Cvičení z optimalizace
Metoda nejmenších čtverců 2

Vojtěch Franc, 2009

1 Úvod

V první časti tohoto cvičení si ukážeme, jak lze použít metodu nejmenších čtverců pro identifikaci (odhad) parametrů autoregresního modelu a jeho použití pro kompresi zvukového signálu. Druhá čast cvičení je věnovaná algoritmu řešení problému nejmenších čtverců pomocí QR rozkladu.

2 Identifikace parametrů autoregresního modelů

Mějme zadanou posloupnost reálných čísel $y_0, y_1, \ldots, y_T$. Posloupnost může například popisovat signál digitalizovaný v diskrétních časových okamžicích $t = 0, 1, \ldots, T$. Naším cílem bude nálezt autoregresní model této posloupnosti, který předpokládá, že

$$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t \;, \tag{1}$$

tj. t-tý prvek posloupnosti je lineární kombinací p předcházejících prvků a náhodné odchylky (šumu) ε_t. Číslo $p \in \mathcal{N}^+$ se nazývá řádem modelu. Autoregresní model p-tého řádu je určen vektorem parametrů $\boldsymbol{a} = (a_0, a_1, \ldots, a_p)^T \in \mathbb{R}^{p+1}$. Odhadování těchto parametrů z dat, tj. ze zadané posloupnosti $y_0, y_1, \ldots, y_T$, se v teorii řízení nazývá identifikací systému. Jedna z metod identifikace je založena na hledání takových parametrů, které zajistí, že model co nejlépe odpovídá zadaným datům ve smyslu minimalizace součtu kvadrátu odchylek ε_t pro $t = p, \ldots, T$. To znamená, že hledáme vektor parametrů $\hat{\boldsymbol{a}}$, který minimalizuje funkci

$$F(\boldsymbol{a}) = \sum_{t=p}^{T} \varepsilon_t^2 = \sum_{t=p}^{T} \left(\left(a_0 + \sum_{i=1}^{p} a_j y_{t-i} \right) - y_t \right)^2 .$$

Řešení tohoto problému vede na řešení soustavy lineárních rovnic $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ metodou nejmenších čtverců, kde

$$\mathbf{A} = \begin{pmatrix} 1 & y_{p-1} & y_{p-2} & \cdots & y_0 \\ 1 & y_p & y_{p-1} & \cdots & y_1 \\ 1 & y_{p+1} & y_p & \cdots & y_2 \\ & & \vdots & & \\ 1 & y_{T-1} & y_{T-2} & \cdots & y_{T-p} \end{pmatrix} , \quad \boldsymbol{x} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} , \quad \boldsymbol{b} = \begin{pmatrix} y_p \\ y_{p+1} \\ y_{p+1} \\ \vdots \\ y_T \end{pmatrix} .$$

Takto získaný model lze použít například pro kompresi posloupnosti $y_0, y_1, \ldots, y_T$. V tomto případě postačí uchovat podsekvenci $y_0, \ldots, y_{p-1}$ a odhadnuté parametry $\hat{\boldsymbol{a}}$. Zbývající členy poslopnosti $y_p, \ldots, y_T$ vygenerujeme rekurzivním použitím (1). Koeficient komprese je tedy $T/(2p+1)$. Podobné myšlenky se využívá při kompresi řečového signálu v reálném čase, která se používá např. v mobilních telefonech.

Úkoly k vypracování

1. Stáhněte si WAV soubor se zvukem gongu `http://cmp.felk.cvut.cz/cmp/courses/OPT/cviceni/07/gong.wav` a nahrajte si ho do Matlabu pomocí příkazu

   ```
   [y,Fs]=wavread('gong.wav');
   ```

   `y` je vektor obsahující hodnoty zvukové sekvence vzorkované s frekvencí `Fs`. Pokud máte počítač se zvukovou kartou, můžete si signál přehrát pomocí příkazu `sound(y,Fs)`.

2. Pro nahranou sekvenci `y` odhadněte autoregresní model pomocí metody nejmenších čtverců. Použijte řád modelu $p = 300$.

3. Použijte odhadnutý model k vygenerování syntetického zvuku gongu. Porovnejte synteticky vygenerovaný a originální zvuk gongu graficky, tj. zobrazte oba průběhy do jednoho grafu pomocí funkce `plot`. Pokud to Váš počítač umožňuje, přehrajte si oba zvukové signály pomocí příkazu `sound(y,Fs)`.

Bonusová úloha Čemu se rovná minimum funkce $F(\boldsymbol{x})$ pokud platí hodnost$(\mathbf{A}) =$ hodnost$([\mathbf{A}\ \boldsymbol{b}])$? Vaši odpověď zdůvodněte.

3 Řešení problému nejmenších čtverců pomocí QR rozkladu

Nechť $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, $m > n$, jsou zadané matice a vektor. Řešení soustavy $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ ve smyslu nejmenších čtverců vede na hledání vektoru $\hat{\boldsymbol{x}}$, který minimalizuje funkci

$$F(\boldsymbol{x}) = \|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|^2 .$$

Existuje několik způsobů jak minimalizovat $F(\boldsymbol{x})$. My se zaměříme na přístup, který se opírá o QR rozklad matice $\mathbf{A}$. QR rozkladem matice $\mathbf{A} \in \mathbb{R}^{m\times n}$, $m > n$, získáme ortonormální matici [1] $\mathbf{Q} \in \mathbb{R}^{m\times m}$ a horní trojúhelníkovou matici [2] $\mathbf{R} \in \mathbb{R}^{n\times n}$ takové, že

$$\mathbf{A} = \mathbf{Q}\begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix} ,$$

kde $\begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix}$ je matice $\mathbb{R}^{m\times n}$, která vznikla vertikálním spojením $\mathbf{R}$ a matice $\mathbf{0} \in \mathbb{R}^{(m-n)\times n}$ se všemi prvky rovnými nule. Rozdělme $\mathbf{Q} = [\mathbf{Q_1}\ \mathbf{Q_2}]$ na submatice $\mathbf{Q_1} \in \mathbb{R}^{m\times n}$ a $\mathbf{Q_2} \in \mathbb{R}^{m\times(m-n)}$. Použitím zavedených matic lze kriterium $F(\boldsymbol{x})$ přepsat do ekvivalentního tvaru

$$\begin{aligned}F(\boldsymbol{x}) &= \|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = \left\|\mathbf{Q}^T\big(\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\big)\right\|^2 = \left\|\mathbf{Q}^T\mathbf{Q}\begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix}\boldsymbol{x} - \mathbf{Q}^T\boldsymbol{b}\right\|^2 \\ &= \left\|\begin{bmatrix}\mathbf{R}\boldsymbol{x}\\ \mathbf{0}\end{bmatrix} - \begin{bmatrix}\mathbf{Q_1}^T\\ \mathbf{Q_2}^T\end{bmatrix}\boldsymbol{b}\right\|^2 = \left\|\begin{bmatrix}\mathbf{R}\boldsymbol{x} - \mathbf{Q_1}^T\boldsymbol{b}\\ \mathbf{Q_2}^T\boldsymbol{b}\end{bmatrix}\right\|^2 ,\end{aligned} \tag{2}$$

kde jsme k odvození druhé rovnosti použili $\|\mathbf{Q}^T\boldsymbol{z}\| = \|\boldsymbol{z}\|$, $\forall \boldsymbol{z} \in \mathbb{R}^m$. Z poslední rovnice v (2) je vidět, že vektor $\hat{x}$, jenž řeší soustavu

$$\mathbf{R}\boldsymbol{x} = \mathbf{Q_1}^T\boldsymbol{b} , \tag{3}$$

[1] Pro ortonormální matici $\mathbf{Q}$ platí $\mathbf{Q}^T\mathbf{Q} = \mathbf{E}$.

[2] Pro horní trojúhelníkovou matici $\mathbf{R}$ platí $R_{i,j} = 0$ pokud $i > j$.

je současně řešením problému nejmenších čtverců, tj. minimalizuje $F(\boldsymbol{x})$. Soustavu (3) lze vyřešit velmi snadno díky tomu, že matice $\mathbf{R}$ je horní trojúhelníková matice.

Úkoly k vypracování

1. Implementujte algoritmus pro řešení problému nejmenších čtverců pomocí QR rozkladu. Vstupem algoritmu nechť je dvojice $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, $m > n$. Výstupem algorimu bude vektor $\hat{\boldsymbol{x}}$ minimalizující $F(\boldsymbol{x})$.

   Matice $\mathbf{Q_1}$ a $\mathbf{R}$ získáte v Matlabu pomocí příkazu

   ```
   [Q1,R] = qr(A,0);
   ```

   Pro řešení soustavy (3) naimplementujte vlastní algoritmus (tj. není povoleno použít \ a podobně).

2. Vámi implementovaný algoritmus použijte pro řešení problému nejmenších čtverců z předchozího příkladu (odstavec 2). Porovnejte výsledek s řešením získaným pomocí `x=A\b`.

Bonusová úloha Dokažte, že platí $\|\mathbf{Q}^T\boldsymbol{z}\| = \|\boldsymbol{z}\|$, $\forall \boldsymbol{z} \in \mathbb{R}^m$ pokud $\mathbf{Q} \in \mathbb{R}^{m\times m}$ je ortonormální matice.