle$ i niech $V \subseteq \mathbb{R}^n$ będzie podprzestrzenią liniową. Wówczas dla dowolnego $x \in \mathbb{R}^n$ istnieją $y \in V$ i $z \in V^\perp$ takie, że

$$x = y + z \tag{4.7}$$

i są one wyznaczone jednoznacznie. Ponadto $y = P_V x$ i $z = P_{V^\perp} x$.

Dowód. Zdefiniujmy $y := P_V x$ i $z := x - P_V x$. Oczywiście $y \in V$. Pokażemy, że $z \in V^\perp$. Z definicji rzutu ortogonalnego, dla dowolnego $v \in V$ mamy

$$\langle z, v \rangle = \langle x - P_V x, v \rangle = 0,$$

a więc $z \in V^\perp$. Jednocześnie, z definicji rzutu ortogonalnego na $V^\perp$ wynika, że dla dowolnego $u \in V^\perp$

$$\langle x - (x - P_V x), u \rangle = \langle P_V x, u \rangle = 0,$$

czyli $z := x - P_V x = P_{V^\perp} x$. Otrzymaliśmy więc

$$x = \underbrace{P_V x}_{\in V} + \underbrace{x - P_V x}_{\in V^\perp}.$$

Pokażemy teraz jednoznaczność. Gdyby istniały $y_1 \in V$ i $z_1 \in V^\perp$ takie, że $y_1 \neq y$ i $z_1 \neq z$ oraz $x = y_1 + z_1$, to odejmując stronami ostatnią równość i (4.7) otrzymalibyśmy $y - y_1 = z - z_1$, a ponieważ $y - y_1 \in V$ i $z - z_1 \in V^\perp$, więc byłoby to możliwe jedynie w przypadku, gdy $y - y_1 = z - z_1 = 0$. Uzyskana sprzeczność dowodzi zatem, że rozkład (4.7) jest jednoznaczny. ■

Twierdzenie 4.4.17 możemy skrótowo zapisać w postaci

$$\mathbb{R}^n = V \oplus V^\perp, \tag{4.8}$$

gdzie $\oplus$ jest symbolem tzw. *sumy prostej*. Równość (4.8) mówi, że przestrzeń $\mathbb{R}^n$ jest sumą prostą podprzestrzeni V i podprzestrzeni do niej ortogonalnej, czyli dowolny element przestrzeni $\mathbb{R}^n$ daje się jednoznacznie zapisać jako sumę elementu podprzestrzeni V i elementu podprzestrzeni $V^\perp$.

Definicja 4.4.18 Niech A będzie macierzą typu $m \times n$. Podprzestrzeń

$$\ker A := \{x \in \mathbb{R}^n : Ax = 0\} \tag{4.9}$$

nazywamy *jądrem* macierzy A lub *podprzestrzenią zerową* tej macierzy. Podprzestrzeń

$$\operatorname{im} A := \{y \in \mathbb{R}^m : \exists_{x \in \mathbb{R}^n} \quad y = Ax\} \tag{4.10}$$

nazywamy *obrazem* macierzy A .

Uwaga 4.4.19 W niektórych podręcznikach używa się symbolu $\mathcal{N}(A)$ na oznaczenie jądra macierzy A i symbolu $\mathcal{R}(A)$ na oznaczenie obrazu macierzy A .

W podobny sposób definiujemy podprzestrzenie $\ker A^T$ i $\operatorname{im} A^T$. Te cztery podprzestrzenie $\ker A$, $\operatorname{im} A$, $\ker A^T$ i $\operatorname{im} A^T$ są podstawowymi podprzestrzeniami związanymi z daną macierzą A , co poniżej uzasadnimy.

Twierdzenie 4.4.20 *Niech A będzie macierzą typu $m \times n$. Wówczas*

$$(\ker A)^\perp = \operatorname{im} A^T \text{ oraz } (\operatorname{im} A)^\perp = \ker A^T. \quad (4.11)$$

Dowód. Niech $y \in \ker A$ i $z \in \operatorname{im} A^T$ będą dowolne. Zatem $Ay = 0$ i $A^T u = z$ dla pewnego $u \in \mathbb{R}^n$. Mamy wówczas

$$\langle y, z \rangle = \langle y, A^T u \rangle = \langle Ay, u \rangle = 0.$$

Zatem $\operatorname{im} A^T = (\ker A)^\perp$. Drugą równość pokazuje się podobnie. ■

W przypadku, gdy $V = \ker A$ bądź $V = \operatorname{im} A$ dla macierzy A typu $m \times n$ równość (4.8) możemy zapisać w postaci

$$\mathbb{R}^n = \ker A \oplus \operatorname{im} A^T$$

bądź w postaci

$$\mathbb{R}^m = \operatorname{im} A \oplus \ker A^T.$$

Ponadto mamy

$$n = \dim(\ker A) + \dim(\operatorname{im} A^T)$$

oraz

$$m = \dim(\operatorname{im} A) + \dim(\ker A^T).$$

Ćwiczenie 4.4.21 Pokazać, że dla macierzy A pełnego rzędu wierszowego zachodzą wzory

$$P_{\ker A}(x) = x - A^T(AA^T)^{-1}Ax \quad (4.12)$$

oraz

$$P_{\operatorname{im} A^T}(x) = A^T(AA^T)^{-1}Ax. \quad (4.13)$$

Porównać równości (4.12) i (4.13) z ćwiczeniem 4.4.13.

Wskazówka. W celu pokazania wzoru (4.12) zauważyć, że $x - A^T(AA^T)^{-1}Ax \in \ker A$, a dalej skorzystać z twierdzenia 4.4.4. W celu pokazania wzoru (4.13) skorzystać z tego, że $A^T(AA^T)^{-1}Ax \in \operatorname{im} A^T = (\ker A)^\perp$ oraz z twierdzenia 4.4.17.

Twierdzenie 4.4.22 *Niech A będzie macierzą typu $m \times n$. Wówczas*

(i)

$$A(\operatorname{im} A^T) = \operatorname{im} A \quad (4.14)$$

oraz

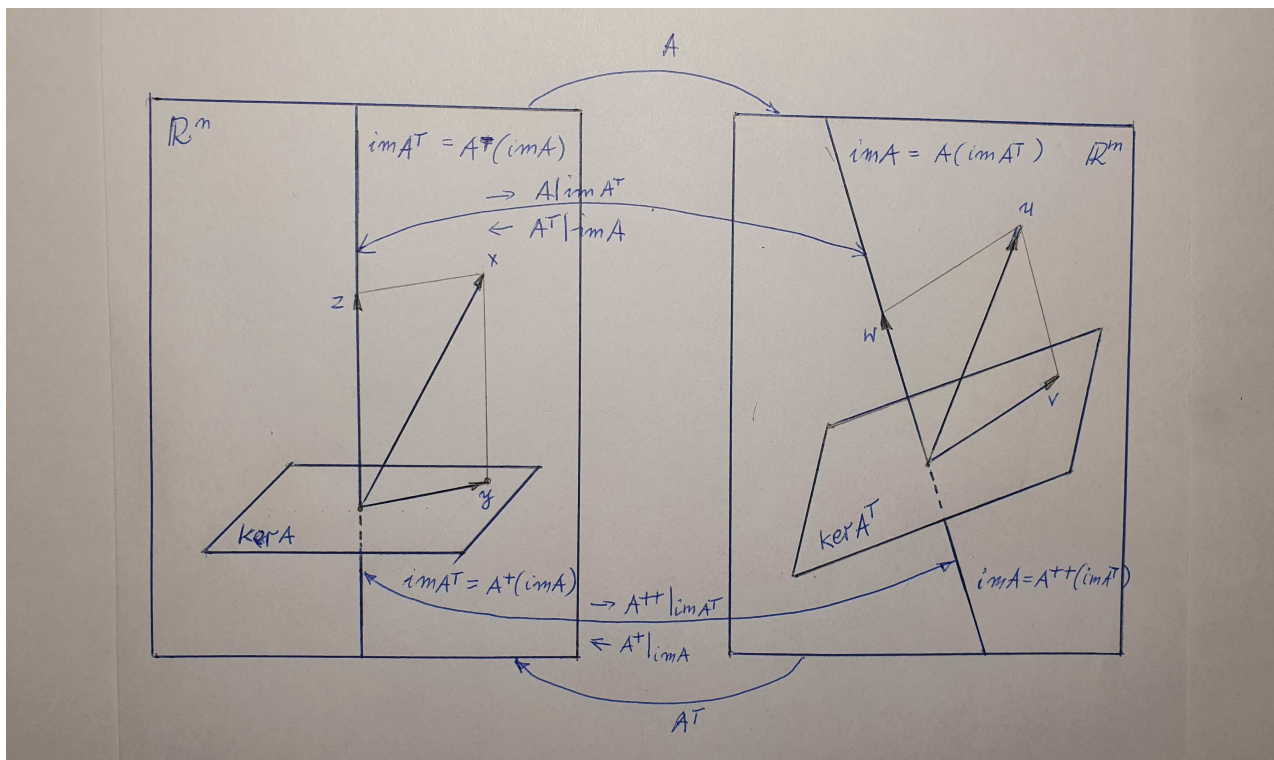
$$A^T(\operatorname{im} A) = \operatorname{im} A^T. \quad (4.15)$$

(ii) Operatory $A : \operatorname{im} A^T \rightarrow \operatorname{im} A$ i $A^T : \operatorname{im} A \rightarrow \operatorname{im} A^T$ są różnowartościowe.

(iii) Zachodzi

$$\dim(\operatorname{im} A) = \dim(\operatorname{im} A^T) = \operatorname{rank}(A), \quad (4.16)$$

gdzie $\operatorname{rank}(A)$ oznacza rząd macierzy A .

Odwzorowania liniowe A i A^T

Dowód. (i) Inkluzja $\subseteq$ w (4.14) jest oczywista. Niech więc $w \in \text{im } A$, czyli $w = Ax$ dla pewnego $x \in \mathbb{R}^n$. Z rozkładu ortogonalnego wiemy, że $x = y + z$, gdzie $y \in \ker A$ i $z \in \text{im } A^T$. Zatem $w = Ax = A(y + z) = Az$, czyli $w \in A(\text{im } A^T)$. Równości (4.15) dowodzi się podobnie.

(ii) Weźmy $x_1, x_2 \in \text{im } A^T$, $x_1 \neq x_2$. Wówczas $x_1 = A^T y_1$ i $x_2 = A^T y_2$ dla pewnych $y_1, y_2 \in \mathbb{R}^m$. Przypuśćmy, że $Ax_1 = Ax_2$. Mamy wówczas

$$0 \neq \|x_1 - x_2\|^2 = \langle x_1 - x_2, x_1 - x_2 \rangle = \langle x_1 - x_2, A^T y_1 - A^T y_2 \rangle = \langle Ax_1 - Ax_2, y_1 - y_2 \rangle = 0.$$

Uzyskana sprzeczność dowodzi tego, że operator $A : \text{im } A^T \rightarrow \text{im } A$ jest różnowartościowy. W podobny sposób dowodzi się różnowartościowości operatora $A^T : \text{im } A \rightarrow \text{im } A^T$. Zatem A przekształca wzajemnie jednoznacznie $\text{im } A^T$ na $\text{im } A$ oraz A^T przekształca wzajemnie jednoznacznie $\text{im } A$ na $\text{im } A^T$.

(iii) Ponieważ podprzestrzeń $\text{im } A$ stanowią wszystkie kombinacje liniowe kolumn macierzy A , więc $\dim(\text{im } A)$ jest liczbą liniowo niezależnych kolumn tej macierzy, czyli jest równa jej rzędowi. Z drugiej strony rząd macierzy A jest również liczbą jej liniowo niezależnych wierszy, co dowodzi prawdziwości (4.16). ■

4.5 Wartości własne i wektory własne macierzy

Wartość własna macierzy kwadratowej jest jednym najważniejszych pojęć algebry liniowej. W tym ustępie przedstawimy własności tego pojęcia. Będą one następnie służyć rozkładowi macierzy symetrycznej na wartości własne i rozkładowi macierzy na wartości osobliwe. Rozkłady te mają ważne zastosowania w wielu dziedzinach w tym w uczeniu maszynowym. Będziemy tu rozważać macierze o elementach rzeczywistych, ale część tych własności przysługuje również macierzom o elementach zespolonych.

Definicja 4.5.1 Liczbę $\lambda \in \mathbb{C}$ nazywamy *wartością własną* macierzy kwadratowej A typu $n \times n$ o wartościach rzeczywistych, jeśli istnieje niezerowy wektor $v \in \mathbb{C}^n$ taki, że

$$Av = \lambda v. \quad (4.17)$$

Dowolny wektor $v \in \mathbb{C}^n$ spełniający równość (4.17) nazywamy *wektorem własnym* macierzy A odpowiadającym wartości własnej λ .

Mimo, że A jest macierzą o wartościach rzeczywistych, zarówno wartość własna może być liczbą zespoloną, jak i wektor własny może mieć zespolone współrzędne. Oczywiście wektor zerowy jest wektorem własnym odpowiadającym dowolnej wartości własnej, gdyż dla dowolnego $\lambda \in \mathbb{C}$ równość (4.17) jest spełniona gdy $x = 0$. W przypadku, gdy wartość własna macierzy A jest liczbą rzeczywistą, to wyznaczenie odpowiadającego jej wektora własnego możemy ograniczyć do $\mathbb{R}^n$.

Definicja 4.5.2 Niech A będzie macierzą kwadratową typu $n \times n$ o elementach rzeczywistych. Wielomian stopnia n postaci $\det(A - \lambda I)$ nazywamy *wielomianem charakterystycznym* macierzy A .

Twierdzenie 4.5.3 Wielkość λ jest wartością własną macierzy A typu $n \times n$ wtedy i tylko wtedy, gdy macierz $A - \lambda I$ jest osobliwa, czyli $\det(A - \lambda I) = 0$. W konsekwencji macierz A posiada n wartości własnych, przy czym niektóre z nich mogą się powtarzać.

Dowód. Równość (4.17) możemy zapisać w postaci $(A - \lambda I)v = 0$. Widzimy, że λ jest wartością własną macierzy A wtedy i tylko wtedy, gdy równanie jednorodne $(A - \lambda I)v = 0$ posiada niezerowe rozwiązanie. Oznacza to, że macierz $A - \lambda I$ jest osobliwa, czyli wszystkie wartości własne macierzy A są rozwiązaniami równania $\det(A - \lambda I) = 0$. Zgodnie z zaadniczym twierdzeniem algebry, pierwiastków tego równania jest n , przy czym niektóre mogą się powtarzać. ■

Mając wyznaczoną wartość własną $\lambda \in \mathbb{C}$ macierzy A , odpowiadający jej wektor własny v możemy wyznaczyć rozwiązując równanie jednorodne $(A - \lambda I)v = 0$. Jak już zauważyliśmy wcześniej, jeśli λ jest liczbą rzeczywistą, to możemy ograniczyć się do wektorów własnych $v \in \mathbb{R}^n$.

Ćwiczenie 4.5.4 Niech λ będzie wartością własną macierzy A typu $n \times n$ i niech C będzie macierzą nieosobliwą tego samego typu. Pokazać, że λ jest wartością własną macierzy $B := CAC^{-1}$. Wywnioskować stąd, że obie macierze A i CAC^{-1} mają te same wartości własne.

Ćwiczenie 4.5.5 Niech A będzie macierzą kwadratową i niech λ będzie jej wartością własną zaś u odpowiadającym jej wektorem własnym. Pokazać, że:

- (a) Dla dowolnego $k \in \mathbb{N}$ wielkość λ^k jest wartością własną macierzy A^k i u jest odpowiadającym jej wektorem własnym;
- (b) Jeśli dodatkowo założymy, że A jest nieosobliwa, to λ^{-1} jest wartością własną macierzy A^{-1} i u jest odpowiadającym jej wektorem własnym.

Twierdzenie 4.5.6 Zbiór wszystkich wektorów własnych odpowiadających ustalonej wartości własnej λ macierzy A jest podprzestrzenią liniową.

Dowód. Niech $\lambda \in \mathbb{C}$ będzie wartością własną macierzy A , zaś $u, v \in \mathbb{C}^n$ będą odpowiadającymi jej wektorami własnymi. Wówczas dla dowolnych $\alpha, \beta \in \mathbb{C}$ zachodzi

$$A(\alpha u + \beta v) = \alpha Au + \beta Av = \alpha \lambda u + \beta \lambda v = \lambda(\alpha u + \beta v),$$

co oznacza, że $\alpha u + \beta v$ jest wektorem własnym odpowiadającym wartości własnej λ . Zatem zbiór wektorów własnych odpowiadających wartości własnej λ jest podprzestrzenią liniową. ■

4.5.1 Wektory własne i wartości własne macierzy symetrycznej i macierzy ortogonalnej

Twierdzenie 4.5.7 *Wszystkie wartości własne macierzy symetrycznej są liczbami rzeczywistymi. W konsekwencji dowolny wektor własny macierzy symetrycznej ma współrzędne rzeczywiste.*

Dowód. Niech $\lambda \in \mathbb{C}$ będzie wartością własną macierzy symetrycznej A (czyli $A = A^T$) i $v \neq 0$ odpowiadającym jej wektorem własnym. Z aksjomatów iloczynu skalarnego w $\mathbb{C}^n$ mamy

$$\lambda \langle v, v \rangle = \langle Av, v \rangle = \langle v, A^T v \rangle = \langle v, Av \rangle = \langle v, \lambda v \rangle = \bar{\lambda} \langle v, v \rangle,$$

a ponieważ $v \neq 0$, więc $\lambda = \bar{\lambda}$, czyli $\lambda \in \mathbb{R}$. ■

Twierdzenie 4.5.8 *Wektory własne odpowiadające różnym wartościom własnym macierzy symetrycznej A są wzajemnie ortogonalne.*

Dowód. Niech $Au = \lambda u$ i $Av = \mu v$ dla $\lambda, \mu \in \mathbb{R}$, $\lambda \neq \mu$ oraz $u, v \in \mathbb{R}^n$. Wówczas

$$\lambda \langle u, v \rangle = \langle Au, v \rangle = \langle u, Av \rangle = \langle u, \mu v \rangle = \mu \langle u, v \rangle.$$

Oznacza to, że $\langle u, v \rangle = 0$. ■

W twierdzeniu 4.5.8 założenie, że macierz jest symetryczna jest istotne. Nawet małe odstępstwo od symetrii macierzy może spowodować dużą zmianę w kącie między wektorami własnymi.

Ćwiczenie 4.5.9 Niech

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

i

$$B = \begin{bmatrix} 1 & \varepsilon \\ 0 & 1 + \varepsilon \end{bmatrix}$$

dla małego $\varepsilon > 0$. Jakie są wartości własne i wektory własne macierzy A i B ? Jaki jest kąt między wektorami własnymi macierzy A i jaki jest kąt między wektorami własnymi macierzy B ?

Twierdzenie 4.5.10 *Wszystkie wartości własne macierzy ortogonalnej U mają moduł równy 1.*

Dowód. Niech U będzie macierzą ortogonalną, $\lambda \in \mathbb{C}$ jej wartością własną i $u \in \mathbb{C}^n$ odpowiadającym jej niezerowym wektorem własnym. Wówczas

$$\langle Uu, Uu \rangle = \langle \lambda u, \lambda u \rangle = \lambda \bar{\lambda} \langle u, u \rangle = |\lambda|^2 \langle u, u \rangle = |\lambda|^2 \langle u, U^T U u \rangle = |\lambda|^2 \langle Uu, Uu \rangle.$$

Stąd $|\lambda|^2 = 1$, czyli $|\lambda| = 1$. ■

Okazuje się, że własność macierzy symetrycznych wyrażona w twierdzeniu 4.5.8 przysługuje również macierzom ortogonalnym.

Twierdzenie 4.5.11 *Wektory własne odpowiadające różnym wartościom własnym macierzy ortogonalnej U są wzajemnie ortogonalne.*

Dowód. Niech $\lambda, \mu \in \mathbb{C}$ będą dwiema różnymi wartościami własnymi macierzy ortogonalnej U i $u, v \in \mathbb{C}^n$ odpowiadającym im wektorom własnym. Najpierw pokażemy, że $\lambda \bar{\mu} \neq 1$. Przypuśćmy bowiem, że jest przeciwnie, czyli $\lambda \bar{\mu} = 1$. Wówczas, na mocy twierdzenia 4.5.10

$$\mu = \lambda \bar{\mu} \mu = \lambda |\mu|^2 = \lambda,$$

co stoi w sprzeczności z założeniem, że $\lambda \neq \mu$. Zatem $\lambda \bar{\mu} \neq 1$. W konsekwencji

$$\lambda \bar{\mu} \langle u, v \rangle = \langle \lambda u, \mu v \rangle = \langle Uu, Uv \rangle = \langle u, U^T U v \rangle = \langle u, v \rangle,$$

co jest możliwe tylko wtedy gdy $\langle u, v \rangle = 0$. ■

4.6 Rozkład macierzy na wartości własne

Rozkład macierzy symetrycznej na wartości własne jest ważnym elementem algebry liniowej zarówno z teoretycznego jak i praktycznego punktu widzenia. Jest on szeroko stosowany w wielu naukach inżynierskich, w tym w uczeniu maszynowym. Rozkład ten jest podstawą do rozkładu macierzy na wartości osobliwe, a ten z kolei ma zastosowanie w tzw. właściwym rozkładzie ortogonalnym, czy w analizie głównych składowych, które mają zastosowanie na przykład w analizie danych czy w statystyce.

Definicja 4.6.1 Niech A będzie macierzą kwadratową. Jeśli istnieje macierz ortogonalna U i macierz diagonalna $D = \text{diag}(d_1, d_2, \dots, d_n)$, gdzie $d_j \in \mathbb{R}, j = 1, 2, \dots, n$, takie, że $A = UDU^T$, to równość tę nazywamy *rozkładem macierzy A na wartości własne*.

Zanim przejdziemy do warunków istnienia takiego rozkładu, podamy jego interpretację geometryczną. Macierz U jako macierz ortogonalna jest macierzą przekształcenia liniowego nie zmieniającego długości wektorów. W istocie takie przekształcenie jest złożeniem obrotów i odbić. W konsekwencji, macierz U jest iloczynem macierzy obrotów Givensa i macierzy odbić Householdera. Jak już wiemy, macierz U^T jest macierzą odwrotną do U , czyli $U^{-1} = U^T$. Z drugiej strony macierz diagonalna D działa na wektory w ten sposób, że przeskalowuje jego współrzędne, a dokładniej przemnaża jego współrzędne przez wartości leżące na głównej przekątnej macierzy D , czyli $Dx = (d_1x_1, d_2x_2, \dots, d_nx_n)$. Zobaczmy więc na czym polega przekształcenie liniowe, którego macierzą jest UDU^T . Przekształcenie to działa na wektor x w ten sposób, że najpierw nakłada na niego przekształcenie ortogonalne U^T (czyli niezmieniające długości wektora x) w wektor $y := U^{-1}x$. Na przykład U^T może być macierzą obrotu o kąt α . Następnie wektor y jest przeskalowany przez macierz diagonalną D , czyli $w = Dy$ (jego współrzędne zostają wydłużone lub skrócone z ewentualną zmianą znaku tej współrzędnej, zgodnie z wartościami na przekątnej macierzy D). Na koniec na wektor w działamy macierzą ortogonalną U , czyli odwrotną do U^T . Ta macierz również nie zmienia długości wektorów. W konsekwencji wynikiem złożenia tych trzech działań jest wektor $z = UDU^Tx$.

Twierdzenie 4.6.2 Niech A będzie rzeczywistą macierzą symetryczną typu $n \times n$. Wówczas istnieje macierz ortogonalna U i macierz diagonalna $D = \text{diag}(d_1, d_2, \dots, d_n)$, gdzie $d_j \in \mathbb{R}, j = 1, 2, \dots, n$, takie, że

$$A = UDU^T, \quad (4.18)$$

przy czym d_j są wartościami własnymi macierzy A , a kolumny u_j macierzy U są odpowiadającymi im wektorami własnymi, $j = 1, 2, \dots, n$.

Dowód. Niech $d_1 \in \mathbb{R}$ będzie wartością własną macierzy A zaś $v_1 \in \mathbb{R}^n$ – odpowiadającym jej unormowanym wektorem własnym, czyli $Av_1 = d_1v_1$ i $\|v_1\| = 1$. Niech $R_1 := R_{v_1-e_1}$ będzie macierzą Householdera przeprowadzającą e_1 w v_1 i jednocześnie v_1 w e_1 (patrz ćwiczenie 2.2.52). Wówczas mamy

$$(R_1^T A R_1)e_1 = R_1^T A v_1 = d_1 R_1^T v_1 = d_1 e_1. \quad (4.19)$$

Zatem pierwsza kolumna macierzy $R_1^T A R_1$ jest wektorem $d_1 e_1$ i z uwagi na symetryczność macierzy A i R_1 , macierz $R_1^T A R_1$ jest symetryczna i

$$R_1^T A R_1 = \begin{bmatrix} d_1 & 0 \\ 0 & A_1 \end{bmatrix}, \quad (4.20)$$

gdzie macierz A_1 typu $(n-1) \times (n-1)$ jest również symetryczna. Mnożąc tę równość z lewej strony przez R_1 i z prawej przez R_1^T i pamiętając, że R_1 jest macierzą ortogonalną otrzymamy

$$A = R_1 \begin{bmatrix} d_1 & 0 \\ 0 & A_1 \end{bmatrix} R_1^T \quad (4.21)$$

Postępując w podobny sposób z macierzą A_1 otrzymamy

$$A_1 = R_2 \begin{bmatrix} d_2 & 0 \\ 0 & A_2 \end{bmatrix} R_2^T, \quad (4.22)$$

gdzie $d_2 \in \mathbb{R}$ jest jej wartością własną, $v_2 \in \mathbb{R}^{n-1}$ odpowiadającym jej unormowanym wektorem własnym i $R_2 := R_{e_1-v_2}$ macierzą Householdera w $\mathbb{R}^{n-1}$ przeprowadzającą $e_1 \in \mathbb{R}^{n-1}$ w v_2 i jednocześnie v_2 w e_1 , zaś A_2 jest macierzą symetryczną typu $(n-2) \times (n-2)$. Oznaczając

$$U_2 = R_1 \begin{bmatrix} 1 & 0 \\ 0 & R_2 \end{bmatrix}$$

otrzymujemy więc

$$\begin{aligned} A &= R_1 \begin{bmatrix} d_1 & 0 \\ 0 & R_2 \begin{bmatrix} d_2 & 0 \\ 0 & A_2 \end{bmatrix} R_2^T \end{bmatrix} R_1^T \\ &= R_1 \begin{bmatrix} 1 & 0 \\ 0 & R_2 \end{bmatrix} \begin{bmatrix} d_1 & 0 \\ 0 & \begin{bmatrix} d_2 & 0 \\ 0 & A_2 \end{bmatrix} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & R_2^T \end{bmatrix} R_1^T \\ &= R_1 \begin{bmatrix} 1 & 0 \\ 0 & R_2 \end{bmatrix} \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & A_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & R_2^T \end{bmatrix} R_1^T \\ &= U_2 \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & A_2 \end{bmatrix} U_2^T \end{aligned}$$

Zauważmy przy tym, że macierz $\begin{bmatrix} 1 & 0 \\ 0 & R_2 \end{bmatrix}$ jest ortogonalna, a więc i macierz U_2 typu $n \times n$ jest ortogonalna jako iloczyn macierzy ortogonalnych. Postępując dalej przez indukcję otrzymamy ostatecznie $A = UDU^T$ gdzie $U = [u_1, \dots, u_n]$ jest macierzą ortogonalną o kolumnach $u_j \in \mathbb{R}^n$, $j = 1, 2, \dots, n$. Mnożąc tę równość obustronnie przez U otrzymamy

$$AU = UD,$$

czyli

$$A[u_1, u_2, \dots, u_n] = [u_1, u_2, \dots, u_n]D = [d_1u_1, d_2u_2, \dots, d_nu_n],$$

co możemy inaczej zapisać $Au_j = d_ju_j$, $j = 1, 2, \dots, n$. Zatem d_j są wartościami własnymi macierzy A zaś u_j odpowiadającymi im wektorami własnymi, $j = 1, 2, \dots, n$, stanowiącymi układ ortogonalny. Ponieważ macierz symetryczna posiada dokładnie n wartości własnych, więc są nimi d_j , $j = 1, 2, \dots, n$. ■

Rozkład macierzy symetrycznej A na wartości własne przedstawiony w twierdzeniu 4.6.2 nazywa się również *rozkładem Schura*.

Uwaga 4.6.3 Niech $\lambda \in \mathbb{R}$ jest wartością własną macierzy A_1 wymiaru $(n-1) \times (n-1)$ występującej w dowodzie twierdzenia 4.6.2 zaś $u \in \mathbb{R}^{n-1}$ będzie odpowiadającym jej wektorem własnym, czyli $A_1u = \lambda u$. Wówczas λ jest wartością własną macierzy A zaś $v = R_1 \begin{bmatrix} 0 \\ u \end{bmatrix}$ jest

odpowiadającym jej wektorem własnym. Istotnie. Na mocy wzoru (4.21) mamy

$$\begin{aligned} Av &= AR_1 \begin{bmatrix} 0 \\ u \end{bmatrix} = R_1 \begin{bmatrix} d_1 & 0 \\ 0 & A_1 \end{bmatrix} R_1^T R_1 \begin{bmatrix} 0 \\ u \end{bmatrix} \\ &= R_1 \begin{bmatrix} d_1 & 0 \\ 0 & A_1 \end{bmatrix} \begin{bmatrix} 0 \\ u \end{bmatrix} = R_1 \begin{bmatrix} 0 \\ A_1 u \end{bmatrix} \\ &= R_1 \begin{bmatrix} 0 \\ \lambda u \end{bmatrix} = \lambda R_1 \begin{bmatrix} 0 \\ u \end{bmatrix} = \lambda v \end{aligned}$$

Definicja 4.6.4 Niech A będzie macierzą kwadratową o elementach rzeczywistych. Krotność pierwiastka λ równania charakterystycznego $\det(A - \lambda I) = 0$ nazywamy *krotnością wartości własnej* λ macierzy A .

Ćwiczenie 4.6.5 Pokazać, że rząd macierzy symetrycznej jest liczbą niezerowych wartości własnych tej macierzy z uwzględnieniem ich krotności.

Wskazówka. Skorzystać z twierdzenia 4.6.2 i ćwiczenia 2.2.8.

Ćwiczenie 4.6.6 Pokazać, że krotność wartości własnej macierzy symetrycznej jest równa wymiarowi podprzestrzeni liniowej odpowiadającym jej wektorów własnych.

Ćwiczenie 4.6.7 Pokazać, że dla niezerowego wektora u macierz uu^T jest macierzą rzędu pierwszego, a więc posiada tylko jedną niezerową i jednokrotną wartość własną $\lambda = \|u\|^2$, zaś u jest odpowiadającym jej wektorem własnym. W szczególności, jeśli u jest wektorem unormowanym, to $\lambda = 1$. Poza tym macierz ta ma $(n - 1)$ -krotną wartość własną $\lambda = 0$, a $(n - 1)$ -wymiarową podprzestrzenią odpowiadających jej wektorów własnych jest $W = \{w \in \mathbb{R}^n : u^T w = 0\}$.

Uwaga 4.6.8 Rozkład (4.18) możemy zapisać w postaci,

$$A = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \dots + \lambda_n u_n u_n^T.$$

Widzimy więc, że macierz symetryczną możemy zapisać jako sumę macierzy rzędu pierwszego $\lambda_j u_j u_j^T$ o wartościach własnych λ_j , $j = 1, 2, \dots, n$.

Ćwiczenie 4.6.9 Pokazać, że dla macierzy symetrycznej G stopnia n o wartościach własnych λ_i , $i = 1, 2, \dots, n$ zachodzi

$$\operatorname{tr}(G) = \sum_{j=1}^n \lambda_j. \quad (4.23)$$

Wskazówka.

- (a) Zauważyć, że dla macierzy kwadratowych A i B stopnia n zachodzi $\operatorname{tr}(AB) = \operatorname{tr}(BA)$;
- (b) Korzystając z (a) pokazać, że dla dowolnej macierzy niesobliwej C zachodzi $\operatorname{tr}(C^{-1}AC) = \operatorname{tr}(A)$;
- (c) Dla dowodu (4.23) skorzystać z (b) i z twierdzenia 4.6.2.

Ćwiczenie 4.6.10 Pokazać, że:

- (a) Jeśli w twierdzeniu 4.6.2 założymy dodatkowo, że macierz A jest dodatnio (nieujemnie) określona, to jej wszystkie wartości własne są dodatnie (nieujemne);

- (b) Prawdziwa jest również implikacja odwrotna, tzn, jeśli wszystkie wartości własne macierzy A są dodatnie (nieujemne), to macierz A jest dodatnio (nieujemnie) określona.

Podsumujemy teraz poznane fakty dotyczące macierzy dodatnio (nieujemnie) określonych.

Twierdzenie 4.6.11 *Dla macierzy symetrycznej A stopnia n następujące warunki są równoważne:*

- (i) A jest określona dodatnio (nieujemnie);
- (ii) wszystkie wartości własne macierzy A są dodatnie (nieujemne);
- (iii) istnieje macierz ortogonalna U i macierz diagonalna D o dodatnich (nieujemnych) elementach na głównej przekątnej takie, że $A = UDU^T$ (tzw. rozkład według wartości własnych, ang. *eigenvalue decomposition*);
- (iv) istnieje określona dodatnio (nieujemnie) macierz G taka, że $A = G^2$.

Równoważność (i) $\iff$ (iv) daje podstawę do określania pierwiastka z macierzy nieujemnie określonej, podobnie jak definiuje się pierwiastek z liczby nieujemnej.

Definicja 4.6.12 Niech A będzie macierzą nieujemnie określoną, macierz G spełniająca równość $A = G^2$ nazywa się *pierwiastkiem z macierzy A* . Piszemy $G = A^{\frac{1}{2}}$.

Uwaga 4.6.13 Dla dodatniej określoności macierzy podaje się czasem również warunek konieczny i wystarczający w postaci:

- (v) wszystkie wiodące minory główne $\Delta_1, \Delta_2, \dots, \Delta_n$ są dodatnie (tzw. twierdzenie Sylwestra).

Twierdzenie to nie przenosi się jednak bezpośrednio na przypadek nieujemnej określoności macierzy, tzn. z nieujemności wiodących minorów głównych nie wynika nieujemna określoność macierzy. Natomiast prawdziwe jest stwierdzenie, że macierz A jest nieujemnie określona wtedy i tylko wtedy, gdy wszystkie jej minory główne (a nie tylko wiodące minory główne $\Delta_1, \Delta_2, \dots, \Delta_n$) są nieujemne. Sprawdzenie tego warunku dla macierzy większego wymiaru jest jednak kłopotliwe (wszystkich minorów głównych macierzy kwadratowej stopnia n jest $2^n - 1$).

Przykład 4.6.14 Dla macierzy symetrycznej

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

wiodące minory główne $\Delta_1, \Delta_2, \Delta_3$ są nieujemne (w istocie $\Delta_i = 0, i = 1, 2, 3$), natomiast macierz ta jest nieokreślona: dla $s = (0, 0, 1)$ mamy $s^T A s > 0$ zaś dla $s = (0, -1, 1)$ mamy $s^T A s < 0$. Zwróćmy uwagę na fakt, że minor główny powstały przez skreślenie pierwszego wiersza i pierwszej kolumny macierzy A jest ujemny.

Twierdzenie 4.6.15 *Niech A będzie macierzą rzeczywistą. Zachodzi równość*

$$\|A\| = \sqrt{\lambda_{\max}(A^T A)},$$

gdzie $\|A\|$ oznacza normę spektralną macierzy A , zaś $\lambda_{\max}(A^T A)$ jest największą wartością własną macierzy $A^T A$.

Dowód. Ponieważ $A^T A$ jest macierzą symetryczną nieujemnie określoną, więc istnieje macierz ortogonalna $U = [u_1, u_2, \dots, u_n]$ i macierz $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ takie, że $A^T A = U D U^T$ i λ_j , $j = 1, 2, \dots, n$, są wartościami własnymi macierzy $A^T A$, zaś u_j odpowiadającymi im wektorami własnymi. Ponieważ $A^T A$ jako macierz Grama jest macierzą nieujemnie określoną, więc $\lambda_j \geq 0$, $j = 1, 2, \dots, n$ (patrz ćwiczenie 4.6.10). Zatem $\lambda_{\max}(A^T A) \geq 0$. Oznaczając $D^{\frac{1}{2}} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n})$ nietrudno zauważyć, że $D = D^{\frac{1}{2}} D^{\frac{1}{2}}$. Ponieważ $y := \|U^T x\| = \|x\|$ (U^T jest macierzą ortogonalną), więc dla dowolnego wektora $x \in \mathbb{R}^n$

$$\begin{aligned} \|Ax\|^2 &= x^T A^T A x = x^T U D U^T x = x^T U D^{\frac{1}{2}} D^{\frac{1}{2}} U^T x \\ &= \|D^{\frac{1}{2}} U^T x\|^2 = \|D^{\frac{1}{2}} y\|^2 = \sum_{j=1}^n \lambda_j y_j^2 \\ &\leq \lambda_{\max}(A^T A) \|y\|^2 = \lambda_{\max}(A^T A) \|x\|^2. \end{aligned}$$

Zatem

$$\|A\|^2 = \sup_{\|x\|=1} \frac{\|Ax\|^2}{\|x\|^2} \leq \lambda_{\max}(A^T A).$$

Z drugiej strony, dla wektora własnego u_i macierzy $A^T A$ odpowiadającego wartości własnej $\lambda_{\max}(A^T A)$ otrzymamy

$$\|Au_i\|^2 = u_i^T A^T A u_i = \lambda_{\max}(A^T A) \|u_i\|^2,$$

czyli

$$\|A\|^2 = \sup_{\|x\|=1} \frac{\|Ax\|^2}{\|x\|^2} \geq \frac{\|Au_i\|^2}{\|u_i\|^2} = \lambda_{\max}(A^T A),$$

a więc $\|A\|^2 = \lambda_{\max}(A^T A)$. ■

Dowód. Przedstawimy teraz inny dowód powyższego twierdzenia oparty o twierdzenie Kuhna–Tuckera, które będzie podane w rozdziale 5.

Zgodnie z definicją normy macierzy, $\|A\| = \sup_{\|x\|=1} \|Ax\|$. Na mocy twierdzenia Weierstrassa ten kres górny jest osiągnięty dla pewnego $x^* \in \mathbb{R}^n$ (funkcja $f(x) := \|Ax\|$ jest ciągła i zbiór $\{x \in \mathbb{R}^n : \|x\| = 1\}$ jest zwarty). W celu wyznaczenia tego kresu górnego wystarczy rozwiązać zadanie

$$\begin{aligned} &\text{maksymalizować} && f(x) = \|Ax\| \\ &\text{przy ograniczeniu} && \|x\| = 1 \end{aligned} \tag{4.24}$$

równoważne zadaniu

$$\begin{aligned} &\text{minimalizować} && h(x) = -\|Ax\|^2 \\ &\text{przy ograniczeniu} && \|x\|^2 = 1. \end{aligned} \tag{4.25}$$

Zauważmy, że w dowolnym punkcie dopuszczalnym spełniony jest warunek regularności LICQ, bo gradient ograniczenia jest niezerowy dla dowolnego punktu dopuszczalnego. Funkcja Lagrange’a dla tego zadania ma postać $L(x, \lambda) := -\|Ax\|^2 + \lambda(\|x\|^2 - 1)$. Na mocy twierdzenia Kuhna–Tuckera istnieje mnożnik Lagrange’a λ^* taki, że para (x^*, λ^*) spełnia układ równań

$$\begin{aligned} \nabla_x L(x, \lambda) &= -2A^T A x + 2\lambda x = 0, \\ \frac{\partial L(x, y)}{\partial \lambda} &= \|x\|^2 - 1 = 0. \end{aligned}$$

Z pierwszego równania wynika, że λ jest wartością własną macierzy $A^T A$. Po pomnożeniu pierwszego równania przez x^T i skorzystaniu z drugiego równania otrzymamy $\|Ax\|^2 = \lambda$. Widzimy więc, że $\sqrt{\lambda}$ jest jednocześnie optymalną wartością funkcji celu $\|Ax\|$ zadania (4.24). Ponieważ niezależnie od x będącego rozwiązaniem zadania (4.25) mnożnik Lagrange’a λ jest określony jednoznacznie (bo spełniony jest warunek LICQ), więc $\|Ax\|^2$ osiąga na zbiorze $\{x \in \mathbb{R}^n : \|x\| = 1\}$ swoje maksimum równe $\|Ax^*\|^2 = \lambda_{\max}(A^T A)$. ■

Wniosek 4.6.16 *Jeśli macierz symetryczna A typu $m \times m$ jest określona nieujemnie, to dla dowolnego $s \in \mathbb{R}^n$ zachodzą nierówności*

$$\lambda_{\min}(A)\|s\|^2 \leq s^T A s \leq \lambda_{\max}(A)\|s\|^2, \quad (4.26)$$

gdzie $\lambda_{\min}(A)$ i $\lambda_{\max}(A)$ są najmniejszą i największą wartością własną macierzy A . Ponadto, pierwsza bądź druga nierówność staje się równością dla s będącego wektorem własnym odpowiadającym wartości własnej $\lambda_{\min}(A)$ bądź $\lambda_{\max}(A)$.

Dowód. Z twierdzenia 4.6.11 i z uwagi 4.6.13 wynika, że $A = UDU^T$, gdzie U jest macierzą ortogonalną, zaś $D = \text{diag}(d)$, przy czym współrzędne d_i , $i = 1, 2, \dots, m$, są nieujemne. Oczywiście $\lambda_{\max}(A) = \max_{i=1,2,\dots,m} d_i$ i $\lambda_{\min}(A) = \min_{i=1,2,\dots,m} d_i$. Oznaczając $u = U^T s$ mamy

$$s^T A s = s^T U D U^T s = u^T D u = \sum_{i=1}^m d_i u_i^2.$$

Ponieważ U jest ortogonalna, więc $\|u\|^2 = \|s\|^2$, a więc

$$\lambda_{\min}(A)\|u\|^2 \leq s^T A s \leq \lambda_{\max}(A)\|u\|^2,$$

co kończy dowód pierwszej części. Drugą część można sprawdzić bezpośrednio. ■

Zdefiniujmy $|A| := \inf_{\|x\|=1} \frac{\|Ax\|^2}{\|x\|^2}$. W podobny jak twierdzenie 4.6.15 sposób można udowodnić następujące

Twierdzenie 4.6.17 *Niech A będzie macierzą rzeczywistą. Zachodzi równość*

$$|A| = \sqrt{\lambda_{\min}(A^T A)},$$

gdzie $\lambda_{\min}(A^T A)$ jest najmniejszą wartością własną macierzy $A^T A$. Jeśli A jest macierzą nieosobliwą, to $|A| > 0$.

Ćwiczenie 4.6.18 Pokazać, że jeśli A jest macierzą nieosobliwą, to $|A| = \frac{1}{\|A^{-1}\|}$.

Wskazówka. Skorzystać z ćwiczenia 4.5.5(b).

Ćwiczenie 4.6.19 Pokazać, że rozkład macierzy symetrycznej A na wartości własne (4.18) można wybrać tak, aby wartości własne leżące na przekątnej macierzy D pojawiały się w zadanej z góry kolejności. W szczególności można rozkład na wartości własne wybrać tak, aby na przekątnej macierzy D leżały wartości własne w malejącej kolejności, czyli

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = \dots = \lambda_n = 0, \text{ gdzie } r = \text{rank}(A).$$

Wskazówka. Zauważyć, że:

- (a) Dla macierzy permutacji P mamy $PP^T = I$, czyli $A = (UP)(P^T DP)(UP)^T$;
- (b) Macierz UP jest ortogonalna jako iloczyn macierzy ortogonalnych, zaś jej kolumny są permutacją kolumn macierzy U ;
- (c) Macierz $P^T DP$ jest diagonalna, a jej elementy są permutacją elementów diagonalnych macierzy D .

Oba twierdzenia 4.6.15 i 4.6.17 wynikają z poniższego ogólniejszego twierdzenia, zwanego *minimaksowym twierdzeniem Couranta–Fischera*.

Twierdzenie 4.6.20 Niech G będzie rzeczywistą macierzą symetryczną typu $n \times n$ o wartościach własnych λ_j , $j = 1, 2, \dots, n$, spełniających nierówności

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n.$$

Wówczas

$$\lambda_k = \max_{\dim S=k} \min_{0 \neq x \in S} \frac{x^T G x}{x^T x},$$

$k = 1, 2, \dots, n$, przy czym symbol $\max_{\dim S=k}$ oznacza, że maksimum jest po wszystkich podprzestrzeniach liniowych S wymiaru k przestrzeni $\mathbb{R}^n$.

Dowód. (według [GvL13]) Niech $G = UDU^T$ będzie rozkładem macierzy symetrycznej G na wartości własne, gdzie U jest macierzą ortogonalną o kolumnach u_j , $j = 1, 2, \dots, n$, zaś $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. Zdefiniujmy $S_k := \text{span}\{u_1, u_2, \dots, u_k\}$, podprzestrzeń rozpiętą przez pierwsze k kolumn macierzy U , $U_k := [u_1, u_2, \dots, u_k]$ i $D_k := \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$. Dowolny element $x \in S_k$ możemy przedstawić w postaci $x = U_k a$, gdzie $a = (a_1, a_2, \dots, a_k) \in \mathbb{R}^k$. Ponieważ podprzestrzeń S_k ma wymiar k , więc oznaczając $y = U^T x$ otrzymamy

$$\begin{aligned} \max_{\dim S=k} \min_{0 \neq x \in S} \frac{x^T G x}{x^T x} &\geq \min_{0 \neq x \in S_k} \frac{x^T G x}{x^T x} = \min_{0 \neq a \in \mathbb{R}^k} \frac{a^T U_k^T U D U^T U_k a}{a^T U_k^T U U^T U_k a} = \min_{0 \neq a \in \mathbb{R}^k} \frac{a^T D_k a}{a^T a} \\ &= \min_{0 \neq a \in \mathbb{R}^k} \frac{\sum_{j=1}^k \lambda_j (a_j)^2}{\sum_{j=1}^k (a_j)^2} \geq \frac{\sum_{j=1}^k \lambda_k (a_j)^2}{\sum_{j=1}^k (a_j)^2} = \lambda_k, \end{aligned}$$

czyli $\lambda_k \leq \max_{\dim S=k} \min_{0 \neq x \in S} \frac{x^T G x}{x^T x}$. Pokażemy, że zachodzi również nierówność przeciwna. Niech S będzie dowolną podprzestrzenią k -wymiarową. Zauważmy, że S musi posiadać niezerowy przekrój z podprzestrzenią $S^k := \text{span}\{u_k, u_{k+1}, \dots, u_n\}$, ponieważ $\dim S + \dim S^k = n + 1$, a więc $\dim(S \cap S^k) \geq 1$. Niech niezerowy wektor $x^* = b_k u_k + b_{k+1} u_{k+1} + \dots + b_n u_n \in S \cap S^k$. Oznaczmy $U^k = [u_k, u_{k+1}, \dots, u_n]$ i $D^k := \text{diag}(\lambda_k, \lambda_{k+1}, \dots, \lambda_n)$ Wówczas $y^* := (U^k)^T x^* \in S^k$ i

$$\min_{0 \neq x \in S} \frac{x^T G x}{x^T x} \leq \frac{x^{*T} G x^*}{x^{*T} x^*} = \frac{x^{*T} U^k D (U^k)^T x^*}{x^{*T} U^k (U^k)^T x^*} = \frac{y^{*T} D^k y^*}{y^{*T} y^*} \leq \lambda_k.$$

Ponieważ nierówności te zachodzą dla dowolnej k -wymiarowej podprzestrzeni S , więc

$$\max_{\dim S=k} \min_{0 \neq x \in S} \frac{x^T G x}{x^T x} \leq \lambda_k.$$

■

Ćwiczenie 4.6.21 Pokazać, że twierdzenia 4.6.15 i 4.6.17 wynikają z twierdzenia Couranta–Fischera.

4.7 Rozkład macierzy na wartości osobliwe

Rozkład macierzy na wartości osobliwe odgrywa ważną rolę w uczeniu maszynowym. Rozkład ten jest podstawą analizy głównych składowych i właściwego rozkładu ortogonalnego, które są wykorzystywane w uczeniu maszynowym. Zagadnienia te przedstawimy w dalszej części tego rozdziału.

Definicja 4.7.1 Niech A będzie rzeczywistą macierzą typu $m \times n$. Jeśli istnieją rzeczywista macierz ortogonalna U typu $m \times m$, rzeczywista macierz ortogonalna V typu $n \times n$ i macierz diagonalna Σ typu $m \times n$ o elementach na przekątnej σ_i , $i = 1, 2, \dots, p = \min\{m, n\}$ spełniających nierówności

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_p = 0 \quad (4.27)$$

i takie, że

$$A = U\Sigma V^T, \quad (4.28)$$

to równość (4.28) nazywamy *rozkładem macierzy A na wartości osobliwe* (ang. *singular value decomposition*), a liczby σ_i , $i = 1, 2, \dots, p$, nazywają się *wartościami osobliwymi* macierzy A (ang. *singular values*). Kolumny macierzy U nazywają się *lewymi wektorami osobliwymi* macierzy A a kolumny macierzy V – jej *prawymi wektorami osobliwymi*.

Zanim przejdziemy do problemu istnienia rozkładu macierzy na wartości osobliwe, poznamy jego pewne własności pod warunkiem, że taki rozkład istnieje.

Zapisując macierze ortogonalne U i V w postaci blokowej tzn. $U = [u_1, u_2, \dots, u_m]$ i $V = [v_1, v_2, \dots, v_n]$, gdzie u_i , $i = 1, 2, \dots, m$, są kolumnami macierzy U , zaś v_j , $j = 1, 2, \dots, n$, są kolumnami macierzy V , i w (4.28) mnożąc macierze blokowo

$$U\Sigma V^T = [u_1, u_2, \dots, u_m] \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_r & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix}$$

równość (4.28) możemy również zapisać w postaci

$$A = \sum_{i=1}^r \sigma_i u_i v_i^T. \quad (4.29)$$

Ponieważ macierz V jest ortogonalna, więc jej kolumny stanowią bazę w $\mathbb{R}^n$, czyli dowolny wektor $x \in \mathbb{R}^n$ możemy zapisać w postaci $x = \sum_{j=1}^n \alpha_j v_j$. Zatem, z uwagi na ortogonalność macierzy U i V ,

$$\|Ax\|^2 = \left\| \sum_{i=1}^r \sigma_i u_i v_i^T \left(\sum_{j=1}^n \alpha_j v_j \right) \right\|^2 = \left\| \sum_{i=1}^r \alpha_i \sigma_i u_i \right\|^2 = \sum_{i=1}^r \alpha_i^2 \sigma_i^2. \quad (4.30)$$

W niektórych zagadnieniach praktycznych, dla danej macierzy A wielkość $\|Ax\|^2$ nazywa się *energiją wektora x* .

Poniższe twierdzenie podaje związek rozkładu macierzy A na wartości osobliwe z rozkładem na wartości własne macierzy $A^T A$ i AA^T . W twierdzeniu tym zakładamy wprowadzić istnienie rozkładu macierzy A na wartości osobliwe, ale później pokażemy, że taki rozkład rzeczywiście istnieje.

Twierdzenie 4.7.2 Niech A będzie rzeczywistą macierzą typu $m \times n$ zaś (4.28) jej rozkładem na wartości osobliwe. Wówczas $AA^T = UDU^T$, gdzie $D = \Sigma\Sigma^T$ jest macierzą diagonalną typu $m \times m$ postaci

$$D = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 & 0 \\ 0 & \sigma_2^2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_r^2 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} \quad (4.31)$$

oraz $A^T A = VDV^T$, gdzie $D = \Sigma^T \Sigma$ jest macierzą diagonalną typu $n \times n$ postaci (4.31).

Dowód. Nierudno zauważyć, że zarówno $\Sigma\Sigma^T$ jak i $\Sigma^T\Sigma$ są postaci (4.31), przy czym pierwsza z nich jest typu $m \times m$, zaś druga – typu $n \times n$. Jest jasne, że $r \leq p = \min\{m, n\}$. Z równości (4.28) i z ortogonalności macierzy U i V mamy

$$AA^T = U\Sigma V^T V\Sigma^T U^T = U\Sigma\Sigma^T U^T = UDU^T$$

oraz

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

■

Uwaga 4.7.3 Mnożąc równość (4.28) prawostronnie przez V otrzymamy jej równoważną postać

$$AV = U\Sigma \quad (4.32)$$

albo po rozpisaniu macierzy U i V na kolumny

$$A[v_1, v_2, \dots, v_n] = [u_1, u_2, \dots, u_m] \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & & \\ 0 & \sigma_2 & \dots & 0 & & \\ \vdots & \vdots & \ddots & \vdots & & \\ 0 & 0 & \dots & \sigma_r & & \\ & & & 0 & & 0 \end{bmatrix} \quad (4.33)$$

czyli

$$Av_i = \sigma_i u_i \text{ dla } i = 1, 2, \dots, r, \quad Av_i = 0 \text{ dla } i = r+1, \dots, n. \quad (4.34)$$

Przejdziemy teraz do problemu istnienia rozkładu macierzy A na wartości osobiwe.

Twierdzenie 4.7.4 Niech A będzie rzeczywistą macierzą typu $m \times n$. Wówczas istnieją rzeczywista macierz ortogonalna U typu $m \times m$, rzeczywista macierz ortogonalna V typu $n \times n$ i macierz diagonalna Σ typu $m \times n$ o elementach rzeczywistych na przekątnej σ_i , $i = 1, 2, \dots, p = \min\{m, n\}$ takie, że zachodzi równość (4.28), przy czym wartości osobiwe σ_i , $i = 1, 2, \dots, p$, spełniają nierówności

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_p = 0. \quad (4.35)$$

Ponadto, $\text{rank}(A) = r$.

Dowód. Ponieważ macierz $A^T A$ jest nieujemnie określona, więc istnieje jej rozkład na wartości własne $A^T A = VDV^T$, gdzie $V = [v_1, v_2, \dots, v_n]$ jest macierzą ortogonalną, której kolumnami są wektory własne macierzy $A^T A$, a $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, przy czym można przyjąć, że wartości własne λ_j , $j = 1, 2, \dots, n$, macierzy A są ustawione w kolejności malejącej, tzn.

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

gdzie $r = \text{rank}(A^T A)$ (patrz ćwiczenie 4.6.19). Niech $\sigma_j = \sqrt{\lambda_j}$, $j = 1, 2, \dots, n$, oraz niech $u_i := \frac{1}{\sigma_i} Av_i$, $i = 1, 2, \dots, r$. Wówczas

$$Av_i = \sigma_i u_i, \quad i = 1, 2, \dots, r. \quad (4.36)$$

Jest oczywiste, że wektory u_i , $i = 1, 2, \dots, r$, są ortogonalne, bo

$$u_i^T u_j = \frac{1}{\sigma_i \sigma_j} (Av_i)^T Av_j = v_i^T v_j = \begin{cases} 1, & \text{jeśli } i = j, \\ 0, & \text{jeśli } i \neq j. \end{cases} \quad (4.37)$$

$i, j = 1, 2, \dots, r$. Jeśli $r = n$, to dowód jest zakończony, bo zgodnie z uwagą 4.7.3, równość (4.36) jest równoważna równości (4.28). Jeśli natomiast $r < n$, to wówczas $\lambda = 0$ jest $(n - r)$ -krotną wartością własną macierzy $A^T A$ i odpowiadającej jej wektory własne tworzą $(n - r)$ -wymiarową podprzestrzeń liniową (patrz ćwiczenie 4.6.6). Wybierając w tej przestrzeni $(n - r)$ liniowo niezależnych wektorów własnych można je poddać ortogonalizacji otrzymując układ ortogonalny $v_{r+1}, v_{r+2}, \dots, v_n$. Podobnie, $m - r$ liniowo niezależnych wektorów własnych macierzy AA^T odpowiadające wartości własnej $\lambda = 0$ można zortogonalizować otrzymując układ ortogonalny $u_{r+1}, u_{r+2}, \dots, u_m$. W tej sytuacji macierze $U = [u_1, u_2, \dots, u_m]$ i $V = [v_1, v_2, \dots, v_n]$ będą ortogonalne i będą spełnione wszystkie równości (4.34), co jest równoważne rozkładowi (4.28). ■

Uwaga 4.7.5 Rozkład na wartości osobliwe nie musi być jednoznacznie określony.

Przykład 4.7.6 Macierz $A := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ma dwa różne rozkłady na wartości osobliwe

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

oraz

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

Ogólnie, założmy dla uproszczenia, że A jest macierzą kwadratową posiadającą przynajmniej dwie identyczne wartości osobliwe σ_i i σ_j , $i \neq j$ i niech $A = U\Sigma V^T$ będzie jej rozkładem na wartości osobliwe. Niech $U_1 = UP$ i $V_1 = VP$, gdzie P jest macierzą permutacji zamieniającą miejscami wyłącznie elementy i i j . Oznacza to, że U_1 powstała przez zamianę w U kolumn i i j , zaś V_1 powstała w wyniku zamiany w V wierszy i i j . Oczywiście U_1 różni się od U i V_1 różni się od V , bo ortogonalne kolumny macierzy U bądź ortogonalne wiersze macierzy V nie mogą się powtarzać. Zauważmy ponadto, że $PP^T = P^T P$, gdyż macierz permutacji jest ortogonalna, oraz $P^T \Sigma P = \Sigma$, gdyż $\sigma_i = \sigma_j$. Mamy więc

$$A = U\Sigma V^T = UPP^T \Sigma PP^T V^T = (UP)(P^T \Sigma P)(VP)^T = U_1 \Sigma V_1^T,$$

czyli A ma dwa różne rozkłady na wartości osobliwe $A = U\Sigma V^T$ i $A = U_1 \Sigma V_1^T$.

Wniosek 4.7.7 Niech $A = U\Sigma V^T$ będzie rozkładem na wartości osobliwe macierzy A typu $m \times n$, przy czym elementy diagonalne macierzy Σ (wartości osobliwe macierzy A) spełniają nierówności (4.35). Wówczas macierz $A^T A$ ma rozkład na wartości własne postaci $A^T A = VDV^T$, gdzie $D = \Sigma^T \Sigma$ jest macierzą diagonalną o elementach na głównej przekątnej (wartościach własnych macierzy $A^T A$) $\lambda_i = \sigma_i^2$, $i = 1, 2, \dots, r$ i $\lambda_i = 0$, $i = r + 1, r + 2, \dots, n$. Podobnie macierz AA^T ma rozkład na wartości własne postaci $AA^T = UDU^T$, gdzie $D = \Sigma \Sigma^T$ jest macierzą diagonalną o elementach na głównej przekątnej (wartościach własnych macierzy AA^T) $\lambda_i = \sigma_i^2$, $i = 1, 2, \dots, r$ i $\lambda_i = 0$, $i = r + 1, r + 2, \dots, m$.

Ćwiczenie 4.7.8 Pokazać, że dla macierzy nieujemnie określonej rozkład na wartości własne i rozkład na wartości osobliwe pokrywają się. Wynika stąd w szczególności, że wartości własne macierzy nieujemnie określonej są jej wartościami osobliwymi i odwrotnie.

Ćwiczenie 4.7.9 Pokazać, że dla macierzy symetrycznej moduł jej wartości własnej jest jej wartością osobliwą i odwrotnie.

Twierdzenie 4.7.10 *Niech A będzie macierzą typu $m \times n$ posiadającą wartości osobliwe σ_i , $i = 1, 2, \dots, p$, spełniające nierówność (4.35). Wówczas*

$$\|A\|_F^2 = \operatorname{tr}(A^T A) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2. \quad (4.38)$$

Dowód. Zgodnie z definicją normy Frobeniusa (2.11) mamy

$$\|A\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2.$$

Z drugiej strony oznaczając kolumny macierzy A przez A_j i jej elementy przez a_{ij} , $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$, mamy

$$\begin{aligned} \operatorname{tr}(A^T A) &= \operatorname{tr} \left(\begin{bmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_n^T \end{bmatrix} \begin{bmatrix} A_1 & A_2 & \cdots & A_n \end{bmatrix} \right) \\ &= \operatorname{tr} \left(\begin{bmatrix} A_1^T A_1 & A_1^T A_2 & \cdots & A_1^T A_n \\ A_2^T A_1 & A_2^T A_2 & \cdots & A_2^T A_n \\ \vdots & \vdots & \ddots & \vdots \\ A_n^T A_1 & A_n^T A_2 & \cdots & A_n^T A_n \end{bmatrix} \right) \\ &= \sum_{j=1}^n A_j^T A_j = \sum_{j=1}^n \sum_{i=1}^m a_{ij}^2, \end{aligned}$$

co dowodzi pierwszej równości. Dla dowodu drugiej równości wystarczy posłużyć się równością (4.23) pamiętając, że wartościami własnymi macierzy $A^T A$ są σ_i^2 , $i = 1, 2, \dots, n$. ■

4.8 Wskaźnik uwarunkowania macierzy

Niech A będzie macierzą nieosobliwą i niech $b \in \mathbb{R}^n$. W konsekwencji zadanie układ równań $Ax = b$ posiada dokładnie jedno rozwiązanie. W praktyce często spotykamy się z sytuacją, gdy prawa strona jest wyznaczona w wyniku pomiarów, zatem może być zaburzona. Przeanalizujemy więc, jak zmienia się rozwiązanie tego układu w zależności od zaburzenia jego prawej strony. Niech x będzie rozwiązaniem tego układu, a $x + \Delta x$ będzie rozwiązaniem układu zaburzonego, czyli

$$A(x + \Delta x) = b + \Delta b.$$

Wówczas oczywiście

$$A(\Delta x) = \Delta b,$$

czyli

$$\Delta x = A^{-1}(\Delta b)$$

Z definicji normy macierzy wynika, że

$$\|b\| = \|Ax\| \leq \|A\| \cdot \|x\|.$$

i

$$\|\Delta x\| = \|A^{-1}(\Delta b)\| \leq \|A^{-1}\| \cdot \|\Delta b\|$$

Zatem

$$\frac{\|\Delta x\|}{\|x\|} \leq \|A\| \cdot \|A^{-1}\| \frac{\|\Delta b\|}{\|b\|}$$

Nierówność ta podaje oszacowanie na błąd względny dla rozwiązania układu równań w zależności od błędu względnego dla prawej strony układu. Wielkość

$$\kappa(A) := \|A\| \cdot \|A^{-1}\| \quad (4.39)$$

nazywamy *wskaznikiem uwarunkowania* (ang. *condition number*) macierzy A .

Ćwiczenie 4.8.1 Niech A będzie macierzą nieosobliwą A mającą wartości osobliwe σ_j , $j = 1, 2, \dots, n$, uporządkowane malejąco.

1. Pokazać, że

$$\kappa(A) = \frac{\sigma_1}{\sigma_n}.$$

2. Pokazać, że $\kappa(A) \geq 1$.

3. Pokazać, że jeśli A jest macierzą ortogonalną, to $\kappa(A) = 1$.

4. Kiedy $\kappa(A) = 1$?

Wskazówka. Skorzystać z twierdzenia 4.6.15.

Jeśli $\kappa(A) \gg 1$, to zadanie wyznaczenia rozwiązania układu $Ax = b$ nazywamy źle uwarunkowanym (ang. *ill conditioned*). Wówczas mała zmiana prawej strony układu $Ax = b$ może spowodować dużą zmianę rozwiązania tego zadania. Ma to miejsce na przykład wtedy, gdy prawa strona jest wynikiem pomiarów, które zazwyczaj obarczone są błędami.

Przykład 4.8.2 Niech A będzie *macierzą Hilberta* stopnia n , czyli $a_{ij} = \frac{1}{i+j-1}$, $i, j = 1, 2, \dots, n$. Rozważmy układ równań $Ax = b$, dla $n = 10$ i $b = (1, 1, \dots, 1)$. Jest to macierz dodatnio określona, a więc jest ona nieosobliwa i układ ten posiada dokładnie jedno rozwiązanie. Rozwiązując to zadanie w pakiecie MATLAB i używając funkcji `inv(A)` otrzymamy

$$x = A^{-1}b = 10^6 * [-0.0000, 0.0010, -0.0238, 0.2397, -1.2772, 3.6450, -7.0979, 6.7051, -3.9883, 0.9237]^T$$

Działając na tak wyliczony wektor x macierzą A otrzymamy

$$Ax = 10^5 * [-1.2208, -1.0692, -0.9514, -0.8572, -0.7800, -0.7157, -0.6612, -0.6145, -0.5739, -0.5384]^T,$$

czyli wektor znacznie różniący się od b . Powodem jest to, że jest to zadanie źle uwarunkowane. Wskaźnik uwarunkowania macierzy A wynosi $\kappa(A) = 1.6025 * 10^{13}$. Rozwiązując układy równań dla źle uwarunkowanych macierzy powinniśmy więc użyć innych technik.

Przypuśćmy, że rozpatrywana norma wektora jest normą euklidesową, a norma macierzy jest normą spektralną i niech $A = U\Sigma V^T$ będzie rozkładem na wartości osobliwe macierzy A i niech σ_j , $j = 1, \dots, n$, będą wartościami osobliwymi tej macierzy spełniającymi $\sigma_1 \geq \dots \geq \sigma_n > 0$. Wówczas $A^T A = V D V^T$ jest rozkładem na wartości własne macierzy $A^T A$, gdzie

$D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ i $\lambda_j = \sigma_j^2$, $j = 1, 2, \dots, n$. Mamy $\|A\| = \sqrt{\lambda_{\max}(A^T A)} = \sigma_1$ i $\|A^{-1}\| = \frac{1}{\sqrt{\lambda_{\min}(A^T A)}} = \frac{1}{\sigma_n}$, zatem

$$\frac{\|\Delta x\|}{\|x\|} \leq \sqrt{\frac{\lambda_{\max}(A^T A)}{\lambda_{\min}(A^T A)}} \cdot \frac{\|\Delta b\|}{\|b\|} = \frac{\sigma_1}{\sigma_n} \frac{\|\Delta b\|}{\|b\|},$$

czyli

$$\frac{\|\Delta x\|}{\|x\|} \leq \frac{\sigma_1}{\sigma_n} \frac{\|\Delta b\|}{\|b\|} \quad (4.40)$$

Okazuje się, że powyższe rozumowanie możemy w części powtórzyć dla dowolnej macierzy A typu $m \times n$. Należy tylko pamiętać, że wówczas mówimy nie o dowolnym rozwiązaniu x układu $Ax = b$ (równym $A^{-1}b$), a o pseudorozwiązaniu $x = A^+b$. Zastępując w powyższym rozumowaniu macierz A^{-1} przez macierz A^+ otrzymamy

$$\frac{\|\Delta x\|}{\|x\|} \leq \|A\| \cdot \|A^+\| \frac{\|\Delta b\|}{\|b\|}.$$

Uwzględniając wyniki ćwiczenia 4.12.16 otrzymamy

$$\frac{\|\Delta x\|}{\|x\|} \leq \frac{\sigma_1}{\sigma_r} \frac{\|\Delta b\|}{\|b\|}$$

Nierówność ta podaje oszacowanie na błąd względny dla pseudorozwiązania $x = A^+b$ układu $Ax = b$ w zależności od błędu względnego dla prawej strony układu.

Definicja 4.8.3 Dla macierzy niezerowej A mającej wartości osobliwe σ_j , $j = 1, \dots, n$, spełniające nierówności $\sigma_1 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_p = 0$, gdzie $p = \min\{m, n\}$, wielkość

$$\kappa(A) = \frac{\sigma_1}{\sigma_r}$$

nazywa się *wskaźnikiem uwarunkowania* macierzy A .

4.9 Twierdzenie Eckarta–Younga

Niektóre praktyczne problemy sprowadzają się do tego, aby dla danej macierzy znaleźć jej najlepsze przybliżenie wśród macierzy "niskiego" rzędu. Bardziej konkretnie, mając daną macierz A rzędu r należy znaleźć macierz B rzędu $k < r$ minimalizującą $\|A - B\|$. Poniższe twierdzenie podaje postać takiej macierzy.

Twierdzenie 4.9.1 (Eckhart–Young) Niech A będzie macierzą rzeczywistą, zaś (4.29) jest rozkładem na wartości osobliwe, i niech $k \leq r$, gdzie r jest rzędem macierzy A . Niech

$$A_k := \sum_{i=1}^k \sigma_i u_i v_i^T. \quad (4.41)$$

Wówczas dla dowolnej macierzy B rzędu co najwyżej k zachodzi

$$\|A - B\| \geq \|A - A_k\|. \quad (4.42)$$

przy czym

$$\|A - A_k\| = \sigma_{k+1}. \quad (4.43)$$

Dowód. Z twierdzenia 4.7.4 mamy $\text{rank}(A_k) = k$. Z równości (4.29) wynika, że

$$B_k := A - A_k = \sum_{i=k+1}^r \sigma_i u_i v_i^T.$$

Ponadto

$$\|B_k\|^2 = \lambda_{\max}(B_k^T B_k) = \sigma_{k+1}^2,$$

a zatem

$$\|A - A_k\| = \left\| \sum_{i=k+1}^r \sigma_i u_i v_i^T \right\| = \sigma_{k+1}.$$

Pokażemy, że dla dowolnej macierzy B rzędu co najwyżej k zachodzi $\|A - B\| \geq \sigma_{k+1}$. Aby to pokazać wystarczy dla dowolnej macierzy B rzędu k znaleźć wektor x taki, że $\|(A - B)x\| \geq \sigma_{k+1}\|x\|$. Dla danej macierzy B rzędu k niech $x \in \mathbb{R}^n$ będzie niezerowym wektorem takim, że

$$x \in \ker(B) \text{ i } x \in V_k := \text{span}\{v_i : i = 1, 2, \dots, k+1\}.$$

Istnienie takiego wektora $x \in \mathbb{R}^n$ wynika z następującego rozumowania. Po pierwsze,

$$\dim(\ker(B)) = n - \dim(\text{im}(B^T)) = n - \text{rank}(B) \geq n - k.$$

Po drugie, $\dim(V_k) = k + 1$. Widzimy więc, że suma wymiarów obu przestrzeni $\ker(B)$ i V_k wynosi co najmniej $n + 1$, a ponieważ są one zawarte w $\mathbb{R}^n$, ich przekrój musi mieć wymiar co najmniej 1. Istnieje więc niezerowy wektor $x \in \ker(B)$ taki, że $x = \sum_{i=1}^{k+1} \alpha_i v_i$ dla pewnych stałych $\alpha_i \in \mathbb{R}$, $i = 1, 2, \dots, k + 1$. Ponieważ $Av_i = \sigma_i u_i$, $i = 1, 2, \dots, k + 1$, i U jest macierzą ortogonalną, więc

$$\begin{aligned} \|(A - B)x\|^2 &= \|Ax\|^2 = \left\| A \left(\sum_{i=1}^{k+1} \alpha_i v_i \right) \right\|^2 = \left\| \sum_{i=1}^{k+1} \alpha_i A v_i \right\|^2 \\ &= \left\| \sum_{i=1}^{k+1} \alpha_i \sigma_i u_i \right\|^2 = \sum_{i=1}^{k+1} \alpha_i^2 \sigma_i^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} \alpha_i^2 = \sigma_{k+1}^2 \|x\|^2. \end{aligned}$$

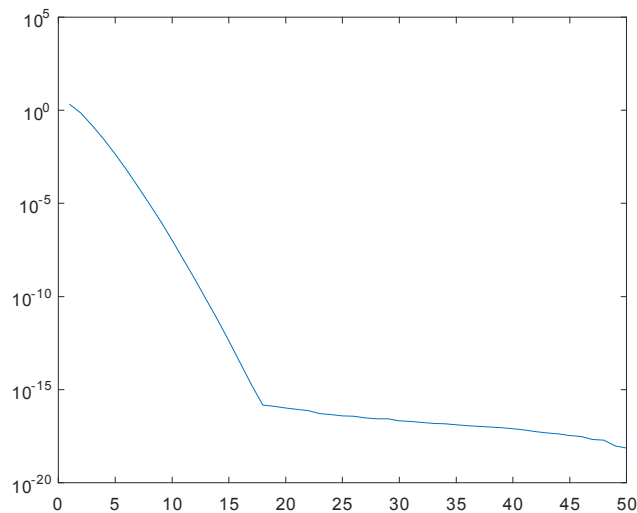
Ostatecznie mamy więc $\|(A - B)x\| \geq \sigma_{k+1}\|x\|$, co kończy dowód. ■

Uwaga 4.9.2 Można pokazać, że jeśli w twierdzeniu 4.9.1 zastąpimy normę spektralną przez normę Frobeniusa, to twierdzenie to będzie nadal prawdziwe o ile równość (4.43) zastąpimy przez

$$\|A - A_k\|_F = \sqrt{\sigma_{k+1}^2 + \sigma_{k+2}^2 + \dots + \sigma_r^2}.$$

Twierdzenie Eckharta–Younga jest używane przez inżynierów, szczególnie w sytuacjach, kiedy operuje się bardzo dużą ilością danych zgromadzonych w macierzy powstałej w wyniku pomiarów. Zazwyczaj dane te tworzą tzw. "chmurę" punktów w przestrzeni pomiarów, tworzących pewną strukturę, gdyż wielkości określające te punkty są często ze sobą skorelowane. Wyznaczając wartości osobliwe można zaobserwować, że mniejsze z tych wartości mają "mały udział" w charakterze tej "chmury". Zastosowanie twierdzenia Eckharta–Younga pozwala zazwyczaj znacznie zmniejszyć wymiar problemu bez znaczącej straty poprawności modelu matematycznego opisującego dane zjawisko. Ma to miejsce na przykład w kompresji danych, w uczeniu maszynowym, czy w dynamice płynów. Mając dane $k + 1$ największych wartości osobliwych macierzy A , twierdzenie Eckharta–Younga pozwala oszacować błąd jaki ma miejsce przy zastąpieniu pełnego modelu modelem zredukowanym do k największych wartości osobliwych i odpowiadających im wektorów osobliwych. W praktyce jest to o tyle korzystne, że zazwyczaj stosunkowo dużo wartości osobliwych jest na tyle małych, że nie mają istotnego wpływu na działanie macierzy A .

Przykład 4.9.3 Niech $A = [a_{ij}]$ będzie macierzą Hilberta typu $n \times n$, tzn. $a_{ij} = \frac{1}{i+j-1}$, $i, j = 1, 2, \dots, n$. Macierz ta ma bardzo duży wskaźnik uwarunkowania $\text{cond}(A) := \|A\| \cdot \|A^{-1}\| = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$ powodujący duże błędy numeryczne dla równań, w których występuje ta macierz. Już dla $n = 10$ mamy $\text{cond}(A) \simeq 1.6 \cdot 10^{13}$. Na poniższym rysunku przedstawiono wartości własne macierzy Hilberta dla $n = 50$, które dla macierzy dodatnio określonej (a taką jest macierz Hilberta) są jednocześnie jej wartościami osobliwymi.



Wartości własne macierzy Hilberta stopnia 50

Weźmy pod uwagę macierz A_{10} powstałą z obcięcia rozkładu na wartości własne macierzy A do 10 największych wartości własnych i odpowiadających im wektorów własnych, tak jak to opisano w równości (4.41). Wówczas dla normy spektralnej zachodzi $\|A - A_{10}\| \simeq 10^{-8}$. Zgodnie z definicją normy macierzy, z nierównością trójkąta i z twierdzeniem Eckharta–Younga otrzymamy dla $x \in \mathbb{R}^{50}$

$$|\|Ax\| - \|A_{10}x\|| \leq \|(A - A_{10})x\| \leq \|A - A_{10}\| \cdot \|x\| \simeq 10^{-8} \cdot \|x\|$$

Oznacza to, że dla wektora unormowanego $x \in \mathbb{R}^{50}$ błąd powstały przez zastąpienie Ax przez $A_{10}x$ będzie wynosić co najwyżej 10^{-8} .

4.10 Analiza głównych składowych

W analizie danych mamy często do czynienia z dużymi macierzami, które gromadzą informacje o obserwowanych zmiennych. Mogą to być na przykład płeć, wiek, wzrost i waga (i ewentualnie inne parametry) poszczególnych ludzi w pewnej bazie danych. Innym przykładem są (zeskanowane i wyskalowane) zestawy cyfr pisanych ręcznie przez pewien zbiór osób zgromadzone w celu stworzenia pewnego algorytmu. Algorytm ten na rozpoznać cyfry na podstawie skanów ręcznie pisanych cyfr nowych osób nieobjętych zgromadzonymi zestawami. Zestawy takie składają się z wielu obserwacji, każda z nich jest wektorem o wielkiej liczbie współrzędnych równej liczbie pikseli obszaru, w którym zapisywane są cyfry. Jeszcze innym przykładem są wyniki obserwacji (próbki) prędkości cieczy obserwowanych w pewnych odstępach czasowych w punktach pewnej siatki. Zwykle liczby obserwacji i liczby pikseli są bardzo duże. Zapisywanie i "obrabianie" tak wielkich zbiorów jest kosztowne. Natomiast zapisanie najbardziej

istotnych informacji wymaga znacznie mniejszej zajętości pamięci. Informacje te należy tylko "wydobyć" z tego ogromnego zbioru danych. Algorytmy "wydobywające" te informacje polegają na odpowiednich transformacjach liniowych i rzutowaniu na przestrzenie dużo niższych wymiarów niż przestrzenie zawierające zgromadzone dane.

Przejdźmy teraz do ogólnego przedstawienia problemu. Danych jest n wektorów $a_j \in \mathbb{R}^m$. Wektory te możemy zapisać w postaci macierzy $A = [a_{ij}]$ typu $m \times n$. Oznaczmy symbolem A^i – i -ty wiersz macierzy A i symbolem A_j – j -tą kolumnę macierzy A . Macierz A możemy więc zapisać blokowo

$$A = \begin{bmatrix} A^1 \\ A^2 \\ \vdots \\ A^m \end{bmatrix}$$

lub

$$A = [A_1, A_2, \dots, A_n].$$

W wierszu A^i macierzy A zapisane są wyniki dla n obserwacji (próbek) tej samej obserwowanej wielkości, np. pierwszej współrzędnej wektora prędkości cieczy w danym punkcie siatki w kolejnych chwilach czasowych. Z kolei kolumna A_j macierzy A zawiera wyniki j -tej próbki wszystkich obserwowanych parametrów, np. pierwszą współrzędną wektora prędkości cieczy w odpowiednio ponumerowanych punktach węzłowych siatki. Najczęściej wiersze macierzy A są ze sobą skorelowane. W przykładzie z obserwacją ruchu cieczy prędkości cieczy w ustalonej chwili czasowej dla poszczególnych punktów siatki są ze sobą "powiązane".

Przedstawimy teraz kolejne czynności wykonywane w analizie głównych składowych.

1. Najpierw wyznaczamy średnią μ_i dla każdego wiersza macierzy A ,

$$\mu_i = \frac{1}{n} \sum_{j=1}^n a_{ij}, i = 1, \dots, m.$$

2. Następnie tworzymy macierz $\bar{A}$ powstałą z macierzy A przez odjęcie od poszczególnych jej wierszy średnich dla tych wierszy, tzn.

$$\bar{A} = A - \mu = \begin{bmatrix} A^1 - \mu_1 \\ A^2 - \mu_2 \\ \vdots \\ A^m - \mu_m \end{bmatrix}, \quad (4.44)$$

gdzie $\mu = [\mu_1, \mu_2, \dots, \mu_m]$. Powstała w ten sposób macierz $\bar{A}$ ma "wycentrowane" wiersze, tzn. ich średnie są zerami.

3. W końcu wyznaczamy tzw. macierz S typu $m \times m$ dla próbek,

$$S = \bar{A}\bar{A}^T, \quad (4.45)$$

która jest w istocie przeskalowaną macierzą kowariancji obserwowanego wektora parametrów

$$\text{cov}(A^1, A^2, \dots, A^m) = \frac{1}{n-1} (A - \mu)(A - \mu)^T = \frac{1}{n-1} S,$$

wyraża sposób "powiązania" ze sobą wielkości poszczególnych obserwowanych parametrów.

Gdy $m \ll n$, macierz S ma dużo mniejszą zajętość pamięci niż A , ale w pewien sposób odzwierciedla zgromadzone informacje. Dokonując rozkładu na wartości własne tej macierzy możemy znaleźć właściwą strukturę i powiązanie zaobserwowanych wielkości. W wielu zastosowaniach liczba m jest również duża, nawet jeśli $m \ll n$. W przykładzie z obserwacją ruchu cieczy m jest liczbą węzłów danej siatki, a więc przy siatce 50×50 mamy $m = 51^2 = 2601$. Jeśli więc mamy 10000 obserwacji, to macierz A ma $26010000 \approx 2.6 \cdot 10^7$ elementów. Czy możemy zaobserwowane dane zapisać bez istotnej straty w dużo mniejszej macierzy? Okazuje się, że można i temu służy analiza głównych składowych.

Przypuśćmy, że wyznaczyliśmy rozkład na wartości własne macierzy $S = UDU^T$, gdzie kolumny u_i , $i = 1, 2, \dots, m$, macierzy $U = [u_1, u_2, \dots, u_m]$ są wektorami własnymi macierzy S typu $m \times m$, a elementy diagonalne d_i , $i = 1, 2, \dots, m$, macierzy $D = \text{diag}(d)$ są odpowiadającymi im wartościami własnymi. Odtworzymy teraz rozkład macierzy $\bar{A}$ na wartości osobliwe,

$$\bar{A} = U\Sigma V^T, \quad (4.46)$$

gdzie Σ jest macierzą diagonalną typu $m \times n$ o elementach na głównej przekątnej równych σ_i , $i = 1, 2, \dots, p$, spełniających nierówności (4.35), zaś $V = [v_1, v_2, \dots, v_n]$ jest macierzą ortogonalną typu $n \times n$. Zgodnie z równością (4.45) i twierdzeniem 4.7.2 mamy bowiem

$$UDU^T = S = \bar{A}\bar{A}^T = U\Sigma\Sigma^T U^T,$$

a więc $d_i = \sigma_i^2$, czyli

$$\sigma_i = \sqrt{d_i}, \quad i = 1, 2, \dots, r.$$

Równość (4.46) możemy przedstawić w postaci

$$U^T \bar{A} = \Sigma V^T.$$

Pamiętając o tym, że $\sigma_i = 0$ dla $i = r + 1, \dots, p$ otrzymamy więc $u_i^T \bar{A} = \sigma_i v_i^T$, $i = 1, 2, \dots, r$. Zatem mamy

$$v_i = \frac{1}{\sigma_i} \bar{A}^T u_i, \quad i = 1, 2, \dots, r. \quad (4.47)$$

W ten sposób wyznaczyliśmy więc macierz V prawych wektorów osobliwych macierzy A .

W praktyce nie musimy pamiętać ani macierzy $\bar{A}$ ani macierzy U . Wystarczy mieć zapamiętane macierze $\bar{A}^T[u_1, u_2, \dots, u_r]$ i $[v_1, v_2, \dots, v_r]$, obie typu $n \times r$, oraz wektor średnich $\mu = (\mu_1, \mu_2, \dots, \mu_m)$. Wówczas macierz $\bar{A}$ jest zapisana w przestrzeni rozpiętej przez wektory własne u_i macierzy S , $i = 1, 2, \dots, r$, zaś zgodnie z równością (4.44), dane wyjściowe A można przedstawić jako $A = \bar{A} + \mu$.

Jeśli wymiar obserwowanych parametrów m dla próbek jest duży, to liczba r niezerowych wartości osobliwych σ_i macierzy $\bar{A}$ może być również duża. W zastosowaniach większość z tych wartości osobliwych jest jednak mała i nie ma istotnego wpływu na operacje wykonywane przy pomocy macierzy A . Zgodnie z twierdzeniem Eckarta–Younga 4.9.1, jeśli w rozkładzie macierzy $\bar{A}$ na wartości osobliwe $\bar{A} = \sum_{i=1}^r \sigma_i u_i v_i^T$ ograniczymy się do pierwszych k składników, czyli do $\bar{A}_k = \sum_{i=1}^k \sigma_i u_i v_i^T$ to

$$\|\bar{A} - \bar{A}_k\| \leq \sigma_{k+1}.$$

Wybierając więc odpowiednio wielkość k tak, aby dla danego $\varepsilon > 0$ zachodziła nierówność $\sigma_{k+1} \leq \varepsilon$ możemy kontrolować błąd

$$\|\bar{A}x - \bar{A}_k x\| \leq \|\bar{A} - \bar{A}_k\| \cdot \|x\| \leq \varepsilon \|x\|.$$

Macierz $\bar{A}_k^T[u_1, u_2, \dots, u_k]$ przedstawia "wycelowane" dane w przestrzeni rozpiętej przez wektory własne u_i macierzy S , $i = 1, 2, \dots, k$. Wymiar macierzy $\bar{A}_k$ wynosi $m \times k$, zaś wymiar macierzy $\bar{A}_k^T[u_1, u_2, \dots, u_k]$ wynosi $n \times k$.

Polecam przejrzeć poniższe strony, gdzie można znaleźć więcej informacji na temat analizy głównych składowych.

https://en.wikipedia.org/wiki/Principal_component_analysis

<https://mst.mimuw.edu.pl/lecture.php?lecture=st2&part=Ch3>

4.11 Właściwy rozkład ortogonalny

Jak już wspomnieliśmy wcześniej, w analizie danych mamy często do czynienia z dużymi macierzami, które gromadzą informacje o obserwowanych zmiennych. W ustępie 4.10 zauważyliśmy, że w rozkładzie na wartości osobliwe macierzy A możemy bez dużej straty dla działania macierzy ograniczyć się do największych wartości osobliwych i związanych z nimi lewych i prawych wektorów osobliwych. Aby to zrobić musimy najpierw znać rozkład na wartości osobliwe macierzy A , a przy dużych macierzach jest to zadanie obliczeniowo trudne. Powstaje naturalne pytanie, czy można po kolei wyznaczać wartości osobliwe w kolejności malejącej i odpowiadające im wektory osobliwe i proces ten przerwać, gdy kolejna wartość osobliwa będzie odpowiednio mała? Okazuje się, że tak. Zanim przejdziemy do konstrukcji, przedstawimy pewne własności.

Twierdzenie 4.11.1 *Niech σ_i , $i = 1, 2, \dots, r$, będą dodatnimi wartościami osobliwymi macierzy A typu $m \times n$ spełniającymi nierówności (4.35) oraz u_i i v_i odpowiadającymi im lewymi i prawymi wektorami osobliwymi. Wówczas*

$$\sigma_j = \max_{x \neq 0, v_i^T x = 0, i=1,2,\dots,j-1} \frac{\|Ax\|}{\|x\|} \quad (4.48)$$

przy czym maximum to jest osiągnięte dla $x = v_j$.

Dowód. Niech $A = U\Sigma V^T$ będzie rozkładem na wartości osobliwe macierzy A , gdzie $U = [u_1, u_2, \dots, u_m]$, $V = [v_1, v_2, \dots, v_n]$ i Σ jest macierzą diagonalną typu $m \times n$ o elementach σ_i , $i = 1, 2, \dots, p = \min\{m, n\}$, na głównej przekątnej, spełniających nierówności (4.35). Dla $K \subseteq \{1, 2, \dots, r\}$ oznaczmy $A^K := \sum_{i \in K} \sigma_i u_i v_i^T$. Dla $J = \{j, j+1, \dots, r\}$ oznaczmy $U_J := [u_j, u_{j+1}, \dots, u_r]$ i $V_J := [v_j, v_{j+1}, \dots, v_r]$. Wówczas $A^J = \sum_{i=j}^r \sigma_i u_i v_i^T x = U_J \Sigma_J V_J^T$ jest rozkładem na wartości osobliwe macierzy A^J , przy czym σ_j jest największą z tych wartości a u_j i v_j są odpowiadającymi jej lewym i prawym wektorem osobliwym, $j = 1, 2, \dots, r$. Zatem z twierdzenia 4.6.15 wynika, że $\|A^J\| = \sqrt{\lambda_{\max}((A^J)^T A^J)} = \sigma_j$. Dla prawego wektora osobliwego v_j macierzy A^J mamy

$$\|A^J v_j\|^2 = v_j^T (A^J)^T A^J v_j = \lambda_{\max}((A^J)^T A^J) \|v_j\|^2 = \sigma_j^2. \quad (4.49)$$

Zatem, zgodnie z równością (4.29), dla $j = 1, \dots, r$ mamy

$$\begin{aligned} \max_{x \neq 0, v_k^T x = 0, k=1,2,\dots,j-1} \frac{\|Ax\|^2}{\|x\|^2} &= \max_{x \neq 0, v_k^T x = 0, k=1,2,\dots,j-1} \frac{\|\sum_{i=1}^r \sigma_i u_i v_i^T x\|^2}{\|x\|^2} \\ &= \max_{x \neq 0} \frac{\|\sum_{i=j}^r \sigma_i u_i v_i^T x\|^2}{\|x\|^2} = \max_{x \neq 0} \frac{\|A^J x\|^2}{\|x\|^2} = \|A^J\|^2 = \sigma_j^2, \end{aligned}$$

co daje równość (4.48). Z równości (4.49) wynika, że ostatnie a więc i pierwsze maksimum w powyższych równościach jest osiągnięte dla $x = v_j$. ■

Przejdziemy teraz do zastosowania powyższego twierdzenia w procedurze *właściwego rozkładu ortogonalnego*, polegającego wyznaczaniu wartości osobliwych macierzy A w kolejności malejącej oraz odpowiadających im lewych i prawych wektorów osobliwych.

Algorytm 4.11.2

Wejście: Macierz rzeczywista A typu $m \times n$, $p = \min\{m, n\}$, $\varepsilon > 0$ – tolerancja optymalności.

Wyjście: Częściowy rozkład na wartości osobliwe macierzy $A \simeq A^K = \sum_{i=1}^k \sigma_i u_i v_i^T x$, gdzie $K = \{1, 2, \dots, k\}$ i $\|A - A^K\| = \sigma_k \leq \varepsilon$.

1. Dla $j = 1, 2, \dots, p$:

(a) rozwiązać zadanie minimalizacji różniczkowalnej wypukłej

$$\begin{array}{ll} \text{maksymalizować} & f_j(x) = \|A^j x\|^2 \\ \text{względem} & x \in \mathbb{R}^n \\ \text{przy ograniczeniach} & \|x\|^2 = 1, \\ & v_i^T x = 0, i = 1, 2, \dots, j-1. \end{array} \quad (4.50)$$

(b) Przyjąć σ_j – maksymalna wartość funkcji celu zadania (4.50), v_j – maksymizator funkcji f_j .

(c) Przyjąć

$$u_j = \frac{1}{\sigma_j} A v_j.$$

2. Jeśli $\sigma_j \leq \varepsilon$ przyjąć $k = j$ i algorytm zatrzymuje się.

https://en.wikipedia.org/wiki/Proper_orthogonal_decomposition

4.12 Macierz pseudoodwrotna Moore'a–Penrose'a

4.12.1 Motywacja – liniowe zadanie najmniejszych kwadratów

Dany jest układ równań liniowych

$$Ax = b,$$

gdzie A jest macierzą typu $m \times n$ i $b \in \mathbb{R}^m$. Jeśli $m = n$ i A jest macierzą nieosobliwą, to istnieje macierz do niej odwrotna A^{-1} i jedynym rozwiązaniem tego układu jest $x = A^{-1}b$. W ogólnym przypadku ilość rozwiązań zależy od rzędu macierzy A i rzędu macierzy poszerzonej $[A, b]$. Dokładniej, istnienie oraz jednoznaczność rozwiązania wynika z twierdzenia Kroneckera–Capellego. W szczególności, jeśli A jest macierzą pełnego rzędu wierszowego, to macierz Grama AA^T układu wierszy macierzy A jest nieosobliwa i układ równań posiada rozwiązanie postaci $x = A^T(AA^T)^{-1}b$, przy czym jeśli $m < n$, to rozwiązań tych jest nieskończenie wiele i tworzą one podprzestrzeń afiniczną wymiaru $n - m$ (patrz ćwiczenie 2.2.24). W praktyce często mamy jednak do czynienia z przypadkiem, gdy $m > n$ i wówczas rozwiązań jest najczęściej brak, chyba że $b \in \text{im } A$. W sytuacji braku rozwiązań poszukujemy takiego $x \in \mathbb{R}^n$, dla którego wielkość $\|Ax - b\|$ jest najmniejsza. Taki punkt jest rozwiązaniem zadania minimalizacji różniczkowalnej wypukłej

$$\begin{array}{ll} \text{minimalizować} & \frac{1}{2} \|Ax - b\|^2 \\ \text{względem} & x \in \mathbb{R}^n. \end{array} \quad (4.51)$$

Z warunków koniecznych i wystarczających minimalizacji funkcji różniczkowalnej wypukłej punkt x jest rozwiązaniem tego zadania wtedy i tylko wtedy, gdy $\nabla \frac{1}{2} \|Ax - b\|^2 = 0$, czyli $A^T(Ax - b) = 0$. Równanie to zwane równaniem normalnym zawsze posiada rozwiązanie. W przypadku, gdy A jest macierzą pełnego rzędu kolumnowego ($\text{rank } A = n$), macierz Grama układu kolumn macierzy A jest nieosobliwa i jedyne rozwiązanie zadania (4.51) jest postaci $x = (A^T A)^{-1} A^T b$. Jeśli

$\text{rank } A < n$, to rozwiązań zadania (4.51) jest nieskończenie wiele i tworzą one podprzestrzeń afiniczną wymiaru $n - \text{rank } A$.

W sposób naturalny pojawia się następujące pytanie. Czy niezależnie od postaci macierzy A istnieje macierz $\bar{A}$ będąca uogólnieniem odwrotności macierzy A spełniająca warunki:

1. Jeśli A jest nieosobliwa (układ równań $Ax = b$ posiada dokładnie jedno rozwiązanie), to $\bar{A} = A^{-1}$.
2. Jeśli układ równań $Ax = b$ posiada rozwiązanie, to jednym z nich jest $x = \bar{A}b$.
3. Jeśli układ równań $Ax = b$ nie posiada rozwiązań, to jednym z rozwiązań zadania (4.51) jest $x = \bar{A}b$.
4. Uogólnioną odwrotnością macierzy $\bar{A}$ jest A , czyli $\overline{(\bar{A})} = A$.

Okazuje się, że taka macierz istnieje. Podamy najpierw formalną definicję uogólnionej odwrotności macierzy A .

4.12.2 Definicja macierzy Moore'a–Penrose'a

Definicja 4.12.1 Niech A będzie macierzą typu $m \times n$. Macierz A^+ typu $n \times m$ spełniającą warunki

- (i) $AA^+A = A$,
- (ii) $A^+AA^+ = A^+$,
- (iii) $(A^+A)^T = A^+A$,
- (iv) $(AA^+)^T = AA^+$.

nazywamy *macierzą pseudoodwrotną Moore'a–Penrose'a* do macierzy A lub krócej *macierzą Moore'a–Penrose'a* albo *macierzą pseudoodwrotną*.

Istnienie macierzy pseudoodwrotnej Moore'a–Penrose'a pokażemy w kolejnym ustępie. Natomiast teraz możemy pokazać, że jeśli taka macierz istnieje, to jest określona jednoznacznie.

Twierdzenie 4.12.2 Niech A będzie macierzą typu $m \times n$. Jeśli istnieje do niej macierz pseudoodwrotna Moore'a–Penrose'a A^+ , to jest ona określona jednoznacznie.

Dowód. Przypuśćmy, że B_1 i B_2 są macierzami pseudoodwrotnymi Moore'a–Penrose'a do macierzy A . Wówczas z definicji 4.12.1 i w własności transpozycji wynika, że

$$\begin{aligned}
 B_1 & \stackrel{(ii)}{=} B_1AB_1 \stackrel{(iv)}{=} B_1(AB_1)^T = B_1B_1^T A^T \stackrel{(i)}{=} B_1B_1^T (AB_2A)^T = B_1B_1^T A^T B_2^T A^T \\
 & \stackrel{(iv)}{=} B_1B_1^T A^T (AB_2^T)^T = B_1B_1^T A^T AB_2 = B_1(AB_1)^T AB_2 \stackrel{(iv)}{=} B_1AB_1AB_2 \stackrel{(ii)}{=} B_1AB_2 \\
 & \stackrel{(ii)}{=} B_1AB_2AB_2 \stackrel{(iii)}{=} B_1A(B_2A)^T B_2 = B_1AA^T B_2^T B_2 \stackrel{(iii)}{=} (B_1A)^T A^T B_2^T B_2 = A^T B_1^T A^T B_2^T B_2 \\
 & \stackrel{(i)}{=} (AB_1A)^T B_2^T B_2 \stackrel{(i)}{=} A^T B_2^T B_2 = (B_2A)^T B_2 \stackrel{(iii)}{=} B_2AB_2 \stackrel{(ii)}{=} B_2.
 \end{aligned}$$

Tak więc $B_1 = B_2$, co kończy dowód. ■

Ćwiczenie 4.12.3 Pokazać, że jeśli $m = n$ i A jest macierzą nieosobliwą, to macierz $A^+ := A^{-1}$ spełnia warunki (i)–(iv) definicji 4.12.1.

Ćwiczenie 4.12.4 Pokazać, że jeśli A jest macierzą pełnego rzędu wierszowego, to macierz $A^+ := A^T(AA^T)^{-1}$ spełnia warunki (i)–(iv) definicji 4.12.1.

Ćwiczenie 4.12.5 Pokazać, że jeśli A jest macierzą pełnego rzędu kolumnowego, to macierz $A^+ := (A^T A)^{-1} A^T$ spełnia warunki (i)–(iv) definicji 4.12.1.

Ćwiczenie 4.12.6 Pokazać, że jeśli Σ jest macierzą diagonalną typu $m \times n$ o elementach na głównej przekątnej σ_i , $i = 1, 2, \dots, p$, gdzie $p = \min\{m, n\}$, spełniających nierówności $\sigma_1 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_p = 0$ dla pewnego $r = 0, 1, 2, \dots, p$, to macierz diagonalna Σ^+ typu $n \times m$ o elementach na głównej przekątnej równych τ_i , $i = 1, 2, \dots, p$, takich, że $\tau_i = \sigma_i^{-1}$ dla $i = 1, 2, \dots, r$ i $\tau_i = 0$ dla $i = r + 1, r + 2, \dots, p$ spełnia warunki (i)–(iv) definicji 4.12.1.

Definicja 4.12.7 Niech A będzie macierzą typu $m \times n$ i $b \in \mathbb{R}^m$. Wektor $x = A^+ b$ nazywamy *pseudorozwiązaniem* układu równań $Ax = b$.

Uwaga 4.12.8 Niech A będzie macierzą typu $m \times n$.

- (a) Jeśli $m = n$ i macierz A jest nieosobliwa, to $x = A^{-1}b$ będące jedynym rozwiązaniem układu równań $Ax = b$ jest również jego pseudorozwiązaniem.
- (b) Jeśli $m > n = \text{rank}(A)$, to $A^+ = (A^T A)^{-1} A^T$, czyli $x = A^+ b$ jest jedynym rozwiązaniem zadania (4.51), zatem w tym przypadku $x = (A^T A)^{-1} A^T b$ jest również pseudorozwiązaniem układu równań $Ax = b$.
- (c) Jeśli $m > n > \text{rank}(A)$, to zadanie (4.51) ma nieskończenie wiele rozwiązań, natomiast pseudorozwiązanie $x = A^+ b$ układu równań $Ax = b$ jest rozwiązaniem zadania (4.51) o najmniejszej normie.
- (d) Jeśli $n > m = \text{rank}(A)$, to $A^+ = A^T (AA^T)^{-1}$. Wówczas układ równań $Ax = b$ ma nieskończenie wiele rozwiązań, natomiast pseudorozwiązanie $x = A^+ b$ układu równań $Ax = b$ jest jego rozwiązaniem o najmniejszej normie.

4.12.3 Macierz Moore'a–Penrose'a, a rozkład na wartości osobliwe

Z samej definicji macierzy pseudodwrotnej nie wynika jeszcze jej istnienie. Poniższe twierdzenie oparte o rozkład na wartości osobliwe macierzy A pokazuje nie tylko istnienie takiej macierzy, ale daje możliwość jej skonstruowania. Z twierdzenia 4.7.4 wynika, że macierz A daje się przedstawić w postaci $A = U\Sigma V^T$, gdzie U jest macierzą ortogonalną typu $m \times m$, Σ jest macierzą diagonalną typu $m \times n$ o nieujemnych elementach σ_j na głównej przekątnej, $j = 1, 2, \dots, p$, gdzie $p = \min\{m, n\}$ oraz V jest macierzą ortogonalną typu $n \times n$. Zazwyczaj rozkład ten przedstawia się w taki sposób, aby $\sigma_1 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_p = 0$.

Twierdzenie 4.12.9 Niech $A = U\Sigma V^T$ będzie rozkładem macierzy A na wartości osobliwe. Macierz

$$A^+ = V\Sigma^+ U^T \quad (4.52)$$

spełnia warunki (i) – (iv) definicji 4.12.1, więc jest macierzą pseudoodwrotną Moore'a–Penrose'a do macierzy A

Dowód. Wystarczy sprawdzić, że macierz A^+ zdefiniowana równością (4.52) spełnia warunki (i)-(iv) definicji 4.12.1.

(i) Ponieważ macierze U i V są ortogonalne i $\Sigma\Sigma^+\Sigma = \Sigma$ (patrz ćwiczenie 4.12.6), więc

$$AA^+A = (U\Sigma V^T)(V\Sigma^+U^T)(U\Sigma V^T) = U\Sigma\Sigma^+\Sigma V^T = U\Sigma V^T = A$$

(ii) Ponieważ macierze U i V są ortogonalne i $\Sigma^+\Sigma\Sigma^+ = \Sigma^+$ (patrz ćwiczenie 4.12.6), więc

$$A^+AA^+ = (V\Sigma^+U^T)(U\Sigma V^T)(V\Sigma^+U^T) = V\Sigma^+\Sigma\Sigma^+U^T = V\Sigma^+U^T = A^+$$

(iii) Ponieważ macierz U jest ortogonalna i macierz $\Sigma^+\Sigma$ jest symetryczna (patrz ćwiczenie 4.12.6), więc

$$\begin{aligned} (A^+A)^T &= [(V\Sigma^+U^T)(U\Sigma V^T)]^T = (V\Sigma^+\Sigma V^T)^T = V\Sigma^+\Sigma V^T \\ &= (V\Sigma^+U^T)(U\Sigma V^T) = A^+A \end{aligned}$$

(iv) Ponieważ macierz V jest ortogonalna i macierz $\Sigma\Sigma^+$ jest symetryczna (patrz ćwiczenie 4.12.6), więc

$$\begin{aligned} (AA^+)^T &= [(U\Sigma V^T)V\Sigma^+U^T]^T = [U\Sigma\Sigma^+U^T]^T = U\Sigma\Sigma^+U^T \\ &= (U\Sigma V^T)(V\Sigma^+U^T) = AA^+ \end{aligned}$$

■

Wniosek 4.12.10 Niech A będzie macierzą typu $m \times n$. Wówczas macierz pseudoodwrotna Moore'a–Penrose'a A^+ istnieje i jest określona jednoznacznie.

Dowód. Wniosek wynika z twierdzeń 4.12.2 i 4.12.9. ■

Uwaga 4.12.11

4.12.4 Własności macierzy Moore'a–Penrose'a

W poniższych ćwiczeniach należy sprawdzić podane tam własności macierzy pseudoodwrotnej Moore'a–Penrose'a.

Ćwiczenie 4.12.12 Niech A będzie macierzą typu $m \times n$. Pokazać, że

1. $(A^+)^+ = A$;
2. $(A^T)^+ = (A^+)^T$;
3. Jeśli $\text{rank}(A) = m$, to $AA^+ = I_m$;
4. Jeśli $\text{rank}(A) = n$, to $A^+A = I_n$;
5. Macierze AA^+ i A^+A są idempotentne;
6. Macierze $I_m - AA^+$ i $I_n - A^+A$ są idempotentne;
7. $\|A^+\| = \sigma_r^{-1}$, gdzie σ_r jest najmniejszą niezerową wartością osobliwą macierzy A .

Twierdzenie 4.12.13 Niech A będzie macierzą typu $m \times n$ i niech $x \in \mathbb{R}^n$. Wówczas

$$A^+Ax \in \text{im } A^T \text{ oraz } x - A^+Ax \in \ker A. \quad (4.53)$$

Dowód. Z definicji macierzy pseudodwrotnej i z własności transpozycji macierzy wynika, że

$$A^+Ax = (A^+A)^Tx = A^T(A^+)^Tx \in \operatorname{im} A^T.$$

Ponadto $x - A^+Ax \in (\operatorname{im} A^T)^\perp$, gdyż dla dowolnego $z \in \operatorname{im} A^T$ mamy $z = A^Tu$ dla pewnego $u \in \mathbb{R}^n$, a zatem z definicji macierzy pseudodwrotnej mamy

$$\langle z, x - A^+Ax \rangle = \langle A^Tu, x - A^+Ax \rangle = \langle u, Ax - AA^+Ax \rangle = \langle u, Ax - Ax \rangle = 0,$$

a więc $x - A^+Ax \in (\operatorname{im} A^T)^\perp = \ker A$. ■

Wniosek 4.12.14 Niech A będzie macierzą typu $m \times n$. Wówczas

(i) Dla dowolnego $x \in \mathbb{R}^n$ zachodzą równości

$$P_{\ker A}(x) = x - A^+Ax \text{ oraz } P_{\operatorname{im} A^T}(x) = A^+Ax. \quad (4.54)$$

W szczególności, jeśli A jest pełnego rzędu wierszowego, to

$$P_{\ker A}x = x - A^T(AA^T)^{-1}Ax \text{ oraz } P_{\operatorname{im} A^T}(x) = A^T(AA^T)^{-1}Ax. \quad (4.55)$$

(ii) Dla dowolnego $y \in \mathbb{R}^m$ zachodzą równości

$$P_{\ker A^T}(y) = y - (AA^+)^Ty \text{ oraz } P_{\operatorname{im} A}(y) = (AA^+)^Ty. \quad (4.56)$$

W szczególności, jeśli A jest pełnego rzędu kolumnowego, to

$$P_{\ker A^T}(y) = y - A(A^TA)^{-1}A^Ty \text{ oraz } P_{\operatorname{im} A}(y) = A(A^TA)^{-1}A^Ty. \quad (4.57)$$

Dowód. (i) Równość (4.54) wynika z twierdzenia o rozkładzie ortogonalnym i z twierdzenia 4.12.13. Równość (4.55) wynika z (4.54) i z postaci macierzy A^+ w przypadku, gdy A jest pełnego rzędu wierszowego (patrz ćwiczenie 4.12.4).

(ii) Ta część wynika z (i) przez zastąpienie macierzy A przez macierz A^T . ■

Ćwiczenie 4.12.15 Niech $u \in \mathbb{R}^n$ będzie wektorem unormowanym i niech $A = u$. Czym jest $\operatorname{im} A$ i $\ker A^T$. Podać postać $P_{\ker A^T}$. Dla operatora $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ i $\lambda \in [0, 2]$ oznaczmy $T_\lambda x := x + \lambda(Tx - x)$ – relaksację operatora T . Dla $\lambda = 2$ operator $T_2 := 2T - I$ nazywamy odbiciem operatora T , gdzie I oznacza identyczność. Podać postać odbicia dla operatora $P_{\ker A^T}$ i porównać z operatorem Householdera. Niech $V \subseteq \mathbb{R}^n$ będzie podprzestrzenią liniową. Sprawdzić geometrycznie, jakie są wartości własne operatora P_V i jakie są wartości własne jego odbicia?

Ćwiczenie 4.12.16 Niech A będzie macierzą typu $m \times n$ i niech $A = U\Sigma V^T$ będzie jej rozkładem na wartości osobliwe. Wyznaczyć $\|A^+\|$.

Wskazówka.

- (a) Zauważyć, że elementy diagonalne σ_i , $i = 1, 2, \dots, r$ macierzy Σ typu $m \times n$ są dodatnie oraz $\sigma_i = 0$ dla $i = r + 1, r + 2, \dots, p = \max\{m, n\}$
- (b) Zauważyć, że macierz $(\Sigma^+)^T \Sigma^+$ jest typu $m \times m$ i

$$(\Sigma^+)^T \Sigma^+ = \begin{bmatrix} \sigma_1^{-2} & 0 & \cdots & 0 & 0 \\ 0 & \sigma_2^{-2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_r^{-2} & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix};$$

(c) Zauważyć, że

$$(A^+)^T A^+ = V(\Sigma^+)^T \Sigma^+ V^T;$$

(d) Zauważyć, że zgodnie z twierdzeniem 4.6.15

$$\|A^+\| = \sqrt{\lambda_{\max}((A^+)^T A^+)};$$

(e) Zauważyć, że $\lambda_{\max}((A^+)^T A^+) = \sigma_r^{-2}$

(f) Wyciągnąć wnioski z (a)-(e).