

Cursul 6

Baze ortonormate, matrici ortogonale

6.1 Aplicații ale algebrei spațiilor euclidiene în Information Retrieval

Motoarele de căutare au la bază un sistem de reprezentare numerică a informației textuale dintr-un corpus (bază de documente în format electronic), o modalitate de căutare a documentelor relevante pentru cererea unui utilizator și de extragere a informației. Sistemele de extragere a informației relevante pentru necesitățile de informare ale utilizatorilor se numesc sisteme IR (Information Retrieval).

Google, Bing, Yahoo au câte un sistem IR. Bibliotecile, marile companii, portalurile diverselor ziare/reviste pun la dispoziție o aplicație de căutare.

Un document poate fi o pagina WEB, un ebook (carte în format electronic), un articol de pe site-ul `http://www.gandul.info`, un raport tehnic al unei companii salvat în format pdf, etc. Pentru a reduce complexitatea unui document și a facilita căutarea rapidă, acestuia i se asociază o reprezentare numerică. Noi discutăm modelul vectorial al unui document.

Un document se descompune într-o listă de termeni, rezultați prin filtrare. Adică nu se constituie reuniunea tuturor termenilor din document, ci se face în prealabil o filtrare, eliminând conjuncții, prepoziții, semne de punctuație. Ceea ce rezultă este un sac de cuvinte (în IR se folosește expresia *bag of words*). Termenii eliminați se numesc *stop words*.

Dacă corpus-ul (baza de texte, documente) conține n documente și m este numărul total de termeni rezultați din filtrarea termenilor celor n documente, atunci fiecare document j din cele $\{1, 2, \dots, n\}$ documente i se asociază un vector

$$d_j = \begin{bmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{mj} \end{bmatrix} \in \mathbb{R}^m,$$

unde p_{ij} reprezintă ponderea termenului i în documentul j . Această pondere cuantifică importanța termenului i pentru documentul j .

Performanța sistemului IR depinde de modul de ponderare a termenilor într-un document. Majoritatea motoarelor de căutare precum și sistemele IR ale marilor companii sau ale bibliotecilor digitale, folosesc ponderea TF-IDF (TF=Term Frequency, IDF=Inverse Document Frequency).

Ponderea TF-IDF conține două componente, una care indică cât de frecvent apare termenul în acel document și alta care indică frecvența termenului în toate documentele.

Dacă un cuvânt apare frecvent într-un document, atunci el este important pentru acel document și deci ar trebui să-i atribuim o pondere mare. Dacă însă acel cuvânt este prezent în foarte multe documente, atunci nu este așa relevant și ar trebui să-i scadă din pondere. De exemplu, într-o bibliotecă digitală destinată matematicii în foarte multe documente apare cuvântul *demonstrație*. Astfel acest termen nu ar trebui să aibă influență mare în alegerea documentelor relevante unei cereri și trebuie să-i scădem ponderea.

Cele două componente ale ponderii sunt cuantificate astfel:

- Frecvența termenului i în documentul j (adică numărul ce indică de câte ori apare termenul i în documentul j) o notăm cu f_{ij} .
- Notând cu n_i numărul de documente ce conțin termenul i , atunci raportul $\frac{n_i}{n}$ reprezintă frecvența cu care termenul i apare în cele n documente din corpus. Dacă apare în multe documente, atunci acest raport este apropiat de 1 și deci definirea ponderii ca fiind, de exemplu $p_{ij} = f_{ij}n_i/n$, nu scade suficient ponderea termenului i . De aceea s-a propus calcularea inversului acestui raport și logaritmul inversului, $\log(n/n_i)$, se numește *inverse document frequency*, adică inversul frecvenței termenului i în documentele corpus-ului.

Ponderea termenului i în documentul j se definește atunci ca fiind:

$$p_{ij} = f_{ij} \log(n/n_i)$$

Să ilustrăm acum că factorul $\log(n/n_i)$ micsorează ponderea termenului i când el apare în multe documente. Presupunem de exemplu că termenii 1=*liniar* și 2=*ortonormat* apar în documentul j (ebook de Algebra) de același număr de ori, egal cu 5, dar *liniar* apare în 15 documente din totalul de 20, iar *ortonormat* în 8 documente din cele 20 ale corpus-ului. Atunci ponderea TF-IDF a termenului 1 este $f_{1j} \log(20/15) = 5 \cdot 0.287 = 1.43$ iar a termenului 2 este $f_{2j} = 5 \log(20/8) = 5 \cdot 0.91 = 4.58$. Prin urmare termenul *ortonormat* are pondere mai mare decât *liniar*.

În general dacă doi termeni i, k apar în n_i , respectiv n_k documente și $n_i < n_k$, atunci $n/n_i > n/n_k$ și $\log(n/n_i) > \log(n/n_k)$, adică coeficientul IDF al termenului k care apare în mai multe documente este mai mic decât al celui ce apare în mai puține documente.

Matricea de m linii și n coloane, ce are drept coloane vectorii reprezentativi ai documentelor din corpus, $D = [d_1 | d_2 | \dots | d_n]$, se numește în IR, *term by document matrix*:

$$D = \begin{array}{l} \text{termen 1} \rightarrow \\ \text{termen 2} \rightarrow \\ \vdots \\ \text{termen } m \rightarrow \end{array} \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \vdots & \vdots & \dots & \vdots \\ p_{m1} & p_{m2} & \dots & p_{mn} \end{bmatrix}$$

Sistemul de căutare asociază cuvintelor cheie, un vector cerere (*query vector*)

$q = (q_1, q_2, \dots, q_m)^T \in \mathbb{R}^m$: $q_i = 1$ dacă stringul de căutare conține termenul i din cei m termeni indexați și $q_i = 0$ dacă stringul nu conține termenul i , $i = \overline{1, m}$.

Pentru a selecta documentele din corpus ce conțin termenii relevanți pentru cerere se calculează similaritatea dintre vectorul cerere și fiecare vector document, d_j , $j = \overline{1, n}$,

folosind similaritatea ”măsurată” de cosinusul unghiului dintre un document și vectorul cerere (în IR se numește *cosine similarity*).

Atât vectorii document, cât și vectorii cerere au coordonate mai mari sau egale ca zero și prin urmare produsul scalar $\langle d_j, q \rangle \geq 0$. Astfel $\cos(\widehat{q, d_j}) = \frac{\langle q, d_j \rangle}{\|d_j\|, \|q\|} \in [0, 1]$. Dacă $\cos(\widehat{q, d_j})$ este apropiat de 1, atunci spunem că între cerere și document există similaritate mare. Dacă însă $\cos(\widehat{q, d_j})$ este mai apropiat de 0, atunci cei doi vectori au o similaritate redusă și deci documentul nu este deloc relevant cererii.

Fiecare sistem IR are fixat un prag al valorii de similaritate, de exemplu sunt returnate userului informațiile din documentele d_j pentru care similaritatea cu vectorul cerere este mai mare decât 0.7,

Dăm acum un exemplu artificial construit dintr-un articol de sport de pe WEB:

Exemplul 1. Presupunem că pe un site sunt 10 documente. În aceste 10 documente avem 12 termeni. În tabelul de mai jos indicăm numărul de apariții al fiecărui termen în documentul D_1, D_2, \dots, D_{10} :

	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	n_i	$\log(10/n_i)$
soccer	0	0	2	0	0	0	0	1	0	0	2	1.6094
Mueller	1	0	1	0	1	0	0	0	0	0	3	1.2040
broken	2	0	1	0	0	0	0	0	1	0	3	1.2040
prediction	1	0	0	0	2	0	1	0	0	0	3	1.2040
football	0	3	0	7	4	0	2	1	0	0	5	0.6931
win	0	0	0	1	0	0	0	0	0	1	2	1.6094
leg	0	0	1	0	0	0	0	0	0	0	1	2.3026
Podolsky	0	5	0	1	0	0	0	0	0	0	2	1.6094
world	3	1	0	0	1	0	1	2	0	0	5	0.6931
cup	4	1	0	0	1	0	0	2	0	0	4	0.9163
victory	0	0	0	0	0	1	0	0	0	1	2	1.6094
match	1	0	0	2	0	0	1	0	1	0	4	0.9163

$n_i, i = \overline{1, 12}$ dă numărul de documente în care apare termenul i . Practic elementele din tabelul de mai sus, exceptând ultimele două coloane, reprezintă f_{ij} , adică frecvența de apariție a termenului i în documentul D_j .

Matricea term by document, $D = (p_{ij}), p_{ij} = f_{ij} \log(10/n_i)$ este:

$$D = \begin{bmatrix} 0 & 0 & 3.2189 & 0 & 0 & 0 & 0 & 1.6094 & 0 & 0 \\ 1.2040 & 0 & 1.2040 & 0 & 1.2040 & 0 & 0 & 0 & 0 & 0 \\ 2.4079 & 0 & 1.2040 & 0 & 0 & 0 & 0 & 0 & 1.2040 & 0 \\ 1.2040 & 0 & 0 & 0 & 2.4079 & 0 & 1.2040 & 0 & 0 & 0 \\ 0 & 2.0794 & 0 & 4.8520 & 2.7726 & 0 & 1.3863 & 0.6931 & 0 & 0 \\ 0 & 0 & 0 & 1.6094 & 0 & 0 & 0 & 0 & 0 & 1.6094 \\ 0 & 0 & 2.3026 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 8.0472 & 0 & 1.6094 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2.0794 & 0.6931 & 0 & 0 & 0.6931 & 0 & 0.6931 & 1.3863 & 0 & 0 \\ 3.6652 & 0.9163 & 0 & 0 & 0.9163 & 0 & 0 & 1.8326 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.6094 & 0 & 0 & 0 & 1.6094 \\ 0.9163 & 0 & 0 & 1.8326 & 0 & 0 & 0.9163 & 0 & 0.9163 & 0 \end{bmatrix}$$

Dacă dăm un search cu cuvintele cheie *Mueller, broken, leg*, să vedem ce documente returnează acest mini-sistem de căutare.

Vectorul cerere este $q = (0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0)^T$.

Calculând similaritățile $\text{sim}_j = \cos(d_j, q)$, $j = \overline{1, 10}$ (unde d_j este coloana j din matricea D , obținem:

$$\text{sim} = \begin{bmatrix} 0.3992 & 0 & 0.6312 & 0 & 0.1724 & 0 & 0 & 0 & 0.4594 & 0 \end{bmatrix}$$

Prin urmare documentele ordonate în ordinea descrescătoare a similarității cu cererea sunt:

$$D_3, D_9, D_1, D_5$$

Restul sunt "perpendiculare" pe cerere, deci irelevante.

6.2 Baze ortonormate într-un spațiu vectorial cu produs scalar

Fie (V, \langle, \rangle) un spațiu cu produs scalar. Doi vectori nenuli, v, w , cu proprietatea că $\langle v, w \rangle = 0$ se numesc vectori ortogonali.

Produsul scalar dintre un vector arbitrar și vectorul nul este egal cu 0, $\langle v, \theta \rangle = 0$, pentru că $\theta = v - v$ și deci $\langle v, \theta \rangle = \langle v, v - v \rangle = \langle v, v \rangle - \langle v, v \rangle = 0$. În acest caz NU spunem că v este ortogonal pe vectorul nul. Ortogonalitatea se definește doar pentru vectori nenuli.

Considerăm un sistem finit de vectori nenuli, $\mathcal{S} = \{v_1, v_2, \dots, v_m\}$.

Definiția 6.2.1 Sistemul \mathcal{S} cu proprietatea că oricare doi vectori distincți sunt ortogonali, adică $\langle v_i, v_j \rangle = 0$, $\forall i, j \in \{1, 2, \dots, m\}, i \neq j$, se numește sistem ortogonal de vectori.

Propoziția 6.2.1 Un sistem ortogonal de vectori este un sistem liniar independent.

Demonstrație: Fie $\mathcal{S} = \{v_1, v_2, \dots, v_m\}$ un sistem ortogonal de vectori, adică produsele scalare a doi câte doi vectori distincți sunt zero:

$$\langle v_i, v_j \rangle = 0, \forall i, j \in \{1, 2, \dots, m\}, i \neq j \quad (6.1)$$

Presupunem că o combinație liniară a acestor vectori este egală cu vectorul nul:

$$\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_i v_i + \dots + \alpha_m v_m = \theta, \quad \alpha_1, \alpha_2, \dots, \alpha_m \in \mathbb{R} \quad (6.2)$$

Înmulțim scalar ambii membri ai relației (6.2) cu vectorul v_i :

$$\langle \alpha_1 v_1 + \dots + \alpha_i v_i + \dots + \alpha_m v_m, v_i \rangle = \langle \theta, v_i \rangle \quad (6.3)$$

Aplicând proprietățile produsului scalar și ortogonalitatea a doi vectori distincți, avem:

$$\alpha_1 \underbrace{\langle v_1, v_i \rangle}_{=0} + \dots + \alpha_i \underbrace{\langle v_i, v_i \rangle}_{\neq 0} + \dots + \alpha_m \underbrace{\langle v_m, v_i \rangle}_{=0} = \langle \theta, v_i \rangle = 0 \quad (6.4)$$

Rezultă deci că

$$\alpha_i \langle v_i, v_i \rangle = 0 \quad (6.5)$$

Dar produsul a două numere reale este zero doar dacă cel puțin unul din factori este zero. Cum $\langle v_i, v_i \rangle \neq 0$ (pentru că $v_i \neq \theta$), rezultă că $\alpha_i = 0$, $i = \overline{1, m}$. Prin urmare vectorii ortogonali sunt liniar independenți. \square

Această particularitate a sistemelor de vectori ortogonali ne asigură că într-un spațiu vectorial euclidian de dimensiune n un sistem ortogonal de n vectori formează o bază.

Definiția 6.2.2 O bază dintr-un spațiu vectorial euclidian formată din vectori ortogonali doi câte doi se numește **bază ortogonală**. O bază $\mathcal{B} = (e_1, e_2, \dots, e_n)$ din (V_n, \langle, \rangle) cu proprietatea că oricare doi vectori distincți sunt ortogonali și norma fiecărui vector este egală cu 1, se numește **bază ortonormată**.

Condițiile ca baza să fie ortonormată se exprimă astfel:

$$\begin{aligned} \langle e_i, e_j \rangle &= 0, \forall i \neq j \quad \text{și} \quad \|e_i\| = 1, \forall i = \overline{1, n} \quad \text{sau} \\ \langle e_i, e_j \rangle &= \delta_{ij} = \begin{cases} 0 & \text{dacă } i \neq j \\ 1 & \text{dacă } i = j \end{cases} \end{aligned} \quad (6.6)$$

Relația $\langle e_i, e_i \rangle = \delta_{ii} = 1$ are loc deoarece $\|e_i\| = 1$, adică $\sqrt{\langle e_i, e_i \rangle} = 1$ sau ridicând la pătrat $\langle e_i, e_i \rangle = 1$.

Diferența dintre o bază ortogonală (u_1, u_2, u_3) și una ortonormată (e_1, e_2, e_3) , din \mathbb{R}^3 este ilustrată în Fig.6.1. Prima bază are vectori de norme (lungimi) diferite, iar a doua are vectorii de aceeași normă (adică 1).

Exemplul 2. Baza canonică din spațiul vectorial \mathbb{R}^n înzestrat cu produsul scalar standard:

$$\langle v, w \rangle = x_1 y_1 + x_2 y_2 + \dots + x_n y_n, \forall v = (x_1, x_2, \dots, x_n)^T, w = (y_1, y_2, \dots, y_n)^T \in \mathbb{R}^n,$$

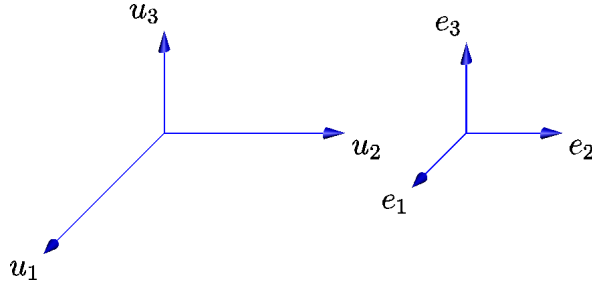


Fig.6.1: Bază ortogonală în stânga și ortonormată în dreapta.

este o bază este o bază ortonormată.

Să verificăm în cazul $n = 3$: $\mathcal{B} = (e_1 = (1, 0, 0)^T, e_2 = (0, 1, 0)^T, e_3 = (0, 0, 1)^T)$.

$$\begin{aligned} \langle e_1, e_2 \rangle &= 1 \cdot 0 + 0 \cdot 1 + 0 \cdot 0 = 0 \\ \langle e_1, e_3 \rangle &= 0 \\ \langle e_2, e_3 \rangle &= 0 \end{aligned}$$

Deci baza canonică este ortogonală. Să arătăm că este și normată, adică fiecare vector are norma 1 (este versor):

$$\|e_1\| = \sqrt{1^2 + 0^2 + 0^2} = 1, \quad \|e_2\| = 1, \quad \|e_3\| = 1.$$

În continuare vom arăta că spre deosebire de bazele arbitrare, coordonatele unui vector într-o bază ortonormată se calculează foarte simplu.

Propoziția 6.2.2 Dacă $\mathcal{B} = (e_1, e_2, \dots, e_n)$ este o bază ortonormată în spațiul V_n și v un vector arbitrar din V_n , atunci coordonatele vectorului v în baza \mathcal{B} sunt

$$x_i = \langle v, e_i \rangle, \quad i = \overline{1, n}$$

Demonstrație: Baza \mathcal{B} fiind ortonormată avem

$$\langle e_i, e_j \rangle = \begin{cases} 0 & \text{dacă } i \neq j \\ 1 & \text{dacă } i = j \end{cases} \quad (6.7)$$

Fie

$$v = x_1 e_1 + x_2 e_2 + \dots + x_i e_i + \dots + x_n e_n \quad (6.8)$$

exprimarea vectorului v în baza ortonormată \mathcal{B} . Înmulțind scalar fiecare membru al relației (6.8) cu vectorul e_i obținem:

$$\langle v, e_i \rangle = x_1 \underbrace{\langle e_1, e_i \rangle}_{=0} + x_2 \underbrace{\langle e_2, e_i \rangle}_{=0} + \dots + x_i \underbrace{\langle e_i, e_i \rangle}_{=1} + \dots + x_n \underbrace{\langle e_n, e_i \rangle}_{=0}, \quad (6.9)$$

adică $\langle v, e_i \rangle = x_i, i = \overline{1, n}$.

□

Observația 6.2.1 Un vector v are următoarea exprimare în baza ortonormată \mathcal{B} :

$$v = \langle v, e_1 \rangle e_1 + \langle v, e_2 \rangle e_2 + \cdots + \langle v, e_n \rangle e_n. \quad (6.10)$$

Această exprimare se mai numește *exprimarea Fourier a vectorului v în baza \mathcal{B}* .

Exemplul 3. În spațiul vectorial euclidian $(\mathbb{R}^3, \langle, \rangle)$ se dă baza

$$\mathcal{B} = \left(u_1 = \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}} \right)^T, u_2 = \left(\frac{1}{\sqrt{18}}, \frac{4}{\sqrt{18}}, \frac{1}{\sqrt{18}} \right)^T, u_3 = \left(\frac{2}{3}, -\frac{1}{3}, \frac{2}{3} \right)^T \right)$$

Să se arate că \mathcal{B} este o bază ortonormată și să se determine coordonatele vectorului $v = (-3, 1, 2)^T$ relativ la baza \mathcal{B} .

Verificăm ortogonalitatea:

$$\langle u_1, u_2 \rangle = u_1^T u_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{18}} \\ \frac{4}{\sqrt{18}} \\ \frac{1}{\sqrt{18}} \end{bmatrix} = \frac{1}{\sqrt{36}} - \frac{1}{\sqrt{36}} = 0$$

$$\langle u_1, u_3 \rangle = u_1^T u_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \\ \frac{2}{3} \end{bmatrix} = \frac{2}{3\sqrt{2}} - \frac{2}{3\sqrt{2}} = 0$$

$$\langle u_2, u_3 \rangle = u_2^T u_3 = \begin{bmatrix} \frac{1}{\sqrt{18}} & \frac{4}{\sqrt{18}} & \frac{1}{\sqrt{18}} \end{bmatrix} \cdot \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \\ \frac{2}{3} \end{bmatrix} = \frac{2}{3\sqrt{18}} - \frac{4}{3\sqrt{18}} + \frac{2}{3\sqrt{18}} = 0.$$

Să calculăm norma fiecărui vector din bază:

$$\|u_1\| = \sqrt{1/2 + 1/2} = 1, \quad \|u_2\| = \sqrt{1/18 + 16/18 + 1/18} = 1, \quad \|u_3\| = \sqrt{4/9 + 1/9 + 4/9} = 1$$

Coordonatele vectorului v relativ la baza \mathcal{B} sunt

$$\begin{aligned} x_1 &= \langle v, u_1 \rangle = (-3, 1, 2)^T \cdot \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}} \right)^T = -\frac{3}{\sqrt{2}} - \frac{2}{\sqrt{2}} = -\frac{5}{\sqrt{2}} \\ x_2 &= \langle v, u_2 \rangle = (-3, 1, 2)^T \cdot \left(\frac{1}{\sqrt{18}}, \frac{4}{\sqrt{18}}, \frac{1}{\sqrt{18}} \right)^T = -\frac{3}{\sqrt{18}} + \frac{4}{\sqrt{18}} + \frac{2}{\sqrt{18}} = \frac{3}{\sqrt{18}} = \frac{1}{\sqrt{2}} \\ x_3 &= \langle v, u_3 \rangle = (-3, 1, 2)^T \cdot \left(\frac{2}{3}, -\frac{1}{3}, \frac{2}{3} \right)^T = -\frac{6}{3} - \frac{1}{3} + \frac{4}{3} = -1. \end{aligned}$$

Deci v se exprimă în baza \mathcal{B} astfel $v = -\frac{5}{\sqrt{2}}u_1 + \frac{1}{\sqrt{2}}u_2 - u_3$.

Pentru a aprecia următoarea proprietate a bazelor ortonormate, observăm că dacă $\mathcal{B} = (f_1, f_2, \dots, f_n)$ este o bază nu neapărat ortonormată și $v = x_1f_1 + x_2f_2 + \dots + x_nf_n$, $w = y_1f_1 + y_2f_2 + \dots + y_nf_n$ sunt doi vectori exprimați în baza \mathcal{B} atunci, produsul lor scalar este:

$$\begin{aligned}
 \langle v, w \rangle &= \langle x_1f_1 + x_2f_2 + \dots + x_nf_n, y_1f_1 + y_2f_2 + \dots + y_nf_n \rangle = \\
 &= x_1y_1 \langle f_1, f_1 \rangle + x_1y_2 \langle f_1, f_2 \rangle + \dots + x_1y_n \langle f_1, f_n \rangle + \\
 &\quad x_2y_1 \langle f_2, f_1 \rangle + x_2y_2 \langle f_2, f_2 \rangle + \dots + x_2y_n \langle f_2, f_n \rangle + \\
 &\quad \vdots \\
 &\quad + x_ny_1 \langle f_n, f_1 \rangle + x_ny_2 \langle f_n, f_2 \rangle + \dots + x_ny_n \langle f_n, f_n \rangle = \\
 &= \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \underbrace{\begin{bmatrix} \langle f_1, f_1 \rangle & \langle f_1, f_2 \rangle & \dots & \langle f_1, f_n \rangle \\ \langle f_2, f_1 \rangle & \langle f_2, f_2 \rangle & \dots & \langle f_2, f_n \rangle \\ \vdots & \vdots & \dots & \vdots \\ \langle f_n, f_1 \rangle & \langle f_n, f_2 \rangle & \dots & \langle f_n, f_n \rangle \end{bmatrix}}_G \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}
 \end{aligned}$$

Precizăm că matricea G de elemente $g_{ij} = \langle f_i, f_j \rangle$ este o matrice simetrică, deoarece produsul scalar a doi vectori este simetric.

Exemplul 4. În spațiul euclidian $(\mathbb{R}^2, \langle, \rangle)$ se dă baza $\mathcal{B} = (f_1 = (-1, 2)^T, f_2 = (3, 1)^T)$ și vectorii $v = 5f_1 + 3f_2$, $w = -f_1 + 4f_2$. Să se calculeze produsul scalar $\langle v, w \rangle$ și cosinusul unghiului dintre cei doi vectori.

Calculăm în prealabil matricea $G = (g_{ij})$:

$$G = \begin{bmatrix} \langle f_1, f_1 \rangle & \langle f_1, f_2 \rangle \\ \langle f_2, f_1 \rangle & \langle f_2, f_2 \rangle \end{bmatrix} = \begin{bmatrix} 5 & -1 \\ -1 & 10 \end{bmatrix}$$

Astfel produsul scalar al celor doi vectori este:

$$\langle v, w \rangle = \langle 5f_1 + 3f_2, -f_1 + 4f_2 \rangle = \begin{bmatrix} 5 & 3 \end{bmatrix} \begin{bmatrix} 5 & -1 \\ -1 & 10 \end{bmatrix} \begin{bmatrix} -1 \\ 4 \end{bmatrix} = 78$$

Propoziția 6.2.3 Fie \mathcal{B} o bază ortonormată în spațiul vectorial euclidian (V_n, \langle, \rangle) :

$$\mathcal{B} = (u_1, u_1, \dots, u_n)$$

Produsul scalar a doi vectori v, w exprimați în baza \mathcal{B} , $v = x_1u_1 + x_2u_2 + \dots + x_nu_n$, $w = y_1u_1 + y_2u_2 + \dots + y_nu_n$, este $\langle v, w \rangle = x_1y_1 + x_2y_2 + \dots + x_ny_n$. În particular norma unui vector v exprimat în baza ortonormată \mathcal{B} este $\|v\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$.

Demonstrație: Baza \mathcal{B} fiind ortonormată produsul scalar a doi vectori din bază este $\langle u_i, u_j \rangle = \delta_{ij}$, adică matricea G de elemente $\langle u_i, u_j \rangle$ este matricea unitate și deci:

$$\begin{aligned} \langle v, w \rangle &= \langle x_1 u_1 + x_2 u_2 + \cdots + x_n u_n, y_1 u_1 + y_2 u_2 + \cdots + y_n u_n \rangle = \\ &= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} I_n \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \end{aligned}$$

Deci proprietatea enunțată are loc. În particular norma unui vector v , este:

$$\sqrt{\langle v, v \rangle} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

□

Simplitatea modului de calcul a produsului scalar a doi vectori exprimați într-o bază ortonormată, ca și modalitatea de determinare a coordonatelor unui vector într-o bază ortonormată, fac ca acest tip de bază să fie foarte util în aplicații. În inginerie se folosesc preponderent baze ortonormate. În procesul de vizualizare și manipulare (rotirea animată, de exemplu) a obiectelor 3D, în grafica pe calculator, în gaming, se construiesc diverse baze ortonormate, care prin calculele mai simple implicate măresc viteza de generare a imaginilor grafice și de simulare a mișcării în gaming.

6.3 Matricea de trecere dintre două baze ortonormate

Pentru implementarea mișcării în animație, robotică și gaming se folosesc diferite baze ortonormate. În continuare arătăm că matricea de trecere între două baze ortonormate este o matrice particulară, care simplifică mult calculele.

Definiția 6.3.1 O matrice pătratică cu elemente reale, $A \in \mathbb{R}^{n \times n}$, cu proprietatea că:

$$A^T A = I_n \tag{6.11}$$

se numește matrice ortogonală.

Exemplul 5. Orice matrice A de forma:

$$A = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}, \quad \alpha \in \mathbb{R} \tag{6.12}$$

este o matrice ortogonală deoarece:

$$\begin{aligned} A^T A &= \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} \cdot \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} = \\ &= \begin{pmatrix} \cos^2 \alpha + \sin^2 \alpha & -\cos \alpha \sin \alpha + \cos \alpha \sin \alpha \\ -\cos \alpha \sin \alpha + \cos \alpha \sin \alpha & \cos^2 \alpha + \sin^2 \alpha \end{pmatrix} = \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2 \end{aligned}$$

Proprietăți ale matricilor ortogonale

1. Condiția $A^T A = I_n$ din definiția matricii ortogonale este echivalentă cu $AA^T = I_n$.

Într-adevăr, înmulțim matricea AA^T , la stânga cu A^T și avem:

$$A^T(AA^T) = (A^T A)A^T = I_n A^T = A^T$$

Deoarece AA^T are "efect neutru" față de înmulțire, rezultă că $AA^T = I_n$. Analog se arată și că $AA^T = I_n$ implică $A^T A = I_n$.

2. O matrice ortogonală are determinantul egal cu 1 sau -1 .

Pentru a demonstra această proprietate reamintim că determinantul produsului a două matrici este egal cu produsul determinantilor:

$$\det(AB) = \det(A)\det(B)$$

și determinantul transpusei unei matrici este egal cu determinantul matricii:

$$\det(A^T) = \det(A)$$

Din definiția unei matrici ortogonale avem:

$$AA^T = I_n$$

și deci

$$\det(AA^T) = \det(I_n) = 1 \Leftrightarrow \det(A)\det(A^T) = 1 \Leftrightarrow (\det(A))^2 = 1 \Leftrightarrow \det(A) = \pm 1$$

3. O matrice ortogonală este inversabilă și inversa sa este egală cu transpusa.

Având determinantul egal cu 1 sau -1, deci nenul, o matrice ortogonală este inversabilă, adică există o unică matrice A^{-1} astfel încât:

$$AA^{-1} = A^{-1}A = I_n$$

Deoarece prin definiție avem că:

$$AA^T = A^T A = I_n,$$

rezultă ca transpusa unei matrici ortogonale coincide cu inversa: $A^T = A^{-1}$.

4. Inversa unei matrici ortogonale este o matrice ortogonală.

Pentru a ilustra această proprietate folosim faptul că inversa unei matrici ortogonale este egală cu transpusa sa, $A^{-1} = A^T$. Prin urmare trebuie să demonstrăm că dacă A este matrice ortogonală atunci și transpusa sa este ortogonală, adică $(A^T)^T A^T = I_n$.

Într-adevăr, din $C^T B^T = (BC)^T$ avem:

$$(A^T)^T A^T = (AA^T)^T = I_n^T = I_n$$

4. Produsul a două matrici ortogonale este o matrice ortogonală.

Fie A, B matrici ortogonale, adică $A^T A = I_n$ și $B^T B = I_n$. Să calculăm produsul $(AB)^T(AB)$:

$$(AB)^T(AB) = (B^T A^T)(AB) = B^T \underbrace{(A^T A)}_{I_n} B = B^T B = I_n$$

În concluzie AB este matrice ortogonală.

Datorită proprietăților 4 și 3 rezultă că mulțimea $O(n)$ a matricilor ortogonale de tip $n \times n$ formează subgrup în grupul multiplicativ al matricilor pătrate de tip $n \times n$, nesingulare.

Fie $(V_n, <, >)$ un spațiu vectorial real de dimensiune n înzestrat cu produsul scalar $<, >$ și $\mathcal{B} = (e_1, e_2, \dots, e_n)$, $\mathcal{B}' = (u_1, u_2, \dots, u_n)$ două baze ortonormate în V_n . Expresând vectorii bazei \mathcal{B}' în funcție de vectorii bazei \mathcal{B} avem:

$$\begin{aligned} u_1 &= a_{11}e_1 + a_{12}e_2 + \dots + a_{1n}e_n \\ u_2 &= a_{21}e_1 + a_{22}e_2 + \dots + a_{2n}e_n \\ &\vdots \\ u_n &= a_{n1}e_1 + a_{n2}e_2 + \dots + a_{nn}e_n \end{aligned} \tag{6.13}$$

Astfel matricea de trecere de la baza \mathcal{B} la baza \mathcal{B}' este:

$$T_{\mathcal{B}\mathcal{B}'} = \begin{bmatrix} \begin{matrix} u_1 \\ \downarrow \\ a_{11} & a_{21} & \dots & a_{n1} \end{matrix} & \begin{matrix} u_2 \\ \downarrow \\ a_{12} & a_{22} & \dots & a_{n2} \end{matrix} & \begin{matrix} u_n \\ \downarrow \\ \vdots & \vdots & \dots & \vdots \end{matrix} \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{bmatrix} \tag{6.14}$$

Propoziția 6.3.1 Matricea de trecere dintre două baze ortonormate este o matrice ortogonală.

Demonstrație: Să arătăm că $T_{\mathcal{B}\mathcal{B}'}^T T_{\mathcal{B}\mathcal{B}'} = I_n$:

$$T_{\mathcal{B}\mathcal{B}'}^T T_{\mathcal{B}\mathcal{B}'} = \begin{matrix} u_1 \rightarrow \\ u_2 \rightarrow \\ \vdots \\ u_n \rightarrow \end{matrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \dots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{bmatrix} \quad (6.15)$$

Elementul din poziția (i, j) a matricii produs este

$$a_{i1}a_{j1} + a_{i2}a_{j2} + \dots + a_{in}a_{jn} = \langle u_i, u_j \rangle,$$

deoarece u_i și u_j sunt exprimați în baza ortonormată \mathcal{B} și deci produsul lor scalar este suma produselor coordonatelor din aceeași poziție (Propoziția 6.1.3). Astfel avem:

$$T_{\mathcal{B}\mathcal{B}'}^T T_{\mathcal{B}\mathcal{B}'} = \begin{bmatrix} \langle u_1, u_1 \rangle & \langle u_1, u_2 \rangle & \dots & \langle u_1, u_n \rangle \\ \langle u_2, u_1 \rangle & \langle u_2, u_2 \rangle & \dots & \langle u_2, u_n \rangle \\ \vdots & \vdots & \dots & \vdots \\ \langle u_n, u_1 \rangle & \langle u_n, u_2 \rangle & \dots & \langle u_n, u_n \rangle \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = I_n,$$

deoarece baza \mathcal{B}' este ortonormată, avem că produsele scalare a doi vectori distincți, $\langle u_i, u_j \rangle = 0$, $i \neq j$ și $\langle u_i, u_i \rangle = 1$. Prin urmare:

$$T_{\mathcal{B}\mathcal{B}'}^T T_{\mathcal{B}\mathcal{B}'} = I_n$$

adică matricea de trecere $T_{\mathcal{B}\mathcal{B}'}$ este o matrice ortogonală. \square

Consecință. Dacă \mathcal{B} , \mathcal{B}' sunt două baze ortonormate în (V_n, \langle, \rangle) atunci matricile de trecere $T_{\mathcal{B}\mathcal{B}'}$, $T_{\mathcal{B}'\mathcal{B}}$ sunt matrici ortogonale și deoarece în general avem $T_{\mathcal{B}'\mathcal{B}} = T_{\mathcal{B}\mathcal{B}'}^{-1}$, în acest caz relația este mai simplă pentru că $T_{\mathcal{B}\mathcal{B}'}^{-1} = T_{\mathcal{B}\mathcal{B}'}^T$:

$$\mathbf{T}_{\mathcal{B}'\mathcal{B}} = \mathbf{T}_{\mathcal{B}\mathcal{B}'}^T \quad (6.16)$$

!!!! Cine uită această proprietate calculează mult inversa, iar cine o aplică pentru orice baze, nu doar ortonormate, GREȘESTE.

Exemplul 6. În spațiul vectorial \mathbb{R}^3 se consideră baza canonică $\mathcal{B} = (e_1, e_2, e_3)$, (care conform celor demonstrate este ortonormată) și baza ortonormată:

$$\mathcal{B}' = \left(u_1 = \left(\frac{1}{\sqrt{10}}, \frac{1}{\sqrt{10}}, \frac{-2}{\sqrt{5}} \right)^T, u_2 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0 \right)^T, u_3 = \left(\frac{2}{\sqrt{10}}, \frac{2}{\sqrt{10}}, \frac{1}{\sqrt{5}} \right)^T \right)$$

- a) Să se determine matricile de trecere între cele două baze.
 b) Să se determine coordonatele vectorului $v = (-2, 0, 1)^T$ relativ la baza \mathcal{B}' .
 c) Fie $w_1 = 2u_1 - 6u_2 + u_3$, $w_2 = 3u_1 - 5u_2$. Să se calculeze produsul scalar $\langle w_1, w_2 \rangle$.

Rezolvare: a) Vectorii bazei \mathcal{B}' sunt exprimați în baza canonică, deci putem constitui direct matricea $T_{\mathcal{B}\mathcal{B}'}$:

$$T_{\mathcal{B}\mathcal{B}'} = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{2}} & \frac{2}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{10}} \\ -\frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \end{bmatrix}$$

Matricea de trecere $T_{\mathcal{B}'\mathcal{B}}$ de la baza \mathcal{B}' la baza \mathcal{B} este în cazul general inversa matricii $T_{\mathcal{B}\mathcal{B}'}$. Cum bazele \mathcal{B} , \mathcal{B}' sunt ambele ortonormate, matricea $T_{\mathcal{B}\mathcal{B}'}$ este ortogonală și deci inversa sa coincide cu transpusa:

$$T_{\mathcal{B}'\mathcal{B}} = T_{\mathcal{B}\mathcal{B}'}^T = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{10}} & -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{2}{\sqrt{10}} & \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{5}} \end{bmatrix}$$

- b) Vectorul v se descompune după vectorii bazei \mathcal{B}' în forma:

$$v = y_1 u_1 + y_2 u_2 + y_3 u_3$$

Coordonatele y_i se pot afla în două moduri:

- 1) $y_i = \langle v, u_i \rangle$, $i = \overline{1, 3}$, așa cum am demonstrat mai sus;
- 2) folosind matricea de trecere:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}_{\mathcal{B}'} = T_{\mathcal{B}'\mathcal{B}} \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{10}} & -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{2}{\sqrt{10}} & \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{2}{\sqrt{10}} - \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{2}} \\ -\frac{4}{\sqrt{10}} + \frac{1}{\sqrt{5}} \end{bmatrix}$$

Modalitatea 1) este convenabilă în calculul "manual" al coordonatelor, iar cea de-a doua modalitate se folosește în implementarea algoritmilor ce fac uz de baze ortonormate. Exploatarea faptului că inversa matricii $T_{\mathcal{B}\mathcal{B}'}$ coincide cu transpusa conduce la eliminarea erorilor ce s-ar produce prin calculul inversei în mod uzual sau folosind reducerea la forma scară redusă.

c) Deoarece baza \mathcal{B}' este o bază ortonormată, produsul scalar a doi vectori $w_1 = x_1u_1 + x_2u_2 + x_3u_3$, $w_2 = y_1u_1 + y_2u_2 + y_3u_3$ este $\langle w_1, w_2 \rangle = x_1y_1 + x_2y_2 + x_3y_3$. În cazul vectorilor dați, produsul scalar devine:

$$\langle w_1, w_2 \rangle = 2 \cdot 3 + (-6) \cdot (-5) + 1 \cdot 0 = 6 + 30 + 0 = 36$$

Exprimarea coordonatelor unui versor într-o bază ortonormată

Propoziția 6.3.2 *Coordonatele unui versor într-o bază ortonormată sunt cosinusurile unghiurilor dintre direcția versorului și vectorii bazei.*

Demonstrație: Fie $\mathcal{B} = (u_1, u_2, \dots, u_n)$ o bază ortonormată în spațiul vectorial euclidian \mathbb{R}^n și $v = x_1u_1 + x_2u_2 + \dots + x_nu_n$ un vector arbitrar din $\mathbb{R}^n \setminus \{\theta\}$, raportat la baza \mathcal{B} . Versorul său este $v^0 = \frac{v}{\|v\|}$. Deoarece baza \mathcal{B} este ortonormată norma $\|v\| = \sqrt{\langle v, v \rangle} = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$. Prin urmare exprimarea versorului v^0 în baza \mathcal{B} este:

$$v^0 = y_1u_1 + y_2u_2 + \dots + y_nu_n = \frac{1}{\sqrt{x_1^2 + x_2^2 + \dots + x_n^2}}(x_1u_1 + x_2u_2 + \dots + x_nu_n$$

și deci coordonata din poziția i este: $y_i = \frac{x_i}{\sqrt{x_1^2 + x_2^2 + \dots + x_n^2}}, i = \overline{1, n}$.

Pe de altă parte, cosinusul unghiului dintre vectorul v și vectorul u_i al bazei este:

$$\cos(\widehat{v, u_i}) = \frac{\langle v, u_i \rangle}{\|v\| \|u_i\|}$$

Dar $\langle v, u_i \rangle = \langle x_1u_1 + x_2u_2 + \dots + x_iu_i + \dots + x_nu_n, 0u_1 + 0u_2 + \dots + 1u_i + \dots + 0u_n \rangle = x_i$, iar $\|v\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$, și $\|u_i\| = 1$ (pentru că baza este ortonormată). Astfel avem:

$$\cos(\widehat{v, u_i}) = \frac{x_i}{\sqrt{x_1^2 + x_2^2 + \dots + x_n^2}} = y_i, \quad i = \overline{1, n}$$

În concluzie, exprimarea unui versor în baza ortonormată \mathcal{B} este:

$$\begin{aligned} v^0 &= \cos(\widehat{v, u_1}) u_1 + \cos(\widehat{v, u_2}) u_2 + \dots + \cos(\widehat{v, u_n}) u_n = \\ &= \cos(\widehat{v^0, u_1}) u_1 + \cos(\widehat{v^0, u_2}) u_2 + \dots + \cos(\widehat{v^0, u_n}) u_n \end{aligned}$$

Ultima egalitate are loc deoarece versorul are aceeași direcție și sens ca și v , deci același unghi cu u_i . \square

Coordonatele versorului unui vector nenul v se mai numesc și **cosinușii directori ai direcției vectorului v** . În fizică și în grafică se spune adeseori: *fie în spațiul 3D, \mathbb{R}^3 , direcția de cosinuși directori $(\cos \beta_1, \cos \beta_2, \cos \beta_3)^T$, ceea ce "s-ar traduce" prin fie direcția de versor $(\cos \beta_1, \cos \beta_2, \cos \beta_3)^T$.*

Exemplul 7. Să se precizeze cosinușii directori ai vectorului $v = (-1, 4, 2)^T$ din \mathbb{R}^3 .

Vectorul v este dat în baza canonică, care este o bază ortonormată. Prin urmare cosinuşii directori ai lui v sunt coordonatele versorului său:

$$v^0 = \frac{v}{\|v\|} = \frac{(-1, 4, 2)^T}{\sqrt{1 + 16 + 4}} = \left(\frac{-1}{\sqrt{21}}, \frac{4}{\sqrt{21}}, \frac{2}{\sqrt{21}} \right)$$

$$\text{Astfel } \cos(\widehat{v, e_1}) = \frac{-1}{\sqrt{21}}, \cos(\widehat{v, e_2}) = \frac{4}{\sqrt{21}}, \cos(\widehat{v, e_3}) = \frac{2}{\sqrt{21}}.$$