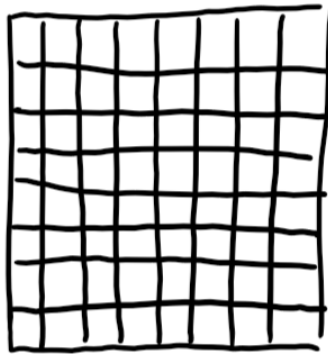


Numerične metode za tenzorje

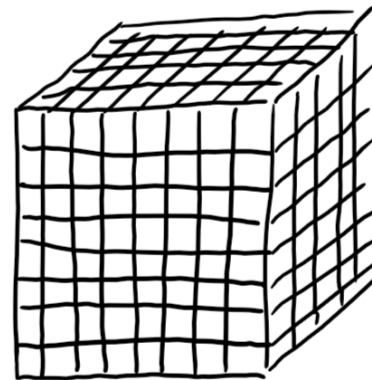
Bor Plestenjak

Matrike so dvodimenzionalne tabele.

Tenzorji so večdimenzionalne tabele.



matrika

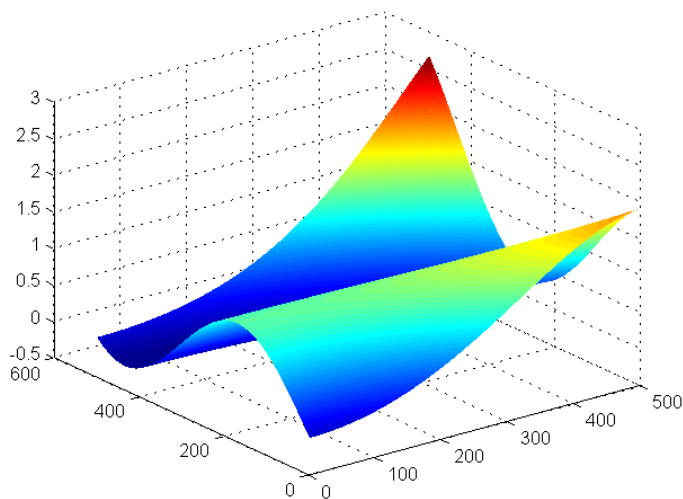


tenzor reda 3

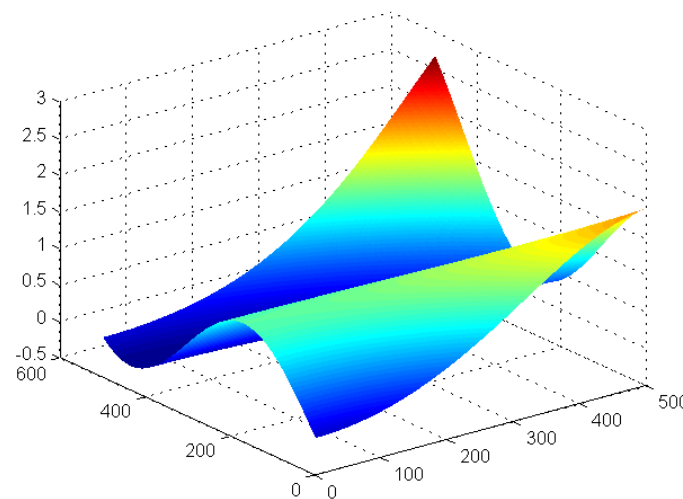
Motivacija

Če so elementi A tabelirane vrednosti $a_{ij} = u(x_i, y_j)$ gladke funkcije u , npr. rešitve PDE, na pravokotni mreži, pogosto obstaja dobra aproksimacija nizkega ranga.

Zgled: Naj bo $a_{ij} = u(x_i, y_j)$ za $u(x, y) = \sin(2x^2 + 5y) + x^2 + y^2$ na $[0, 1] \times [0, 1]$. Za 500×500 matriko dobimo dobro aproksimacijo že z matriko ranga 4.



polna matrika



matrika ranga 4

$$A \approx \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \sigma_3 u_3 v_3^T + \sigma_4 u_4 v_4^T.$$

Podoben kompakten zapis bi radi imeli za tenzorje.

Oznake

$\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ je **tenzor reda d** oziroma d -razsežna matrika (tabela) z elementi

$$a_{i_1 i_2 \dots i_d} \in \mathbb{R}.$$

Za $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ lahko definiramo skalarni produkt

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \dots \sum_{i_d=1}^{n_d} a_{i_1 i_2 \dots i_d} b_{i_1 i_2 \dots i_d}$$

in posplošitev Frobeniusove norme

$$\|\mathcal{A}\| = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}.$$

Tenzor $\mathcal{A}$ ima rang 1, če obstajajo taki vektorji $a^{(i)} \in \mathbb{R}^{n_i}$ za $i = 1, \dots, d$, da je

$$a_{i_1 i_2 \dots i_d} = a_{i_1}^{(1)} a_{i_2}^{(2)} \dots a_{i_d}^{(d)}$$

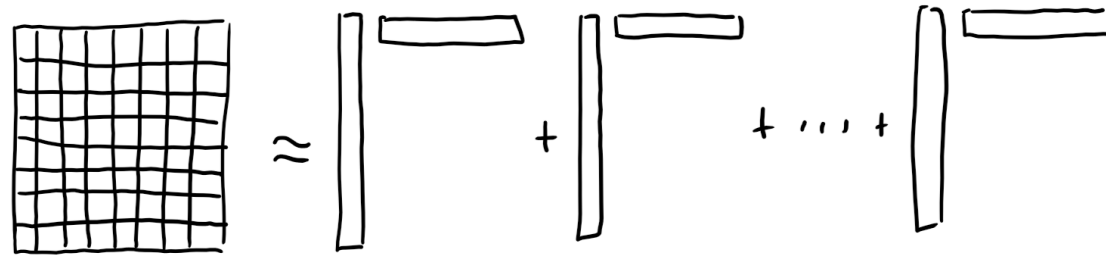
za vse $i_1, \dots, i_d$. Pišemo

$$\mathcal{A} = a^{(1)} \circ a^{(2)} \circ \dots \circ a^{(d)}$$

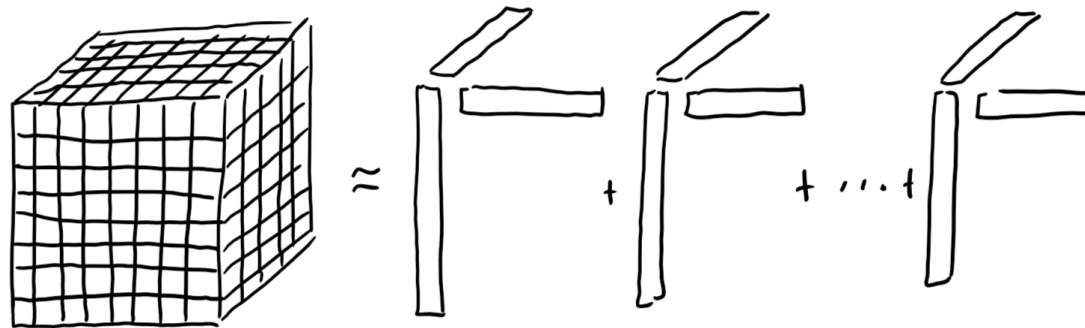
pri čemer je $a^{(1)} \circ a^{(2)} \circ \dots \circ a^{(d)}$ **zunanji produkt vektorjev** $a^{(1)}, a^{(2)}, \dots, a^{(d)}$.

Ideja

Za matriko dobimo aproksimacijo v obliki



Posplošitev za tenzor reda 3 bi bila oblika



oziroma

$$\mathcal{A} \approx \sum_{i=1}^k \lambda_i u_i \circ v_i \circ w_i.$$

Tako bi npr. namesto $\mathcal{O}(n^3)$ podatkov lahko uporabili le $\mathcal{O}(kn)$ podatkov.

Kroneckerjev in Khatri–Raojev produkt

Kroneckerjev produkt matrik $A \in \mathbb{R}^{m \times n}$ in $B \in \mathbb{R}^{p \times q}$ je

$$\begin{aligned} A \otimes B &= \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix} \in \mathbb{R}^{mp \times nq} \\ &= [a_1 \otimes b_1 \quad a_1 \otimes b_2 \quad \cdots \quad a_1 \otimes b_q \quad a_2 \otimes b_1 \quad \cdots \quad a_m \otimes b_q]. \end{aligned}$$

Za vektorja $x \in \mathbb{R}^m$ in $y \in \mathbb{R}^n$ je

$$x \otimes y = \begin{bmatrix} x_1 y \\ \vdots \\ x_n y \end{bmatrix} \in \mathbb{R}^{mn}.$$

Khatri–Raojev produkt matrik $A \in \mathbb{R}^{m \times n}$ in $B \in \mathbb{R}^{p \times n}$ je

$$A \odot B = [a_1 \otimes b_1 \quad a_2 \otimes b_2 \quad \cdots \quad a_n \otimes b_n] \in \mathbb{R}^{mp \times n}.$$

Vektorizacija

Iz vseh elementov tenzorja lahko sestavimo vektor. Tu je pomembno, kakšen vrstni red uporabljamo, da je usklajen med vsemi operacijami.

Običajno uporabljamo obratno leksikografsko urejenost, kjer je

$$i_1 i_2 \dots i_d \prec j_1 j_2 \dots j_d \iff i_k < j_k \text{ in } i_\ell = j_\ell \text{ za } \ell = k + 1, \dots, d \text{ za nek } k.$$

Zgled: če imamo $\mathcal{A} \in \mathbb{R}^{2 \times 3 \times 2}$, kjer je

$$\mathcal{A}(:, :, 1) = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}, \quad \mathcal{A}(:, :, 2) = \begin{bmatrix} 7 & 9 & 11 \\ 8 & 10 & 12 \end{bmatrix},$$

potem je

$$\text{vec}(\mathcal{A}) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ \vdots \\ 11 \\ 12 \end{bmatrix} = \begin{bmatrix} a_{111} \\ a_{211} \\ a_{121} \\ \vdots \\ a_{132} \\ a_{232} \end{bmatrix}.$$

Če je $\mathcal{A} = a \circ b \circ c$ tenzor ranga 1, je to ekvivalentno $\text{vec}(\mathcal{A}) = c \otimes b \otimes a$.

Če je $A = uv^T$ matrika ranga 1 je to ekvivalentno $\text{vec}(A) = v \otimes u$.

Matrifikacija

Če smeri razdelimo na dve množici in eno uporabimo za vrstice, drugo pa za stolpce, dobimo **matrifikacijo** tenzorja $\mathcal{A}$.

Zapišemo $\{1, \dots, d\} = r \cup c$, kjer je $r = \{r_1, \dots, r_k\}$ in $c = \{c_1, \dots, c_{d-k}\}$.

Tako dobimo matriko velikosti $n_{r_1} \dots n_{r_k} \times n_{c_1} \dots n_{c_k}$, označimo jo z $\mathcal{A}^{(r)}$. Velja

$$(\mathcal{A}^{(r)})_{(i_{r_1} \dots i_{r_k}), (i_{c_1} \dots i_{c_{d-k}})} = \mathcal{A}_{i_1 \dots i_d},$$

kjer je

$$\begin{aligned} (i_{r_1} \dots i_{r_k}) &\in \{1, \dots, n_{r_1}\} \times \dots \times \{1, \dots, n_{r_k}\} \\ (i_{c_1} \dots i_{c_{d-k}}) &\in \{1, \dots, n_{c_1}\} \times \dots \times \{1, \dots, n_{c_{d-k}}\} \end{aligned}$$

Pri tem spet upoštevamo npr. obratno leksikografsko urejenost.

Pri **matrifikaciji v smeri j** (razprtju) za $j = 1, \dots, d$, izberemo $r = \{j\}$ in dobimo

$$\mathcal{A}_{(j)} \in \mathbb{R}^{n_j \times n_1 \dots n_{j-1} n_{j+1} \dots n_d}.$$

Stolpci matrike $\mathcal{A}_{(j)}$ so vlakna (fibers) v smeri j .

Matrifikacije po smereh

Zgled: če imamo $\mathcal{A} \in \mathbb{R}^{2 \times 3 \times 2}$, kjer je

$$\mathcal{A}(:, :, 1) = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}, \quad \mathcal{A}(:, :, 2) = \begin{bmatrix} 7 & 9 & 11 \\ 8 & 10 & 12 \end{bmatrix},$$

potem je

$$\mathcal{A}_{(1)} = \begin{bmatrix} 1 & 3 & 5 & 7 & 8 & 11 \\ 2 & 4 & 6 & 8 & 10 & 12 \end{bmatrix},$$

$$\mathcal{A}_{(2)} = \begin{bmatrix} 1 & 2 & 7 & 8 \\ 3 & 4 & 9 & 10 \\ 5 & 6 & 11 & 12 \end{bmatrix},$$

$$\mathcal{A}_{(3)} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 7 & 8 & 9 & 10 & 11 & 12 \end{bmatrix}.$$

Vlakna v smeri 1 so stolpci, v smeri 2 so vrstice, v smeri 3 pa cevi (tubes).

Razprtja tenzorjev ranga 1

Naj bo

$$\mathcal{A} = a \circ b \circ c = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \circ \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \circ \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

Potem je

$$\mathcal{A}(:, :, 1) = \begin{bmatrix} a_1 b_1 c_1 & a_1 b_2 c_1 & a_1 b_3 c_1 \\ a_2 b_1 c_1 & a_2 b_2 c_1 & a_2 b_3 c_1 \end{bmatrix}, \quad \mathcal{A}(:, :, 2) = \begin{bmatrix} a_1 b_1 c_2 & a_1 b_2 c_2 & a_1 b_3 c_2 \\ a_2 b_1 c_2 & a_2 b_2 c_2 & a_2 b_3 c_2 \end{bmatrix}$$

$$\mathcal{A}_{(1)} = \begin{bmatrix} a_1 b_1 c_1 & a_1 b_2 c_1 & a_1 b_3 c_1 & a_1 b_1 c_2 & a_1 b_2 c_2 & a_1 b_3 c_2 \\ a_2 b_1 c_1 & a_2 b_2 c_1 & a_2 b_3 c_1 & a_2 b_1 c_2 & a_2 b_2 c_2 & a_2 b_3 c_2 \end{bmatrix} = a (c \otimes b)^T,$$

$$\mathcal{A}_{(2)} = \begin{bmatrix} a_1 b_1 c_1 & a_2 b_1 c_1 & a_1 b_1 c_2 & a_2 b_1 c_2 \\ a_1 b_2 c_1 & a_2 b_2 c_1 & a_1 b_2 c_2 & a_2 b_2 c_2 \\ a_1 b_3 c_1 & a_2 b_3 c_1 & a_1 b_3 c_2 & a_2 b_3 c_2 \end{bmatrix} = b (c \otimes a)^T$$

$$\mathcal{A}_{(3)} = \begin{bmatrix} a_1 b_1 c_1 & a_2 b_1 c_1 & a_1 b_2 c_1 & a_2 b_2 c_1 & a_1 b_3 c_1 & a_2 b_3 c_1 \\ a_1 b_1 c_2 & a_2 b_1 c_2 & a_1 b_2 c_2 & a_2 b_2 c_2 & a_1 b_3 c_2 & a_2 b_3 c_2 \end{bmatrix} = c (b \otimes a)^T.$$

$$\mathcal{A} = z^{(1)} \circ \dots \circ z^{(d)} \implies \mathcal{A}_{(k)} = z^{(k)} \left(z^{(n)} \otimes \dots \otimes z^{(k+1)} \otimes z^{(k-1)} \otimes \dots \otimes z^{(1)} \right)^T.$$

Kontrakcije tenzorjev

Za $\mathcal{A} \in \mathbb{R}^{4 \times 2 \times 4 \times 3}$ in $\mathcal{B} \in \mathbb{R}^{4 \times 3 \times 5 \times 4}$ lahko definiramo $\mathcal{C} \in \mathbb{R}^{2 \times 3 \times 3 \times 5}$ kot

$$\mathcal{C}(i_1, i_2, i_3, i_4) = \sum_{\ell_1=1}^4 \sum_{\ell_2=1}^4 A(\ell_1, i_1, \ell_2, i_2) B(\ell_1, i_3, i_4, \ell_2).$$

Temu pravimo **kontrakcija tenzorjev** in je posplošitev matričnega množenja.

Za $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_c}$, $\mathcal{B} \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_d}$ izberemo p indeksov $i_1, \dots, i_p$ za $\mathcal{A}$ in p indeksov $j_1, \dots, j_p$ za $\mathcal{B}$, pri čemer je $n_{i_\ell} = m_{j_\ell}$ za $\ell = 1, \dots, p$, in dobimo

$$\mathcal{C} \in \mathbb{R}^{n_{k_1} \times \cdots \times n_{k_c-p} \times m_{\ell_1} \cdots m_{\ell_{d-p}}},$$

kjer nastopa preostalih $c - p$ indeksov za $\mathcal{A}$ in $d - p$ indeksov za $\mathcal{B}$.

Zgornji primer ustreza izbirama indeksov (1, 3) za $\mathcal{A}$ in (1, 4) za $\mathcal{B}$.

Množenje v smeri j

To je poseben primer kontrakcije.

Tenzor $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ množimo v smeri j z matriko $A \in \mathbb{R}^{m_j \times n_j}$ in dobimo

$$\mathcal{X} \times_j A \in \mathbb{R}^{n_1 \times \cdots \times n_{j-1} \times m_j \times n_{j+1} \times \cdots \times n_d},$$

kjer je

$$(\mathcal{X} \times_j A)_{i_1 \cdots i_{j-1} k i_{j+1} \cdots i_d} = \sum_{i_j=1}^{n_j} x_{i_1 \cdots i_d} a_{k i_j}.$$

To je ekvivalentno

$$(\mathcal{X} \times_j A)_{(j)} = A \mathcal{X}_{(j)}$$

oziroma

$$\text{vec}(\mathcal{X} \times_j A) = (I_{n_d} \otimes \cdots \otimes I_{n_{j+1}} \otimes A \otimes I_{n_{j-1}} \otimes \cdots \otimes I_{n_1}) \text{vec}(\mathcal{X}).$$

Velja

$$(\mathcal{X} \times_j A) \times_k B = (\mathcal{X} \times_k B) \times_j A, \quad j \neq k$$

in

$$(\mathcal{X} \times_j A) \times_j B = \mathcal{X} \times_j (BA).$$

Rang tenzorja

Pri **kanoničnem poliadičnem (CP) razcepu** tenzor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ zapišemo kot vsoto tenzorjev ranga 1 v obliki

$$\mathcal{X} = \sum_{\ell=1}^r \lambda_{\ell} \mathbf{a}_{\ell} \circ \mathbf{b}_{\ell} \circ \mathbf{c}_{\ell}.$$

Skalarje λ_{ℓ} dodamo zato, da za vektorje $\mathbf{a}_{\ell}, \mathbf{b}_{\ell}, \mathbf{c}_{\ell}$ potem zahtevamo, da so normirani. Krajši zapis je

$$X = \llbracket \boldsymbol{\lambda}; A, B, C \rrbracket,$$

ker je $A = [\mathbf{a}_1 \cdots \mathbf{a}_r]$, $B = [\mathbf{b}_1 \cdots \mathbf{b}_r]$, $C = [\mathbf{c}_1 \cdots \mathbf{c}_r]$, $\boldsymbol{\lambda} = [\lambda_1 \cdots \lambda_r]^T$.

Rang tenzorja $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ je najmanjši r , za katerega je možno $\mathcal{X}$ zapisati kot vsoto r tenzorjev ranga 1

$$\mathcal{X} = \sum_{\ell=1}^r \lambda_{\ell} \mathbf{a}_{\ell}^{(1)} \circ \cdots \circ \mathbf{a}_{\ell}^{(d)} = \llbracket \boldsymbol{\lambda}; A^{(1)}, \dots, A^{(d)} \rrbracket,$$

kjer je $\boldsymbol{\lambda} = [\lambda_1 \cdots \lambda_r]^T$ in $A^{(j)} = \begin{bmatrix} \mathbf{a}_1^{(j)} & \cdots & \mathbf{a}_r^{(j)} \end{bmatrix} \in \mathbb{R}^{n_j \times r}$ za $j = 1, \dots, d$.

Razlike med rangom tenzorja ($d > 2$) in matrike (1)

Rang matrike je neodvisen od obravnave nad $\mathbb{R}$ ali $\mathbb{C}$.

Pri tenzorjih to ni res. Rang $2 \times 2 \times 2$ tenzorja

$$\mathcal{X}(:, :, 1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathcal{X}(:, :, 2) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

je 3 nad $\mathbb{R}$ in 2 nad $\mathbb{C}$, saj je $\mathcal{X} = \llbracket A, B, C \rrbracket$ za

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}, \quad (\mathbb{R})$$

$$A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}, \quad B = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}. \quad (\mathbb{C})$$

Razlike med rangom tenzorja ($d > 2$) in matrike (2)

Obstajajo tenzorji, ki jih lahko poljubno dobro aproksimiramo s tenzorji nižjega ranga, kar pri matrikah ni možno.

Zgled: Tenzor ranga 3

$$\mathcal{X} = a_1 \circ b_1 \circ c_2 + a_1 \circ b_2 \circ c_1 + a_2 \circ b_1 \circ c_1$$

lahko poljubno dobro aproksimiramo s tenzorjem ranga 2

$$\mathcal{Z}_\alpha = \alpha \left(a_1 + \frac{1}{\alpha} a_2 \right) \circ \left(b_1 + \frac{1}{\alpha} b_2 \right) \circ \left(c_1 + \frac{1}{\alpha} c_2 \right) - \alpha a_1 \circ b_1 \circ c_1,$$

saj je

$$\mathcal{X} - \mathcal{Z}_\alpha = \frac{1}{\alpha} \left(a_2 \circ b_2 \circ c_1 + a_2 \circ b_1 \circ c_2 + a_1 \circ b_2 \circ c_2 + \frac{1}{\alpha} a_2 \circ b_2 \circ c_2 \right),$$

torej je $\lim_{\alpha \rightarrow \infty} \|\mathcal{X} - \mathcal{Z}_\alpha\| = 0$.

Razlike med rangom tenzorja ($d > 2$) in matrike (3)

Rang $m \times n$ matrike ne more biti večji od $\min(m, n)$.

Rang generične $m \times n$ matrike je $\min(m, n)$.

Za tenzor kaj podobnega ne velja. Za splošen tenzor $\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ ni formule, ki bi točno omejila njegov maksimalen rang. To je znano le za nekaj posebnih primerov, npr. za $d = 3$, če je ena dimenzija največ 3.

Tako je za $2 \times 2 \times 2$ tenzor maksimalen rang 3, za $3 \times 3 \times 3$ tenzor pa 5.

Tenzorji nimajo vedno samo en generičen rang. Če npr. vzamemo naključen realen $2 \times 2 \times 2$ tenzor, je 79% verjetnost, da ima rang 2, in 21% verjetnost, da ima rang 3.

Določiti rang tenzorja je NP težak problem.

Aproksimacija s CP razcepom

Za tenzor $\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ iščemo dobro aproksimacijo s tenzorjem ranga r v obliki

$$\mathcal{X} \approx \sum_{\ell=1}^r \lambda_{\ell} a_{\ell}^{(1)} \circ \dots \circ a_{\ell}^{(d)} = \llbracket \boldsymbol{\lambda}; A^{(1)}, \dots, A^{(d)} \rrbracket.$$

Direkten algoritem, kot je npr. SVD v primeru matrike, ne obstaja.

Pri reševanju fiksiramo r in uporabimo **alternirajoče najmanjše kvadrate (ALS)**.

Bistvena ideja: če fiksiramo vse matrike $A^{(j)}$ razen $A^{(k)}$, lahko iskanje matrike B , ki minimizira (vse λ_{ℓ} nastavimo na 1)

$$\min_{B \in \mathbb{R}^{n_k \times r}} \left\| \mathcal{X} - \llbracket A^{(1)}, \dots, A^{(k-1)}, B, A^{(k+1)}, \dots, A^{(d)} \rrbracket \right\|,$$

prevedemo na navaden problem najmanjših kvadratov.

Pomagamo si z zvezo, da za $\mathcal{Y} = \llbracket A^{(1)}, \dots, A^{(k-1)}, B, A^{(k+1)}, \dots, A^{(d)} \rrbracket$ velja

$$\mathcal{Y}_{(k)} = B \left(A^{(d)} \odot \dots \odot A^{(k+1)} \odot A^{(k-1)} \odot \dots \odot A^{(1)} \right)^T$$

in iščemo minimum

$$\left\| \mathcal{X}_{(k)} - B \left(A^{(d)} \odot \dots \odot A^{(k+1)} \odot A^{(k-1)} \odot \dots \odot A^{(1)} \right)^T \right\|_F.$$

Algoritem ALS-CP

Za dani $\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ in r iščemo dobro aproksimacijo v obliki

$$\mathcal{X} \approx \tilde{\mathcal{X}} = \sum_{\ell=1}^r \lambda_{\ell} a_{\ell}^{(1)} \circ \dots \circ a_{\ell}^{(d)} = \llbracket \boldsymbol{\lambda}; A^{(1)}, \dots, A^{(d)} \rrbracket,$$

kjer so stolpci matrik $A^{(j)}$ normirani za $j = 1, \dots, d$.

izberi začetne matrike z normiranimi stolpci $A^{(j)} \in \mathbb{R}^{n_j \times r}$ za $j = 1, \dots, d$
ponavljaj (dokler ni več izboljšanja ali do maksimalnega števila korakov)

$k = 1, \dots, d$

poišči $B \in \mathbb{R}^{n_k \times r}$, ki minimizira $\left\| \mathcal{X} - \llbracket A^{(1)}, \dots, A^{(k-1)}, B, A^{(k+1)}, \dots, A^{(d)} \rrbracket \right\|$

zapiši $A^{(k)} = B \text{diag}(\boldsymbol{\lambda})$, da so stolpci $A^{(k)}$ normirani.

vrni $\boldsymbol{\lambda}, A^{(1)}, \dots, A^{(d)}$

Algoritem izvajamo dokler se $\|\mathcal{X} - \tilde{\mathcal{X}}\|$ ne zmanjšuje več. Konvergenca do globalnega minimuma ni zagotovljena, niti ne do lokalnega, prav tako je lahko zelo počasna. Po drugi strani metoda velikokrat vrne čisto zadovoljive rezultate.

Kako rešimo LS problem v algoritmu ALS-CP

Pri iskanju matrike $B \in \mathbb{R}^{n_k \times r}$, ki minimizira

$$\min \left\| \mathcal{X} - \llbracket A^{(1)}, \dots, A^{(k-1)}, B, A^{(k+1)}, \dots, A^{(d)} \rrbracket \right\|,$$

uporabimo zvezo, da za $\mathcal{Y} = \llbracket A^{(1)}, \dots, A^{(k-1)}, B, A^{(k+1)}, \dots, A^{(d)} \rrbracket$ velja

$$\mathcal{Y}_{(k)} = B \left(A^{(d)} \odot \dots \odot A^{(k+1)} \odot A^{(k-1)} \odot \dots \odot A^{(1)} \right)^T$$

in iščemo minimum

$$\left\| \mathcal{X}_{(k)} - B \left(A^{(d)} \odot \dots \odot A^{(k+1)} \odot A^{(k-1)} \odot \dots \odot A^{(1)} \right)^T \right\|_F,$$

kar je ekvivalentno

$$\left\| \left(A^{(d)} \odot \dots \odot A^{(k+1)} \odot A^{(k-1)} \odot \dots \odot A^{(1)} \right) B^T - \mathcal{X}_{(k)}^T \right\|_F.$$

Če je $N = n_1 \dots n_d$, ima predoločen sistem matriko $N/n_k \times r$. Normalni sistem je

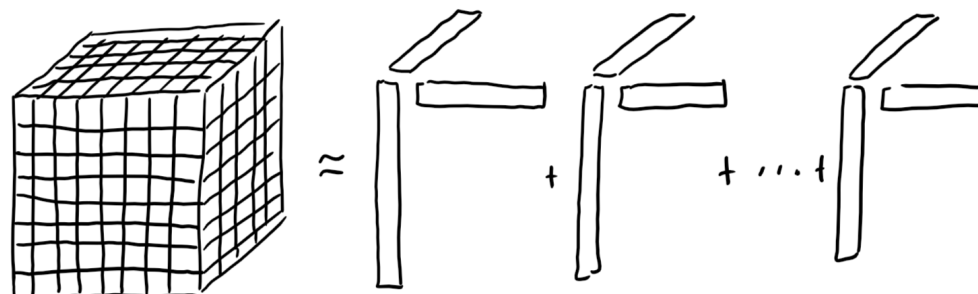
$$G B^T = \left(A^{(d)} \odot \dots \odot A^{(k+1)} \odot A^{(k-1)} \odot \dots \odot A^{(1)} \right)^T \mathcal{X}_{(k)}^T,$$

kjer je G matrika velikosti $r \times r$ oblike ($*$ je Hadarmardov produkt)

$$G = (A^{(d)T} A^{(d)}) * \dots * (A^{(k+1)T} A^{(k+1)}) * (A^{(k-1)T} A^{(k-1)}) * \dots * (A^{(1)T} A^{(1)}).$$

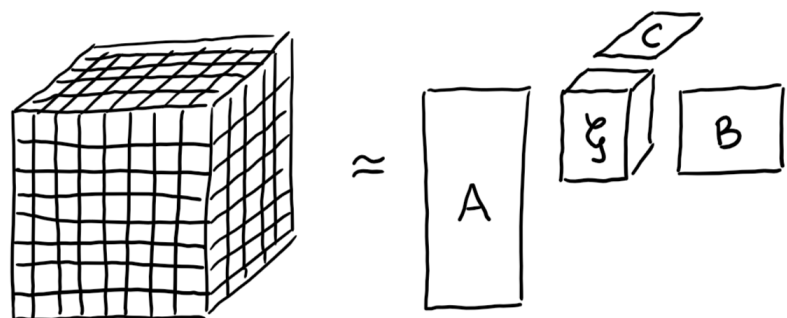
Tuckerjev razcep

Pri CP razcepu iščemo dobro aproksimacijo oblike



$$\approx \text{rank-1 tensor} + \text{rank-1 tensor} + \dots + \text{rank-1 tensor} = \sum_{i=1}^r \lambda_i a_i \circ b_i \circ c_i = [[\lambda; A, B, C]]$$

Pri **Tuckerjevem razcepu** iščemo dobro aproksimacijo oblike



$$\approx \begin{matrix} & C \\ \begin{matrix} \mathcal{G} \\ A \end{matrix} & B \end{matrix} = \mathcal{G} \times_1 A \times_2 B \times_3 C = [[\mathcal{G}; A, B, C]],$$

kjer je $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ **jedro** in $A \in \mathbb{R}^{n_1 \times r_1}$, $B \in \mathbb{R}^{n_2 \times r_2}$, $C \in \mathbb{R}^{n_3 \times r_3}$.

$$\mathcal{Y} = [[\mathcal{G}; A, B, C]] \text{ pomeni } y_{ijk} = \sum_{p=1}^{n_1} \sum_{q=1}^{n_2} \sum_{s=1}^{n_3} g_{pqs} a_{ip} b_{jq} c_{ks}.$$

CP je posebna oblika Tuckerja z $\mathcal{G} = \text{diag}(\lambda_1, \dots, \lambda_r)$ (superdiagonalni tenzor).

Drugačen pogled na singularni razcep

Če je $A = U\Sigma V^T$ SVD, je

$$U^T A = \Sigma V^T = \begin{bmatrix} \sigma_1 v_1^T \\ \sigma_2 v_2^T \\ \vdots \\ \sigma_n v_n^T \end{bmatrix}.$$

Vrstice matrike $U^T A$ so paroma pravokotne z monotono padajočimi normami.

HOSVD (high order SVD)

Izrek (HOSVD): Za vsak $\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ obstaja razcep

$$\mathcal{X} = \llbracket \mathcal{G}; U^{(1)}, \dots, U^{(d)} \rrbracket,$$

kjer imajo matrike $U^{(\ell)} \in \mathbb{R}^{n_\ell \times r_\ell}$ ortonormirane stolpce, $\mathcal{G} \in \mathbb{R}^{r_1 \times \dots \times r_d}$, in v vsaki smeri $\ell = 1, \dots, d$ za rezine $\mathcal{G}$ velja

- a) rezini z različnima indeksoma sta pravokotni: $\langle \mathcal{G}_{i_\ell=p}, \mathcal{G}_{i_\ell=q} \rangle = 0$ za $p \neq q$.
- b) norme rezin monotonno padajo: $\|\mathcal{G}_{i_\ell=1}\| \geq \|\mathcal{G}_{i_\ell=2}\| \geq \dots \geq \|\mathcal{G}_{i_\ell=r_\ell}\|$.

Pri tem je $r_\ell = \text{rang}(\mathcal{X}_{(\ell)})$ za $\ell = 1, \dots, d$.

Pravimo, da je $(r_1, \dots, r_d)$ **multilinearni rang** tenzorja $\mathcal{X}$.

$$\ell = 1, \dots, d$$

izračunaj singularni razcep $\mathcal{X}_{(\ell)} = U \Sigma V^T$

za $U^{(\ell)}$ vzemi prvih r_ℓ stolpcev U , kjer je $r_\ell = \text{rang}(\mathcal{X}_{(\ell)})$

$$\mathcal{G} = \mathcal{X} \times_1 U^{(1)T} \times_2 U^{(2)T} \dots \times_d U^{(d)T}$$

Aproksimacija v Tuckerjevi obliki

Za dani $\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ in $(r_1, \dots, r_d)$ iščemo dobro aproksimacijo $\mathcal{X}$ s tenzorjem z multilinearim rangom $(r_1, \dots, r_d)$. Uporabimo postopek kot za HOSVD in v vsaki smeri vzamemo le r_ℓ stolpcev:

$$\begin{aligned} \ell &= 1, \dots, d \\ &\text{izračunaj singularni razcep } \mathcal{X}_{(\ell)} = U \Sigma V^T \\ &\text{za } U^{(\ell)} \text{ vzemi prvih } r_\ell \text{ stolpcev } U \\ \mathcal{G} &= \mathcal{X} \times_1 U^{(1)T} \times_2 U^{(2)T} \dots \times_d U^{(d)T} \end{aligned}$$

Za razliko od matrik in singularnega razcepa ne velja, da je zgornji Tuckerjev razcep optimalna aproksimacija z multilinearim rangom $(r_1, \dots, r_d)$. Lahko pa ga uporabimo kot začetni približek za metodo **HOOI (high order orthogonal iteration)**, kjer uporabimo ALS in po smereh minimiziramo

$$\|\mathcal{X} - \llbracket \mathcal{G}; U^{(1)}, \dots, U^{(d)} \rrbracket\|.$$

Za fiksne $U^{(1)}, \dots, U^{(d)}$ z ortonormiranimi stolpci je optimalno jedro projekcija

$$\mathcal{G} = \mathcal{X} \times_1 U^{(1)T} \times_2 U^{(2)T} \dots \times_d U^{(d)T}.$$

HOOI (high order orthogonal iteration)

Iz $\mathcal{G} = \mathcal{X} \times_1 U^{(1)T} \times_2 \cdots \times_d U^{(d)T}$, označimo $U = U^{(d)} \otimes \cdots \otimes U^{(1)}$, dobimo

$$\begin{aligned}\|\mathcal{X} - \llbracket \mathcal{G}; U^{(1)}, \dots, U^{(d)} \rrbracket\|^2 &= \|\text{vec}(\mathcal{X}) - U \text{vec}(\mathcal{G})\|_2^2 = \|(I - UU^T) \text{vec}(\mathcal{X})\|_2^2 \\ &= \|\text{vec}(\mathcal{X})\|_2^2 - \|U^T \text{vec}(\mathcal{X})\|_2^2.\end{aligned}$$

Iščemo torej maksimum $\|U^T \text{vec}(\mathcal{X})\|_2$. Pri fiksni matriki $U^{(j)}$ za $j \neq k$ iz

$$\|U^T \text{vec}(\mathcal{X})\|_2 = \left\| U^{(k)T} \mathcal{X}_{(k)} \left(U^{(d)} \otimes \cdots \otimes U^{(k+1)} \otimes U^{(k-1)} \otimes \cdots \otimes U^{(1)} \right) \right\|_2.$$

sledi, da je za $U^{(k)}$ najbolje vzeti prvih r_k stolpcev iz singularnega razcepa.

začetni približek: ortonormirane matrike $U^{(\ell)} \in \mathbb{R}^{n_\ell \times r_\ell}$ za $\ell = 1, \dots, d$
ponavljaj (dokler ni več izboljšanja ali do maksimalnega števila korakov)

$k = 1, \dots, d$

izračunaj SVD $\mathcal{X}_{(k)} \left(U^{(d)} \otimes \cdots \otimes U^{(k+1)} \otimes U^{(k-1)} \otimes \cdots \otimes U^{(1)} \right) = \tilde{U}_k \Sigma V^T$

za $U^{(k)}$ vzemi prvih r_k stolpcev $\tilde{U}_k$

$$\mathcal{G} = \mathcal{X} \times_1 U^{(1)T} \times_2 U^{(2)T} \times_3 \cdots \times_d U^{(d)T}.$$

Naj bo $a_{ijk} = u(x_i, y_j, z_k)$ za

$$u(x, y, z) = \sin(2x^2 + 5y - 3z^2 + 5xy^2z) + x^2 + y^2 - z^2$$

na $[0, 1] \times [0, 1] \times [0, 1]$. Za tenzor velikosti $200 \times 200 \times 200$ dobimo dobro aproksimacijo za $\mathcal{A}$ že s tenzorjem nizkega ranga. Uporabljen je Matlabov paket Tensor Toolbox (<http://www.sandia.gov/~tgkolda/TensorToolbox/>).

Aproksimacijo v formatu CP dobimo z ukazom `parafac_als`. Za izbiro $r = 10$ po 39 iteracijah dobimo tenzor $\mathcal{A}_{10}$ ranga 10, da je

$$\|\mathcal{A} - \mathcal{A}_{10}\|_F = 1.36 \cdot 10^{-2} \|\mathcal{A}\|_F.$$

Aproksimacijo v Tuckerjevem formatu dobimo z `tucker_als`. Za izbiro $(r_1, r_2, r_3) = (6, 6, 6)$ dobimo po 3 iteracijah tenzor $\mathcal{B}$, da je

$$\|\mathcal{A} - \mathcal{B}\|_F = 4.60962 \cdot 10^{-3} \|\mathcal{A}\|_F.$$

Z odrezanim HOSVD z izbiro $(r_1, r_2, r_3) = (6, 6, 6)$ dobimo tenzor $\mathcal{C}$, da je

$$\|\mathcal{A} - \mathcal{C}\|_F = 4.60965 \cdot 10^{-3} \|\mathcal{A}\|_F.$$

Vidimo, da odrezani HOSVD sicer ne vrne najboljše aproksimacije z izbranim multirangom, je pa v tem primeru razlika zelo majhna.