Odhad skrytých periodicit metodami autoregresního modelování

Petr Tichavský, Ústav teorie informací a automatizace,
Pod vodárenskou věží 4, 182 00 Praha 8

Odhad parametrů sinusoidálních signálů z dat zatížených aditivním šumem není úloha nová a kdysi patrně motivovala celé odvětví spektrální analýzy časových řad.

Předpokládáme model dat ve tvaru

$$x(t) = \sum_{i=1}^{M} \alpha_i \sin (w_i t + \Phi_i) \qquad t=1,\dots,N, \tag{1}$$

kde α_i, w_i, Φ_i, $i=1,\dots,M$, jsou neznámé amplitudy, frekvence a fáze.
Pozorovaná realizace časové řady nechť se od modelu (1) liší o aditivní /náhodný/ šum

$$\hat{x}(t) = x(t) + e(t), \qquad t=1,\dots,N. \tag{2}$$

O náhodných veličinách $e(t)$ zpravidla předpokládáme, že jsou vzájemně nezávislé, mají nulovou střední hodnotu a konečný rozptyl σ^2, případně navíc, že mají normální rozložení s těmito parametry.

Tradičně používaným prostředkem pro odhad parametrů modelu (1) je Schusterův periodogram, doplněný o Fisherův test významnosti jednotlivých složek. Dá se ovšem ukázat, viz [7], že v některých případech, zvláště při malé délce řady N a vyšším podílu signálu vůči šumu $\alpha^2/2\sigma^2$, nebo při výskytu sinusoid s blízkými frekvencemi $|w_i - w_j| \le \pi/N$, je klasický postup málo efektivní ve srovnání s metodami založenými na autoregresnim modelování.

Dříve, než přistoupíme k jejich výkladu, si povšimneme skutečnosti, že model (1) závisí na parametrech $\alpha_i \sin \Phi_i$, $\alpha_i \cos \Phi_i$, $i = ,\dots,M$, lineárně.
Tento fakt umožňuje odhad amplitud a fází modelu (1) při pevných hodnotách frekvencí metodou nejmenších čtverců pomocí vzorce v uzavřeném tvaru. Nelinearita dané regresní úlohy tedy spočívá pouze v nelinearitě modelu vzhledem k parametrům frekvencí.
Metody, založené na autoregresním modelování, využívají tvrzení následující věty :

Posloupnost $\{x(t)\}$ tvaru (1) vyhovuje následující homogenní diferenční rovnici :

$$x(t) + a_1\, x(t-1) + \dots + a_{2M}\, x(t-2M) = 0, \qquad t=2M+1,\dots,N. \tag{3}$$

Přitom reálná čísla $a_1, \dots, a_{2M}$ mají tu vlastnost, že polynom

$$A(z) = z^{2M} + a_1 z^{2M-1} + \dots + a_{2M-1} z + a_{2M} \tag{4}$$

má za své kořeny komplexní jednotky $\exp(\pm j w_k)$, $k = 1, \dots, M$.

Diferenční rovnice (3) platí i v případě, že model $\{x(t)\}$ je součtem M exponenciálně tlumených sinusoid s tím rozdílem, že kořeny polynomu $A(z)$ už nemají jednotkové moduly.

Odhad frekvencí pak probíhá ve dvou krocích:

1) odhad koeficientů rekurentní diferenční rovnice (3)
2) nalezení kořenů polynomu $A(z)$ a jejich argumentů.

Odhad koeficientů $a_1,\dots,a_{2M}$ se dá provádět různými způsoby. Například Pronyho metoda používá tzv. kovariančních rovnic, jak je obvyklé při autoregresním modelování. Při formulaci úlohy pomocí kovarianční funkce procesu řešíme Yule-Walkerovy rovnice. Časté je též využití jisté symetrie koeficientů, nebo též rovnic pro tzv. zpětnou predikci, viz. [2]. Ke konzistentnímu odhadu frekvencí vede Pisarenkova metoda [3], při které se za odhad vektoru $(1,a_1,\dots,a_{2M})$ bere vlastní vektor výběrové kovarianční matice daného procesu řádu 2M+1, který odpovídá jejímu nejmenšímu vlastnímu číslu.

Podstatné zlepšení statistických vlastností citovaných odhadů přinesla myšlenka Tuftse a Kumaresana v článku [4] použít autoregresního modelu s vyšším řádem než 2M. Jejich metodu dále modifikovali Rahman a Yu [5]. Jejich metoda se opírá o následující zobecnění předchozí věty:

Nechť $N-2M \geq L \geq 2M$ a vektor $a_{(L+1)} = (1,a_1^{(L)},\dots,a_L^{(L)})$ vyhovuje soustavě

$$X_{(L+1)}\, a_{(L+1)} = 0, \tag{5}$$

kde

$$X_{(L+1)} = \begin{pmatrix} x(L+1) & x(L) & \dots & x(1) \\ x(L+2) & x(L+1) & \dots & x(2) \\ \vdots & & & \\ x(N) & x(N-1) & \dots & x(N-L) \end{pmatrix} \tag{6}$$

a posloupnost $\{x(t)\}$ je dána vztahem (1). Pak čísla $\exp(\pm jw_i)$, $i = 1,\dots,M$, jsou kořeny polynomu

$$A^{(L)}(z) = \sum_{i=0}^{L} a_i^{(L)} z^{L-i}. \tag{7}$$

Vhodný odhad vektoru $a_{(L+1)}$, který už nyní není určen jednoznačně, se se provádí pomocí singulárního rozkladu příslušné matice $\hat{X}_{(L+1)}$, nebo vlastního rozkladu výběrové kovarianční matice

$$\hat{R}_{(L+1)} = \frac{1}{N-L}\, \hat{X}^T_{(L+1)} \hat{X}_{(L+1)}. \tag{8}$$

Doporučuje se odhad

$$\hat{a}_{(L+1)} = \left(\sum_{k=2M+1}^{L+1} (\hat{v}_k)_1\, \hat{v}_k \right) \Big/ \left(\sum_{i=2M+1}^{L+1} |(\hat{v}_i)_1|^2 \right), \tag{9}$$

kde $\hat{v}_{2M+1},\dots,\hat{v}_{L+1}$ jsou vlastní vektory matice $\hat{R}_{(L+1)}$, odpovídající nejmenším L+1-2M vlastním číslům a $(\hat{v}_k)_1$ je první složka vektoru $\hat{v}_k$, $k = 2M+1,\dots,L+1$.

Odhad počtu složek M v modelu (1) je možné provádět na základě určitého testování rovnost těchto nejmenších vlastních čísel , viz [6].

Z kořenů odhadnutého polynomu $\hat{A}^{(L)}(z)$, kterých máme při L>2M pro odhad frekvencí nadbytek, vybíráme ty, které v komplexní rovině leží nejblíže jednotkové kružnici. Alternativní možností je vyčíslení polynomu $\hat{A}^{(L)}(z)$ na jednotkové kružnici (s výhodou lze použít Fourierovy transformace) a za odhad frekvencí vzít ty body, ve kterých $|\hat{A}^{(L)}(z)|^2$ nabývá nejmenších lokálních minim.

LITERATURA

[1] G.R.B. Prony "Essai experimental et analytique", Paris, Journal de L' Ecole Polytechnique, vol. 1, cahier 2, pp. 24-76, 1795.

[2] S.M. Kay, S.L. Marple, "Spectrum analysis - a modern perspective" Proceedings of the IEEE, vol. 69, No. 11, pp. 1380 - 1409, November 1981.

[3] V.F. Pisarenko, "The retrieval of harmonics from a covariance function", Geophysical J. Royal Astronomical Soc., vol. 33, pp. 347 - 366, 1973

[4] D.W. Tufts, R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood", Proceedings of the IEEE, vol. 70, No. 9, pp. 975 - 989, September 1982.

[5] A. Rahman, K. Yu, "Total least squares approach for frequency estimation using linear prediction", IEEE Transaction on Acoust., Speech and Signal Processing, vol. ASSP - 35, No. 9, pp. 1440 - 1454, October 1987.

[6] J.J. Fuchs, "Estimating the number of sinusoids in additive white noise", IEEE Trans. on Acoust. Speech, Signal Processing, vol. 36, No. 12, pp. 1846 - 1853, December 1988.

[7] P. Tichavský, "Odhad skrytých periodicit," výzkumná zpráva ÚTIA č.1601, říjen 1989

[8] P. Tichavský, "Programový systém KECAL pro odhad skrytých periodicit", výzkumná zpráva ÚTIA č. 1637, prosinec 1989.