Kapitola 8

Gram-Schmidtova ortogonalizace

V této kapitole budeme řešit následující úlohu.

Je dána posloupnost $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_l$ prvků lineárně nezávislé množiny $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_l\}$ vektorů z prostoru $\mathbf{C}^n$ (nebo z $\mathbf{R}^n$). Naším úkolem je najít posloupnost $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_l$ vektorů z $\mathbf{C}^n$ (nebo z $\mathbf{R}^n$) takovou, že množina $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_l\}$ je ortonormální a navíc pro každé $k = 1, \ldots, l$ platí $\mathbf{L}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\} = \mathbf{L}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$.

Při konstrukci ortonormální množiny $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_l\}$ budeme postupovat indukcí podle k.

Pro $k = 1$ zvolíme

$$\mathbf{u}_1 = \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}.$$

Jmenovatel zlomku je nenulový, neboť $\mathbf{x}_1 \neq \mathbf{0}$, protože lineárně nezávislá množina nemůže obsahovat nulový vektor. Dále $\|\mathbf{u}_1\| = 1$ podle Cvičení 7.2.2.

Předpokládejme nyní, že pro nějaké $k \geq 1$ a $k < l$ už máme sestrojenou ortonormální množinu $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, pro kterou platí $\mathbf{L}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_i\} = \mathbf{L}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i\}$ pro každé $i = 1, \ldots, k$. Budeme hledat nutné podmínky, které musí vektor $\mathbf{u}_{k+1}$ splňovat. Vektor $\mathbf{u}_{k+1}$ musíme zvolit tak, aby platila rovnost $\mathbf{L}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k+1}\} = \mathbf{L}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k+1}\}$. Protože $\mathbf{x}_{k+1} \in \mathbf{L}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k+1}\} = \mathbf{L}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k+1}\}$ a $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k+1}\}$ má být ortonormální báze podprostoru $\mathbf{L}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k+1}\}$, musí mít vektor $\mathbf{x}_{k+1}$ Fourierův rozklad podle Tvrzení 7.5, tj.

$$\mathbf{x}_{k+1} = \sum_{j=1}^{k+1} < \mathbf{u}_j, \mathbf{x}_{k+1} > \mathbf{u}_j.$$

Musí být $< \mathbf{u}_{k+1}, \mathbf{x}_{k+1} > \neq 0$, v opačném případě by totiž platilo

$$\mathbf{x}_{k+1} \in \mathbf{L}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\} = \mathbf{L}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\},$$

což by bylo ve sporu s lineární nezávislostí množiny vektorů $\{\mathbf{x}_1, \ldots, \mathbf{x}_{k+1}\}$. Je tedy $< \mathbf{u}_{k+1}, \mathbf{x}_{k+1} > \neq 0$ a z rovnosti pro $\mathbf{x}_{k+1}$ můžeme tedy vypočítat

$$\mathbf{u}_{k+1} = \frac{\mathbf{x}_{k+1} - \sum_{j=1}^{k} < \mathbf{u}_j, \mathbf{x}_{k+1} > \mathbf{u}_j}{< \mathbf{u}_{k+1}, \mathbf{x}_{k+1} >}.$$

Dále má být $\|\mathbf{u}_{k+1}\| = 1$, tj. musí být

$$| < \mathbf{u}_{k+1}, \mathbf{x}_{k+1} > | = \|\mathbf{x}_{k+1} - \sum_{j=1}^{k} < \mathbf{u}_j, \mathbf{x}_{k+1} > \mathbf{u}_j\|.$$

Všimněme si, že uvedenou normu známe, neboť závisí pouze na vektorech $\mathbf{u}_1, \ldots, \mathbf{u}_k$, které všechny známe na základě indukčního předpokladu. Označíme tuto normu ν_{k+1}. Dostali jsme tedy nutnou podmínku pro vyjádření vektoru $\mathbf{u}_{k+1}$:

$$\mathbf{u}_{k+1} = \frac{\mathbf{x}_{k+1} - \sum_{j=1}^{k} < \mathbf{u}_j, \mathbf{x}_{k+1} > \mathbf{u}_j}{\nu_{k+1}}.$$

Skutečnost, že tato podmínka je také postačující, je obsahem následující věty.

Věta 8.1 *Buď $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_l\}$ lineárně nezávislá množina vektorů z $\mathbf{C}^n$ (nebo z $\mathbf{R}^n$). Položme*

$$\mathbf{u}_1 = \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}$$

a pro každé $k = 1, \ldots, l-1$

$$\mathbf{u}_{k+1} = \frac{\mathbf{x}_{k+1} - \sum_{j=1}^{k} < \mathbf{u}_j, \mathbf{x}_{k+1} > \mathbf{u}_j}{\nu_{k+1}},$$

kde $\nu_{k+1} = \|\mathbf{x}_{k+1} - \sum_{j=1}^{k} < \mathbf{u}_j, \mathbf{x}_{k+1} > \mathbf{u}_j\|$.

Potom je $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_l\}$ ortonormální množina a platí

$$\mathbf{L}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\} = \mathbf{L}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$$

pro každé $k = 1, \ldots, l$.

Důkaz. Zřejmě $\|\mathbf{u}_i\| = 1$ pro každé $i = 1, \ldots, l$. Indukcí podle k dokážeme, že pro každé $i < k$ platí $< \mathbf{u}_i, \mathbf{u}_k >= 0$. Pro $k = 1$ toto tvrzení platí, neboť žádný vektor $\mathbf{u}_i$ pro $i < 1$ neexistuje. Předpokládejme tedy, že pro nějaké $k \geq 1$ a $k < l$ platí $< \mathbf{u}_i, \mathbf{u}_k >= 0$ pro každé $i < k$. Zvolme libovolné $i < k + 1$. Pak platí

$$\begin{aligned} < \mathbf{u}_i, \mathbf{u}_{k+1} > &= < \mathbf{u}_i, \nu_{k+1}^{-1}(\mathbf{x}_{k+1} - \sum_{j=1}^{k} < \mathbf{u}_j, \mathbf{x}_{k+1} > \mathbf{u}_j) > \\ &= \nu_{k+1}^{-1}(< \mathbf{u}_i, \mathbf{x}_{k+1} > - \sum_{j=1}^{k} < \mathbf{u}_j, \mathbf{x}_{k+1} > \cdot < \mathbf{u}_i, \mathbf{u}_j >) \\ &= \nu_{k+1}^{-1}(< \mathbf{u}_i, \mathbf{x}_{k+1} > - < \mathbf{u}_i, \mathbf{x}_{k+1} >) = 0. \end{aligned}$$

Tím je dokázáno, že $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_l\}$ je ortonormální množina. Zbývá dokázat, že

$$\mathbf{L}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\} = \mathbf{L}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$$

pro každé $k = 1, \ldots, l$. Opět budeme postupovat indukcí podle k. Pro $k = 1$ tato rovnost zřejmě platí, neboť $\mathbf{u}_1$ je nenulovým skalárním násobkem $\mathbf{x}_1$. Předpokládejme, že pro nějaké $k \geq 1$ a $k < l$ platí $\mathbf{L}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\} = \mathbf{L}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$. Protože

$$\mathbf{u}_{k+1} = \frac{\mathbf{x}_{k+1} - \sum_{j=1}^{k} < \mathbf{u}_j, \mathbf{x}_{k+1} > \mathbf{u}_j}{\nu_{k+1}},$$

platí $\mathbf{u}_{k+1} \in \mathbf{L}\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{x}_{k+1}\} \subseteq= -\mathbf{L}\{\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{x}_{k+1}\}$ a tedy

$$\mathbf{L}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k+1}\} \subseteq \mathbf{L}\{\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{x}_{k+1}\}.$$

Opačnou inkluzi dokážeme podobně pomocí indučního předpokladu a vyjádření

$$\mathbf{x}_{k+1} = \nu_{k+1}\mathbf{u}_{k+1} + \sum_{j=1}^{k} < \mathbf{u}_j, \mathbf{x}_{k+1} > \mathbf{u}_j.$$

□

Konstrukce množiny $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_l\}$ vede k následujícímu *klasickému Gram-Schmidtovu algoritmu.*

Algoritmus 8.1

input $\mathbf{x}_1, \ldots, \mathbf{x}_l,$
output $\mathbf{u}_1, \ldots, \mathbf{u}_l,$

$\mathbf{u}_1 \leftarrow \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}$,

for $k = 2, \ldots, l$,

$\mathbf{u}_k \leftarrow \mathbf{x}_k - \sum_{j=1}^{k-1} < \mathbf{u}_j, \mathbf{x}_k > \mathbf{u}_j$,

$\mathbf{u}_k \leftarrow \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|}$.

Následující *věta o QR-rozkladu* je maticovou formulací Gram-Schmidtova algoritmu.

Věta 8.2 *Buď $\mathbf{A} = (a_{ij})$ matice typu $m \times n$ nad $\mathbf{C}$ (nebo nad $\mathbf{R}$) s lineárně nezávislými sloupci. Pak existuje matice $\mathbf{Q}$ typu $m \times n$ s ortonormálními sloupci a horní trojúhelníková matice $\mathbf{R}$ řádu n s kladnými prvky na hlavní diagonále takové, že platí $\mathbf{A} = \mathbf{QR}$.*

Důkaz. Označme $\mathbf{a}_j = \mathbf{A}_{*j}$ sloupcové vektory matice $\mathbf{A}$. Potom $\mathbf{a}_1, \ldots, \mathbf{a}_n$ je lineárně nezávislá množina vektorů z $\mathbf{C}^m$ (nebo z $\mathbf{R}^m$). Podle Věty 8.1 tvoří vektory $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ definované jako

$$\mathbf{q}_1 = \frac{\mathbf{a}_1}{\nu_1}, \quad \nu_1 = \|\mathbf{a}_1\|$$

a

$$\mathbf{q}_k = \frac{\mathbf{a}_k - \sum_{i=1}^{k-1} < \mathbf{q}_i, \mathbf{a}_k > \mathbf{q}_i}{\nu_k}, \quad \nu_k = \|\mathbf{a}_k - \sum_{i=1}^{k-1} < \mathbf{q}_i, \mathbf{a}_k > \mathbf{q}_i\|,$$

pro $k = 2, \ldots, n$, ortonormální množinu v $\mathbf{C}^m$ ($\mathbf{R}^m$).

Vzorce definující $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ přepíšeme ve tvaru

$$\begin{aligned} \mathbf{a}_1 &= \nu_1 \mathbf{q}_1, \\ \mathbf{a}_k &= \nu_k \mathbf{q}_k + < \mathbf{q}_1, \mathbf{a}_k > \mathbf{q}_1 + \cdots + < \mathbf{q}_{k-1}, \mathbf{a}_k > \mathbf{q}_{k-1} \end{aligned}$$

pro $k = 2, \ldots, n$.

Označíme-li $\mathbf{Q} = (\mathbf{q}_1 | \mathbf{q}_2 | \cdots | \mathbf{q}_n)$ a

$$\mathbf{R} = \begin{pmatrix} \nu_1 & < \mathbf{q}_1, \mathbf{a}_2 > & < \mathbf{q}_1, \mathbf{a}_3 > & \cdots & < \mathbf{q}_1, \mathbf{a}_n > \\ 0 & \nu_2 & < \mathbf{q}_2, \mathbf{a}_3 > & \cdots & < \mathbf{q}_2, \mathbf{a}_n > \\ 0 & 0 & \nu_3 & \cdots & < \mathbf{q}_3, \mathbf{a}_n > \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & < \mathbf{q}_{k-1}, \mathbf{a}_n > \\ 0 & 0 & 0 & \cdots & \nu_n \end{pmatrix},$$

pak poslední vztahy vyjadřující $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ znamenají, že $\mathbf{A} = \mathbf{QR}$. □

Klasický Gram-Schmidtův algoritmus není numericky stabilní. Jeho stabilitu lze vylepšit (i když ne úplně zaručit) pomocí jiného uspořádání jednotlivých algebraických operací při výpočtu ortonormální množiny $\{\mathbf{u}_1, \ldots, \mathbf{u}_l\}$. Numericky stabilní procesy ortogonalizace si ukážeme v příští kapitole.

Formule

$$\mathbf{u}_1 = \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}$$

$$\mathbf{u}_k = \frac{\mathbf{x}_k - \sum_{j=1}^{k-1} < \mathbf{u}_j, \mathbf{x}_k > \mathbf{u}_j}{\|\mathbf{x}_{k+1} - \sum_{j=1}^{k-1} < \mathbf{u}_j, \mathbf{x}_k > \mathbf{u}_j\|},$$

pro každé $k = 2, \ldots, l$, si přepíšeme následovně.

Označíme $\mathbf{U}_1 = \mathbf{0}_{n\times 1}$ nulový vektor dimenze n a $\mathbf{U}_k = (\mathbf{u}_1|\mathbf{u}_2|\cdots|\mathbf{u}_{k-1})$ matici typu $n \times (k-1)$ pro $k = 2, \ldots, l$. Potom platí pro každé $k > 1$

$$\mathbf{U}_k^*\mathbf{x}_k = \begin{pmatrix} \mathbf{u}_1^*\mathbf{x}_k \\ \mathbf{u}_2^*\mathbf{x}_k \\ \vdots \\ \mathbf{u}_{k-1}^*\mathbf{x}_k \end{pmatrix} = \begin{pmatrix} < \mathbf{u}_1, \mathbf{x}_k > \\ < \mathbf{u}_2, \mathbf{x}_k > \\ \vdots \\ < \mathbf{u}_{k-1}, \mathbf{x}_k > \end{pmatrix}$$

a tedy

$$\mathbf{U}_k\mathbf{U}_k^*\mathbf{x}_k = \sum_{j=1}^{k-1} < \mathbf{u}_j, \mathbf{x}_k > \mathbf{u}_j,$$

a

$$\mathbf{x}_k - \sum_{j=1}^{k-1} < \mathbf{u}_j, \mathbf{x}_k > \mathbf{u}_j = \mathbf{x}_k - \mathbf{U}_k\mathbf{U}_k^*\mathbf{x}_k = (\mathbf{I}_n - \mathbf{U}_k\mathbf{U}_k^*)\mathbf{x}_k.$$

Při započtení toho, že $\mathbf{U}_1 = \mathbf{0}_{n\times 1}$, platí poslední rovnost pro každé $k \geq 1$. To znamená, že

$$\mathbf{u}_k = \frac{(\mathbf{I}_n - \mathbf{U}_k\mathbf{U}_k^*)\mathbf{x}_k}{\|(\mathbf{I}_n - \mathbf{U}_k\mathbf{U}_k^*)\mathbf{x}_k\|}$$

pro každé $k = 1, \ldots, l$.

Matici $\mathbf{I}_n - \mathbf{U}_k\mathbf{U}_k^*$ vyjádříme jako součin jednodušších matic. Označíme $\mathbf{E}_1 = \mathbf{I}_n$ a $\mathbf{E}_i = \mathbf{I}_n - \mathbf{u}_{i-1}\mathbf{u}_{-1}^*$ pro $i > 1$. Potom platí

$$\begin{aligned} \mathbf{E}_2\mathbf{E}_1 &= \mathbf{I}_n - \mathbf{u}_1\mathbf{u}_1^* \\ \mathbf{E}_3\mathbf{E}_2\mathbf{E}_1 &= \mathbf{I}_n - \mathbf{u}_1\mathbf{u}_1^* - \mathbf{u}_2\mathbf{u}_2^* \\ &\vdots \\ \mathbf{E}_k\cdots\mathbf{E}_3\mathbf{E}_2\mathbf{E}_1 &= \mathbf{I}_n - \mathbf{u}_1\mathbf{u}_1^* - \mathbf{u}_2\mathbf{u}_2^* - \cdots - \mathbf{u}_{k-1}\mathbf{u}_{k-1}^*, \end{aligned}$$

při výpočtu jsem použili ortogonalitu vektorů $\mathbf{u}_i$.

Vektory $\mathbf{u}_i$ získané klasickým Gram-Schmidtovým algoritmem tak můžeme vyjádřit ve tvaru

$$\mathbf{u}_i = \frac{\mathbf{E}_k \cdots \mathbf{E}_2 \mathbf{E}_1 \mathbf{x}_k}{\|\mathbf{E}_k \cdots \mathbf{E}_2 \mathbf{E}_1 \mathbf{x}_k\|}$$

pro $k = 1, 2, \ldots, l$.

Uvedené vyjádření vektorů $\mathbf{u}_i$ vede k následujícímu *modifikovanému Gram-Schmidtovu algoritmu.*

Algoritmus 8.2

input $\mathbf{x}_1, \ldots, \mathbf{x}_l$,
output $\mathbf{u}_1, \ldots, \mathbf{u}_l$,

$\mathbf{u}_1 \leftarrow \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}$,

for $k = 2, \ldots, l$,

for $j = k, \ldots, l$,

$\mathbf{u}_j \leftarrow \mathbf{E}_k \mathbf{u}_j$,

$\mathbf{u}_k \leftarrow \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|}$.

Důkaz posledního tvrzení této kapitoly vyplývá z výpočtů před formulací modifikovaného Gram-Schmidtova algoritmu.

Tvrzení 8.3 *Modifikovaný Gram-Schmidtův algoritmus vede ke stejné ortonormální posloupnosti* $\mathbf{u}_1, \ldots, \mathbf{u}_l$ *jako klasický Gram-Schmidtův algoritmus.*

Ani modifikovaný Gram-Schmidtův algoritmus není numericky stabilní pro všechny typy úloh. Nicméně je o něco stabilnější než klasický algoritmus. Můžete si to ověřit v násleující úloze.

Úloha 8.1

Použijte aritmetiku s plovoucí čárkou, která zaokrouhluje na tři platná místa a obě varianty Gram-Schmidtova algoritmu k ortogonalizaci posloupnosti vektorů $\mathbf{x}_1 = (1, 10^{-3}, 10^{-3})^T$, $\mathbf{x}_2 = (1, 10^{-3}, 0)^T$ a $\mathbf{x}_3 = (1, 0, 10^{-3})^T$. Zatímco modifikovaný Gram-Schmidtův algoritmus dává tři vektory, které jsou tak ortonormální, jak jen to je v aritmetice zaokrouhlující na tři platná místa možné, u klasického Gram-Schmidtova algoritmu druhý a třetí vektor nejsou příliš kolmé.