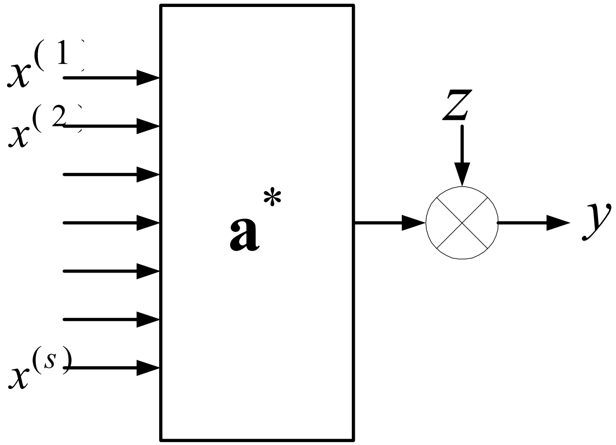
1. Przedmiot identyfikacji

System



$$y = F^*(\mathbf{x}, z) = F(\mathbf{x}, z, \mathbf{a}^*), \text{ gdzie } \mathbf{x} = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(s)} \end{bmatrix} \text{ – mierzone, } \mathbf{a}^* = \begin{bmatrix} a_1^* \\ a_2^* \\ \vdots \\ a_s^* \end{bmatrix} \text{ – zestaw współczynników konkretyzujących } F()$$

$F()$ – znane, $\mathbf{a}^*$ – nieznane

informacja aprioryczna

$$F^* \in \{F(\mathbf{x}, z, \mathbf{a}); \mathbf{a} \in R^s\}$$

istnieje dokładnie jedno $\mathbf{a}^*$, takie że $F^*(\mathbf{x}, z) = F(\mathbf{x}, z, \mathbf{a}^*)$ dla każdej pary $(\mathbf{x}, z)$ (**identyfikowalność**)

informacja pomiarowa

$$\{\mathbf{x}_k, y_k\}_{k=1}^N$$

cel

znaleźć przepis $\Psi()$ (algorytm identyfikacji)

$$\Psi(\text{inf . aprioryczna, inf. pomiarowa}) = \mathbf{a}_N \sim \mathbf{a}^*$$

założenia o zakłóceniach

$$\begin{aligned}\text{jest addytywne, tj. } y &= F(\mathbf{x}, \mathbf{a}^*) + z \\ & \quad \{z_k\}_{k=1}^N \text{ – ciąg i.i.d.} \\ \mathbf{E}z &= 0 \\ var z &< \infty \\ & \quad z, \mathbf{x} \text{ – niezależne}\end{aligned}$$

natura procesów

$$z \text{ – losowe, } \mathbf{x} \text{ – deterministyczne lub losowe, } y \text{ –losowe, } \mathbf{a}_N \text{ –losowe}$$

własności $\mathbf{a}_N$

– praktyczne: wartości $\mathbf{E}\mathbf{a}_N[i]$ i $var\mathbf{a}_N[i]$

– asymptotyczne: fakt i typ zbieżności $\mathbf{a}_N \to \mathbf{a}^*$

2. Metoda najmniejszych kwadratów (Gauss 1809)

wskaźnik teoretyczny oceniający jakość dopasowania

$$Q_J(\mathbf{a}) = \mathbf{E}\{y - F(\mathbf{x}, \mathbf{a})\}^2 \to \min_{\mathbf{a}}$$

$$\begin{aligned}Q_J(\mathbf{a}) &= \mathbf{E}\{F(\mathbf{x}, \mathbf{a}^*) + z - F(\mathbf{x}, \mathbf{a})\}^2 = \mathbf{E}\{[F(\mathbf{x}, \mathbf{a}^*) - F(\mathbf{x}, \mathbf{a})]^2 + 2[F(\mathbf{x}, \mathbf{a}^*) - F(\mathbf{x}, \mathbf{a})]z + z^2\} = \\ &= \mathbf{E}\{F(\mathbf{x}, \mathbf{a}^*) - F(\mathbf{x}, \mathbf{a})\}^2 + var z\end{aligned}$$

konkluzja

$$\mathbf{a}^* = \arg\min_{\mathbf{a}} Q_J(\mathbf{a})$$

w praktyce nie jest możliwe obliczenie $Q_J(\mathbf{a})$ (brak znajomości odpowiednich rozkładów), proponuje się estymację wartości oczekiwanej za pomocą wartości średniej

$$\widehat{Q}_{J,N}(\mathbf{a}) = \frac{1}{N}\sum_{k=1}^{N}\{y_k - F(\mathbf{x}_k, \mathbf{a})\}^2 \to \min_{\mathbf{a}}$$

uśredniane składniki $\{y_k - F(\mathbf{x}_k, \mathbf{a})\}^2$ tworzą ciąg i.i.d., zatem na podstawie *MPWL*

$$\widehat{Q}_{J,N}(\mathbf{a}) \stackrel{p.1}{\to} Q_J(\mathbf{a}), \text{ gdy } N \to \infty$$

pytanie: czy prawdziwe jest następujące wnioskowanie

$$\widehat{Q}_{J,N}(\mathbf{a}) \stackrel{p.1}{\to} Q_J(\mathbf{a}) \Longrightarrow \arg\min_{\mathbf{a}} \widehat{Q}_{J,N}(\mathbf{a}) \stackrel{p.1}{\to} \arg\min_{\mathbf{a}} Q_J(\mathbf{a})$$

odpowiedź: tylko wtedy, gdy

$$\sup_{\mathbf{a}}(\widehat{Q}_{J,N}(\mathbf{a}) - Q_J(\mathbf{a})) \stackrel{p.1}{\to} 0$$

3. Liniowy system statyczny *MIMO*

$$\begin{aligned} y &= \mathbf{a}^{*T}\mathbf{x} + z = \mathbf{x}^T\mathbf{a}^* + z \\ y_k &= \mathbf{x}_k^T\mathbf{a}^* + z_k, \qquad k = 1, 2, ..., N \end{aligned}$$

opis macierzowo–wektorowy

$$X_N = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix} = \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & .. & x_1^{(s)} \\ x_2^{(1)} & x_2^{(2)} & .. & x_2^{(s)} \\ .. & .. & .. & .. \\ x_N^{(1)} & x_N^{(2)} & .. & x_N^{(s)} \end{bmatrix}, \qquad Y_N = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \qquad Z_N = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix}$$

równanie pomiarów

$$Y_N = X_N\mathbf{a}^* + Z_N$$

wektor wyjść modelu

$$\overline{Y}_N = X_N\mathbf{a}$$

empiryczne kryterium jakości

$$\begin{aligned} \overline{Q}_{J,N}(\mathbf{a}) &= \sum_{k=1}^{N}\{y_k - \mathbf{x}_k^T\mathbf{a}\}^2 = (Y_N - X_N\mathbf{a})^T(Y_N - X_N\mathbf{a}) = \|Y_N - X_N\mathbf{a}\|_2^2 = ... \text{ – kwadrat normy euklidesowej} \\ ... &= (Y_N^T - \mathbf{a}^T X_N^T)(Y_N - X_N\mathbf{a}) = Y_N^T Y_N - \mathbf{a}^T X_N^T Y_N - Y_N^T X_N\mathbf{a} + \mathbf{a}^2 X_N^T X_N \to \min_{\mathbf{a}} \end{aligned}$$

gradient

$$\nabla_{\mathbf{a}}\overline{Q}_{J,N}(\mathbf{a}) = -2X_N^T Y_N + 2X_N^T X_N\mathbf{a} = 0$$

równanie normalne

$$X_N^T X_N\mathbf{a} = X_N^T Y_N$$

$\dim X_N^T X_N = s \times s$, $\dim X_N^T Y_N = s \times 1$, $\dim \mathbf{a} = s \times 1$

- kwestia isnienia rozwiązania

$$\mathrm{lin}col X_N^T X_N = \mathrm{lin}col X_N^T$$

- kwestia jednoznaczności rozwiązania

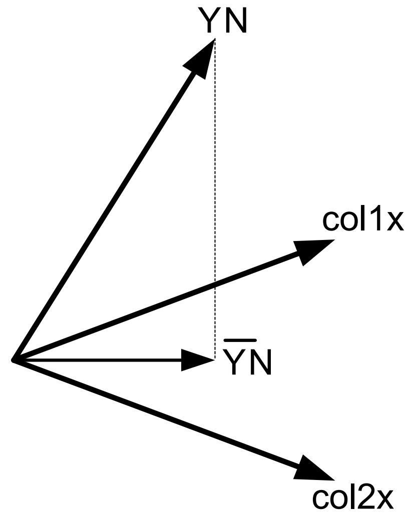
$$\begin{aligned} \det X_N^T X_N &> 0 \qquad (\neq 0) \\ \text{inaczej } rank X_N^T X_N &= s \text{ (nieosobliwa)} \end{aligned}$$

Fakt: $rank X_N^T X_N = rank X_N^T = rank X_N$.
Warunek jednoznaczności: $rank X_N = s$

1) $N \geqslant s$ – dostatecznie dużo pomiarów (warunek konieczny, ale nie wystarczający)
2) wśród N pomiarów musi się znaleźć s niezeleżnych liniowo wektorów
Gdy $rank X_N < s$, wtedy rozwiązanie równania normalnego nie jest jednoznaczne.

- kwestia nazwy równania (normalność-prostopadłość)

$$\begin{aligned}
X_N^T(Y_N - X_N\mathbf{a}) &= 0 \\
X_N^T(Y_N - \overline{Y}_N) &= 0, \text{ gdzie } \overline{Y}_N \text{ – wyjście modelu o parametrach } \mathbf{a} \\
X_N^T &= [col_1x, col_2x, ..., col_sx] \\
\overline{Y}_N &= a_1col_1x + a_2col_2x + ... + a_scol_sx \text{ – liniowa kombinacja wektorów kolumnowych} \\
\text{wniosek} &: \ \overline{Y}_N \in \text{lin}col X_N^T \\
\text{natomiast w ogólności} &: \ Y_N \notin \text{lin}col X_N^T
\end{aligned}$$



$$X_N^T(Y_N - \overline{Y}_N) = 0 \Rightarrow (col_xi)^T(Y_N - \overline{Y}_N) = 0 \text{ dla każdego } i = 1, 2, ..., s$$

wniosek: wektor różnicy $Y_N - \overline{Y}_N$ jest prostopadły do każdej z kolumn col_xi (rzut prostopadły jest "najkrótszy")

1. Identyfikacja liniowych systemów dynamicznych z czasem dyskretnym

$$x_k + \alpha_1 x_{k-1} + ... + \alpha_n x_{k-n} = \beta_0 u_k + \beta_1 u_{k-1} + ... + \beta_m u_{k-m}, \tag{1}$$

gdzie wartości rzędów n i m są znane.
Po wprowadzeniu operatora $q = z^{-1}$ przesuwającego wstecz (tzn. $qx_k = x_{k-1}$, $q^2 x_k = x_{k-2}$ itd.) oraz wielomianów

$$\begin{aligned} A(q) &= 1 + \alpha_1 q + \alpha_2 q^2 + ... + \alpha_n q^n, \\ B(q) &= \beta_0 + \beta_1 q + ... + \beta_m q^m, \end{aligned}$$

równanie (1) upraszcza się do postaci

$$A(q)x_k = B(q)u_k \qquad \rightarrow \qquad x_k = \frac{B(q)}{A(q)} u_k.$$

2. Identyfikacja w warunkach bez zakłóceń pomiarowych

$$\theta = (\alpha_1, \alpha_2, ..., \alpha_n, \beta_0, \beta_1, ..., \beta_m)^T$$

$$r_k = (-x_{k-1}, -x_{k-2}, ..., -x_{k-n}, u_k, u_{k-1}, ..., u_{k-m})^T,$$

$$x_k = r_k^T \theta.$$

Nieznane parametry zawarte w wektorze θ identyfikuje się na podstawie par pomiarów (u_k, x_k).
Zapiszmy

$$R_N = \begin{bmatrix} r_1^T \\ r_2^T \\ \vdots \\ \vdots \\ r_N^T \end{bmatrix}, \qquad X_N = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_N \end{bmatrix}.$$

$$X_N = R_N \theta, \tag{2}$$

$$R_N^T X_N = R_N^T R_N \theta, \qquad (\theta \text{ – niewiadome}), \tag{3}$$

warunek jednoznaczności rozwiązania

$$\det R_N^T R_N \neq 0$$

$$\text{rank} R_N = \dim \theta.$$

warunek konieczny

$$N \geq \dim \theta. \tag{4}$$

$$\theta = (R_N^T R_N)^{-1} R_N^T X_N.$$

3. Identyfikacja w obecności zakłóceń pomiarowych

Cel: estymacja wektora parametrów $\theta = (\alpha_1, \alpha_2, ..., \alpha_n, \beta_0, \beta_1, ..., \beta_m)^T$ na podstawie pomiarów we–wy $\{(u_k, y_k)\}_{k=1}^N$, gdzie $y_k = x_k + \varepsilon_k$. Do prawdziwego (niedostępnego dla pomiarów) wyjścia obiektu x_k dodają się przypadowe zakłócenia ε_k

- o zerowej wartości oczekiwanej $\mathbf{E}\varepsilon_k = 0$,
- skończonej wariancji $var\varepsilon_k < \infty$,
- niezależne od procesu wejściowego u_k

Zapisujemy

$$x_k = y_k - \varepsilon_k = \frac{B(q)}{A(q)} u_k \qquad / \cdot A(q)$$

i otrzymujemy równanie różnicowe o zmiennych y i u

$$A(q)y_k = B(q)u_k + z_k.$$

$$z_k = A(q)\varepsilon_k = \varepsilon_k + \alpha_1\varepsilon_{k-1} + \alpha_2\varepsilon_{k-2} + ... + \alpha_n\varepsilon_{k-n},$$

z_k nie jest szumem białym, gdyż jego autokorelacja dla $|\tau| \leq n$ jest niezerowa

$$\mathbf{E}z_k z_{k-\tau} \neq 0.$$

Po wprowadzeniu rzeczywistego regresora (uogólnionego wektora wejść)

$$\phi_k = (-y_{k-1}, -y_{k-2}, ..., -y_{k-n}, u_k, u_{k-1}, ..., u_{k-m})^T$$

oraz macierzy uogólnionych wejść, wektora wyjść i zakłóceń

$$\Phi_N = \begin{bmatrix} \phi_1^T \\ \phi_2^T \\ \vdots \\ \vdots \\ \phi_N^T \end{bmatrix}, \qquad Y_N = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_N \end{bmatrix}, \qquad Z_N = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ \vdots \\ z_N \end{bmatrix},$$

równanie pomiarów przyjmuje postać

$$Y_N = \Phi_N\theta + Z_N,$$

gdzie θ jest niewiadomym wektorem stałych parametrów, a Z_N – nieznanym zakłóceniem losowym.

$$Y_N = \Phi_N\theta \text{ – w ogólności sprzeczny}$$

estymator metodą najmniejszych kwadratów

$$\widehat{\theta} = (\Phi_N^T\Phi_N)^{-1}\Phi_N^T Y_N,$$

błąd estymacji

$$\begin{aligned}
\Delta &= \widehat{\theta} - \theta = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y_N - \theta = \\
&= (\Phi_N^T \Phi_N)^{-1} \Phi_N^T \Phi_N \theta + (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Z_N - \theta = \\
&= (\frac{1}{N} \Phi_N^T \Phi_N)^{-1} \frac{1}{N} \Phi_N^T Z_N.
\end{aligned}$$

Ponieważ proces wyjściowy jest ergodyczny

$$\frac{1}{N} \Phi_N^T Z_N \to \mathbf{E} \phi_k z_k = \mathbf{E} \begin{bmatrix} -y_{k-1} \\ \vdots \\ -y_{k-n} \\ u_k \\ u_{k-1} \\ \vdots \\ u_{k-m} \end{bmatrix} z_k$$

z prawdopodobieństwem 1, gdy $N \to \infty$. Skorelowanie procesu $\{z_k\}$ powoduje, że

$$\mathbf{E} \phi_k z_k = \begin{bmatrix} -\mathbf{E} y_{k-1} z_k \\ \vdots \\ -\mathbf{E} y_{k-n} z_k \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \neq 0$$

Zwiększanie liczby pomiarów nie doprowadzi nas do prawdziwych wartości parametrów. W celu pokonania tej trudności, stosuje się techniki oparte na filtracji zakłóceń lub tzw. metodę zmiennych instrumentalnych

9.
Wersja rekurencyjna algorytmu najmniejszych kwadratów

wersja off-line (standardowa)

$$a_N = (X_N^T X_N)^{-1} X_N^T Y_N$$

wersja on-line (rekurencyjna)

$$a_N = f(a_{N-1}, x_N, y_N) = a_{N-1} + \delta(a_{N-1}, x_N, y_N)$$

odwracana macierz (w wersji off-line)

$$X_N^T X_N = \sum_{k=1}^{N} x_k x_k^T = \sum_{k=1}^{N-1} x_k x_k^T + x_N x_N^T = X_{N-1}^T X_{N-1} + x_N x_N^T$$

oznaczmy

$$P_N = (X_N^T X_N)^{-1}$$

$$cov(a_N) = P_N \sigma_z^2$$

możemy zapisać

$$\begin{aligned} a_N &= P_N X_N^T Y_N \\ a_{N-1} &= P_{N-1} X_{N-1}^T Y_{N-1} \end{aligned}$$

$$P_N = \left(P_{N-1}^{-1} + x_N x_N^T\right)^{-1}, \text{ gdzie } P_{N-1}^{-1} = X_{N-1}^T X_{N-1}$$

Lemat 1 *(o odwracaniu macierzy)*

$$\left(A + uu^T\right)^{-1} = A^{-1} - \frac{1}{1 + u^T A^{-1} u} A^{-1} u u^T A^{-1}$$

Przyjmując $A = P_{N-1}^{-1}$ i $u = x_N$ otrzymujemy

$$\begin{aligned} P_N &= P_{N-1} - \frac{1}{1 + x_N^T P_{N-1} x_N} P_{N-1} x_N x_N^T P_{N-1} = P_{N-1} - \varkappa_N P_{N-1} x_N x_N^T P_{N-1} \\ \text{gdzie } \varkappa_N &= \frac{1}{1 + x_N^T P_{N-1} x_N} \end{aligned}$$

$$\begin{aligned} a_N &= \left(P_{N-1} - \varkappa_N P_{N-1} x_N x_N^T P_{N-1}\right) \left(X_{N-1}^T Y_{N-1} + x_N y_N\right) = \\ &= P_{N-1} X_{N-1}^T Y_{N-1} + P_{N-1} x_N y_N - [\varkappa_N P_{N-1} x_N](x_N^T P_{N-1} X_{N-1}^T Y_{N-1} - x_N^T P_{N-1} x_N y_N) = \\ &= a_{N-1} + [\varkappa_N P_{N-1} x_N]\{\frac{1}{\varkappa_N} y_N - x_N^T a_{N-1} - x_N^T P_{N-1} x_N y_N\} \end{aligned}$$

wstawiając

$$\frac{1}{\varkappa_N} = 1 + x_N^T P_{N-1} x_N$$

otrzymujemy

$$\{...\} = y_N + x_N^T P_{N-1} x_N y_N - x_N^T a_{N-1} - x_N^T P_{N-1} x_N y_N = y_N - x_N^T a_{N-1}$$

wniosek

$$\begin{aligned} a_N &= a_{N-1} + \varkappa_N P_{N-1} x_N (y_N - x_N^T a_{N-1}) \\ P_N &= P_{N-1} - \varkappa_N P_{N-1} x_N x_N^T P_{N-1} \quad / \cdot x_N \end{aligned} \tag{1}$$

$$\begin{aligned} P_N x_N &= P_{N-1} x_N - \varkappa_N P_{N-1} x_N x_N^T P_{N-1} x_N = \varkappa_N P_{N-1} x_N \{\frac{1}{\varkappa_N} - x_N^T P_{N-1} x_N\} \\ \{...\} &= 1 \\ P_N x_N &= \varkappa_N P_{N-1} x_N \end{aligned}$$

wstawiamy do (1)

$$
\begin{aligned}
a_N &= a_{N-1} + P_N x_N (y_N - x_N^T a_{N-1}) \\
P_N &= P_{N-1} - \frac{P_{N-1} x_N x_N^T P_{N-1}}{1 + x_N^T P_{N-1} x_N}
\end{aligned}
$$

warunki początkowe

$$
a_0 = 0,\ P_0 = diag[10^3 \div 10^5]
$$

zaleta: algorytm pracuje bez odwracania macierzy