# ESS101- Modeling and Simulation
# Lecture 10-12

***Paolo Falcone***

*Department of Signals and Systems*

*Chalmers University of Technology*

*Göteborg, Sweden*

# Today (Chapter 9)

- Parametric identification
  - Tailor-made models
  - Ready-made models
    - Model structures (ARMAX, ARX, OE)
  - Prediction
  - Prediction Error Method
    - Least squares method
    - Convergence analysis
    - Excitation

# Parametric system identification

***Problem statement***. Given a system in the form

$$\dot{x} = f\left(x,u,\theta\right)$$
$$y = g\left(x,u,\theta\right)$$

find the vector $\theta$ based on measurements of $\underline{u}$ and $y$.

***Linear parametric identification***. The function $f$ is linear. Possible forms

$$\begin{aligned}\dot{x} &= A(\theta)x + B(\theta)u \\ y &= C(\theta)x + D(\theta)u\end{aligned} \qquad \textbf{\textit{or}} \qquad Y(s) = G(s,\theta)U(s)$$

# Parametric system identification. Models types

- ***Tailor-made models***
  - obtained as the results of physical modeling
  - the parameters have a physical meaning

$$\dot{x} = f(x,u,\theta) \qquad \dot{x} = A(\theta)x + B(\theta)u$$

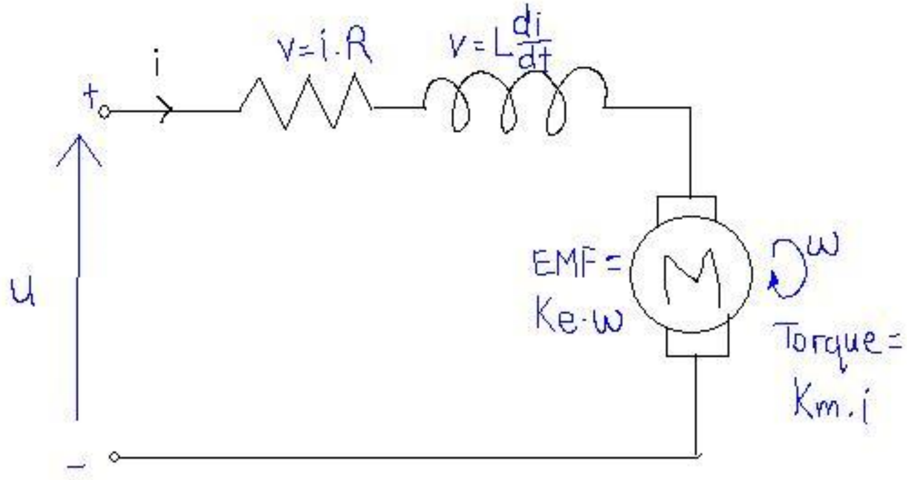
$$y = g(x,u,\theta) \qquad y = C(\theta)x + D(\theta)u$$

$$Y(s) = G(s,\theta)U(s)$$

- ***Ready-made models*** (or “black-box” models)
  - describe the input-output relationships
  - no physical interpretation

$$Y(z) = G(z,\theta)U(z)$$

# Tailor-made models. Example




$$\frac{di(t)}{dt} = -\frac{R}{L}i(t) - \frac{k}{L}\omega(t) + \frac{1}{L}V(t)$$

$$\frac{d\omega(t)}{dt} = \frac{k}{J}i(t) - \frac{b}{J}\omega(t)$$

$$\dot{x} = A(\theta)x + B(\theta)u$$

$$y = Cx$$

$$x = \begin{bmatrix} i \\ \omega \end{bmatrix}, \quad u = V$$

$$A(\theta) = \begin{bmatrix} -\frac{\theta_1}{\theta_2} & -\frac{\theta_3}{\theta_2} \\ \frac{\theta_3}{\theta_5} & -\frac{\theta_4}{\theta_5} \end{bmatrix} \quad B(\theta) = \begin{bmatrix} \frac{1}{\theta_2} \\ 0 \end{bmatrix}$$

$$y = \omega$$

$$\theta = \begin{bmatrix} R & L & k & b & J \end{bmatrix}^T$$

# Linear ready-made models

Consider the ***parametrized***, ***linear discrete time*** model

$$y(t) = G(q,\theta)u(t) + w(t)$$

where

- $G(q,\theta)$ is a parametrized transfer function

$$G(q,\theta) = \frac{B(q)}{F(q)} = \frac{b_1 q^{-nk} + b_2 q^{-nk-1} + \cdots + b_{n_b} q^{-nk-n_b+1}}{1 + f_1 q^{-1} + \cdots + f_{n_f} q^{-n_f}}$$

$$\theta = \begin{bmatrix} b_1 & \cdots & b_{n_b} & f_1 & \cdots & f_{n_f} & c_1 & \cdots & c_{n_c} & d_1 & \cdots & d_{n_d} \end{bmatrix}^T$$

- $w$ is filtered white noise

$$w(t) = H(q,\theta)e(t) \qquad H(q,\theta) = \frac{C(q)}{D(q)} = \frac{1 + c_1 q^{-1} + \cdots + c_{n_c} q^{-n_c}}{1 + d_1 q^{-1} + \cdots + d_{n_d} q^{-n_d}}$$

# Model structures

The most general model structure is

$$y(t) = \frac{B(q,\theta)}{F(q,\theta)}u(t) + \frac{C(q,\theta)}{D(q,\theta)}e(t)$$

- ***A***uto ***R***egressive ***M***oving ***A***verage with e***X***ogenous input (ARMAX)

$$F(q,\theta) = D(q,\theta) = A(q,\theta), \quad \underbrace{A(q,\theta)y(t)}_{\text{Auto Regression}} = \underbrace{B(q,\theta)u(t)}_{\text{eXogenous input}} + \underbrace{C(q,\theta)e(t)}_{\text{Moving Average}}$$

- ***A***uto ***R***egressive with e***X***ogenous input (ARX)

$$F(q,\theta) = D(q,\theta) = A(q,\theta), \qquad \underbrace{A(q,\theta)y(t)}_{\text{Auto Regression}} = \underbrace{B(q,\theta)u(t)}_{\text{eXogenous input}} + e(t)$$

$$C(q,\theta) = 1$$

# Model structures

The most general model structure is

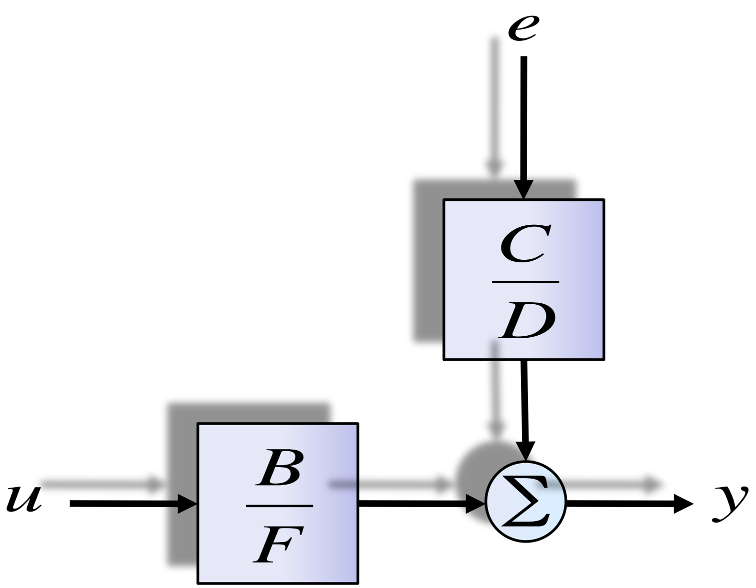
$$y(t)=\frac{B(q,\theta)}{F(q,\theta)}u(t)+\frac{C(q,\theta)}{D(q,\theta)}e(t)$$

- ***O**utput **E**rror* (OE)

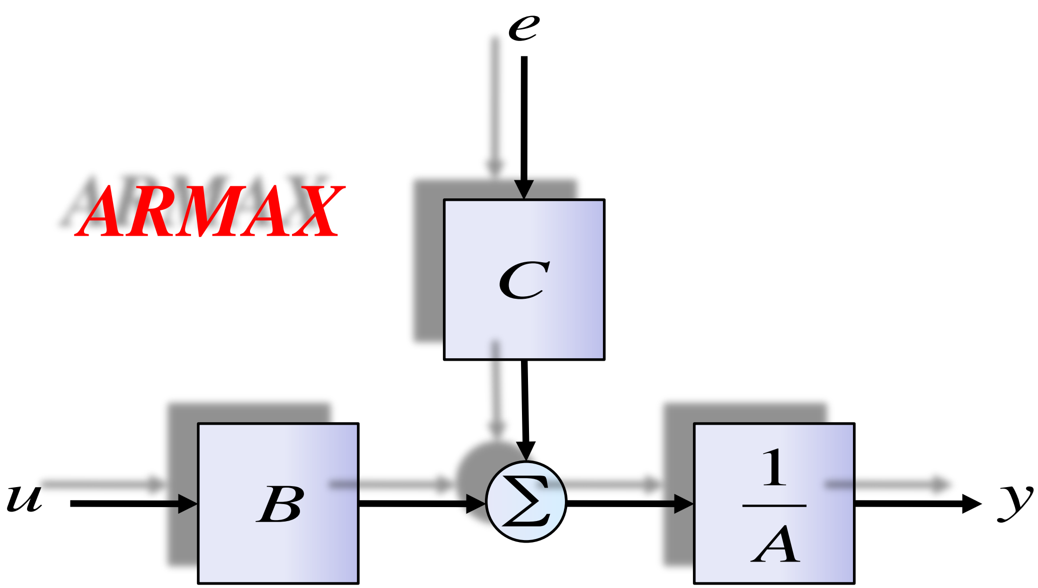
$$\frac{C(q,\theta)}{D(q,\theta)}=1,\quad y(t)=\frac{B(q,\theta)}{A(q,\theta)}u(t)+e(t)$$

***Observe that*** in the *ARX* models the noise passes through the model dynamics while in *OE* model it does not

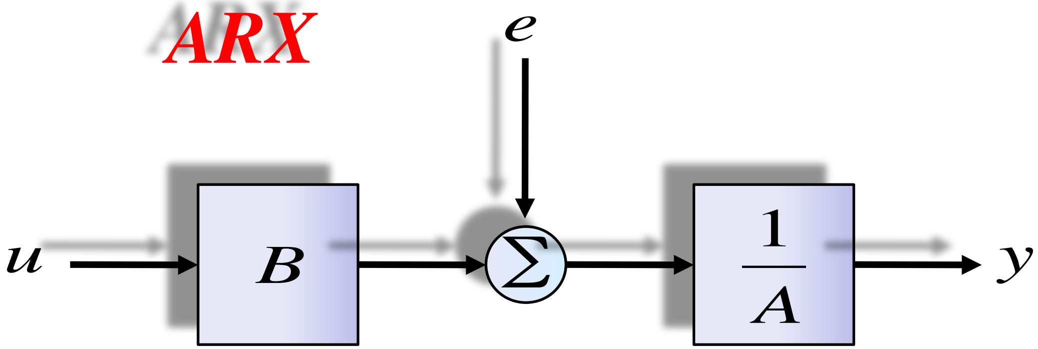
# Model structures



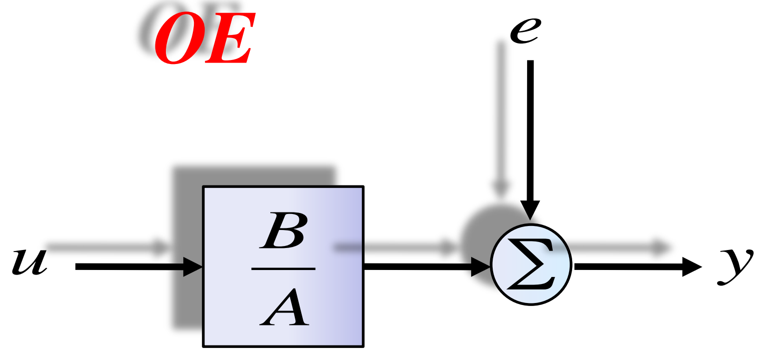

ARMAX




ARX




OE



# Notation

### *ARMAX($n_a$,$n_b$,$n_c$)*

$$\underbrace{A(q,\theta)}_{n_a = \#\,\text{coeff}} y(t) = \underbrace{B(q,\theta)}_{n_b\ \#\,\text{coeff}} u(t) + \underbrace{C(q,\theta)}_{n_c\ \#\,\text{coeff}} e(t)$$

### *ARX($n_a$,$n_b$)*

$$\underbrace{A(q,\theta)}_{n_a\ \#\,\text{coeff}} y(t) = \underbrace{B(q,\theta)}_{n_b\ \#\,\text{coeff}} u(t) + e(t)$$

# Prediction

***Definition***. The ***predictor*** is a mathematical object we use to calculate *y(t)*, based on measurements *y(s)* and *u(s)* with $s<t$.

Obtain the predictor by ***removing the noise***. Example the OE model

OE ***model***

$$y(t)=\frac{B(q,\theta)}{A(q,\theta)}u(t)+e(t)$$

OE ***predictor***

$$\hat{y}(t\,|\,\theta)=\frac{B(q,\theta)}{A(q,\theta)}u(t)$$

# Prediction

## In general

$$y(t) = \underbrace{\frac{B(q,\theta)}{F(q,\theta)}}_{G(q,\theta)} u(t) + \underbrace{\frac{C(q,\theta)}{D(q,\theta)}}_{H(q,\theta)} e(t)$$

$$H^{-1}(q,\theta)y(t) = H^{-1}(q,\theta)G(q,\theta)u(t) + e(t)$$

$$y(t) + H^{-1}(q,\theta)y(t) = y(t) + H^{-1}(q,\theta)G(q,\theta)u(t) + e(t)$$

$$y(t) = \left[1 - H^{-1}(q,\theta)\right]y(t) + H^{-1}(q,\theta)G(q,\theta)u(t) + e(t)$$

$$\hat{y}(t \mid \theta) = \left[1 - H^{-1}(q,\theta)\right]y(t) + H^{-1}(q,\theta)G(q,\theta)u(t)$$

# Prediction

## In general

$$y(t) = \underbrace{\frac{B(q,\theta)}{F(q,\theta)}}_{G(q,\theta)} u(t) + \underbrace{\frac{C(q,\theta)}{D(q,\theta)}}_{H(q,\theta)} e(t)$$

$$\hat{y}(t \mid \theta) = \left[1 - H^{-1}(q,\theta)\right] y(t) + H^{-1}(q,\theta) G(q,\theta) u(t)$$

## Observe that

$$1 - H^{-1}(q,\theta) = 1 - \frac{D(q,\theta)}{C(q,\theta)}$$

$$= \frac{(c_1 - d_1)q^{-1} + \cdots + (c_{n_c} - d_{n_c})q^{-n_c}}{1 + c_1 q^{-1} + \cdots + c_{n_c} q^{-n_c}}$$

$\hat{y}(t \mid \theta)$ only depends on $y(s)$, $s<t$

# Prediction

Consider the ***ARX*** model

$$y(t) = \frac{B(q,\theta)}{A(q,\theta)}u(t) + \frac{1}{A(q,\theta)}e(t)$$

The corresponding predictor is obtained by observing that

$$A(q,\theta)y(t) = B(q,\theta)u(t) + e(t)$$

$$\Downarrow$$

$$\left(1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}\right)y(t) = B(q,\theta)u(t) + e(t)$$

$$\Downarrow$$

$$\hat{y}(t \mid \theta) = -a_1 y(t-1) - \cdots - a_{n_a} y(t-n_a) + B(q,\theta)u(t)$$

# Prediction error method (PEM)

1. Calculate the ***predictor*** of the system output

$$\hat{y}(t \mid \theta)$$

2. Calculate the ***prediction error***

$$\varepsilon(t \mid \theta) = y(t) - \hat{y}(t \mid \theta)$$

3. Calculate the ***criterion (loss or cost) function***

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t \mid \theta)$$

4. Calculate the ***“best” vector of parameters***

$$\theta_N^* = \arg\min_{\theta} V_N(\theta)$$

# Iterative search for minimum

In general the solution of the minimization problem is found numerically, through iterative methods

Solve the *necessary* optimality condition

$$\frac{d}{d\theta}V_N(\theta)=0$$

$$\hat{\theta}^{i+1}=\hat{\theta}^{i}-\mu_i M_i V_N^{'}(\hat{\theta}^{i})$$

$$M_i=\left(V_N^{''}(\hat{\theta}^{i})\right)^{-1}$$

Assuming $V_N$ twice differentiable

# PEM for ARX models

Consider the ***ARX*** model

$$y(t) = \frac{B(q,\theta)}{A(q,\theta)}u(t) + \frac{1}{A(q,\theta)}e(t)$$

and its predictor

$$\hat{y}(t\,|\,\theta) = -a_1 y(t-1) - \cdots - a_{n_a} y(t-n_a) + B(q,\theta)u(t)$$

It can be rewritten as

$$\hat{y}(t\,|\,\theta) = \theta^T \varphi(t), \quad \theta = \begin{bmatrix} a_1 \\ \vdots \\ a_{n_a} \\ b_1 \\ \vdots \\ b_{n_b} \end{bmatrix}, \quad \varphi(t) = \begin{bmatrix} -y(t-1) \\ \vdots \\ -y(t-n_a) \\ u(t) \\ \vdots \\ u(t-n_b) \end{bmatrix},$$

# PEM for ARX models

The predictor is linear in the parameters vector

Apply the PEM method. The prediction error is

$$\varepsilon(t \mid \theta) = y(t) - \theta^T \varphi(t)$$

The criterion function becomes

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left(y(t) - \theta^T \varphi(t)\right)^2$$
$$= \frac{1}{N}\sum_{t=1}^{N} y^2(t) - 2\theta^T f_N + \theta^T R_N \theta$$

with

$$\begin{cases} f_N = \dfrac{1}{N}\displaystyle\sum_{t=1}^{N}\varphi(t)y(t) \\ R_N = \dfrac{1}{N}\displaystyle\sum_{t=1}^{N}\varphi(t)\varphi^T(t) \end{cases}$$

$$\boxed{\theta_N^* = \arg\min_{\theta} V_N(\theta) = R_N^{-1} f_N}$$

# PEM for ARX models

For ***ARX*** models the PEM gives

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left(y(t) - \theta^T\varphi(t)\right)^2$$

$$= \frac{1}{N}\sum_{t=1}^{N} y^2(t) - 2\theta^T f_N + \theta^T R_N \theta$$

with

$$\begin{cases} f_N = \dfrac{1}{N}\displaystyle\sum_{t=1}^{N}\varphi(t)y(t) \\ R_N = \dfrac{1}{N}\displaystyle\sum_{t=1}^{N}\varphi(t)\varphi^T(t) \end{cases}$$

$$\theta_N^* = \arg\min_{\theta} V_N(\theta) = R_N^{-1} f_N$$

***Observe that*** in $f_N$ and $R_N$ are estimates of the covariance and cross covariance of $y$ and $u$

$$R_u^N(\tau) = \frac{1}{N}\sum_{t=1}^{N} u(t-i)u(t-j) \qquad R_{yu}^N(\tau) = \frac{1}{N}\sum_{t=1}^{N} y(t-i)u(t-j)$$

# Model properties

- ***Bias error***.
  - The chosen model structure is not capable of describing the system
  - The parameters change when identification is made in different operating conditions

- ***Variance errors***.
  - Parameters change when identification is repeated
  - Use longer measurement sequences

# Convergence of the estimate

1. Select the family of models

2. Minimize the criterion function

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\varepsilon(t \mid \theta)^2$$

$\theta_N^*$ is the optimum

# Convergence of the estimate

Observations

- $\theta_N^*$ is the minimum of $V_N(\theta)$
- $V_N(\theta)$ depends on data (stochastic processes)

Hence

- $\theta_N^*$ is a stochastic process

**Q: *what happens to* $\theta_N^*$ *when* $N \to \infty$ *?***

# Convergence of the estimate

***Observe that***

$\hat{y}(t \mid \theta)$ is a stationary process, then $\varepsilon(\cdot \mid \theta) = y(\cdot) - \hat{y}(\cdot \mid \theta)$ and $\varepsilon(\cdot, \theta)^2$ are a stationary process as well

$V_N(\theta)$ tends to the variance of the prediction error as $N \to \infty$

$$V_N(\theta) \to E\left[\varepsilon(\cdot \mid \theta)^2\right] = \hat{V}(\theta)$$

NOTE. $E\left[\varepsilon(\cdot \mid \theta)^2\right]$ depends on $\theta$

# Convergence of the estimate

Denote by

$$\Delta = \left\{ \theta \mid \hat{V}(\hat{\theta}) \leq \hat{V}(\theta),\ \forall \theta \in \Theta \right\}$$

the set of minimum points (in general multiple minima)

Since $V_N(\theta) \to E\left[\varepsilon(\cdot \mid \theta)^2\right] = \hat{V}(\theta)$ we expect the convergence to uniformly hold in the parameters space as well

$$\theta_N^* \to \Delta$$

***Conclusion***. The estimated vector of parameters converges to a value ***minimizing the variance of the prediction error***

# Convergence of the estimate

Assume the data has been generated by the model $M(\theta^0)$

**Q: *does*** $\theta_N^*$ ***converge to*** $\theta^0$ ***?***

Consider the prediction error $\varepsilon(t \mid \theta) = y(t) - \hat{y}(t \mid \theta)$

$$\varepsilon(t \mid \theta) = e(t) + \left(\hat{y}\left(t \mid \theta^0\right) - \hat{y}(t \mid \theta)\right)$$

with $e(t) = y(t) - \hat{y}\left(t \mid \theta^0\right)$

# Convergence of the estimate

***Observe that***

$\hat{y}(t \mid \theta^0) - \hat{y}(t \mid \theta)$ depends on the past values of $u(\cdot), y(\cdot)$

while $e(t)$ does not.

$\hat{y}(t \mid \theta^0) - \hat{y}(t \mid \theta)$ and $e(t)$ are uncorrelated. Hence:

$$\underbrace{Var\left[\varepsilon(t \mid \theta)\right]}_{\hat{V}_N(\theta)} = \underbrace{Var\left[e(t)\right]}_{\lambda\ (e(t)\ \text{white noise})} + Var\left[\hat{y}(t \mid \theta^0) - \hat{y}(t \mid \theta)\right]$$

***Conclusion.*** $\hat{V}_N(\theta)$ (the rhs) is minimized by setting $\theta = \theta^0$

***Observe that*** the prediction error is a white noise

# Variance error

Assume the ***bias is zero***. I.e.:

$$\theta_N^* \to \theta^0, \text{ as } N \to \infty$$

then, it can be shown that

$$E\left[\left(\theta_N^* - \theta^0\right)\left(\theta_N^* - \theta^0\right)^T\right] \approx \frac{1}{N} \underbrace{\lambda}_{\text{noise variance}} R^{-1}$$

$$R = E\left[\left(\frac{d}{d\theta}\hat{y}(t \mid \theta^0)\right)\left(\frac{d}{d\theta}\hat{y}(t \mid \theta^0)\right)^T\right]$$

***Conclusion.*** The variance of the estimated parameters can be decreased by increasing the size of the data set.

# Excitation

Consider an ***ARX(1,1)*** model

$$\boxed{\theta_N^* = R_N^{-1} f_N} \qquad \text{with} \begin{cases} f_N = \dfrac{1}{N}\displaystyle\sum_{t=1}^{N} \varphi(t) y(t) \\ R_N = \dfrac{1}{N}\displaystyle\sum_{t=1}^{N} \varphi(t)\varphi^T(t) \end{cases}$$

***Least Squares (LS) formula***

***Requirement.*** $R_N$ invertible

$$R_N = \begin{bmatrix} \dfrac{1}{N}\displaystyle\sum_{t=1}^{N} y^2(t-1) & \dfrac{1}{N}\displaystyle\sum_{t=1}^{N} y(t-1)u(t-1) \\ \dfrac{1}{N}\displaystyle\sum_{t=1}^{N} u(t-1)y(t-1) & \dfrac{1}{N}\displaystyle\sum_{t=1}^{N} u^2(t-1) \end{bmatrix} = \begin{bmatrix} R_y(0) & R_{yu}(0) \\ R_{uy}(0) & R_u(0) \end{bmatrix}$$

# Excitation

## In general

$$R_N = \begin{bmatrix} R_y^{n_a-1} & R_{yu} \\ R_{uy} & R_u^{n_b-1} \end{bmatrix} \quad \text{where}$$

$$R_y^{n_a-1} = E\left[\begin{bmatrix} y(t-1) \\ y(t-2) \\ \vdots \\ y(t-n_a) \end{bmatrix}\begin{bmatrix} y(t-1) & y(t-2) & \cdots & y(t-n_a) \end{bmatrix}\right]$$

$$R_u^{n_b-1} = E\left[\begin{bmatrix} u(t-1) \\ u(t-2) \\ \vdots \\ u(t-n_b) \end{bmatrix}\begin{bmatrix} u(t-1) & u(t-2) & \cdots & u(t-n_b) \end{bmatrix}\right]$$

# Excitation

A ***necessary condition**** on the regularity of $R_N$ is that

$R_u^{n_b-1}$ ***must*** be invertible

**deriving from the Sylverster criterion of the positive definiteness of a matrix*

$$R_u^{n_b-1} = \begin{bmatrix} R_u(0) & R_u(1) & \cdots & R_u(n_b-1) \\ R_u(1) & R_u(0) & \cdots & R_u(n_b-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_u(n_b-1) & R_u(n_b-2) & \cdots & R_u(0) \end{bmatrix}$$

In this case $u$ is said to be ***persistently exciting***

# Excitation

***Definition.*** A signal $w$ is persistently exciting of order $n$ if $R_w^n$ is invertible.

***Q:*** Which signal is persistently exciting of every order?

***Q:*** Which signal would you ***NOT*** use in the LS formula?

***Note that*** singularity of $R_N$ may occur for other reasons as well (system structure)