MI
Metody Identyfikacji

wykład #6

1. *Sygnały pseudolosowe*
2. *Identyfikacja modeli parametrycznych*

Biały szum jako wejście

- Funkcja autokorelacji

$$R_{uu}(\tau) = S_{u0}\,\delta(\tau)$$

- Funkcja korelacji wzajemnej

$$R_{uy}(\tau) = S_{u0}\int_0^{\infty} g(t')\delta(\tau - t')\mathrm{d}t' = S_{u0}\,g(\tau)$$

- Odpowiedź impulsowa

$$g(\tau) = \frac{1}{S_{u0}} R_{uy}(\tau)$$

Niestety biały szum jest nierealizowalny!!!

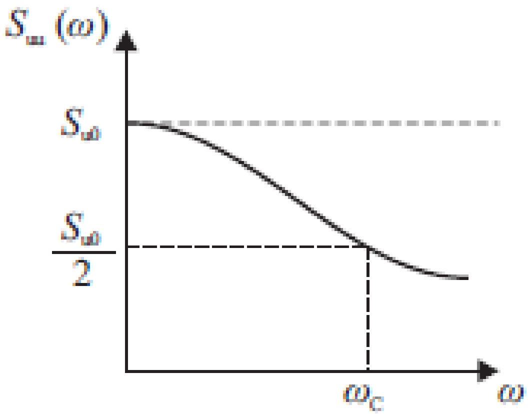
Szum szerokopasmowy

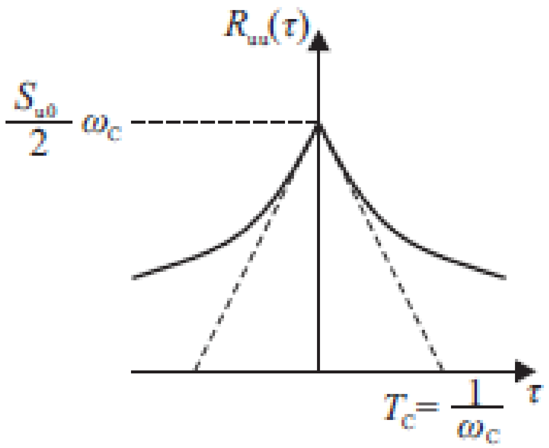
- Filtrowanie *białego* szumu

$$S_{uu}(\omega) = |G_F(i\omega)|^2 S_{u0}$$

- Dla filtru jedno-inercyjnego

$$\begin{aligned} R_{uu}(\tau) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} |G_F(i\omega)|^2 S_{u0} e^{i\omega\tau} d\omega \\ &= \frac{1}{\pi}\int_0^{\infty} \frac{S_{u0}}{1+T_C^2\omega^2} \cos\omega\tau d\omega \\ &= \frac{1}{2} S_{u0}\omega_C e^{-\omega_C|\tau|} . \end{aligned}$$





- Błąd wynikający ze skończonego pasma przenoszenia maleje wraz ze wzrostem pasma (częstotliwości granicznej)
- Niestety jednocześnie rośnie błąd wynikający z zakłócenia
- Tym samym pasmo nie może być za duże

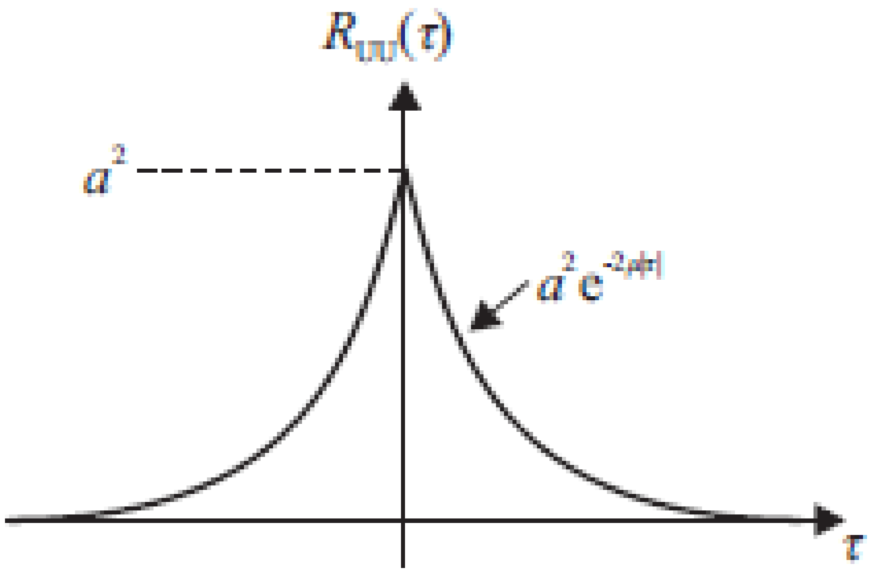
Wykorzystanie naturalnych szumów jako wejść

- Muszą być spełnione następujące warunki:
  - Stacjonarny
  - Pasmo przenoszenia musi być wyższe od największej interesującej częstotliwości
  - Gęstość widmowa musi być większa niż zakłócenia na wyjściu – aby skrajnie nie wydłużać eksperymentu
  - Nie może być skorelowany z innymi zakłóceniami
  - Otwarta pętla
  - Nie w trybie ręcznym

Binarne sygnały losowe

- Ciągły sygnał binarny – RBS
  Continuous-time Random Binary Signal
- Dwa stany: +a i –a
- Zmiana występuje w dowolnie wybranym czasie
- Zalety:
  - Prosta generacja sygnału
  - Proste wyznaczenie funkcji korelacji wzajemnej
  - **Wysoki stosunek gęstości widmowej w zależności od amplitudy sygnału**

$$
\begin{aligned}
\mathrm{E}\{u(t)u(t+\tau)\} &= a^2(P(0)+P(2)+\ldots) - a^2(P(1)+P(3)+\ldots)\\
&= a^2 \mathrm{e}^{-\mu|\tau|}\left(1 - \frac{\mu\tau}{1!} + \frac{(\mu\tau)^2}{2!} \pm \ldots\right)\\
&= a^2 \mathrm{e}^{-2\mu|\tau|} .
\end{aligned}
$$

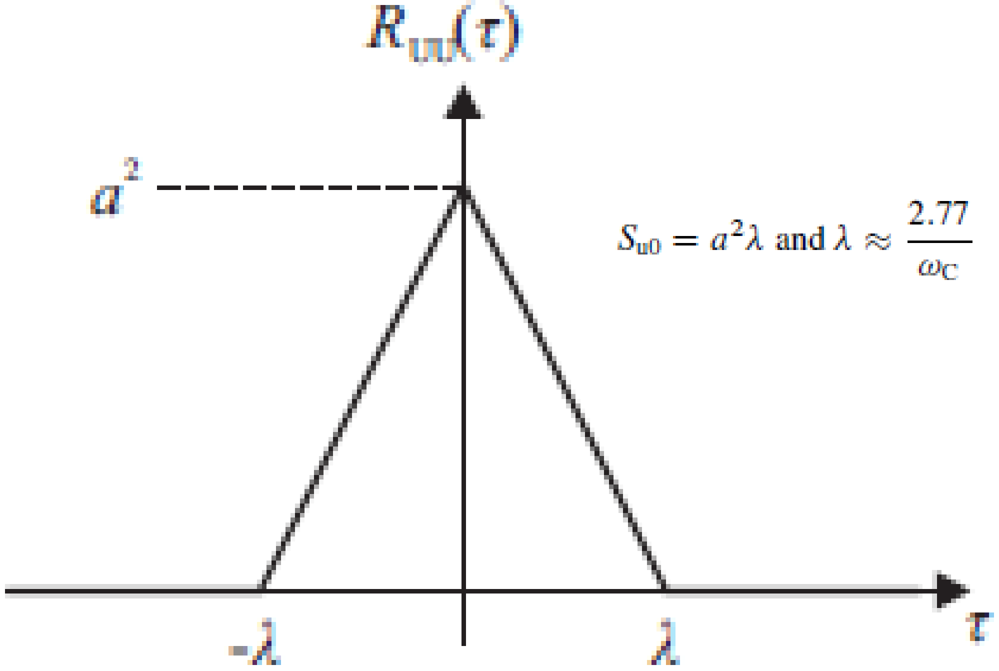


Binarne sygnały losowe

- Dyskretny sygnał binarny – DRBS
  Discrete Random Binary Signal
- Dwa stany: +a i –a
- Zmiana występuje z określonym okresem próbkowania
- Zalety:
  - **Nad wyraz prosta** generacja sygnału
  - Proste wyznaczenie funkcji korelacji wzajemnej
  - **Wysoki stosunek gęstości widmowej w zależności od amplitudy sygnału**

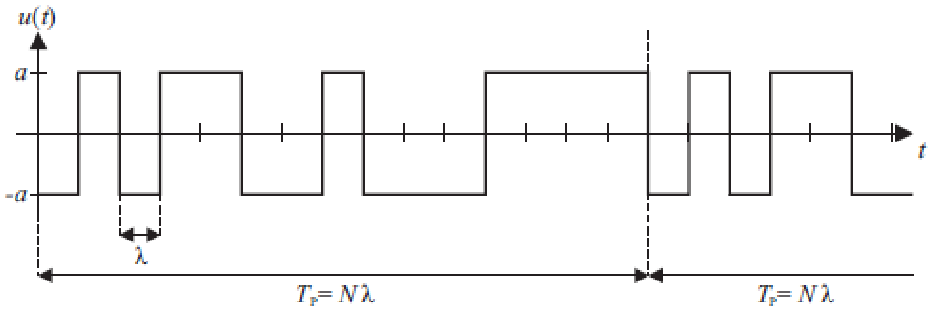
$$R_{uu}(\tau) = \begin{cases} a^2\left(1 - \frac{|\tau|}{\lambda}\right) \text{ for } |\tau| < \lambda \\ 0 \text{ for } |\tau| \geq \lambda \ . \end{cases}$$

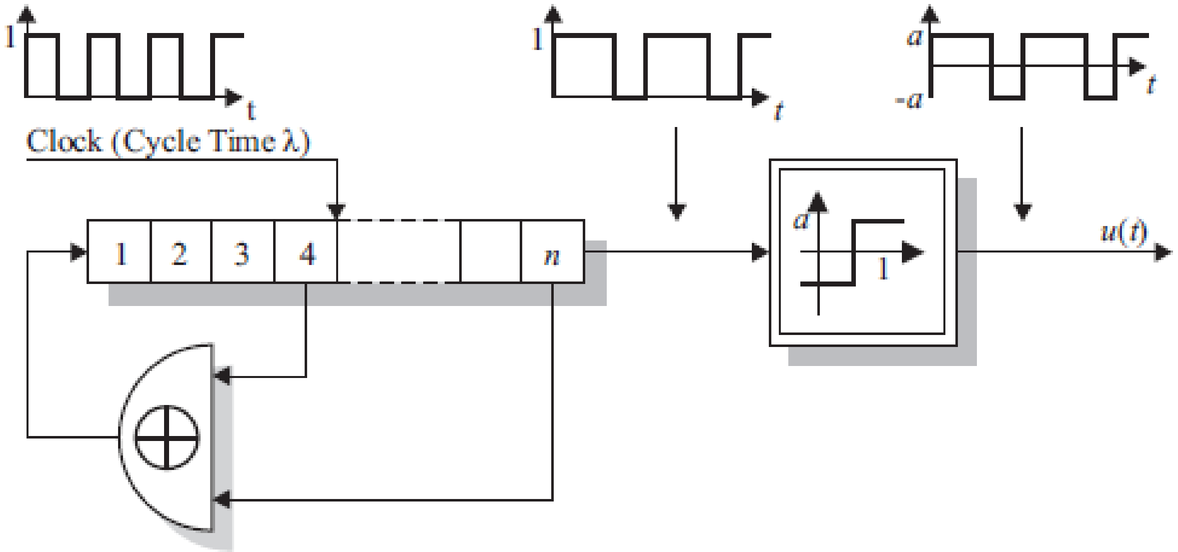
λ - okres próbkowania

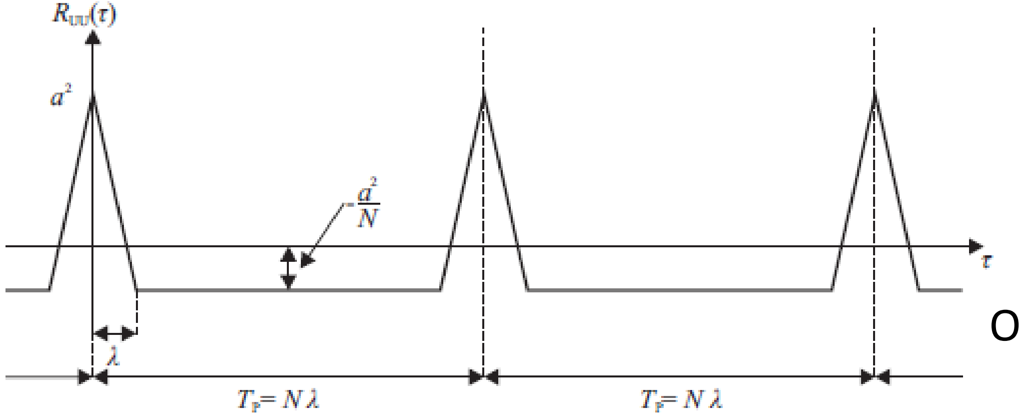


PRBS – pseudolosowy sygnał binarny

- PRBS – Pseudo Random Binary Signal
- Długość: $N\lambda = (2^n\text{-}1)\lambda$
- Średnia $= a/N$







Okresowa funkcja autokorelacji

PRBS – przykładowe realizacje

No. of Stages	Feedback Law	Length
2	1 XOR 2	3
3	1 XOR 3 or 2 XOR 3	7
4	1 XOR 4 or 3 XOR 4	15
5	2 XOR 5 or 3 XOR 5	31
6	1 XOR 6 or 5 XOR 6	63
7	1 XOR 7 or 3 XOR 7 or 4 XOR 7 or 6 XOR 7	127
8	1 XOR 2 XOR 7 XOR 8	255
9	4 XOR 9 or 5 XOR 9	511
10	3 XOR 10 or 7 XOR 10	1023
11	2 XOR 11 or 9 XOR 11	2047

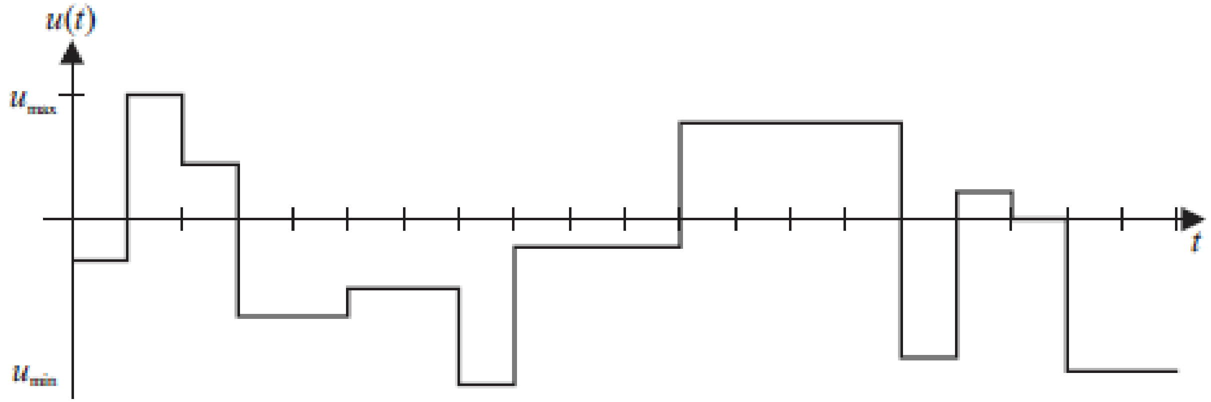
PRBS - dobór

- Zasady wyboru parametrów:
  - Amplituda największa z możliwych, tak aby ograniczyć wpływ zakłóceń, ale trzeba uwzględnić ograniczenia procesowe/technologiczne na sterowanie i wyjście
  - Im większy okres próbkowania tym lepiej, aby częstotliwość odcięcia $\omega_c = 1/\lambda$ nie była za mała

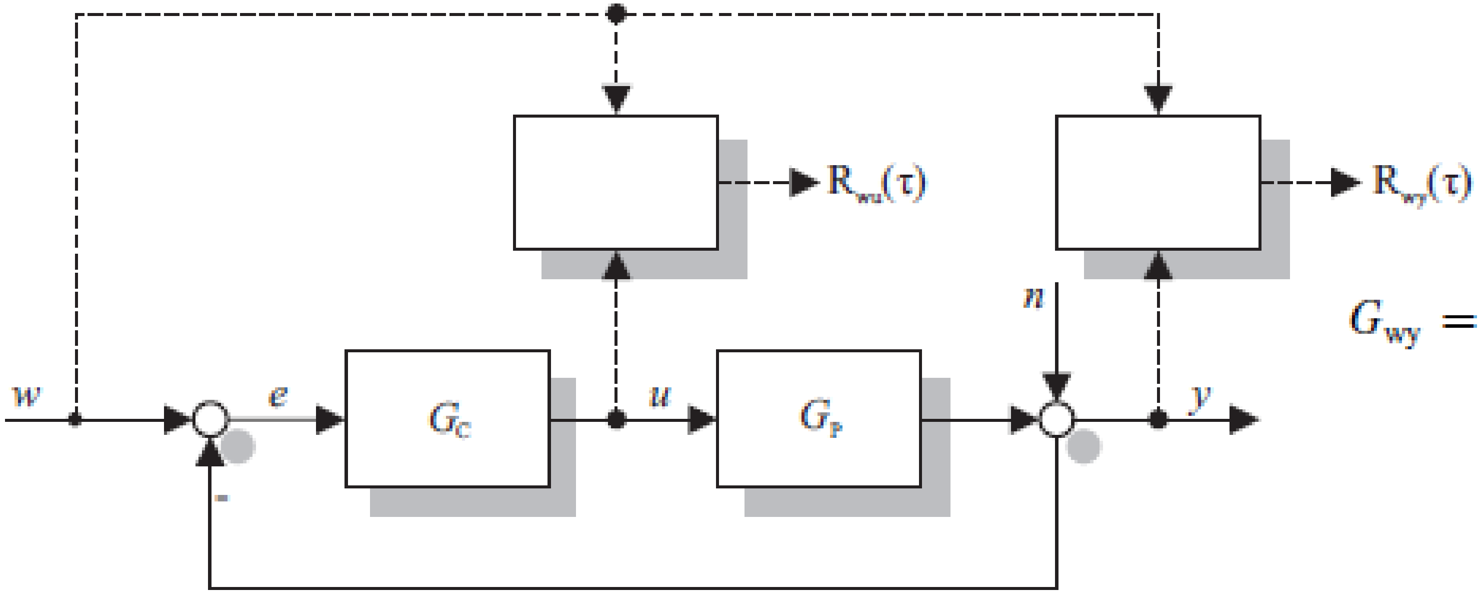
    Sugerowany wybór:
    - $\lambda \leq T_i/5$ (T_i – najmniejsza stała czasowa procesu)
    - $\lambda \leq T_{dom}/10$ (T_{dom} – dominująca stała czasowa procesu)
  - Długość okresu PRBS nie powinna być dłuższa od czasu ustalenia procesu T_{95}
    - $T_p \approx 1.5\ T_{95}$

PRBS zmodulowany amplitudowo

- Dla obiektów nieliniowych
- Wystarczająco pobudzony sygnał wejściowy



Analiza korelacyjna w pętli zamkniętej



$$G_{wy} = \frac{G_C G_P}{1 + G_C G_P}$$

$$G_{uw} = \frac{u(s)}{w(s)} = \frac{G_C}{1 + G_C G_P}$$

$$G_P = \frac{G_{wy}}{G_{wu}}$$

MNK - historia

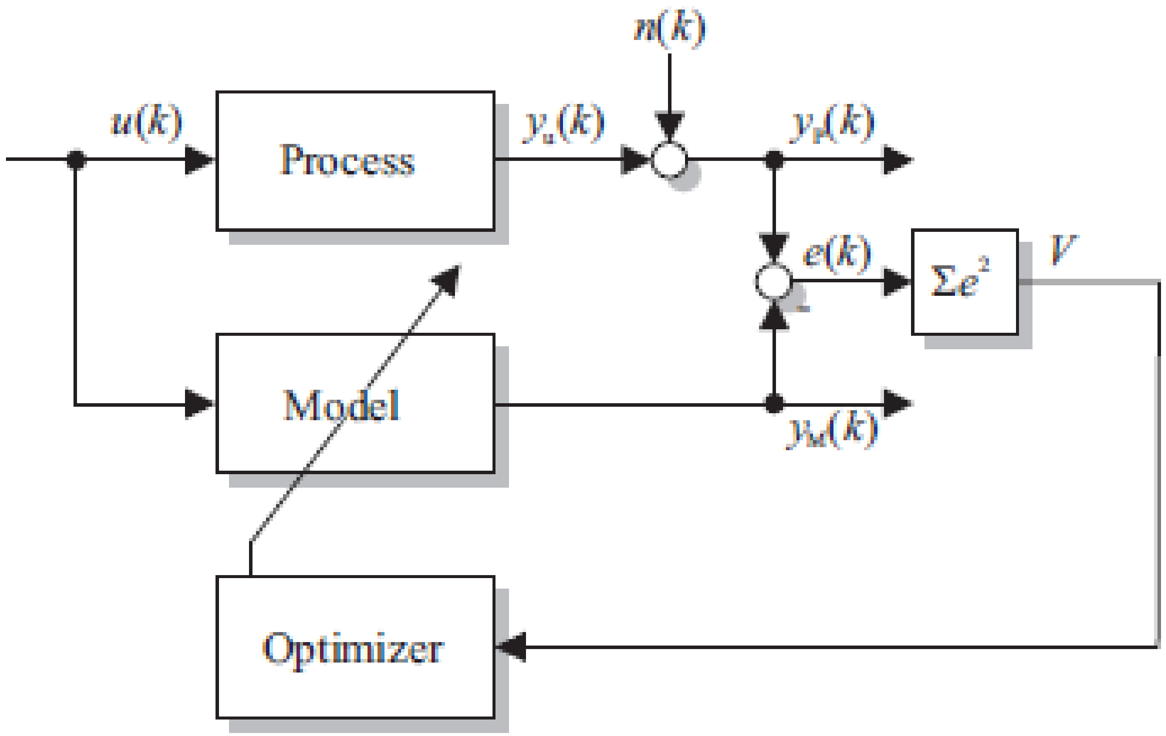
- Boscowich's (1711-1787) method and a minimization of the maximum deviation consisted of a minimization of the sum of absolute values of the residuals under the condition that the sum of the residuals should be equal to zero.
- An algorithm for finding the minimax residual was given in 1783 by Laplace, and in 1789 he simplified his earlier procedure.
- Another method was proposed by Euler and Lambert, according to which the estimates should be the quantities which minimize the absolute value of the largest deviation
- Legendre (1752-1833) published in 1805 a memoir, *Nouvelles méthodes pour la determination des cometes*, in which he introduced and named the method of least squares.
- Gauss (1777-1855) published in 1809 a book, *Theoria motus corporum coelestium in sectionibus conicis solem ambientium* (Gauss, 1809), where he discussed the method of least squares and, mentioning Legendre's work, stated that he himself had used the method since 1795
- Legendre was offended by Gauss's statemen
- In the following years, Gauss tried to produce evidence for his claim but had only little success
- The astronomer Olbers included in a paper in 1816 a footnote asserting that Gauss had shown him the method of least squares in 1802. (Bessel published a similar note in a report in 1832.)

Metoda Najmniejszych Kwadratów – MNK

- Ogólny schemat stosowania metody
- Minimalizacja błędu obserwacji *e(k)*
- Stosujemy kryterium oceny w postaci sumy kwadratów błędu

$$V = e^2(1) + e^2(2) + \ldots + e^2(N) = \sum_{k=1}^{N} (e(k))^2$$

- Dlaczego suma kwadratów?
  - Prostsza do minimalizacji
  - Przy założeniu rozkładu normalnego zakłócenia prowadzi do nieobciążonej estymaty parametrów
  - wada: nazbyt uwzględnia efekt pojedynczych dużych błędów w stosunku do małych ale częstych



Metoda Najmniejszych Kwadratów – MNK

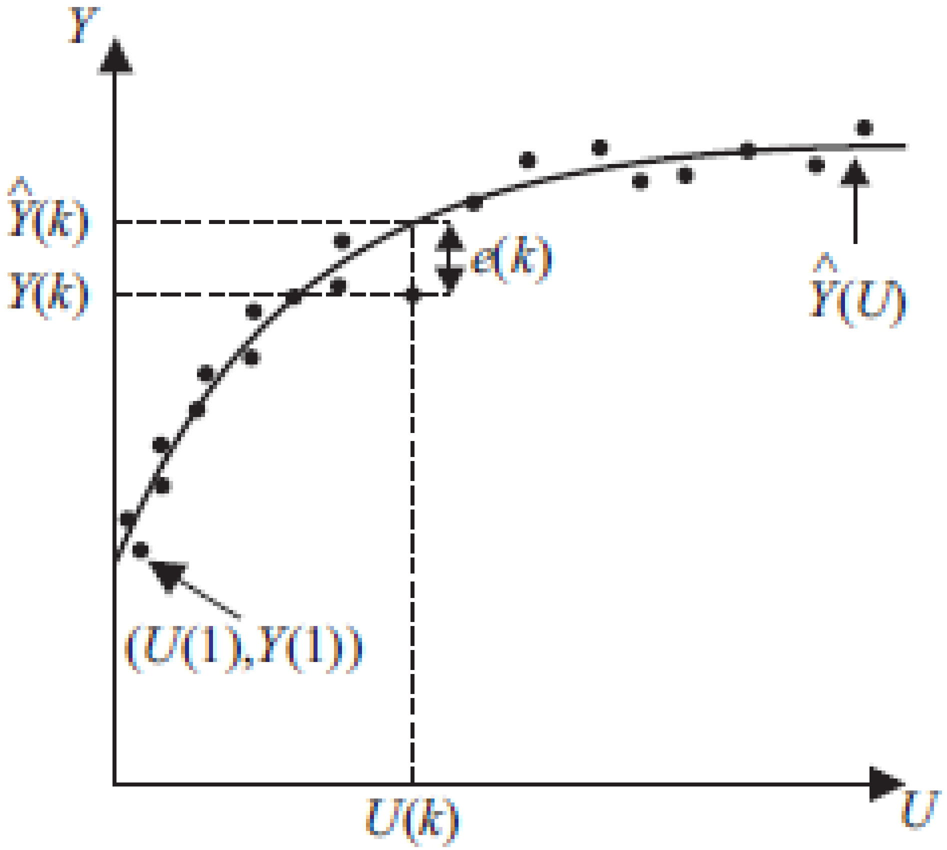
- Rozwiązanie:

$$u^{\mathrm{T}} u \hat{K} = u^{\mathrm{T}} y_{\mathrm{P}} \Leftrightarrow \hat{K} = \left(u^{\mathrm{T}} u\right)^{-1} u^{\mathrm{T}} y_{\mathrm{P}}$$

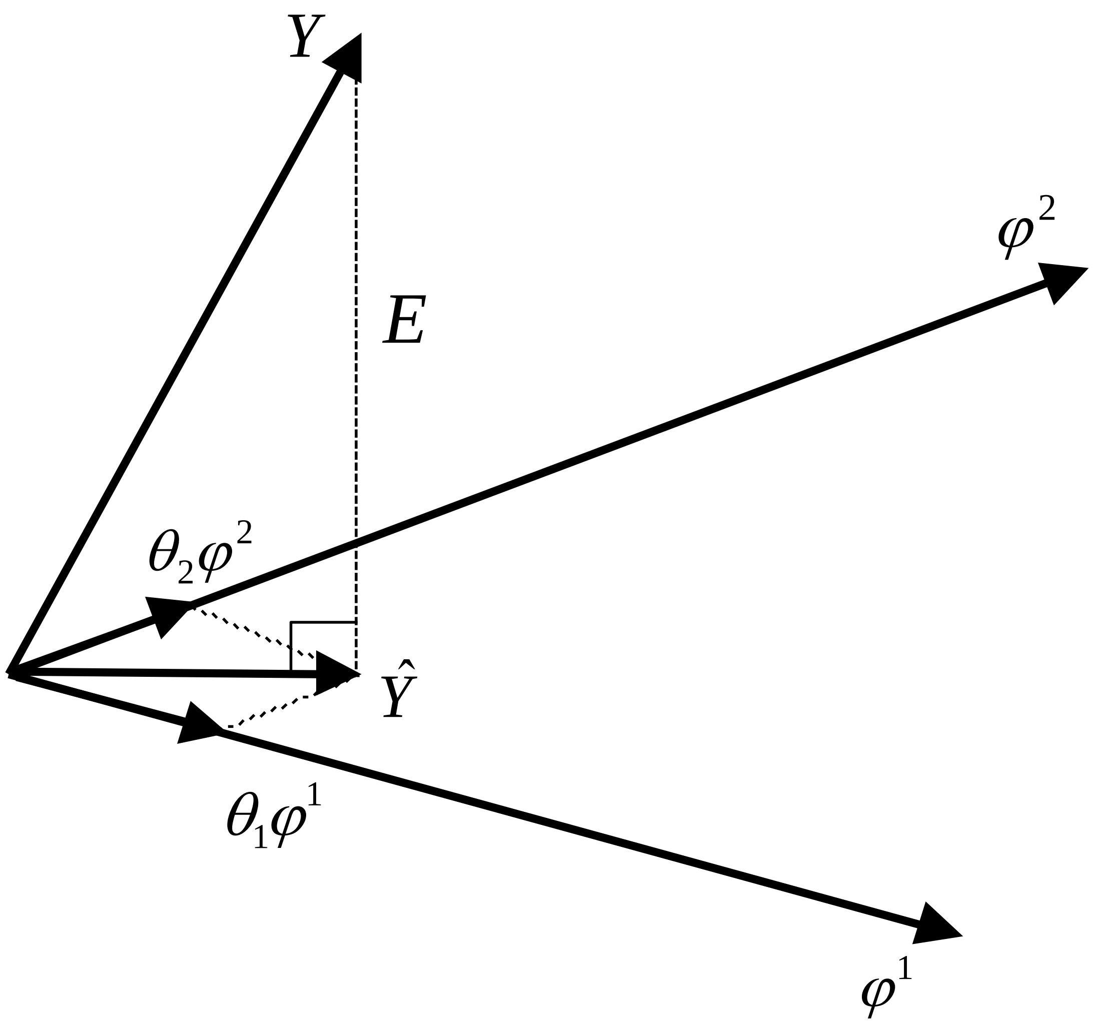
- Błąd:

$$e(k) = Y_{\mathrm{P}}(k) - Y_{\mathrm{M}}(k) = y_{\mathrm{P}}(k) - y_{\mathrm{M}}(k) = y_{\mathrm{P}}(k) - K_{\mathrm{M}}\big(U(k) - U_{00}\big)$$

- Komentarze:
  - Sygnał wejściowy jest dokładnie mierzony jak i jego wartość średnia
  - Obiekt jest wystarczająco pobudzony
  - Zakłócenie n(k) jest stacjonarne
  - Wejście jest nieskorelowane z zakłóceniem
  - Wartość oczekiwana albo wejścia, albo zakłócenia jest zerowa



MNK – interpretacja geometryczna



Metoda największej wiarygodności

- Metoda opiera się na warunkowym prawdopodobieństwie pomiaru $p_y(y|u,\theta)$ zwanym funkcja wiarygodności
- Idea polega na poszukiwaniu $\hat{\theta}$, która maksymalizuje tak zdefiniowaną wiarygodność

$$p_y(y|\theta)|_{\theta=\hat{\theta}} \rightarrow max$$

- Maksimum znajdujemy w klasyczny sposób poprzez różniczkowanie

$$\left.\frac{\partial p_y(y|\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} \stackrel{!}{=} 0$$

- Teraz będziemy poszukiwać statycznej nieliniowości.

Metoda największej wiarygodności

- Pomiar wyjścia jest dany:
  $y = \Psi\Theta + n$
- Każdy próbka zakłócenia ma rozkład normalny opisany funkcją gęstości (przy czym zakładamy zerową wartość oczekiwaną $\mu = 0$

$$p(n(k)) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(n(k)-\mu)^2}{2\sigma_n^2}\right)$$

- Dla białego szumu poszczególne próbki są nieskorelowane na całej długości N, zatem funkcja gęstości dla całego ciągu n jest iloczynem gęstości poszczególnych próbek.

$$p(\boldsymbol{n}) = \prod_{k=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(n(k))^2}{2\sigma_n^2}\right) = \frac{1}{(2\pi)^{\frac{N}{2}}\sqrt{\det \boldsymbol{\Sigma}}} e^{-\frac{1}{2}\boldsymbol{n}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{n}}$$

- Co daje N-wymiarową funkcję gęstości, gdzie Σ jest macierzą kowariancji.

$$\Sigma = \sigma_n^2 I \quad det\Sigma = N\sigma_n^2$$

Metoda największej wiarygodności

- Zatem rozkład prawdopodobieństwa pomiaru ma postać

$$p(y|\theta) = \frac{1}{(2\pi)^{\frac{N}{2}}\sqrt{N}\sigma_{\mathrm{n}}}\exp\left(-\frac{1}{2\sigma_{\mathrm{n}}^{2}}(y-\Psi\theta)^{\mathrm{T}}(y-\Psi\theta)\right)$$

- Teraz możemy wyznaczyć maksimum. Wiedząc że $p(x)$ i $log\ p(x)$ maja maksimum dla tego samego x

$$\left.\frac{\partial \log f(y|\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}} = \frac{1}{2\sigma_{\mathrm{N}}^{2}}\left((y-\Psi\hat{\theta})^{\mathrm{T}}(y-\Psi\hat{\theta})\right) \stackrel{!}{=} 0$$

- Otrzymujemy:

$$\widehat{\Theta} = (\Psi^{T}\Psi)^{-1}\Psi^{T}y$$

- Zatem estymator MNK i największej wiarygodności dają takie same rozwiązanie

Metoda największej wiarygodności

- Można wyznaczyć dolną granicę (Cramer-Rao) wariancji estymaty

$$\mathrm{E}\{(\hat{\theta}-\theta)^2\} \geq \frac{1}{\mathrm{E}\left\{\left(\frac{\partial}{\partial\theta}\log p_y(y|\theta)\right)^2\right\}}$$

$$\mathrm{E}\{(\hat{\theta}-\theta)^2) \geq \frac{-1}{\mathrm{E}\left\{\frac{\partial^2}{\partial\theta^2}\log p_y(y|\theta)\right\}}.$$

- Estymator jest nazywany wydajnym, jeśli osiąga dolną granicę wariancji estymaty.
- BLUE (Best Linear Unbiased Estimator) osiąga minimalną wariancje wszystkich nieobciążonych estymatorów

Ograniczenia

- Równościowe

$$C\theta = d$$

- Może być spełnione i otrzymujemy rozwiązanie:

$$\tilde{\theta} = \hat{\theta} - (\Psi^{\mathrm{T}}\Psi)^{-1} C \left(C (\Psi^{\mathrm{T}}\Psi)^{-1} C^{\mathrm{T}} \right)^{-1} (C\hat{\theta} - d)$$

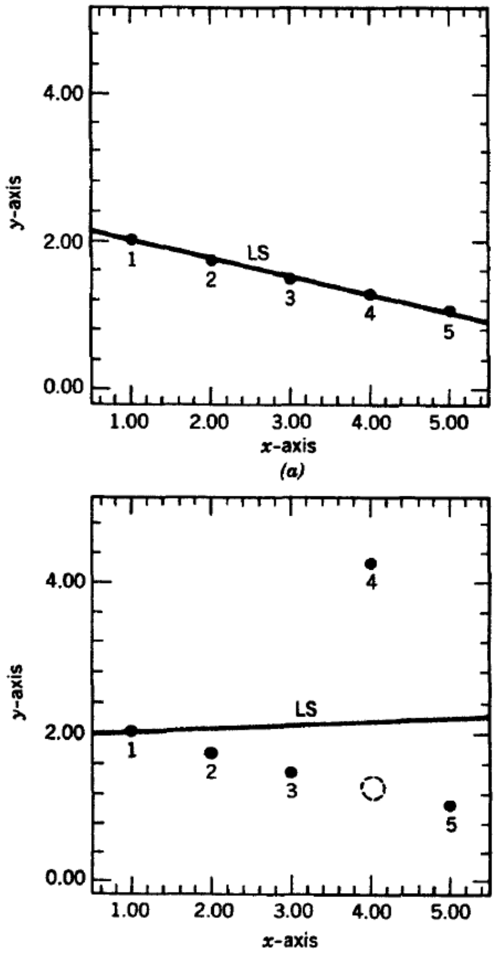
- Nierównościowe

$$\mathrm{A}\theta \leq b$$

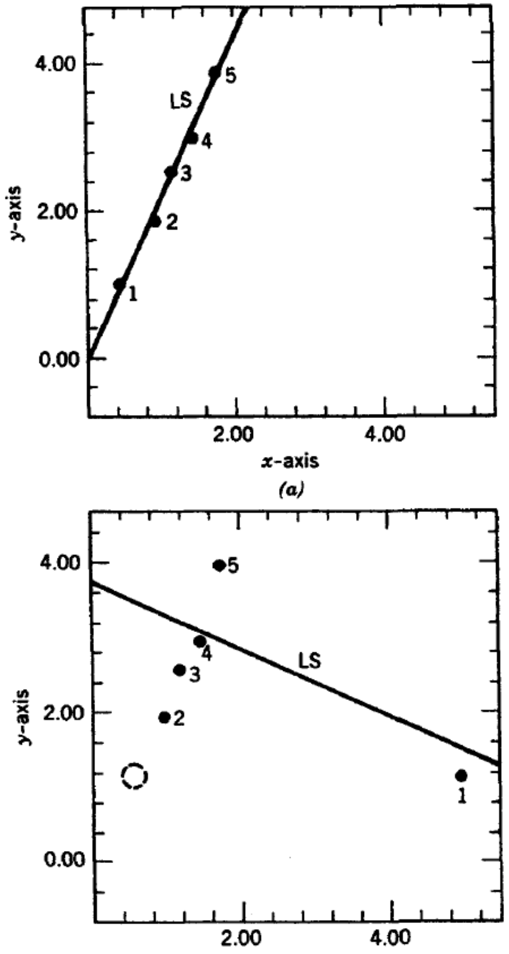
- Rozwiązujemy metodami optymalizacji nieliniowej

Outliers i ich wpływ na MNK

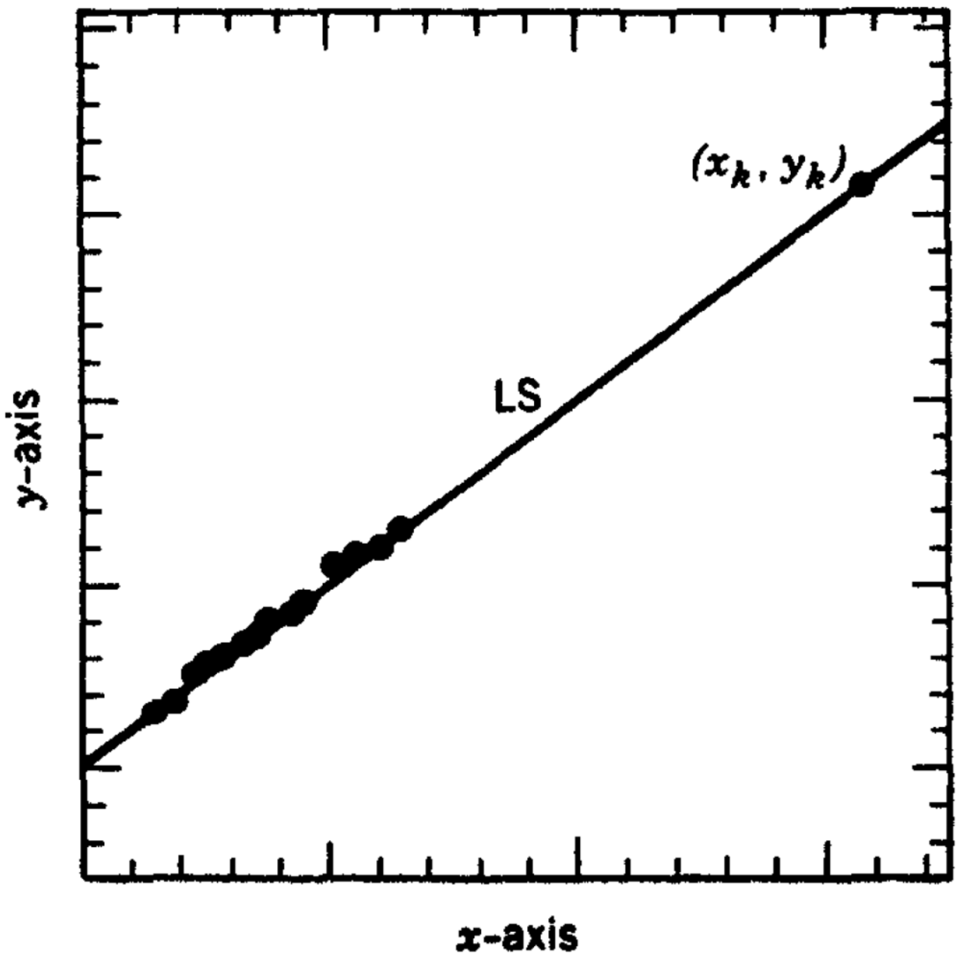
Outlier wg osi OY



Outlier wg osi OX

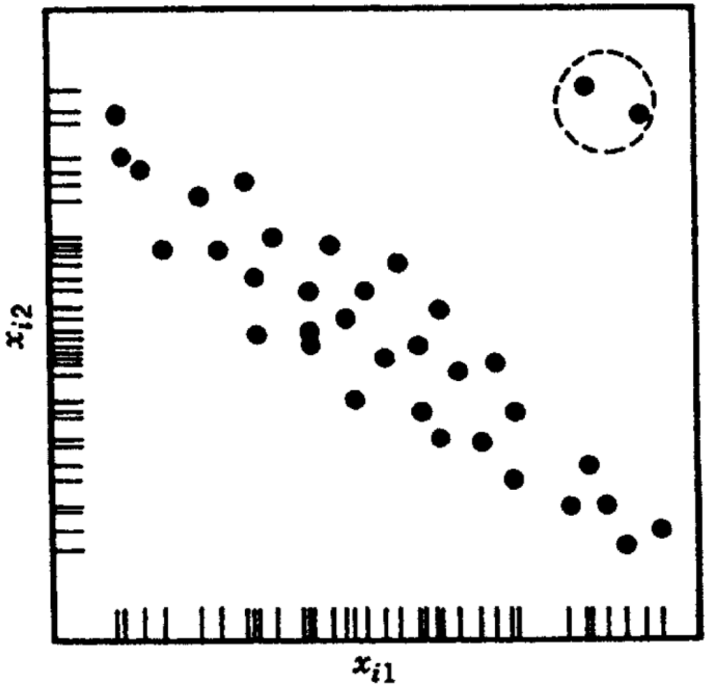


Outliers i ich wpływ na MNK



- Leverege point – punkt dźwigniowy
- Outlier jednocześnie w obu osiach
- Brak wpływu na regresję

Outliers i ich wpływ na MNK



- Dwa punkty *dźwigniowe*
- Nie są outlierem w żadnej z osi

Outliers

- Definicja:

an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism

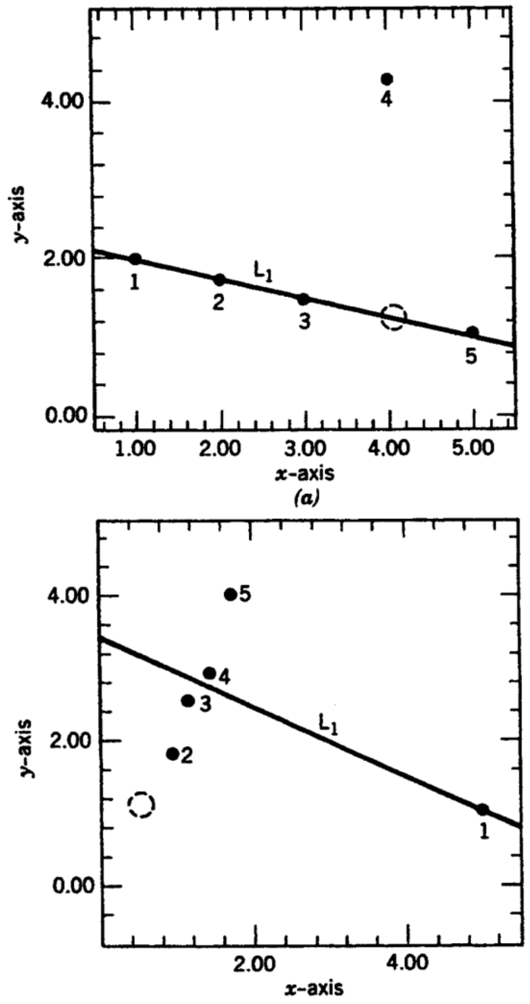
- Breakdown:

the smallest fraction of contamination that can cause the estimator T to take on values arbitrarily far from T(2)

  - Dla MNK wynosi 0%

- Przyczyny:
  - The data come from some heavy tailed distribution such as Student's t. There is no question that any observation is in any way erroneous.
  - The data arise from two distributions. One of these, the 'basic distribution', generates 'good' observations, while another the 'contaminating distribution', generates 'contaminants'. If the contaminating distribution has tails which are heavier than those of the basic distribution, then there will be a tendency for the contaminants to be outliers - that is, to separate visibly from the good observations, which will then constitute the inliers.

Outliers i ich wpływ na MNK



- Wskaźnik jakości Laplace’a, czyli
- IAE – błąd sumy wartości bezwzględnych
- Breakdown wynosi też 0%

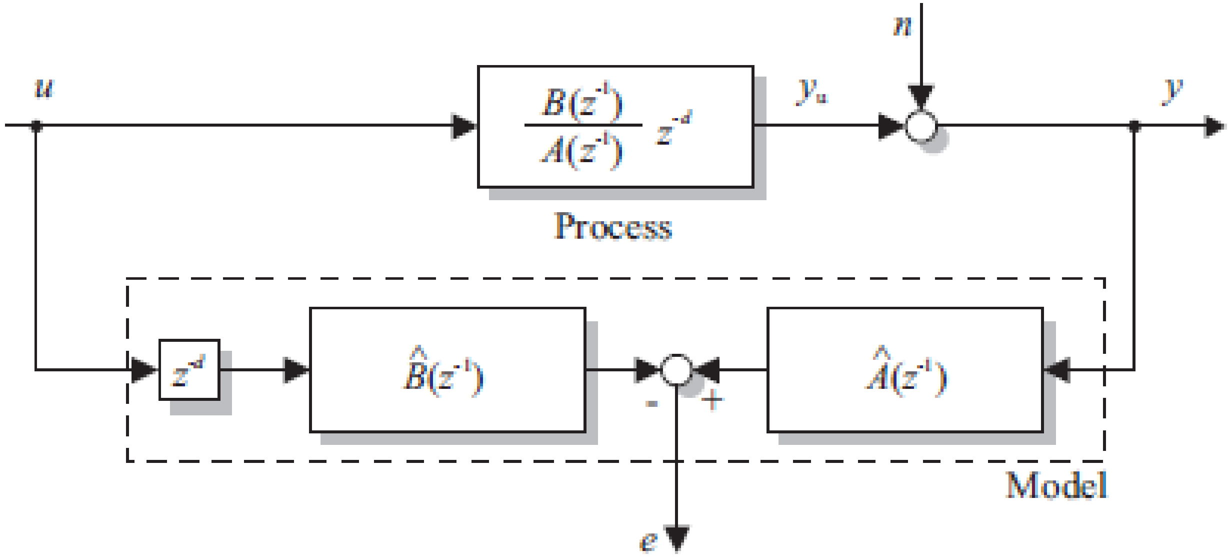
- Estymatory odporne prowadząca do wartości breakdown sięgających 50%
  - LMS – Least Median of Squares
  - LTS - Least Trimmed Squares

MNK dla układów dynamicznych

- Model ARX

$$\hat{\Theta} = \left[\frac{1}{n}\sum_{t=1}^{n} \varphi\,(t) \cdot \varphi^T(t)\right]^{-1} \cdot \left[\frac{1}{n}\sum_{t=1}^{n} \varphi\,(t) \cdot y(t)\right]$$

- Statystyczne własności estymaty są inne niż w przypadku statycznym, gdyż $\varphi\,(t)$ jest pewną realizacją procesu stochastycznego a nie ustaloną statyczną wielkością $\varphi\,(u)$

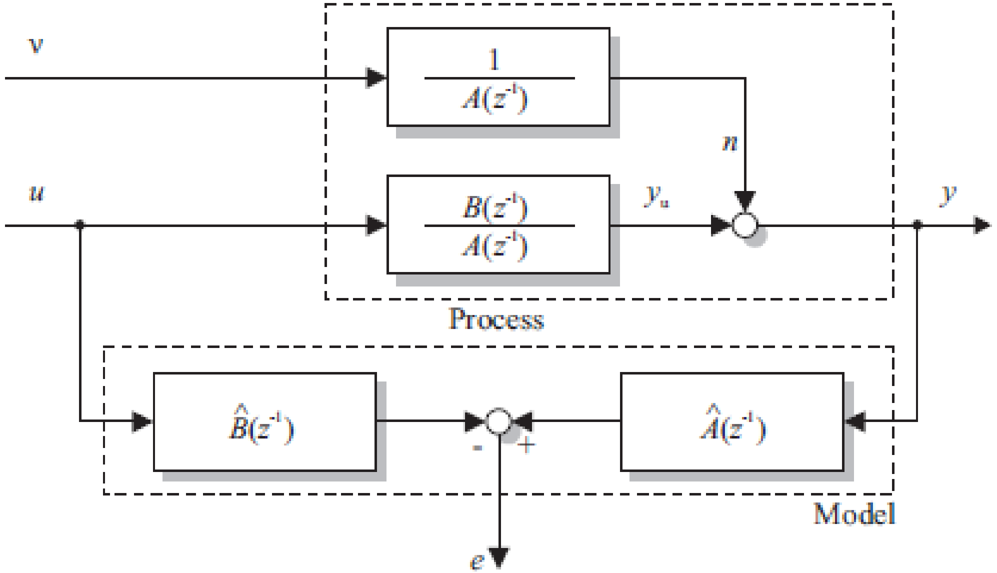


MNK: własności estymaty

- Stąd $\widehat{\Theta} \xrightarrow{n\to\infty} \Theta_0$ jeśli
  - $E\{\varphi\,(t) \cdot \varphi^T(t)\}$ nieosobliwa.
    Na ogół spełnione, z wyjątkiem:
    - Sygnał wejściowy nie jest dostatecznie dobrze pobudzony,
    - $v(t) = 0$ i rząd modelu za wysoki, tzn. $A_o(q^{-1})$ i $B_o(q^{-1})$ mają wspólne człony
    - $u(t)$ generowana przez liniowe sprzężenie zwrotne od wyjścia
  - $E\{\varphi\,(t) \cdot v(t)\} = 0$
  - Jedynie gdy $v(t)$ jest białym szumem.
    - Jeśli nie jest białym szumem, to jest skorelowana z poprzednimi wejściami, gdyż $y(t)$ zależy od $v(s)$ $s<t$ poprzez równanie procesu.
      wtedy $E\{\varphi\,(t) \cdot v(t)\} \neq 0$

MNK

- Metoda łatwa w użyciu, ale daje estymator nieobciążony przy ostrych założeniach
  - Odchyłkę można czasem tolerować, np. gdy stosunek sygnału do szumu jest wysoki
  - Model może nie być zbyt dokładny (stosowany jako model wewnętrzny w układzie regulacji)
- Modyfikacje służące otrzymaniu estymatorów **zgodnych**:
  - Modyfikacje układu równań normalnych, tzw. metody zmiennej instrumentalnej IV (*ang.* Instrumental Variable)
  - Przyjęcie innej, bardziej precyzyjnej postaci modelu. Niekoniecznie liniowy względem parametrów i stosujemy wtedy metodę błędu predykcji



Identyfikowalność parametrów

- Identyfikowalność jest uzależniona od:
  - Proces S
  - Zdefiniowanie eksperymentu X
  - Struktura modelu M
  - Metoda identyfikacji
- Różne definicje:
  - Funkcja celu ma jednoznaczne rozwiązanie (Bellmann i Astrom, 1970)
  - Spójność estymatora, tzn. jeśli estymaty zbiegają do rzeczywistych pomiarów

$$\widehat{\Theta}(N) \xrightarrow{N\to\infty} \Theta_o$$

$$\lim_{N\to\infty} \mathrm{E}\{\hat{\theta}(N)\} = \theta_0$$

$$\lim_{N\to\infty} \operatorname{cov} \Delta\hat{\theta}(N) = 0$$

  - przy czym zbieżność określona jest w różnym sensie, np. prawdopodobieństwa, średnich kwadratów, itp..

Uwarunkowania

- Sygnał wejściowy musi być znany i mierzony tak jak i jego wartość średnia
- Sygnał wejściowy musi być wystarczająco pobudzony.
  - Dla sygnału wejściowego $u(k) = U(k) - \overline{U}$ powinien być spełniony warunek, że

$$\overline{U} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=m+d}^{m+d+N-1} U(k)$$

  - oraz

$$R_{uu}(\tau) = \lim_{N\to\infty} (U(k) - \overline{U})(U(k-\tau) - \overline{U})$$

  - Istnieją a macierz R_{22} jest dodatnio określona.
- Analogiczne warunki są dla metod korelacyjnych.
- Przykłady:
  - Biały szum,
  - Sygnał ruchomej średniej rzędu m
  - PRBS

Uwarunkowania

- Rzędy wielomianów oraz opóźnienie są znane
- Proces powinien spełniać następujące warunki:
  - Stabilny, czyli bieguny mianownika są w kole jednostkowym,
  - Nie wszystkie parametry b_i, $i=0,1,2,\ldots,m$ są zerowe.
  - Licznik i mianownik nie ma wspólnych zer/biegunów.
- Jeśli minimalny rząd m jest znany, wtedy identyfikowalność jest zapewniona poprzez stabilność, sterowalność oraz obserwowalność.
- Zakłócenie losowe musi być stacjonarne
- Błąd nie może być skorelowany oraz jego wartość oczekiwana musi wynosić zero

Nieznana wartość stała wejścia

- Różnicowanie

$$U(k) - U(k-1) = u(k) - u(k-1) = \Delta u(k)$$
$$Y(k) - Y(k-1) = y(k) - y(k-1) = \Delta y(k)$$

- Uśrednianie

$$\hat{Y}_{00} = \frac{1}{N}\sum_{k=0}^{N-1} Y(k)$$

uśrednianie rekurencyjne

$$\hat{Y}_{00} = \hat{Y}_{00}(k-1) + \frac{1}{k}\left(Y(k) - \hat{Y}_{00}(k-1)\right)$$

uśrednianie rekurencyjne
z zapominaniem

$$\hat{Y}_{00} = \lambda \hat{Y}_{00}(k-1) + (1-\lambda)Y(k)$$

- Włączenie wartości średniej do identyfikowanego wektora parametrów

$$Y(k) = -a_1 Y(k-1) - \ldots - a_m Y(k-m) + b_1 U(k-d-1) + \ldots + b_m U(k-d-m) + C\,,$$
$$C = (1 + a_1 + \ldots + a_m)Y_{00} - (b_1 + \ldots + b_m)U_{00}$$