Cursul 7

Proiecţia ortogonală a unui vector pe un subspaţiu vectorial. Soluţia celor mai mici pătrate. Procedura Gramm-Schmidt

Noţiunile şi rezultatele pe care le prezentăm în continuare sunt valabile în orice spaţiu vectorial real V_n înzestrat cu un produs scalar, dar discutăm acum doar cazul concret al spaţiului $(\mathbb{R}^n, <>)$.

7.1 Complementul ortogonal $v^\perp$

În $\mathbb{R}^n$ fixăm un vector nenul $v = (a_1, a_2, \dots, a_n)^T$ şi notăm cu $v^\perp$ mulţimea vectorilor din $\mathbb{R}^n$ ce sunt perpendiculari pe v:

$$v^\perp = \{w = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n, w \perp v \Leftrightarrow < v, w >= 0 \Leftrightarrow a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0\}$$

Observăm că $v^\perp$ este mulţimea soluţiilor sistemului liniar şi omogen:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0$$

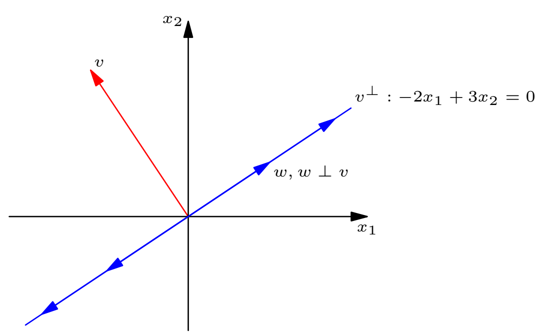
şi deci $v^\perp$ este un subspaţiu vectorial al lui $\mathbb{R}^n$. Matricea "sistemului" este:

$$A = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$$

şi deoarece $v \neq \theta$, rang(A)=1 (cel puţin o coordonată $a_i \neq 0$). Deci dimensiunea spaţiului Null(A)=$v^\perp$ este $n - 1 =$ numărul de coloane în A minus rang(A).

Subspaţiul vectorial $v^\perp$, al lui $\mathbb{R}^n$ se numeşte **complementul ortogonal** al vectorului v.

Exemplul 1. Fie $v = (-2, 3)^T \in \mathbb{R}^2$. Complementul său ortogonal, $v^\perp$, este subspaţiul lui $\mathbb{R}^2$ de ecuatie $-2x_1 + 3x_2 = 0$. Din punct de vedere geometric aceasta este o dreaptă ce trece prin origine. Toţi vectorii "cu suportul pe această dreaptă" sunt din $v^\perp$, adică sunt ortogonali pe v:

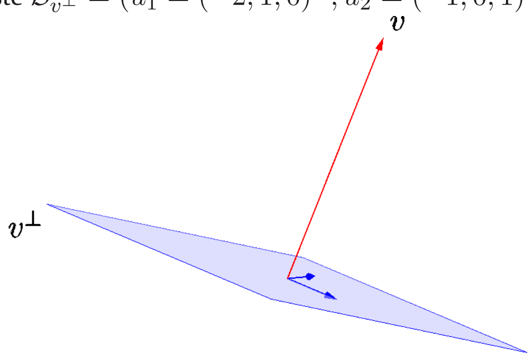


Exemplul 2. Să se determine o bază în $v^{\perp}$ unde $v = (1, 2, 1)^T \in \mathbb{R}^3$.

Complementul ortogonal al vectorului v are ecuaţia $x_1 + 2x_2 + x_3 = 0$. Conform celor deduse mai sus, dimensiunea lui $v^{\perp}$ este în acest caz $3 - 1 = 2$. Deci o bază în $v^{\perp}$ conţine doi vectori. Deoarece $v^{\perp}$ este mulţimea soluţiilor sistemului $x_1 + 2x_2 + 3x_3 = 0$, determinăm această mulţime. Alegem x_1 necunoscuta principală şi $x_2 = \alpha$, $x_3 = \beta$, necunoscute secundare. Astfel $x_1 = -2\alpha - \beta$ şi deci:

$$v^{\perp} = \{w = \begin{bmatrix} -2\alpha - \beta \\ \alpha \\ \beta \end{bmatrix} = \alpha \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \alpha, \beta \in \mathbb{R}\}$$

Astfel o bază în $v^{\perp}$ este $\mathcal{B}_{v^{\perp}} = (u_1 = (-2, 1, 0)^T, u_2 = (-1, 0, 1)^T)$.



Exemplul 3. În $\mathbb{R}^4$ se dă subspaţiul vectorial, S, de ecuaţie

$$-7x_1 + 3x_2 + x_3 - 2x_4 = 0$$

Subspaţiul este complementul ortogonal al vectorului $v = (-7, 3, 1, -2)^T \in \mathbb{R}^4$. v este ortogonal pe orice vector din subspaţiul S sau simplu, v este ortogonal pe S.

7.2 Proiecţia ortogonală a unui vector pe un subspaţiu

Fie S un subspaţiu vectorial al lui $\mathbb{R}^n$, $S \neq \{\theta\}$ şi v un vector nenul din $\mathbb{R}^n$.

Definiţia 7.2.1 *Proiecţia ortogonală a vectorului v pe subspaţiul S este vectorul $s \in S$ cu proprietatea că $v - s \perp S$, adică vectorul $v - s$ este ortogonal pe orice vector din subspaţiu, în particular şi pe s, $< v - s, s >= 0$ (Fig.7.1)*

Se poate demonstra că proiecţia ortogonală a unui vector v pe subspaţiul S este unică.

Cazul 1. Proiecţia ortogonală a unui vector v pe un vector nenul w este proiecţia ortogonală a vectorului v pe subspaţiul $1D$ generat de vectorul $w \neq 0$: $S = \{s \in V \mid s = \alpha w, \alpha \in \mathbb{R}\} = \text{span}(w)$. Conform definiţiei, proiecţia ortogonală a lui v pe S este un vector $s = aw$, $a \in \mathbb{R}$, astfel încât $< v - aw, s >= 0$. Dar dacă un vector este ortogonal pe $s = aw$, atunci el este ortogonal şi pe w şi deci avem $< v - aw, w >= 0$, adică $< v, w > -a < w, w >= 0$. Rezultă astfel că

$$a = \frac{< v, w >}{< w, w >},$$

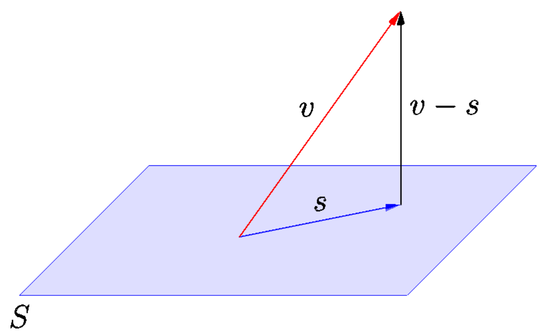


Fig.7.1: Proiecţia ortogonală a unui vector v pe un subspaţiu S.

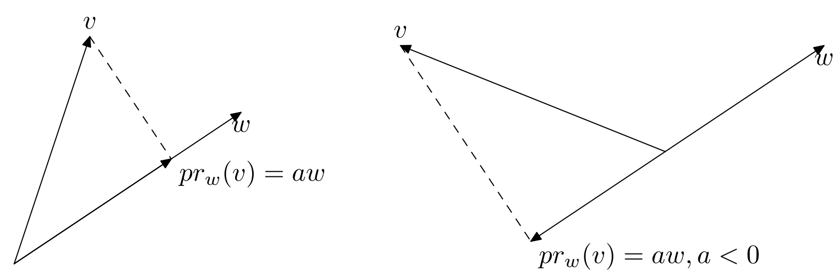


Fig.7.2: Ilustrarea proiecţiei ortogonale a unui vector nenul, v, pe un vector $w \neq \theta$.

iar vectorul proiecţie este

$$\mathbf{s} := \mathbf{pr_w(v)} = \frac{<\mathbf{v}, \mathbf{w}>}{<\mathbf{w}, \mathbf{w}>}\mathbf{w}$$

Dacă vectorul w este un versor, adică $\|e\| = 1 \quad \Leftrightarrow \quad <e, e>= 1$, atunci proiecţia ortogonală a lui v pe versorul e este:

$$pr_e(v) = \frac{<v, e>}{<e, e>}e = (<v, e>)e$$

Consecinţă: Deoarece coordonatele unui vector v într-o bază ortonormată, $\mathcal{B} = (e_1, e_2, \ldots, e_n)$, sunt $x_i =<v, e_i>$, rezultă că exprimarea vectorului v în baza $\mathcal{B}$ este suma proiecţiilor ortogonale ale vectorului v pe vectorii bazei:

$$v =<v, e_1> e_1+ <v, e_2> e_2 + \cdots + <v, e_n> e_n = pr_{e_1}(v) + pr_{e_2}(v) + \cdots + pr_{e_n}(v)$$

Această proprietate generalizează la orice spaţiu euclidian, modalitatea de a descompune un vector după două direcţii ortogonale, aşa cum se proceda la fizică cu o forţă ce acţionează în plan (Fig.7.3).

Cazul 2. Proiecţia ortogonală a unui vector $v \neq \theta$ pe un subspaţiu vectorial $S \subset \mathbb{R}^n$ de dimensiune mai mare ca 1 (Fig.7.1).

Să determinăm acum o modalitate de a găsi proiecţia ortogonală a unui vector $v \in \mathbb{R}^n$, pe un subspaţiu liniar, S **raportat la o bază ortonormată**.

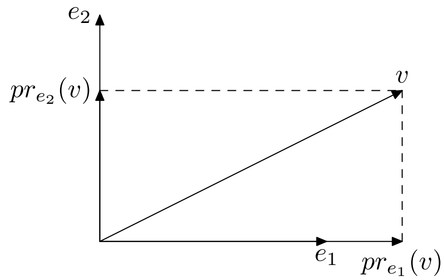


Fig.7.3: Proiecţia ortogonală a vectorului v pe vectorii bazei ortonormate $\mathcal{B} = (e_1, e_2)$. $v = pr_{e_1}(v) + pr_{e_2}(v)$.

Propoziţia 7.2.1 *Fie S un subspaţiu vectorial de dimensiune m al lui $\mathbb{R}^n$, $m < n$, şi $\mathcal{B}_S = (u_1, u_2, \ldots, u_m)$ o bază ortonormată în S. Atunci vectorul $s \in S$, care este suma proiecţiilor ortogonale ale lui v pe vectorii bazei din S,*

$$s = pr_{u_1}(v) + pr_{u_2}(v) + pr_{u_m}(v) = < v, u_1 > u_1 + < v, u_2 > u_2 + \cdots < v, u_m > u_m$$

este proiecţia ortogonală a vectorului v pe subspaţiul S.

Demonstraţie: Pentru a demonstra că $pr_S(v)$ este vectorul $s = < v, u_1 > u_1 + < v, u_2 > u_2 + \cdots + < v, u_m > u_m$, arătăm că $v - s \perp s$, adică $< v - s, s > = 0$ sau echivalent, $< v, s > = < s, s >$. Cum proiecţia ortogonală a lui v pe s este unicul vector cu această proprietae, va rezulta că $pr_S(v) = s$.

Aplicând proprietăţile produsului scalar avem:

$$\begin{aligned} < v, s > &= < v, < v, u_1 > u_1 + < v, u_2 > u_2 + \cdots + < v, u_m > u_m > = \\ &= < v, u_1 >< v, u_1 > + < v, u_2 >< v, u_2 > + \cdots + < v, u_m >< v, u_m > \\ &= < v, u_1 >^2 + < v, u_2 >^2 + \cdots + < v, u_m >^2 \end{aligned}$$

Pentru a calcula $< s, s >$ ţinem seama că s este exprimat într-o bază ortonormată şi deci $< s, s > = ||s||^2$ =suma coordonatelor lui s în baza ortonormată, la pătrat, adică:

$$< s, s > = < v, u_1 >^2 + < v, u_2 >^2 + \cdots + < v, u_m >^2 = < v, s >$$

□

Exemplul 4. În $\mathbb{R}^3$ considerăm subspaţiul liniar generat de vectorii ortonormaţi $u_1 = \dfrac{1}{\sqrt{11}}(1, -3, 1)^T$, $u_2 = \dfrac{1}{\sqrt{6}}(2, 1, 1)^T$ si vectorul $v = (-5, 1, 3)^T \in \mathbb{R}^3$. Să se determine coordonatele în baza canonică din $\mathbb{R}^3$ a proiecţiei ortogonale a vectorului v pe subspaţiul S.

Vectorii u_1, u_2 fiind ortogonali sunt liniar independenţi si pentru că generează subspaţiul S, ei formează o bază ortonormată în S. Conform Propoziţiei precedente, proiecţia ortogonală a lui v pe S este:

$$s = pr_S(v) = < v, u_1 > u_1 + < v, u_2 > u_2$$

Dar,

$$< v, u_1 >= \frac{1}{\sqrt{11}}(-5+3-3) = -\frac{5}{\sqrt{11}}, \quad < v, u_2 >= -\frac{6}{\sqrt{6}}$$

şi deci $s = pr_S(v) = -\frac{5}{\sqrt{11}}u_1 - \frac{6}{\sqrt{6}}u_2$. Dar vectorul proiecţie este exprimat în baza ortonormată din S şi în enunt se cere să determinăm coordonatele lui s în baza canonică din $\mathbb{R}^3$. Pentru aceasta înlocuim vectorii u_1, u_2 din expresia lui s prin coordonatele lor, care sunt evident, date în baza canonică şi avem:

$$s = -\frac{5}{\sqrt{11}}\frac{1}{\sqrt{11}}(1,-3,1)^T - \frac{6}{\sqrt{6}}\frac{1}{\sqrt{6}}(2,1,1)^T) = (-5,15,-5)^T + (-12,-6,-6)^T = (-17,9,-11)^T$$

7.3 Metoda Gramm-Schmidt de ortonormare a unei baze

În spaţiul $(\mathbb{R}^n, <,>$ baza canonică este o bază ortonormată. În acest spaţiu însă există o infinitate de alte baze ortonormate. Cum construim o bază ortonormată dintr-una neortonormată? Folosind metoda Gramm-Schmidt, care este un algoritm de construcţie recursivă a unei baze ortonormate $\mathcal{B}' = (q_1, q_2, \ldots, q_n)$ într-un spaţiu euclidian arbitrar $(V_n, <,>)$, pornind de la o bază oarecare $\mathcal{B} = (v_1, v_2, \ldots, v_n)$. Procedura recursivă pe care o prezentăm se bazează pe proprietatea proiecţiei ortogonale a unui vector v pe un subspaţiu liniar S în care cunosc o bază ortonormată, $(q_1, q_2, \ldots, q_m)$, şi anume că $v - pr_S(v)$ este ortogonal pe orice vector din subspaţiu, deci în particular pe fiecare vector al bazei: $v - pr_S(v) \perp q_i, \forall\, i$.

`Etapa 1:` În etapa 1 se defineşte vectorul q_1 ca fiind versorul vectorului v_1:

$$\mathtt{q_1} \leftarrow \mathtt{v_1}/\|\mathtt{v_1}\|$$

`Etapa 2:` Din q_1 şi v_2 se construieşte un vector o_2 ortogonal pe q_1 şi anume luîm $o_2 = v_2 - pr_{q_1}(v_2) = v_2 - \underbrace{< v_2, q_1 > q_1}_{pr_{q_1}(v_2)}$,

Notând versorul lui o_2 cu q_2:

$$\mathtt{q_2} \leftarrow \frac{\mathtt{v_2} - < \mathtt{v_2}, \mathtt{q_1} > \mathtt{q_1}}{\|\mathtt{v_2} - < \mathtt{v_2}, \mathtt{q_1} > \mathtt{q_1}\|}$$

avem construit sistemul ortonormat $\mathcal{O}_2 = (\mathtt{q_1}, \mathtt{q_2})$.

`Etapa k:` Presupunem că am construit deja un sistem ortonormat $\mathcal{O}_k = (\mathtt{q_1}, \mathtt{q_2}, \ldots, \mathtt{q_k})$, $3 \leq k < n$.

`Etapa k+1:` Definim vectorul o_{k+1} care să fie ortogonal pe fiecare vector $q_1, q_2, \ldots, q_k$, ca fiind diferentă dintre vectorul v_{k+1} şi proiecţia sa ortogonală pe subspaţiul liniar S_k, generat de vectorii ortonormaţi construiţi deja, $q_1, q_2, \ldots, q_k$:

$$o_{k+1} = v_{k+1} - pr_{S_k}(v_{k+1} = v_{k+1} - \underbrace{(< v_{k+1}, q_1 > q_1 + < v_{k+1}, q_2 > q_2 + \cdots + < v_{k+1}, q_k > q_k)}_{pr_{S_k}(v_{k+1})}$$

Din definiţia proiecţiei ortogonale a vectorului v_{k+1} pe subspaţiul S_k rezultă că o_{k+1} este ortogonal pe fiecare vector $q_1, q_2, \ldots, q_k$. Notând versorul lui o_{k+1} cu q_{k+1} avem:

$$\mathtt{q_{k+1}} \leftarrow \frac{\mathtt{v_{k+1}} - \sum_{\mathtt{i=1}}^{\mathtt{k}} < \mathtt{v_{k+1}}, \mathtt{q_i} >}{\|\mathtt{v_{k+1}} - \sum_{\mathtt{i=1}}^{\mathtt{k}} < \mathtt{v_{k+1}}, \mathtt{q_i} > \|}$$

şi deci sistemul $\mathcal{O}_{k+1} = (\mathtt{q_1}, \mathtt{q_2}, \ldots, \mathtt{q_k}, \mathtt{q_{k+1}})$ este un sistem ortonormal construit recursiv.

Observaţia 7.3.1 *Procedura Gramm–Schmidt se poate aplica nu neapărat unei baze arbitrare dintr-un spaţiu euclidian $(V_n, <,>)$ ci şi unui sistem liniar independent $(v_1, v_2, \ldots, v_m)$, $m < n$ căruia i se asociază un sistem ortonormat $\mathcal{O}_m = (q_1, q_2, \ldots, q_m)$. Evident că subspaţiul liniar S al lui V_n generat de vectorii $(v_1, v_2, \ldots, v_m)$ coincide cu subspaţiul liniar generat de sistemul $\mathcal{O}_m$. Cu alte cuvinte $(v_1, v_2, \ldots, v_m)$ este o bază arbitrară în S, iar $(q_1, q_2, \ldots, q_m)$ este o bază ortonormată în S.*

Exemplul 5. În spaţiul euclidian $(\mathbb{R}^3, <,>)$ se dă baza $\mathcal{B} = (v_1 = (0,1,1)^T, v_2 = (1,0,1)^T, v_3 = (1,1,0)^T)$. Aplicând procedura Gramm-Schmidt să se construiască o bază ortonormată în $\mathbb{R}^3$ pornind de la baza $\mathcal{B}$.

`Etapa 1:` Luăm

$$q_1 = \frac{v_1}{\|v_1\|} = \frac{(0,1,1)^T}{\sqrt{2}} = \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$$

`Etapa 2:`

$$o_2 = v_2 - < v_2, q_1 > q_1$$

Dar produsul scalar

$$< v_2, q_1 >= \frac{1}{\sqrt{2}}$$

şi deci:

$$o_2 = v_2 - \frac{1}{\sqrt{2}} q_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 \\ -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \parallel \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

În calculul manual este util să înlocuim vectorii cu coordonate raţionale cu vectori coliniari, cu coordonate întregi (ceea ce nu e cazul într-o implementare a algoritmului Gramm-Schmidt!). Astfel luăm în cazul nostru

$$o_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

Înlocuind o_2 cu versorul său $q_2 = o_2/\|o_2\|$ avem:

$$\mathtt{q}_2 \leftarrow \frac{1}{\sqrt{6}} \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

`Etapa 3:` Construim vectorul:

$$o_3 = v_3 - [< v_3, q_1 > q_1 + < v_3, q_2 > q_2]$$

Calculăm:

$$< v_3, q_1 >= 1/\sqrt{2} \quad \text{şi} \quad < v_3, q_2 >= \frac{1}{\sqrt{6}}$$

Astfel

$$o_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{\sqrt{2}} \underbrace{\begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}}_{q_1} - \frac{1}{\sqrt{6}} \underbrace{\begin{bmatrix} \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}}_{q_2} = \begin{bmatrix} \frac{2}{3} \\ \frac{2}{3} \\ -\frac{2}{3} \end{bmatrix}$$

$\|o_3\| = 2\sqrt{3}/3$ şi deci versorul q_3 al lui o_3 este:

$$q_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

Exemplul 6. a) Să se determine o bază în subspaţiul vectorial S al lui $\mathbb{R}^3$,

$$S = \{v = (x, y, z)^T \mid -3x + 5y + 2z = 0\}$$

şi apoi aplicând procedura Gramm-Schmidt să se ortonormeze baza determinată.

b) Să se determine proiecţia ortogonală a vectorului $v = (-1, 1, 3)$ pe subspaţiul S.

Subspaţiul S este mulţimea soluţiilor sistemului: $-3x + 5y + 2z = 0$. Rangul matricii A a sistemului este 1 deci dimensiunea subspaţiului S este $3 - 1 = 2$. Pentru a determina o bază arbirară rezolvăm sistemul în raport cu necunoscuta principală z şi avem:

$$S = \left\{ v = \begin{bmatrix} x \\ y \\ \frac{3}{2}x - \frac{5}{2}y \end{bmatrix}, x, y \in \mathbb{R} \right\}$$

Orice vector din S se exprimă astfel:

$$v = x \begin{bmatrix} 1 \\ 0 \\ \frac{3}{2} \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ -\frac{5}{2} \end{bmatrix} = \frac{x}{2} \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} + \frac{y}{2} \begin{bmatrix} 0 \\ 2 \\ -5 \end{bmatrix}$$

Deci o bază arbitrară în subspaţiul vectorial S este

$$\mathcal{B}_S = \left(v_1 = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix}, v_2 = \begin{bmatrix} 0 \\ 2 \\ -5 \end{bmatrix} \right),$$

care evident nu este ortonormată.

Luăm

$$q_1 = \frac{v_1}{\|v_1\|} = \frac{(2, 0, 3)^T}{\sqrt{13}} = \left(\frac{2}{\sqrt{13}}, 0, \frac{3}{\sqrt{13}} \right)^T$$

şi

$$o_2 = v_2 - < v_2, q_1 > q_1 = \begin{bmatrix} 0 \\ 2 \\ -5 \end{bmatrix} + \frac{15}{\sqrt{13}} \begin{bmatrix} \frac{2}{\sqrt{13}} \\ 0 \\ \frac{3}{\sqrt{13}} \end{bmatrix} = \begin{bmatrix} \frac{30}{13} \\ 2 \\ -\frac{20}{13} \end{bmatrix}$$

Luând $q_2 = o_2/\|o_2\|$ avem baza ortonormată (q_1, q_2) în S.

b) Proiecţia ortogonală a vectorului v pe subspaţiul S este vectorul

$$s = \mathrm{pr}_{q_1}(v) + \mathrm{pr}_{q_2}(v) =< v, q_1 > q_1 + < v, q_2 > q_2$$

7.4 Cea mai bună aproximaţie printr-un vector dintr-un subspaţiu

Într-un spaţiul vectorial $(\mathbb{R}^n, <,>)$ se defineşte distanţa dintre doi vectori prin: $d(v, w) = \|v - w\|$.

Considerăm un subspaţiu vectorial, S, de dimensiune m al spaţiului vectorial $(\mathbb{R}^n, <,>)$.

În continuare vom arăta că dintre toate distanţele, $d(v, t)$, unde $t \in S$, distanţa cea mai mică este distanţa dintre v şi proiecţia lui ortogonală, $s = pr_S(v)$, pe S. Cu alte cuvinte, cel mai apropiat vector din S, de vectorul $v \in \mathbb{R}^n$ este vectorul $s = pr_S(v)$.

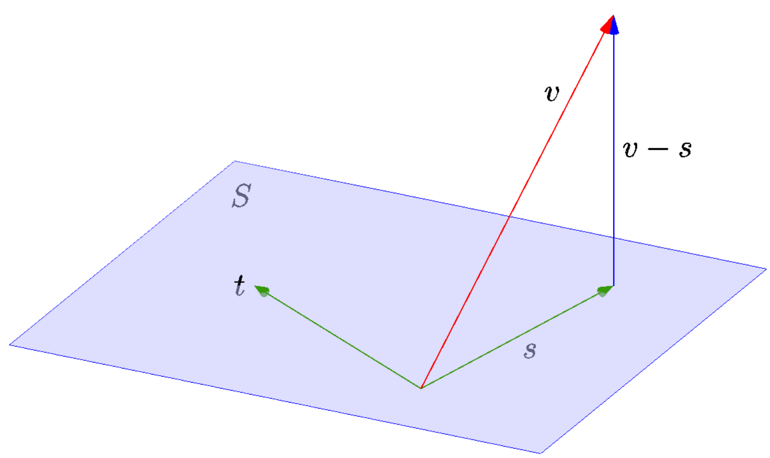


Fig.7.4: Proiecţia ortogonală a vectorului v pe subspaţiul S.

Mai precis, demonstrăm:

Propoziţia 7.4.1 *Dacă v este un vector fixat din $\mathbb{R}^n$ şi $s = pr_S(v)$, proiecţia ortogonală a vectorului v pe S, atunci oricare ar fi t, un vector din subspaţiul S, avem că:*

$$\|v - pr_S(v)\| \leq \|v - t\|$$

Cu alte cuvinte distanţa de la vectorul v la proiecţia sa ortogonală pe un subspaţiu, este mai mică decât distanţa lui v la oricare alt vector din subspaţiu.

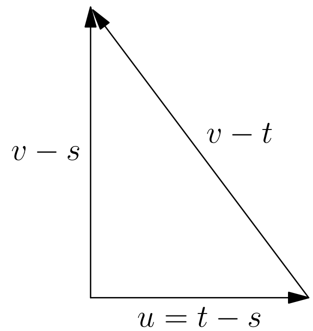


Fig.7.5: Poziţia relativă a vectorilor implicaţi în demonstraţia Propoziţiei 7.4.1.

Demonstraţie: Fie t un vector din subspaţiul S, $t \neq s$. Deoarece $s, t \in S$ rezultă că şi $u = t - s$ este din S. Dar conform din definiţia proiecţiei ortogonale ştim că $v - s$ este ortogonal pe orice vector din S deci şi pe u. Datorită acestei ortogonalităţi ($v - s \perp u = t - s$) şi pentru că diferenţa a doi vectori efectuată cu regula triunghiului dă $v - s - u = v - s - t + s = v - t$, putem scrie relaţia lui Pitagora în triunghiul din Fig. 7.5:

$$\|v - t\|^2 = \|v - s\|^2 + \|t - s\|^2$$

Dacă $t \neq s$, atunci $\|t - s\|^2 > 0$ şi deci $\|v - t\|^2 > \|v - s\|^2$, adică $\|v - s\| < \|v - t\|$, iar dacă $t = s$, avem evident $||v - s|| = ||v - t||$. $\square$

Datorită acestei proprietăţi de monstrate, proiecţia ortogonală a unui vector v pe un susbspaţiu vectorial $S \subset \mathbb{R}^n$, se numeşte **cea mai bună aproximaţie a vectorului v printr-un vector din S**, iar norma diferenţei $v - s$, err $= ||v - pr_S(v)||$, se numeşte eroarea aproximării.

Cea mai bună aproximaţie a unui vector are numeroase aplicaţii în *Machine Learning*, în problemele de recunoaştere a formelor (feţelor umane, a scrisului de mână, etc).

Reţinem că: Având dat un subspaţiu vectorial S al lui $\mathbb{R}^n$, orice vector $v \in \mathbb{R}^n$ se exprimă unic, ca suma a doi vectori ortogonali, unul din S şi celălat ortogonal pe S:

$$v = s + \underbrace{v - s}_{u} = pr_S(v) + u, \quad pr_S(v) \in S, u = v - s \perp S$$

7.5 Soluţia celor mai mici pătrate pentru un sistem $Ax = b$

Deşi prin definiţie un sistem liniar $Ax = b$, $A \in \mathbb{R}^{m \times n}$, incompatibil, NU are soluţie, în învăţărea supervizată, ca metodă de învăţăre în *machine learning* se foloseşte un vector x^* care într-un anume sens este ”cel mai aproape” de a fi soluţie a sistemului.

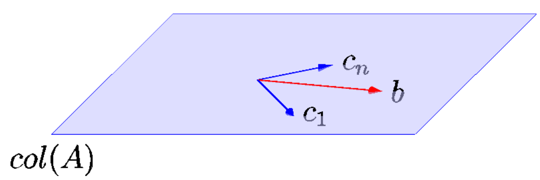
Fie $A \in \mathbb{R}^{m \times n}$ o matrice de m linii şi n coloane, având coloanele notate $c_1, c_2, \ldots, c_n$, $A = [c_1|c_2| \ldots |c_n]$, iar $b \in \mathbb{R}^m$. Considerăm sistemul liniar $Ax = b$ sau detaliat:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Propoziţia 7.5.1 *Sistemul $Ax = b$ este compatibil dacă şi numai dacă b este un vector din subspaţiul generat de coloanele matricii A, adică b se poate exprima ca o combinaţie liniară a vectorilor coloană, $c_1, c_2, \dots, c_n$.*

Demonstraţie: Sistemul $Ax = b$ este compatibil dacă şi numai dacă există $x = (x_1, x_2, \dots, x_n)^T$ astfel încât $Ax = b$, adică:

$$b = Ax = [c_1|c_2|\dots|c_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1c_1 + x_2c_2 + \dots + x_nc_n$$



□

Consecintă. Sistemul $Ax = b$, unde $A \in \mathbb{R}^{m\times n}$, este incompatibil (nu are soluţie) dacă şi numai dacă vectorul termenilor liberi, $b \in \mathbb{R}^m$, nu aparţine subspaţiului vectorial generat de coloanele matricii A:

$$b \notin \text{col}(A)$$

.

În continuare considerăm sisteme incompatibile, supradeterminate, adică sisteme ce conţin mai multe ecuaţii decât necunoscute sau echivalent, matricea sistemului $A \in \mathbb{R}^{m\times n}$ are mai multe linii decât coloane, $m > n$.

În acest caz, subspaţiul coloanelor, $col(A)$, poate avea cel mult dimensiunea $n < m$, şi deci acesta este un subspaţiu liniar al lui $\mathbb{R}^m$. Pentru orice vector $x \in \mathbb{R}^n$, notăm $r(x) = b - Ax$. $r(x)$ se numeşte abaterea lui Ax de la b sau reziduul sistemului.

Pentru un sistem compatibil, reziduul asociat unei soluţii este este vectorul nul, pentru că dacă $x \in \mathbb{R}^n$ este o soluţie, atunci $Ax = b$ şi deci $r(x) = b - Ax = \theta$. În cazul sistemelor supradeterminate şi incompatibile se caută un vector x^* pentru care Ax^* este cel mai apropiat de b sau echivalent reziduul $r(x^*)$ este cel mai apropiat de 0.

Definiţia 7.5.1 *Vectorul $x^* \in \mathbb{R}^n$ care minimizează distanţa de la b la Ax, adică minimizează norma $||r(x)|| = ||b - Ax||$:*

$$||b - Ax^*|| \leq ||b - Ax||, \quad \forall\, x \in \mathbb{R}^n$$

sau echivalent:

$$||b - Ax^*||^2 \leq ||b - Ax||^2, \quad \forall\, x \in \mathbb{R}^n$$

se numeşte soluţie în sensul celor mai mici pătrate pentru sistemul incompatibil $Ax = b$ (în limba engleză este **least squares solution***).*

Daţi search *least square solution in machine learning* şi vedeţi ce număr mare de rezultate este afisat!!!!

Soluţia se numeste a celor mai mici pătrate pentru că norma la pătrat a vectorului reziduu, $r(x) = b - Ax$, se exprimă ca o sumă de patrate. Într-adevăr, :

$$r(x) = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} - \begin{bmatrix} a_{11}x_1 + a12x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a22x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + am2x_2 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} b_1 - (a_{11}x_1 + a12x_2 + \cdots + a_{1n}x_n) \\ b_2 - (a_{21}x_1 + a22x_2 + \cdots + a_{2n}x_n) \\ \vdots \\ b_m - (a_{m1}x_1 + am2x_2 + \cdots + a_{mn}x_n) \end{bmatrix}$$

şi deci

$$\begin{aligned} ||r(x)||^2 &= \\ &= (b_1 - (a_{11}x_1 + a12x_2 + \cdots + a_{1n}x_n))^2 + (b_2 - (a_{21}x_1 + a22x_2 + \cdots + a_{2n}x_n))^2 + \cdots \\ &+ (b_m - (a_{m1}x_1 + am2x_2 + \cdots + a_{mn}x_n))^2 \end{aligned} \tag{7.1}$$

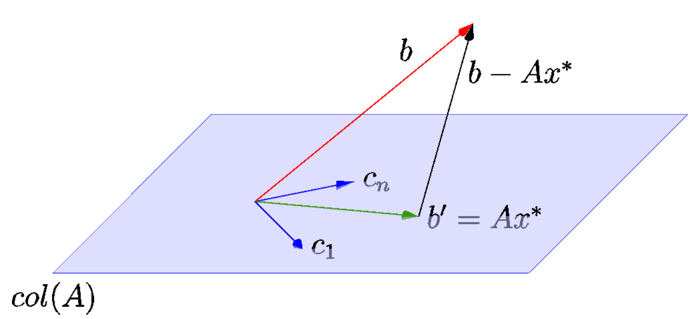
Prin urmare soluţia celor mai mici pătrate este vectorul x ale cărui coordonate minimizează suma de pătrate din (7.1).

Să deducem o metodă de a afla soluţia celor mai mici pătrate pentru un sistem incompatibil, $Ax = b$, cu A de tip $m \times n$, $m > n$.

Sistemul fiind incompatibil, b nu aparţine subspaţiului generat de coloanele matricii A, adică b ”este exterior acestui subspaţiu”.

Considerăm proiecţia sa ortogonală pe subspaţiul $S = \text{col}(A)$, $b' = pr_{col(A)}b$. Deoarece b' este din col(A), rezultă că există scalarii $x_1^*, x_2^*, \ldots, x_n^*$ astfel încât b' să se exprime ca o combinaţie liniară a coloanelor matricii A:

$$b' = x_1^* c_1 + x_2^* c_2 + \cdots + x_n^* c_n = [c_1|c_2|\ldots|c_n] \begin{bmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{bmatrix} = Ax^*$$



Orice alt vector $t \in col(A)$ este de forma $t = x_1c_1 + x_2c_2 + \cdots x_nc_n = Ax$, unde $x = (x_1, x_2, \ldots, x_n)^T$, adică orice alt vector din $col(A)$ se exprimă ca o combinaţie liniară a coloanelor. Conform Propoziţiei 7.4.1 avem că $||b - \underbrace{Ax^*}_{b'}|| \leq ||b - \underbrace{Ax}_{t}||$, $\forall\, x \in \mathbb{R}^n$ şi deci x^* este soluţia celor mai mici pătrate pentru sistemul incompatibil $Ax = b$.

Cum aflăm efectiv soluţia x^*?

Pentru aceasta dăm încă o modalitate de a exprima produsul Dy al unei matrici $D \in \mathbb{R}^{m\times n}$, cu un vector $y = (y_1, y_2, \ldots, y_n)^T \in \mathbb{R}^n$. Notând liniile din D cu $\ell_1, \ell_2, \ldots, \ell_m$, avem :

$$Dy = \begin{bmatrix} d_{11} & d_{12} & \ldots & d_{1n} \\ d_{21} & d_{22} & \ldots & d_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ d_{m1} & d_{m2} & \ldots & d_{mn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} <\ell_1, y> \\ <\ell_2, y> \\ \vdots \\ <\ell_m, y> \end{bmatrix} \tag{7.2}$$

Cu alte cuvinte dacă notăm produsul $Dy = w$, atunci coordonata i a lui w este:

$$w_i = d_{i1}y_1 + d_{i2}y_2 + \cdots + d_{in}y_n = <\ell_i, y>, \quad i = \overline{1, m}$$

Din definiţia proiecţiei ortogonale a unui vector pe un subspaţiu rezultă că $b - b' = b - Ax^*$ este ortogonal pe orice vector din subspaţiul $S = col(A)$, deci în particular şi pe fiecare vector coloană $c_1, c_2, \ldots, c_n$:

$$<b - Ax^*, c_i>= 0, \forall\, i = 1, 2, \ldots, n$$

Notăm $y = b - Ax^* \in \mathbb{R}^m$. Atunci $<c_i, b - Ax^*>= 0$ înseamnă că $<c_i, y>= 0$, $\forall\, i = \overline{1, n}$.

Din $<c_i, y>= 0$, $\forall\, i = \overline{1, n}$, rezultă că:

$$\begin{bmatrix} <c_1, y> \\ <c_2, y> \\ \vdots \\ <c_n, y> \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Cum $c_1, c_2, \ldots, c_n$ sunt coloane în A, ele vor fi linii în A^T şi deci conform (7.2), avem:

$$\begin{bmatrix} <c_1, y> \\ <c_2, y> \\ \vdots \\ <c_n, y> \end{bmatrix} = A^T y = 0$$

Din relaţia $A^T y = 0$ rezultă revenind la $y = b - Ax^*$ că $A^T(b - Ax^*) = 0$ sau echivalent:

$$A^T Ax^* = A^T b$$

Cu alte cuvinte, soluţia celor mai mici pătrate pentru sistemul incompatibil supradeterminat $Ax = b$, dacă există, este soluţia sistemului

$$A^T Ax^* = A^T b$$

Sistemul liniar $A^T Ax = A^T b$ se numeşte sistemul normal asociat celui iniţial, $Ax = b$. Remarcăm că matricea $A^T A \in \mathbb{R}^{n\times n}$ a sistemului normal este o matrice pătratică şi simetrică deoarece $(A^T A)^T = A^T A$. Deci acest sistem este compatibil determinat, dacă şi numai dacă rangul matricii $A^T A$ este egal cu n.

Propoziţia 7.5.2 *Rangul matricii A^TA coincide cu rangul matricii $A \in \mathbb{R}^{m\times n}$.*

Demonstraţie: Arătăm mai întâi că $Null(A) = Null(A^TA)$. Efectuăm demonstraţia prin dublă incluziune. Dacă $v \in Null(A)$, atunci $Av = \theta$ şi deci şi $A^T(Av) = A^T(\theta) = \theta$, adică $v \in Null(A^TA$ şi deci $Null(A) \subseteq Null(A^TA)$. Reciproc, fie $v \in Null(A^TA)$, adică $A^TAv = \theta$. Înmulţind această ultimă relaţie la stânga cu v^T obţinem:

$$(v^TA^T)(Av) = 0 \;\Leftrightarrow (Av)^T(Av) = 0 \;\; < Av, Av >= 0 \;\Rightarrow Av = \theta \;\Rightarrow v \in Null(A)$$

Prin urmare am arătat şi că $Null(A^TA) \subseteq Null(A)$ şi deci $Null(A^TA) = Null(A)$. Ambele matrici A^TA şi A au n coloane. Dimensinunea$(Null(A)) = n - r$ şi deci $rang(A) = n - dim(Null(A)) = n - dim(Null(A^TA)) = rang(A^TA)$. □

Dacă rangul matricii $A \in \mathbb{R}^{m\times n}$ a sistemului incomaptibil, supradeterminat $Ax = b$ este egal cu n (numărul coloanelor sale), atunci şi rangul lui A^TA este n şi deci sistemul normal $A^TAx = A^Tb$ este compatibil determinat şi unica sa soluţie x^*, este soluţia celor mai mici pătrate a sistemului $Ax = b$.

7.5.1 Aplicarea soluţiei celor mai mici pătrate la aflarea dreptei celor mai mici pătrate

În analiza datelor ce constau din perechi de puncte (x_i, y_i), $i = \overline{1,m}$ suntem uneori interesaţi dacă nu cumva "norul" de puncte este distribuit în jurul unei drepte. Geometric se ştie că o dreaptă în plan este perfect determinată de 2 puncte. Având un set de m puncte, necoliniare, evident că nu există o dreaptă care să le conţină pe toate. Pentru a determina ceea ce se numeşte în machine learning, dreapta celor mai mici pătrate se impune ca datele (x_i, y_i) să verifice ecuaţia unei drepte $y = \alpha x + \beta$. Astfel obţinem m ecuaţii cu 2 necunoscute α şi β:

$$\begin{array}{rcl} \alpha x_1 + \beta & = & y_1 \\ \alpha x_2 + \beta & = & y_2 \\ \vdots & = & \vdots \\ \alpha x_m + \beta & = & y_m \end{array}$$

adică un sistem supradeterminat şi incompatibil. Matricial sistemul se scrie:

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

Pentru date foarte generale rangul matricii sistemului este 2, iar al matricii prelungite 3, deci sistemul este incompatibil (deduceţi în ce caz rangul lui $\overline{A}$ ar putea fi tot 2!!!).

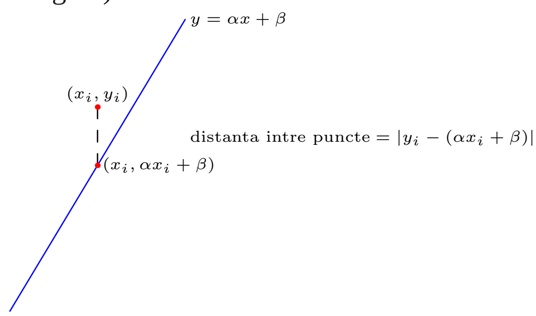
Asociind sistemul normal $A^TAx^* = A^Tb$ se obţine soluţia unică $x^* = (\alpha^*, \beta^*)^T$ şi dreapta $y = \alpha^*x + \beta^*$ se numeşte dreapta celor mai mici pătrate, pentru că reziduul $r(x)$, $x = (\alpha, \beta)^T$, asociat sistemului $Ax = b$ este în acest caz:

$$r(x) = \begin{bmatrix} y_1 - (\alpha x_1 + \beta) \\ y_2 - (\alpha x_2 + \beta) \\ \vdots \\ y_m - (\alpha x_m + \beta) \end{bmatrix}$$

iar

$$||r(x)||^2 = \sum_{i=1}^{m} (y_i - (\alpha x_i + \beta))^2$$

Deci problema aflării dreptei $y = \alpha^* x + \beta^*$ constă în a afla dintre toate dreptele din plan pe cea ai cărei parametri α, β, minimizează suma pătratelor din expresia lui $||r(x)||^2$. Dar $(y_i - (\alpha x_i + \beta))^2$ nu este altceva decât distanta la pătrat ”măsurată pe verticală” dintre punctul (x_i, y_i) şi dreapta $y = \alpha x + b$ (vezi figura).



Exemplul 7. Numărul de vizitatori unici ai site-ului `http://www.exemplu.com` (e doar un exemplu, nu daţi click pe link!), în patru zile consecutive, este respectiv, 34, 21, 30, 18. Pentru a face predicţii privind tendinţele surferilor în perioada următotoare să determinăm dreapta celor mai mici pătrate asociate perechilor de puncte:

$$(1, 34), (2, 21), (3, 30), (4, 18)$$

unde $1, 2, 3, 4$ reprezintă ziua monitorizării site-ului.

Impunem ca cele 4 patru puncte să verifice ecuaţia unei drepte $y = \alpha x + \beta$ şi obţinem:

$$\begin{array}{lcl} \alpha + \beta & = & 34 \\ 2\alpha + \beta & = & 21 \\ 3\alpha + be & = & 30 \\ 4\alpha + \beta & = & 18 \end{array}$$

Notând cu A matricea sistemului, calculăm $A^T A$:

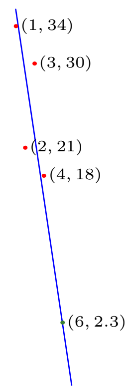
$$A^T A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 30 & 10 \\ 10 & 4 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 34 \\ 21 \\ 30 \\ 18 \end{bmatrix} = \begin{bmatrix} 224 \\ 103 \end{bmatrix}$$

Sistemul normal este $A^T Ax^* = A^T b$, adică în cazul nostru:

$$\begin{bmatrix} 30 & 10 \\ 10 & 4 \end{bmatrix} \begin{bmatrix} \alpha^* \\ \beta^* \end{bmatrix} = \begin{bmatrix} 224 \\ 103 \end{bmatrix}$$

Rezolvand obţinem soluţia $\alpha^* = -6.7$, $\beta^* = 42.5$ Drepta celor mai mici pătrate, adică dreapta cea mai apropiată de toate punctele (în sensul precizat) este $y = -6.7x + 42.5$. Având panta negativă ar trebui să ne întristăm pentru că pe măsură ce zilele trec, site-ul e vizitat din ce în ce mai rar. De exemplu pentru ziua a a şasea numărul vizitelor este predictionat la $y = -6.7 * 6 + 42.5 = 2.3$, adică în jur de 2 :(Prin urmare predicţiile pentru o zi i se obţin calculând ordonata punctului de pe dreapta $y = \alpha x + \beta$, corespunzătoare lui $x = i$.



(În semestrul 2 învăţăm cum se fac predicţii cu acurateţe mai mare şi nivel de încredere ridicat).