Ortogonalitate. Transformări și proiecții ortogonale. Aplicații în metode de regresie.

Cuprins

1 Obiective laborator

În urma parcurgerii acestui laborator, studentul va fi capabil să:

- definească noțiunile de vectori ortogonali și matrice ortogonală;
- aplice metode de transformare ortogonală: Householder și Givens;
- implementeze procesul Gram-Schmidt;
- folosească polinoame ortogonale.

2 Noțiuni teoretice

2.1 Norme

Considerând un spațiu vectorial V peste un corp $\mathbb{K}$, o normă pe V este o funcție $||\cdot|| : V \to \mathbb{R}$ care satisface următoarele proprietăți pentru orice $x, y \in V$ și $\alpha \in \mathbb{K}$:

- $||x|| \geq 0$ și $||x|| = 0 \Leftrightarrow x = 0$ (pozitiv definită);
- $||\alpha x|| = |\alpha| \cdot ||x||$;
- $||x + y|| \leq ||x|| + ||y||$ (inegalitatea triunghiului).

2.1.1 Norme vectoriale

- **Valoarea absolută.** Este o normă pe $\mathbb{R}$ sau $\mathbb{C}$. Numerele complexe formează un spațiu uni-dimensional peste $\mathbb{C}$ și unul bi-dimensional peste $\mathbb{R}$.
- **Distanța Manhatten.** $||\mathbf{x}||_1 \coloneqq \sum_i |x_i|$.
- **Norma euclidiană.** Pe $\mathbb{R}^n$, norma euclidiană este definită ca $||\mathbf{x}||_2 \coloneqq \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$. Luând în considerare numerele complexe, acestea se identifică cu $\mathbb{R}^2$.
- **Norma infinit.** $||\mathbf{x}||_\infty \coloneqq \max_i |x_i|$.
- **Norma *p*.** $||\mathbf{x}||_p \coloneqq (\sum_i |x_i|^p)^{1/p}$. Normele de mai sus sunt particularizări ale normei p pentru diferite valori ale lui p.

2.1.2 Norme matriceale

Multe norme matriceale mai au proprietatea de a fi *submultiplicative*:

$$||AB|| \leq ||A|| \cdot ||B||$$

- **Norma *p* matriceală.** Ea este indusă de norma p a vectorilor.
  $||A||_p \coloneqq \max_{x\neq 0} \frac{||Ax||_p}{||x||_p} = \max_{||x||=1} ||Ax||_p$.
  - **p = 1.** $||A||_1 \coloneqq \max_j \sum_i |a_{ij}|$. Este suma maximă a valorilor absolute de pe coloane.

- **p = 2. Norma/Raza spectrală.** $||A||_2 := \sqrt{\lambda_{max}(A^*A)}$. Este rădăcina patrată a celei mai mari valori proprii a matricei A^*A. Este egala cu cea mai mare valoare singulară a matricei A.
  Demonstrație. Fie $B = A^*A$. Atunci B este simetrică și din teorema spectrală avem o bază ortonormată de vectori proprii v_i și valori proprii λ_i. Fie $v = \sum_i \alpha_i v_i$ și $||v|| = 1$. Atunci:

  $$||Av||_2^2 = \langle Av, Av\rangle = \langle v, A^*Av\rangle = \langle \sum_i \alpha_i v_i, \sum_i \alpha_i \lambda_i v_i\rangle = \sum_i \lambda_i \alpha_i^2$$

  Având constrângerea $||v|| = 1$, $\sum_i \alpha_i^2 = 1 \implies ||A||_2 = \lambda_{max}(A)$. ■
- **p = ∞.** $||A||_\infty := \max_i \sum_j |a_{ij}|$. Este suma maximă a valorilor absolute de pe rânduri.

- **Norma Frobenius.** $||A||_F := \sqrt{\sum_{i,j} |a_{ij}|^2} = \sqrt{trace(A^*A)}$.

Teorema Gelfand. Pentru orice normă matriceală avem:

$$\lim_{k\to\infty} ||A^k||^{1/k} = \rho(A)$$

Mai mult, $\rho(A) \leq ||A||$ pentru orice normă matriceală.

Demonstrație. Fie λ valoarea proprie cea mai mare a lui A și v un vector propriu asociat. Atunci:

$$||A|| \geq \frac{||Av||}{||v||}(\forall v) = \frac{||\lambda v||}{||v||} = |\lambda| \Rightarrow \rho(A) \leq ||A||$$

Ce ne indică normele matriceale induse de vectori? Ele ne dau o măsură a cât de mult se dilată un vector atunci când este aplicată o anumită transformare liniară. În 1 avem o reprezentare a vectorilor unitari.

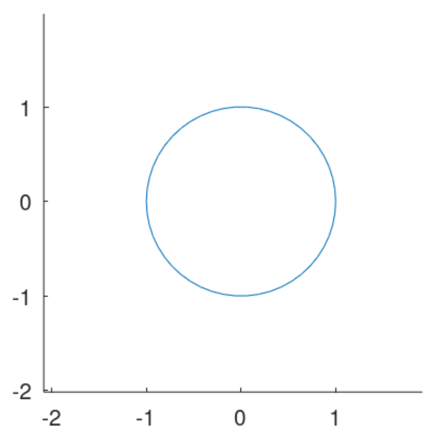


Figura 1: Vectori unitate

Ce se întâmplă dacă aplicăm transformarea $A1 = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}$?

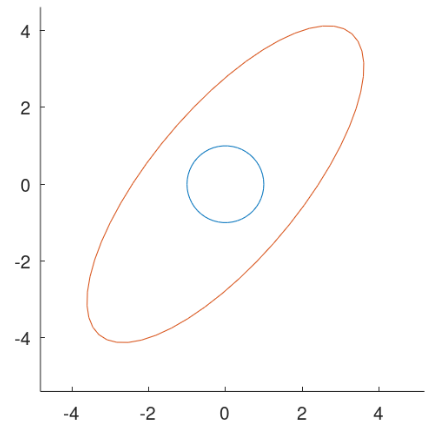


Figura 2: Transformarea $A1$

2.2 Produs scalar

Produsul scalar al unui spațiu vectorial V peste F este o funcție $\langle \cdot, \cdot \rangle : V \times V \to F$ care satisface următoarele proprietăți pentru orice $x, y, z \in V$ și $\alpha \in F$:

- $\langle x, y \rangle = \overline{\langle y, x \rangle}$ (conjugare simetrică);
- $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$ (liniaritate);
- $\langle x, x \rangle \geq 0$ și $\langle x, x \rangle = 0 \Leftrightarrow x = 0$ (pozitivitate).

Din aceastea rezultă și altele:

- $\langle x, \alpha y + \beta z \rangle = \overline{\alpha} \langle x, y \rangle + \overline{\beta} \langle x, z \rangle$;
- $\langle x + y, x + y \rangle = \langle x, x \rangle + 2\Re(\langle x, y \rangle) + \langle y, y \rangle$;

Orice produs scalar induce o normă pe spațiul vectorial V prin $||x|| = \sqrt{\langle x, x \rangle}$.

Într-un spațiu euclidian, produsul scalar este definit ca $\langle x, y \rangle = x^T y$.

2.2.1 Proiecții

Teoremă. $\langle x, y \rangle = u^T v = ||u||||v|| \cos(\theta)$, unde θ este unghiul dintre cei doi vectori.

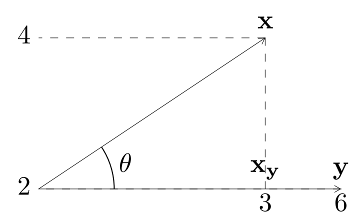
Demonstrație. Fie $\mathbf{r} = \mathbf{u} - \mathbf{v}$. Atunci, din teorema cosinusului avem:

$$||\mathbf{r}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2 - 2||\mathbf{u}||||\mathbf{v}|| \cos(\theta)$$
$$||\mathbf{u} - \mathbf{v}||^2 - ||\mathbf{u}||^2 - ||\mathbf{v}||^2 = -2||\mathbf{u}||||\mathbf{v}|| \cos(\theta)$$
$$-2 \sum_i u_i v_i = -2||\mathbf{u}||||\mathbf{v}|| \cos(\theta)$$
$$\sum_i u_i v_i = ||\mathbf{u}||||\mathbf{v}|| \cos(\theta)$$

Astfel, putem scrie produsul scalar ca $\langle x, y\rangle = u^T v = ||u||||v|| \cos(\theta)$. ■

Fie doi vectori $\mathbf{x}$ și $\mathbf{y}$, iar proiecția lui $\mathbf{x}$ pe $\mathbf{y}$ este $\mathbf{x_y}$. Pentru a îl găsi pe $\mathbf{x_y}$, ne gândim astfel:

- În primul rând, ne trebuie norma lui $\mathbf{x_y}$: $||\mathbf{x}|| \cos(\theta)$.
- Având norma, trebuie să avem și o direcție. Proiecția fiind pe $\mathbf{y}$, o putem găsi prin *normalizare*: $\frac{\mathbf{y}}{||\mathbf{y}||}$.



Avem și o interpretare geometrică a produsului scalar: produsul dintre norma proiecției pe un vector și norma vectorului pe care se proiectează.

Definim operatorul de proiecție astfel: $proj_y x = \frac{||x|| \cos(\theta)}{||y||} y = \frac{\langle x,y\rangle}{\langle y,y\rangle} y$.

$$proj_y x = \frac{||x|| \cos(\theta)}{||y||} y = \frac{||y||||x|| \cos(\theta)}{||y||^2} y = \frac{\langle x, y\rangle}{\langle y, y\rangle} y$$

2.3 Vectori ortogonali. Matrice unitară/ortogonală.

Doi vectori $x, y \in \mathbb{R}^n$ sunt ortogonali dacă produsul lor scalar este zero, adică $x^T y = 0$. Cu alte cuvinte, direcțiile lor sunt perpendiculare. În plus, dacă $||x||_2 = ||y||_2 = 1$, atunci cei doi vectori sunt *ortonormați.*

O bază a unui spațiu vectorial se numește ortogonală, respectiv ortonormată, dacă vectorii acesteia sunt ortogonali, respectiv ortonormați.

O matrice $A \in \mathbb{C}^{n\times n}$ se numește *unitară* dacă $A^*A = AA^* = I_n$, unde A^* este conjugata transpusă a lui A. Dacă A este reală, atunci matricea se numește *ortogonală* și putem scrie $A^T A = AA^T = I_n$. Ele sunt foarte utilizate în diverse aplicații, precum descompunerea QR sau descompunerea valorilor singulare.

O matrice $A \in \mathbb{C}^{n\times n}$ sau $\mathbb{R}^{n\times n}$ unitară/ortogonală are următoarele proprietăți:

- coloanele (rândurile) sale formează o bază ortonormată a spațiului vectorial $\mathbb{C}^n$ sau $\mathbb{R}^n$;
- norma vectorilor coloană (rând) este 1;
- $A^{-1} = A^*$;
- este normală, adică $A^*A = AA^*$;
- valorile proprii se află pe cercul unitate;
- vectorii proprii sunt ortogonali;
- $det(A) = \pm 1$;
- $||A||_2 = 1$;
- conservă produsul scalar: $(Ax)^*(Ay) = x^*A^*Ay = x^*y$;

Astfel, matricile ortogonale se pot interpreta geometric ca fiind rotații, reflecții, permutări, identități sau combinații ale acestora.

Ce se întâmplă dacă aplicăm o matrice ortogonală asupra vectorilor unitate?
Fie $A2 = \begin{bmatrix} \cos(\frac{\pi}{7}) & -sin(\frac{\pi}{7}) \\ sin(\frac{\pi}{7}) & cos(\frac{\pi}{7}) \end{bmatrix}$.

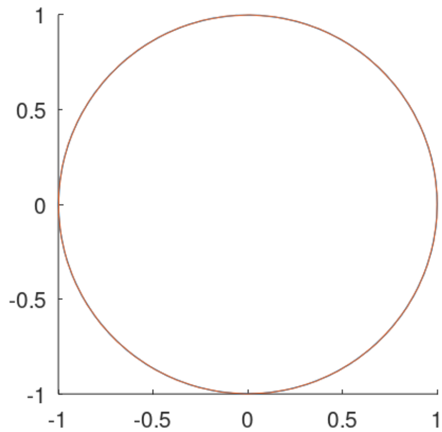


Figura 3: Transformarea $A2$

Toți vectorii au fost rotiți cu un unghi de $\frac{\pi}{7}$. Nu s-a modificat nimic altceva! Norma vectorilor a rămas la fel și deci graficele coincid. În următoarea figură, aplicăm $A3 = 2 * A2$, matrice care nu mai este ortogonală.

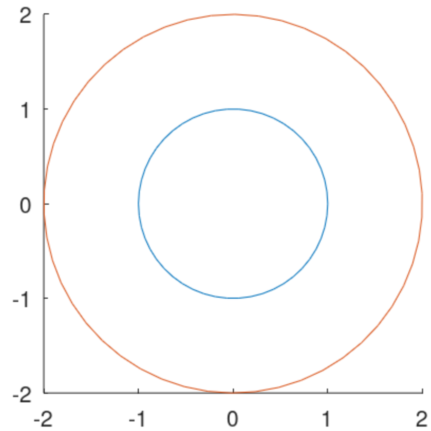


Figura 4: Transformarea $A2$

2.4 Transformări ortogonale. Descompunerea QR.

Definiție. Fie $T : V \to V$ o transformare liniară.

$$T \text{ - ortogonală} \equiv \langle T(x), T(y) \rangle = \langle x, y \rangle$$

Utilitatea transformărilor ortogonale în cazul sistemelor liniare constă în faptul că putem aplica o serie de astfel transformări pentru a introduce 0-uri în matricea sistemului. La final, aflarea soluției va consta în rezolvarea unui sistem triunghiular. Matricea va avea forma $A = QR$, unde Q este o matrice ortogonală și R este o matrice superior triunghiulară.

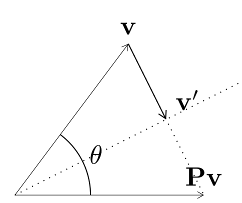
$$Ax = b \Leftrightarrow QRx = b \Leftrightarrow Rx = Q^*b$$

Cu aceste transformări, ne dorim să aducem vectori de la forma $\begin{bmatrix} x \\ y \end{bmatrix}$ la forma $\begin{bmatrix} x' \\ 0 \end{bmatrix}$.

2.4.1 Reflexii. Transformarea Householder.

Căutăm o transformare P astfel încât $Pv = ||v||e$, unde e este un vector din baza canonică.

Pentru reflexie, ne alegem un vector d care ne va da *direcția de reflexie*, $||d||_2 = 1$



$$\begin{aligned} v' &= proj_d(-v) = \frac{\langle v, d\rangle}{\langle d, d\rangle} d \\ v' &= -v^* dd = -dd^* v \implies \end{aligned}$$

$$\begin{aligned} Pv &= v - 2v' = v - 2dd^* v \\ P &= I - 2dd^* \end{aligned}$$

Iar în cazul în care d nu are norma 1, ajungem la forma generală a reflectorului Householder, prin normalizare:

$$P = I - 2\frac{dd^T}{d^T d}$$

Afirmație. P este ortogonală.

Demonstrație. $P^T P = (I - 2dd^*)^*(I - 2dd^*) = I - 2dd^* - 2dd^* + 4dd^*dd^* = I - 4dd^* + 4dd^* = I$. ■

Cum găsim d pentru a introduce 0-uri?

Cum P este ortognală, știm că $||Pv||_2 = ||v||_2$. Astfel, ne dorim ca $Pv = \pm||v||_2 e_1$.

$$\begin{aligned} v + d &= Pv \\ v + d &= \pm||v||_2 e_1 \\ d &= \pm||v||_2 e_1 - v, \text{cum semnul lui } d \text{ nu contează, alegem} \\ d &= v \pm ||v||_2 e_1 \end{aligned}$$

Plus sau minus? Răspunsul îl putem găsi efectuând puțină analiză numerică, fără a demonstra nimic formal de data asta. Plecăm de la următoarea întrebare: Este bine ca Pv și v să fie apropiate?

Știind că calculul numeric nu este perfect, putem presupune că nici reflexia nu va fi perfectă. Problema este că dacă **v** este deja foarte aproape de axe, e foarte posibil ca **Pv** să fie chiar mai departe de aceasta.

În schimb, dacă **v** și reflexia acestuia sunt depărtate, eroare poate fi neglijabilă.

Ne dorim ca $||v - \alpha||v||_2 e_1||_2$ să fie maximă, unde $\alpha = \pm 1$. Considerăm doar cazul numerelor reale.

$$\begin{aligned} ||v - \alpha||v||_2||_2 = (v - \alpha||v||_2 e_1)^T (v - \alpha||v||_2 e_1) &= (v - \alpha||v||_2 e_1)^T (v - \alpha||v||_2 e_1) \\ &= v^T v - 2\alpha||v||_2 v^T e_1 + \alpha^2 ||v||_2^2 e_1^T e_1 \\ &= v^T v - 2\alpha||v||_2 v_1 + ||v||_2^2 \end{aligned}$$

Cum $v^T v$ și $||v||_2^2$ sunt constante, trebuie să găsim maximul termenului $-\alpha||v||_2 v_1$, mai precis $-\alpha v_1$. Cum α poate fi doar 0 sau 1, rezultă imediat că $\alpha = -sign(v_1)$.

În figura de mai jos (figura 5), avem un grafic al $||A - QR||_2/||A||_2$. Au fost generate 1000 de teste, deci avem 1000 de puncte pe grafic. În fiecare test s-a generat o matrice 3x3 aleatorie iar prima sa coloană a fost înlocuită cu $\begin{bmatrix}1 & \delta & 0\end{bmatrix}^T$, unde δ lua 17 valori între 10^{-16} și 1. Cu albastru avem cazul când alegem semnul ca mai sus, iar cu roșu când alegem semnul opus.

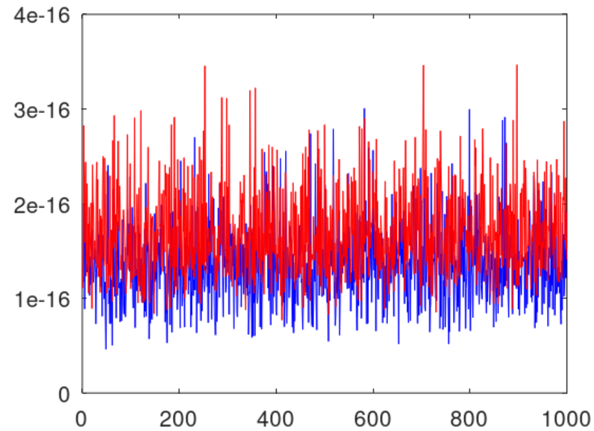


Figura 5: Eroarea relativă

Se poate observa că eroarea relativă, în medie, mai mică atunci când alegem semnul bine.

Anulare catastrofală. Având o precizie limitată în calculul numeric, se pare că diferența a două aproximari a unor numere foarte apropiate poate duce la o aproximare foarte rea.

Demonstrație. Fie aproximările $\overline{x}$ și $\overline{y}$, cu erorile relative $\epsilon_x = \frac{x-\overline{x}}{x}$ și $\epsilon_y = \frac{y-\overline{y}}{y}$.

$$\overline{x} = x(1+\epsilon_x)$$
$$\overline{y} = y(1+\epsilon_y)$$

Atunci,

$$\begin{aligned}\overline{x} - \overline{y} &= x(1+\epsilon_x) - y(1+\epsilon_y) = x - y + x\epsilon_x - y\epsilon_y \\ &= x - y + (x-y)\epsilon_{xy}, \text{unde } \epsilon_{xy} = \frac{x\epsilon_x - y\epsilon_y}{x-y} \\ &= (x-y)(1+\epsilon_{xy})\end{aligned}$$

Numitorul lui ϵ_{xy} este foarte mic daca $x \approx y$, deci eroarea devine foarte mare. ■

În concluzie, obținem următoarea formulă pentru d: $d = v + sign(v_1)||v||_2 e_1$.

Pentru a oferi un exemplu practic a descompunerii QR cu Householder, mai e utilă următoarea informație: putem "umple" vectorul d cu 0-uri în locurile unde ne dorim ca vectorii să nu fie afectați. De exmeplu, dacă avem 3 dimensiuni și vrem să punem 0 **doar** pe poziția 3, atunci $d = \begin{bmatrix} 0 \\ v2 - sign(v_2)||v'||_2 \\ v3 \end{bmatrix}$ sau $d = \begin{bmatrix} v1 - sign(v_1)||v'||_2 \\ 0 \\ v3 \end{bmatrix}$, unde $v' = \begin{bmatrix} v2 \\ v3 \end{bmatrix}$, respectiv $v' = \begin{bmatrix} v1 \\ v3 \end{bmatrix}$

Exemplu. Fie matricea $A_1 = \begin{bmatrix} 2 & 4 & 5 \\ 1 & -1 & 1 \\ 2 & 1 & -1 \end{bmatrix}$. Ne dorim să găsim matricea Q și R astfel încât $A = QR$

și Q să fie ortogonală.

La prima iterație ne dorim să punem 0-uri pe pozițiile (2, 1) și (3, 1).

$$v = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \quad ||v||_2 = 3, \quad d = \begin{bmatrix} 2 - sign(2) * 3 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, \quad ||d||_2^2 = 6$$

$$H_1 = I_3 - 2\frac{dd^T}{d^T d} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}\begin{bmatrix} -1 & 1 & 2 \end{bmatrix}$$

$$H_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 1 & -1 & -2 \\ -1 & 1 & 2 \\ -2 & 2 & 4 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 & 1 & 2 \\ 1 & 2 & -2 \\ 2 & -2 & -1 \end{bmatrix}$$

$$A_2 = H_1 A_1 = \frac{1}{3}\begin{bmatrix} 2 & 1 & 2 \\ 1 & 2 & -2 \\ 2 & -2 & -1 \end{bmatrix}\begin{bmatrix} 2 & 4 & 5 \\ 1 & -1 & 1 \\ 2 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 3 & 3 \\ 0 & 0 & 3 \\ 0 & 3 & 3 \end{bmatrix}$$

La a doua iterație ne dorim să punem 0 pe poziția (2, 3).

$$v = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \quad ||v||_2 = 3, \quad d = \begin{bmatrix} 0 \\ -sign(0) * 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ -3 \\ 3 \end{bmatrix}, \quad ||d||_2^2 = 18$$

$$H_2 = I_3 - 2\frac{dd^T}{d^T d} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{9}\begin{bmatrix} 0 \\ -3 \\ 3 \end{bmatrix}\begin{bmatrix} 0 & -3 & 3 \end{bmatrix}$$

$$H_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{9}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 9 & -9 \\ 0 & -9 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

$$A_3 = H_2 A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix}\begin{bmatrix} 3 & 3 & 3 \\ 0 & 0 & 3 \\ 0 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 3 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & -3 \end{bmatrix}$$

A_3 este matricea R, iar $Q = H_1^T H_2^T = H_1 H_2$ este matricea ortogonală.

2.4.2 Rotații Givens

Pentru demonstrarea acestor rotiri ne putem referi foarte simplu la numerele complexe. Fără a demonstra, putem identifica orice număr complex de forma $a + bi$ ca fiind un vector din $\mathbb{R}^2$, $\begin{bmatrix} a \\ b \end{bmatrix}$. Înmulțirea acestui număr cu un alt număr complex $c + di$ este echivalentă cu înmulțirea vectorului cu matricea $\begin{bmatrix} c & -d \\ d & c \end{bmatrix}$.

Știind că $(\cos\alpha + i\sin\alpha)(\cos\beta + i\sin\beta) = \cos(\alpha+\beta) + i\sin(\alpha+\beta)$, și că orice număr complex poate fi scris sub forma $r(\cos\beta + i\sin\beta)$, putem deduce că înmulțirea cu matricea $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ rotește vectorii cu un unghi θ, în sens trigonometric. Această matrice este una ortognală.

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = I$$

Cum $-\sin\theta = \sin(-\theta)$, putem vedea prima matrice din ecuația de mai sus ca o rotație cu $-\theta$.

Prin aceste rotații, ne dorim același lucru ca și la Householder, să aducem vectori de la forma $\begin{bmatrix} x \\ y \end{bmatrix}$ la forma $\begin{bmatrix} x' \\ 0 \end{bmatrix}$. Fiind o transformare ortogonală, norma vectorilor se păstrează și am avea că $x' = \sqrt{x^2+y^2}$. Pentru a calcula matricea, nu avem nevoie de unghi, ci doar de cos și sin.

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x\cos\theta - y\sin\theta \\ y\cos\theta + x\sin\theta \end{bmatrix} = \begin{bmatrix} \sqrt{x^2+y^2} \\ 0 \end{bmatrix}$$

Rezolvând acest sistem găsim că $\cos\theta = \frac{x}{\sqrt{x^2+y^2}}$ și $\sin\theta = -\frac{y}{\sqrt{x^2+y^2}}$.

Pentru matrici mai mari, de dorim să găsim o matrice G astfel încât

$$G \begin{bmatrix} a \\ \vdots \\ x \\ \vdots \\ y \\ \vdots \\ z \end{bmatrix} = \begin{bmatrix} a \\ \vdots \\ x' \\ \vdots \\ 0 \\ \vdots \\ z \end{bmatrix}$$

G seamănă foarte mult cu matricea identitate și se numește matricea Givens.

$$G = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & c & \cdots & -s & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & s & \cdots & c & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix}$$

Cu alte cuvinte, matricea G este plină de 0 mai puțin următoarele elemente: $G(i,i) = G(j,j) = c$, $G(i,j) = s$ și $G(j,i) = -s$, unde i și j sunt pozițiile pe care se găsește y, respectiv x. Asta înseamnă că i va indica mereu poziția elementului în vector pe care vrem să îl facem 0. O regulă bună la descompunerea QR este ca (i,j) să fie fix poziția elementului din matrice pe care vrem sa îl facem 0.

Exemplu. Fie $A_1 = \begin{bmatrix} 0 & 1 & 2 \\ 3 & 2 & 0 \\ 4 & 1 & 5 \end{bmatrix}$. Ne dorim să găsim matricea Q și R astfel încât $A = QR$ și Q să fie ortogonală.

Facem 0 pe poziția $(3,1)$.

$$(i,j) = (3,1), \quad x = 0, \quad y = 4, \quad r = \sqrt{0^2+4^2} = 4, \quad c = \frac{x}{r} = 0, \quad s = -\frac{y}{r} = -1$$

$$G_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}$$

$$A_2 = G_1 A_1 = \begin{bmatrix} 4 & 1 & 5 \\ 3 & 2 & 0 \\ 0 & -1 & -2 \end{bmatrix}$$

Facem 0 pe poziția $(2, 1)$.

$$(i, j) = (2, 1), \quad x = 4, \quad y = 3, \quad r = \sqrt{4^2 + 3^2} = 5, \quad c = \frac{x}{r} = \frac{4}{5}, \quad s = -\frac{y}{r} = -\frac{3}{5}$$

$$G_2 = \begin{bmatrix} \frac{4}{5} & 0 & \frac{3}{5} \\ 0 & 1 & 0 \\ -\frac{3}{5} & 0 & \frac{4}{5} \end{bmatrix}$$

$$A_3 = G_2 A_2 = \begin{bmatrix} 5 & 2 & 4 \\ 0 & 1 & 0 \\ 0 & -1 & -2 \end{bmatrix}$$

Facem 0 pe poziția $(3, 2)$.

$$(i, j) = (3, 2), \quad x = 1, \quad y = -1, \quad r = \sqrt{1^2 + (-1)^2} = \sqrt{2}, \quad c = \frac{x}{r} = \frac{\sqrt{2}}{2}, \quad s = -\frac{y}{r} = \frac{\sqrt{2}}{2}$$

$$G_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

$$A_4 = G_3 A_3 = \begin{bmatrix} 5 & 2 & 4 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & -\sqrt{2} \end{bmatrix}$$

Am adus matricea la forma superior triunghiulară. Așadar, $R = A_4$ și $Q = G_1^T G_2^T G_3^T$.

Se observă că față de Householder, am folosit mai multe matrici pentru a obține descompunerea. În medie, numărul de reflexii pe care trebuie să le facem este egal cu numărul de coloane, iar numărul de rotații este cu un ordin de mărime mai mare, mai ales dacă matricea inițială nu este pătratică.

Avantajul rotațiilor Givens este că față de reflexii, este ca sunt mult mai paralelizabile. O rotație $G(i, j)$ afectează doar liniile i și j ale matricei, deci am putea calcula mai multe rotații în paralel. Ca și regulă, Householder se folosește pentru matrici dense, iar Givens pentru matrici rare.

2.5 Polinoame ortogonale

Polinoamele pot defini și ele un spațiu vectorial. Folosind baza $1, x, x^2, \ldots, x^n$, putem genera orice polinom. Acestui spațiu vectorial îi putem atașa un produs scalar. Pe când la vectori produsul scalar era definit ca o sumă a produselor coomponentelor, analog în cest caz ar fi integrala. Considerăm polinoame definite pe $[-1, 1]$.

$$\langle f, g \rangle = \int_{-1}^{1} f(x) g(x) dx$$

Vom aplica Gram-Schmidt pe baza polinoamelor de mai sus, pentru a obține *polinoame ortogonale* (în special polinoame Legendre). Notăm baza de polinoame cu $e_i = x^i, \forall i = 0, 1, \ldots$.

$$u_i(x) = e_i(x) - \sum_{j=0}^{i-1} \frac{\langle e_i, u_j \rangle}{\langle u_j, u_j \rangle} u_j(x)$$

Pentru normalizare, impunem ca $u(1) = 1$.

$$u(x) = \frac{u(x)}{u(1)}$$

2.6 Tehnici pentru sisteme inconsistente. Regresii liniare în sens CMMP.

Sunt cazuri în care sistemul $\mathbf{Ax} = \mathbf{b}$ nu are soluție, chiar din modul în care a fost construit (A nepătratică). Totuși, ne punem următoarea problemă:

Găsește $\mathbf{x}$ astfel încât $||\mathbf{Ax} - \mathbf{b}||_2$ să fie minim.

Explicație analitică. Propoziția de mai sus este echivalentă cu găsirea minimului lui $||\mathbf{Ax} - \mathbf{b}||_2^2$. În cazul acesta putem face referire la produsul scalar:

$$\begin{aligned} ||\mathbf{Ax} - \mathbf{b}||_2^2 &= \langle Ax - b, Ax - b\rangle \\ &= \langle Ax, Ax\rangle - 2\langle Ax, b\rangle + \langle b, b\rangle \\ &= x^T A^T Ax - 2x^T A^T b + b^T b \end{aligned}$$

Astfel, transformăm problema în una de optimizare, unde ne dorim să găsim $\mathbf{x}$ astfel încât funcția

$$\mathbf{f(x) = x^T A^T A x - 2x^T A^T b + b^T b}$$

să fie minimă.

Gradientul și hessiana funcției sunt:

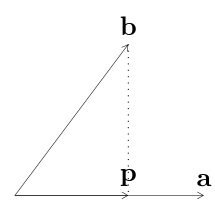
$$\begin{aligned} &\nabla \mathbf{f(x) = 2A^T Ax - 2A^T b = 2(A^T Ax - A^T b)} \\ &\mathbf{H(x) = 2A^T A} \end{aligned}$$

Hessiana este o matrice semi-pozitiv definită, având toate valorile proprii mai mari sau egale cu 0. Deci minimul este ar fi atins când gradientul este 0, adică

$$\mathbf{A^T Ax - A^T b = 0}$$

Practic, rămâne să rezolvăm sistemul $A^T Ax = A^T b$. În cazul regresiei liniare, sistemul nu se poate rezolva atunci cand nu avem un set de date cu cel puțin 2 puncte (trebuie să trasăm o dreapta iar dreapta este definită de două) puncte.

Explicație geometrică. În cazul cel mai simplu, ne putem gândi la distanța de la un punct la o dreaptă. Distanța este dată de segmentul perpendicular de la punct la dreaptă. În cazul a doi vectori, distanța este dată de diferența dintre un vector și proiecția sa pe celălalt vector.



Practic, vectorul $\mathbf{p}$ este vectorul cel mai apropiat de $\mathbf{b}$, din span-ul lui $\mathbf{a}$. Fie $||a||_2 = 1$, atunci $\mathbf{p}$ este

$$\begin{aligned} \mathbf{p} &= \frac{\langle \mathbf{b}, \mathbf{a}\rangle}{\langle \mathbf{a}, \mathbf{a}\rangle}\mathbf{a} \\ &= \mathbf{a}\langle \mathbf{b}, \mathbf{a}\rangle \\ &= \mathbf{aa^T b} \end{aligned}$$

Definim $P = aa^T$ ca fiind *matricea de proiecție* pe a. Dacă a nu are norma 1, atunci $P = \frac{aa^T}{a^T a}$.

Cum proiectăm pe un plan? Contstruim matricea $A = \begin{bmatrix} a_1 & a_2 \end{bmatrix}$, care reprezintă baza planului. Proiectția pe plan este dată de suma proiecțiilor pe baza planului, adică $\mathbf{p} = x_1\mathbf{a_1} + x_2\mathbf{a_2} = A\mathbf{x}$. Cum $b - p = b - Ax$ este perpendicular pe plan, din construcție, avem următoarele relații:

$$\begin{aligned} \mathbf{a_1^T}(\mathbf{b} - \mathbf{Ax}) &= \mathbf{0} \\ \mathbf{a_2^T}(\mathbf{b} - \mathbf{Ax}) &= \mathbf{0} \end{aligned}$$

Ecvhivalent, în forma matriceală, avem $A^T(b - Ax) = 0$. Rezultă că $A^TAx = A^Tb$. Soluția este dată de $x = (A^TA)^{-1}A^Tb$. De aici putem obține matricea de proiecție:

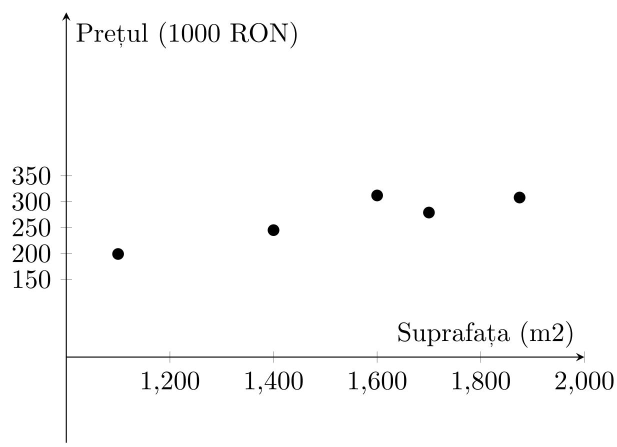
$$\begin{aligned} \mathbf{x} =& (\mathbf{A^TA})^{-1}\mathbf{A^Tb} \\ \mathbf{p} = \mathbf{Ax} =& = \mathbf{A}(\mathbf{A^TA})^{-1}\mathbf{A^Tb} \\ \mathbf{P} =& \mathbf{A}(\mathbf{A^TA})^{-1}\mathbf{A^T} \end{aligned}$$

Orice matrice de proiecție are proprietatea că $P^2 = P$. Mai mult, în acest caz avem că $P^T = P$.

Exemplu. Într-un cartier de case încercăm să prezicem prețul unei case în funcție de suprafața casei. Avem următoarele date:

Suprafața (m2)	Prețul (1000 RON)
1400	245
1600	312
1700	279
1875	308
1100	199

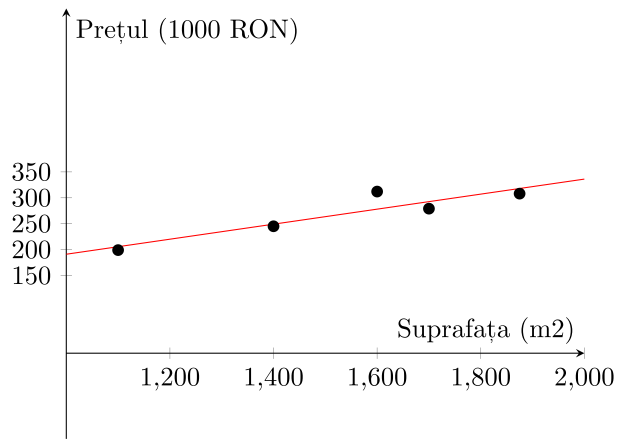
Punem pe grafic acest set de date:



Se obervă o relație liniară între suprafață și preț, dată de $y = a_1x + a_2$. Formulăm această problemă ca un sistem de ecuații liniare:

$$\begin{bmatrix} 1400 & 1 \\ 1600 & 1 \\ 1700 & 1 \\ 1875 & 1 \\ 1100 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 245 \\ 312 \\ 279 \\ 308 \\ 199 \end{bmatrix}$$

Soluția este dată de rezolvarea sistemului $A^T Ax = A^T b$. După rezolvare, găsim că $a_1 = 0.1450$ și $a_2 = 46.0575$. Graficul dreptei de regresie este:



3 Probleme

1. Script MATLAB pentru descompunerea QR a unei matrice folosind reflexii Householder.
2. Script MATLAB pentru descompunerea QR a unei matrice folosind rotații Givens.
3. Demonstrați că matricea Householder este simetrică.
4. Aduceți o matrice 3x3 Householder la forma diagonală folosind rotații Givens.
5. Să se determine descompunerea QR pentru matricea $A = \begin{bmatrix} 3 & 1 & -2 \\ 1 & 3 & 1 \\ -2 & 1 & 3 \end{bmatrix}$ folosind transformarea Givens.
6. Regresie liniară în sens CMMP pe propriul set de date. Trasați graficele punctelor și a dreptei. Hint: `hold on`

Referințe

[1] Michael L. Overton, Pinze Yu. On the choice of sign defining Householder transformations. Numerical Algebra, Control and Optimization, 2025, 15(2): 502-505. doi: 10.3934/naco.2023025