АЛГОРИТМ АДАПТИВНОГО СТИСНЕННЯ ЧАСОВИХ РЯДІВ ДАНИХ

Сільвестров А.М., професор, Зіменков Д.К., старший викладач, Вещиков Г.В., магістрант
КПІ ім. Ігоря Сікорського, кафедра теоретичної електротехніки

Вступ. Сучасні інформаційні системи комп'ютерного моніторингу енергетичних, екологічних, економічних, технологічних та інших змінних у часі величин широко використовують первинні перетворювачі фізичних величин в електричні, електричні в цифрові. Останні поступають у відповідні цифрові бази даних інформаційних систем. З метою не втрати якихось поточних даних частота дискретизації зашумлених вимірювальних величин може бути набагато вищою від вимог теореми Котельникова, щодо корисної складової сигналу. Це призводить до суттєвого зростання об'єму цифрових даних і переповнення їх баз.

Постановка задачі. У зв'язку з цим актуальною є задача стиснення дискретної у часі інформації без трати її корисної складової. Отже, необхідно розробити алгоритм адаптивний як до співвідношення «шум-корисний сигнал» та і до темпу зміни корисної інформації.

Матеріали і результати досліджень. *Апріорна інформація*: записи часового (чи просторового) ряду $x(\kappa)$, $k = 0,1,..,N$.

Крок 1. Робастне згладжування методом Тьюкі [1] по семи точкам з ліквідацією одиночних (подвійних) можливих аномальних даних:

$$\hat{x}(\kappa) = med\left\{x(\kappa \pm e)\right\},\ e = 0,1,2,3, \tag{1}$$

де med - медіана (середнє значення між ранжованими за величиною в семи поточних точках).

Крок 2. Оцінювання середньоквадратичної похибки ряду:

$$\sigma_{\hat{x}} = \sqrt{\frac{1}{N}\sum_{\kappa=0}^{N}\left(x(\kappa) - \hat{x}(\kappa)\right)^2} \tag{2}$$

Крок 3. Апроксимація регуляризованим методом найменших квадратів (МНК) перших $8 \div 12$ точок ряду $\hat{x}(\kappa)$ степеневим поліномом:

$$x_{м}(k) = \hat{x}(0) + \sum_{i=1}^{n} \alpha_i \cdot k^i, \tag{3}$$

де, за відомого $\hat{x}(0)$, визначаються n коефіцієнтів α_i полінома n-ї степені за умови мінімуму регуляризованого [2] функціонала:

$$J = \frac{1}{2}\sum_{\kappa=0}^{N_1}\left\{\left(\hat{x}(\kappa) - x_{м}(\kappa)\right)^2 + \frac{1}{2}\gamma\sum_{i=1}^{n}\alpha_i^2\right. . \tag{4}$$

Тут γ – регалізуючий за Тихоновим параметр: якщо $\gamma = 0$, то буде звичайний МНК; якщо $\gamma \to \infty$, то всі $\alpha_i \to 0$. Беремо γ в межах $0{,}01 \div 0{,}1$. Завдяки тому, що $\gamma > 0$, інформаційна матриця МНК буде позитивно обумовлена (її визначник завжди більше нуля) і розкид оцінок α_i буде не великий.

Крок 4. Отримавши для першого інтервалу ($k = 0,1,..,N_1$), ($N_1 = 8 \div 12$) де-що зміщені але стабільні оцінки параметрів α_i; далі в алгоритмі рекурентного МНК (РМНК) використовуємо їх, як апріорні. Аналогічно і матриці P_m, P_{m+1} в РМНК:

$$\alpha_{i+1} = \alpha_1 + P_{m+1} \cdot \varphi_{m+1} \left(\hat{x}_{m+1} - \varphi_{m+1}^{\mathrm{T}} \cdot \alpha_m \right) \tag{5}$$

де α - вектор α_i, $i = \overrightarrow{I,n}$;

$$P_m = \left[\Phi_m^{\mathrm{T}} \cdot \Phi_m \right]^{-1}; \; P_{m+1} = \left[\Phi_{m+1}^{\mathrm{T}} \cdot \Phi_{m+1} \right]^{-1}; \; P_{m+1} = \left[\mathrm{P}_m^{-1} + \varphi_{m+1} \cdot \varphi_{m+1}^{\mathrm{T}} \right]^{-1};$$

$$\Phi_m = \begin{bmatrix} \varphi_1(1) & \varphi_2(1) \dots \varphi_n(1) \\ \varphi_1(2) & \varphi_2(2) \dots \varphi_n(2) \\ \dots & \dots \\ \varphi_1(m) & \varphi_2(m) \dots \varphi_n(m) \end{bmatrix}; \; \Phi_{m+1} = \begin{bmatrix} \varphi_1(1) & \varphi_2(1) \dots \varphi_n(1) \\ \varphi_1(2) & \varphi_2(2) \dots \varphi_n(2) \\ \dots & \dots \\ \varphi_1(m+1) & \varphi_2(m+1) \dots \varphi_n(m+1) \end{bmatrix};$$

$\varphi_i(j)$ – це $k^i(j), j = 1, \dots, N; i = 1, \dots, n.$

Значення m рекуретно розширяються до величини $N_{\max}$ за якої максимальне значення похибки апроксимації ряду $\hat{x}(\kappa)$ моделлю $\hat{x}_м(\kappa)$ не сягне $3\sigma_{\hat{x}}$:

$$Sup \, |\hat{x}(k) - x_{\text{м}}(k)| \approx 3\sigma_{\hat{x}}. \tag{6}$$

Крок 5 і наступні повторюють кроки $1 \div 4$ для наступних інтервалів.

Таким чином, перебираючи степінь n полінома від 3 до 5, отримуємо модель (3) ряду і приймаємо за оптимальну ту, за якої буде найменша похибка апроксимації на заданому інтервалі ряду або допустима ($3\sigma_{\hat{x}}$) на максимально довгому інтервалі.

Приклад. Необхідно стиснути дані про вологість повітря, подані рядом в 100-120 відліків часу. За допустимої похибки (2) в 0,17 одиниць достатньо одного поліному 3-го порядку (рис.1)

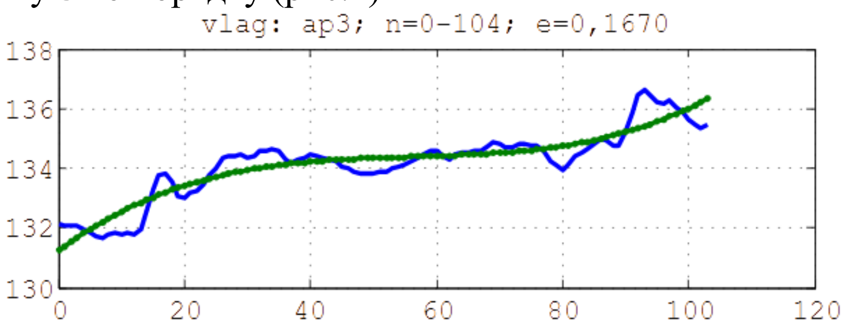

Рисунок 1 – Апроксимація одним поліномом третього порядку ($3\sigma_{\hat{x}} = 0{,}167$)

За похибки в 0,06 одиниць-двох поліномів 4-го порядку (рис.2) на інтервалах $[0,70],[70,100]$

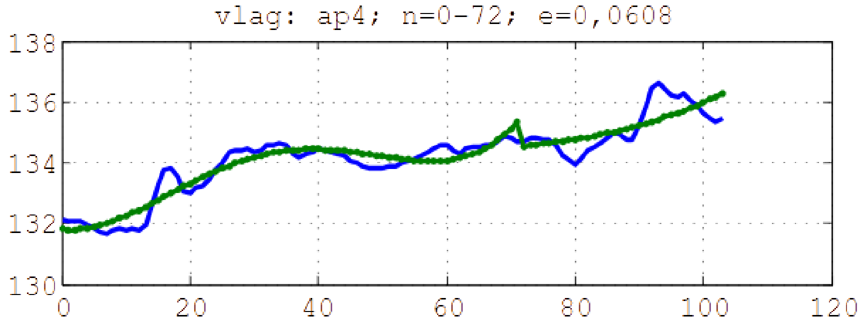

Рисунок 2 – Апроксимація двома поліномами четвертого порядку ($3\sigma_{\hat{x}} = 0{,}06$)

За похибки 0,03- двох поліномів 5-го порядку (рис.3) на інтервалах $[0,48]$, $[48,100]$

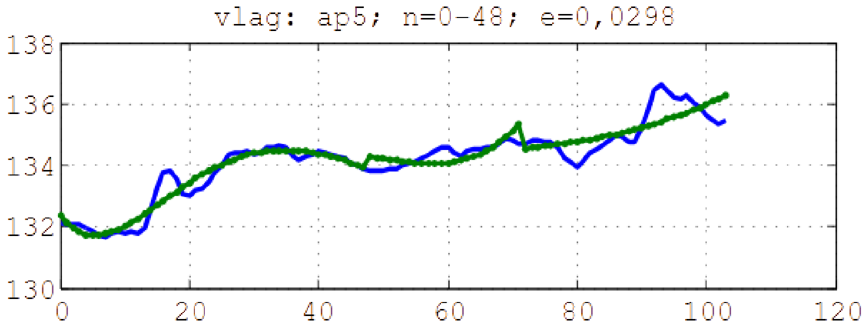

Рисунок 3 – Апроксимація двома поліномами п'ятого порядку ($3\sigma_{\hat{x}} = 0{,}03$)

Висновки. Таким чином адаптивна до $\sigma_{\hat{x}}$ структурно-параметрична апроксимація РМНК дозволяє значно (на 1 – 2 порядки) зменшити об'єм інформації без втрати корисної складової.

Перелік посилань

1. P.I. Bidyuk, V.D. Romanenko and O.L. Timoshchuk. Time Series Analysis/ Kyiv: Polytechnika, NTUU «KPI», 2013. – 280 с.

2. Островерхов М.Я. Методи дослідження електротехнічних комплексів і систем: монографія/М.Я. Островерхов, А.М. Сільвестров, К.Х. Зеленський//К: Талком, 2019. – 300 с.