# System Identification
# Lecture 6

# PE Method - Linear Regression and Approximate Modelling

Paul Van den Hof

Control Systems Group
Department of Electrical Engineering
Eindhoven University of Technology

# Contents

# Linear regression

What are the attractive properties of model structures that are linear-in-the-parameters?

## Linear regression

### Particular model structures: ARX, FIR

Prediction:

$$\begin{aligned}\hat{y}(t|t-1) &= H^{-1}Gu(t) + [1 - H^{-1}]y(t) \\ &= B(q^{-1})u(t) + [1 - A(q^{-1})]y(t)\end{aligned}$$

Consequence:

$$\hat{y}(t|t-1;\theta) = \varphi^T(t)\theta$$

with

$$\begin{aligned}\varphi(t) &= [-y(t-1) \ \cdots \ -y(t-n_a) \ u(t) \ \cdots u(t-n_b+1)]^T \\ \theta &= [a_1 \ a_2 \ \cdots a_{n_a} \ b_0 \ \cdots b_{n_b-1}]^T\end{aligned}$$

Predictor is linear-in-the-parameters $\theta$.

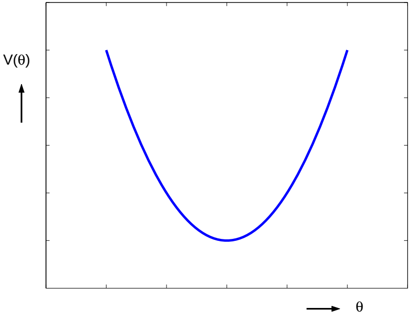
Criterium:

$$V_N(\theta) = \frac{1}{N}\sum_{t=0}^{N-1} \varepsilon(t,\theta)^2$$

with

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

is quadratic in $\theta$.




(for scalar $\theta$)

Minimum is uniquely achieved in $\hat{\theta}_N$ by solving

$$\left.\frac{\partial V_N(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_N} = 0$$

$$V_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N}[y(t) - \varphi^T(t)\theta]^2$$

$$\begin{aligned}
\frac{\partial V_N(\theta, Z^N)}{\partial\theta} &= 2\frac{1}{N}\sum_{t=1}^{N}[y(t) - \varphi^T(t)\theta]\frac{-\partial\varphi^T(t)\theta}{\partial\theta} \\
&= -2\frac{1}{N}\sum_{t=1}^{N}\underbrace{[y(t) - \varphi^T(t)\theta]}_{scalar}\cdot\underbrace{\varphi(t)}_{vector} \\
&= -2\frac{1}{N}\sum_{t=1}^{N}[\varphi(t)y(t) - \varphi(t)\varphi^T(t)\theta]
\end{aligned}$$

Setting derivative to 0 in $\theta = \hat{\theta}_N$ delivers:

$$\left[\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t)\right]\hat{\theta}_N = \frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t)$$

As a consequence:

$$\hat{\theta}_N = \Bigg[\underbrace{\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t)}_{R(N)}\Bigg]^{-1} \cdot \underbrace{\frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t)}_{f(N)}$$

- Analytical solution through simple matrix operations.
- All terms on the right hand side are provided by the data
- Requirement for unique solution: $R(N)$ is invertible.
  This comes down to a sufficient excitation of the input signal.

Analysis:
If the data generating system is given by

$$y(t) = \varphi^T(t)\theta_0 + w_0(t)$$

which is equivalent to

$$y(t) = \frac{B_0(q^{-1})}{A_0(q^{-1})}u(t) + \frac{1}{A_0(q^{-1})}w_0(t)$$

Then

$$\hat{\theta}_N = \theta_0 + R(N)^{-1} \cdot \frac{1}{N}\sum_{t=1}^{N}\varphi(t)w_0(t).$$

The second (bias) term disappears if $w_0(t)$ uncorrelated to the components in $\varphi(t)$.

$$\bar{\mathbb{E}}\{\varphi(t)w_0(t)\} = \bar{\mathbb{E}} \begin{bmatrix} -y(t-1) \\ \vdots \\ -y(t-n_a) \\ u(t) \\ \vdots \\ u(t-n_b+1) \end{bmatrix} w_0(t)$$

For the asymptotic bias to be 0, $w_0(t)$ has to satisfy

$$\bar{\mathbb{E}}\{\varphi(t)w_0(t)\} = 0.$$

For ARX models ($y$- and $u$-part of regressor are present):
Asymptotically unbiased if $w_0$ is white noise.

This condition is satisfied if $\mathcal{S} \in \mathcal{M}$.

For FIR models ($y$-part of regressor is not present):
Asymptotically unbiased if $w_0$ uncorrelated to $u$

This condition is satisfied if $G_0 \in \mathcal{G}$.

**Note** the relation with the consistency results of Lecture 5:

- In an ARX model structure $G_0$ can be estimated consistently if $\mathcal{S} \in \mathcal{M}$;
- In an FIR model structure $G_0$ can be estimated consistently if $G_0 \in \mathcal{G}$
  (because of the independent parametrization of $G$ and $H$)

## Summary

Model structures (ARX,FIR) leading to linear regression problems have become very popular, because of their computational simplicity.

# Approximate Modelling

What can be said about

$$G(q, \theta^*) \quad H(q, \theta^*)$$

if $\mathcal{S} \notin \mathcal{M}$, and even $G_0 \notin \mathcal{G}$?

Can we characterize the approximative properties of identified models if they can not capture all dynamics of the data generating system?

We know (convergence result) that:

$$\theta^* = \arg\min_{\theta} \bar{V}(\theta)$$

and

$$\begin{aligned} \bar{V}(\theta) &= \bar{\mathbb{E}}\varepsilon(t,\theta)^2 \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_\varepsilon(\omega) d\omega \end{aligned}$$

(Parsseval; both expressions are equal to $R_\varepsilon(0)$)

## f-domain expression for the limit model

Combine:

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

with

$$y(t) = G_0(q)u(t) + v(t)$$

then follows:

$$\varepsilon(t,\theta) = H(q,\theta)^{-1}\left[[G_0(q) - G(q,\theta)]u(t) + v(t)\right]$$

Consequence of the identification criterion:

$$\bar{V}(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_{\varepsilon}(\omega)d\omega =$$

$$\bar{V}(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{|G_0(e^{i\omega}) - G(e^{i\omega},\theta)|^2\Phi_u(\omega) + \Phi_v(\omega)}{|H(e^{i\omega},\theta)|^2} d\omega$$

This criterion plays the role of approximation criterion.

## Alternative form

Write

$$\varepsilon(t,\theta) = e(t) + \frac{G_0(q) - G(q,\theta)}{H(q,\theta)} u(t) + \frac{H_0(q) - H(q,\theta)}{H(q,\theta)} e(t)$$

then

$$\theta^* = \arg\min_{\theta}$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \left[ \frac{|G_0(e^{i\omega}) - G(e^{i\omega},\theta)|^2}{|H(e^{i\omega},\theta)|^2} \Phi_u(\omega) + \frac{|H_0(e^{i\omega}) - H(e^{i\omega},\theta)|^2}{|H(e^{i\omega},\theta)|^2} \sigma_e^2 \right] d\omega$$

Expression shows how $\theta^*$ is obtained.

Two mechanisms:

- Minimization of $\dfrac{|G_0(e^{i\omega}) - G(e^{i\omega}, \theta)|^2 \Phi_u(\omega)}{|H(e^{i\omega}, \theta)|^2}$
- Minimization of $\dfrac{|H_0(e^{i\omega}) - H(e^{i\omega}, \theta)|^2 \sigma_e^2}{|H(e^{i\omega}, \theta)|^2}$

Problems are coupled if $H(q, \theta)$ is parametrized.

Observation:

Type of approximation of $G_0$ by $G(q, \hat{\theta}_N)$ is dependent on noise model $H(q, \theta)$.

Special case: fixed noise model $H(q,\theta) = H_*(q)$ (e.g. OE). Then

$$\theta^* = \arg\min_{\theta} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|G_0(e^{i\omega}) - G(e^{i\omega},\theta)|^2 \Phi_u(\omega)}{|H_*(e^{i\omega})|^2} d\omega$$

(second term in integrand is $\theta$-independent).
$\theta^*$ is determined by minimizing the integrated quadratic error $G_0 - G(\theta)$ with weighting function

$$\frac{\Phi_u(\omega)}{|H_*(e^{i\omega})|^2}$$

At those frequencies where the weighting function is large, the model error will be small.

## Example

$$y(t) = G_0(q)u(t)$$

$G_0$ 5th order; $\Phi_u(\omega) = 1$ (white noise).

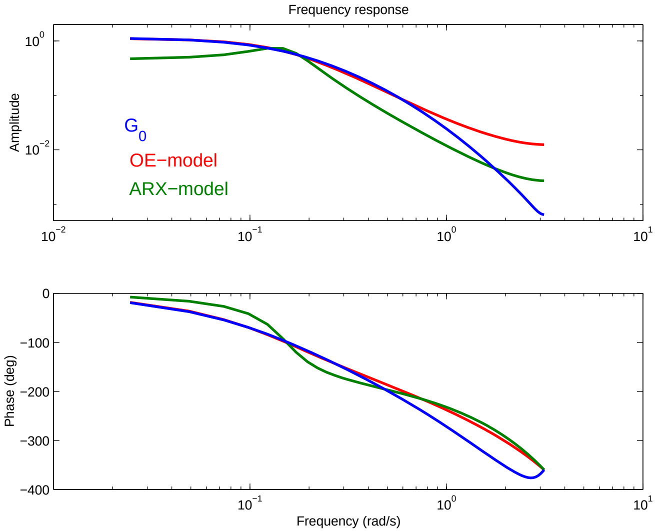
To illustrate the effect of approximation 2nd order models will be estimated:

- OE-model, 2nd order

$$y(t) = \frac{b_1 q^{-1} + b_2 q^{-2}}{1 + f_1 q^{-1} + f_2 q^{-2}} u(t) + e(t)$$

- ARX-model, 2nd order

$$(1 + a_1 q^{-1} + a_2 q^{-2}) y(t) = (b_1 q^{-1} + b_2 q^{-2}) u(t) + e(t)$$


Frequency response
Amplitude
$10^0$
$10^{-2}$
$G_0$
OE−model
ARX−model
$10^{-2}$
$10^{-1}$
$10^0$
$10^1$
Phase (deg)
0
−100
−200
−300
−400
$10^{-1}$
$10^0$
$10^1$
Frequency (rad/s)

Comparison of situations:

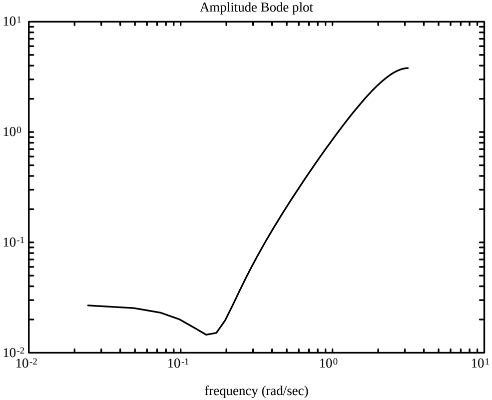
OE-model

$$\min \frac{1}{2\pi}\int_{-\pi}^{\pi} |G_0 - G(\theta)|^2 \Phi_u d\omega$$

ARX-model

$$\min \frac{1}{2\pi}\int_{-\pi}^{\pi} |G_0 - G(\theta)|^2 \Phi_u \cdot |A(e^{i\omega},\theta)|^2 d\omega$$

Additional weighting with (a priori unknown) function.




Weighting function $|A(e^{i\omega}, \hat{\theta}_N)|$ in ARX-case.

General situation:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{|G_0(e^{i\omega})-G(e^{i\omega},\theta)|^2\Phi_u(\omega)+\Phi_v(\omega)}{|H(e^{i\omega},\theta)|^2}d\omega$$

In case of independent parametrization or fixed noise model $H = H_*$:
$G(q,\hat{\rho}_N)$ is obtained through

$$\hat{\rho}_N = \arg\min \frac{1}{2\pi}\int_{-\pi}^{\pi}|G_0 - G(\rho)|^2\frac{\Phi_u}{|H_*|^2}d\omega$$

This holds for OE, BJ, FIR.

In other cases: compromise in which a.o.$\Phi_v$ is playing a role in the approximation of $G_0$.

## Prefiltering of data

Prefiltering of the data with filter $L(q)$, leads to

$$\varepsilon_F(t,\theta) = L(q)\varepsilon(t,\theta)$$

and

$$\Phi_{\varepsilon_F}(\omega) = |L(e^{i\omega})|^2 \cdot \Phi_\varepsilon(\omega)$$

For a fixed noise model the weighting function becomes:

$$\frac{\Phi_u(\omega)|L(e^{i\omega})|^2}{|H_*(e^{i\omega})|^2}$$

Influencing the approximation criterion
(and so the resulting model)
by

- Choice of input spectrum $\Phi_u$
- Choice of prefilter $L$
- Choice of noise model $H$

## Example

$$\mathcal{S}: \quad y(t) = G_0(q)u(t) + e(t)$$

with $G_0$ 4th order with three delays.

We have to use a given set of data $Z^N$ ($N = 5000$) for the identification where $u$ is the sum of a white noise of variance 5 and four high-frequency sinusoids with amplitude 10, with frequencies: 1, 1.3, 1.5 and 2.0 rad/sec.

Objective: Using the given data, identify a good model $G(q, \hat{\theta}_N)$ for $G_0$ in the frequency range [0 0.7] $rad/sec$ in the reduced order model structure:

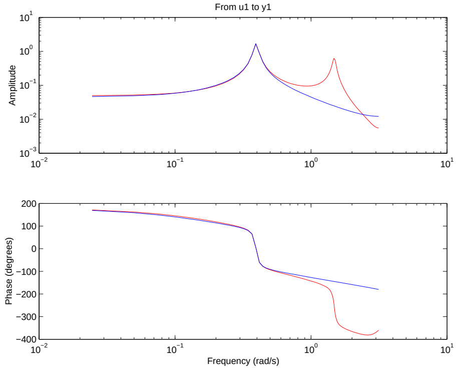
$$\mathcal{M} = OE(n_b = 2, n_f = 2, n_k = 3)$$

Since $Z^N$ is given, the only degree of freedom we have is to use a pre-filter $L$ to shape the bias error.

We want a small bias error in the frequency range [0 0.7] $\Longrightarrow$ choose $L$ such that $|L(e^{i\omega})|^2\Phi_u(\omega)$ is relatively (much) larger in the frequency range [0 0.7] than in [0.7 $\pi$].

$\Longrightarrow$ $L$ Butterworth low pass filter of order 7 and cut-off frequency $0.7 rad/s$
We filter $u$ and $y$ collected from $\mathcal{S}$ by this $L$ and we obtain filtered data with which we perform the identification in $\mathcal{M}$

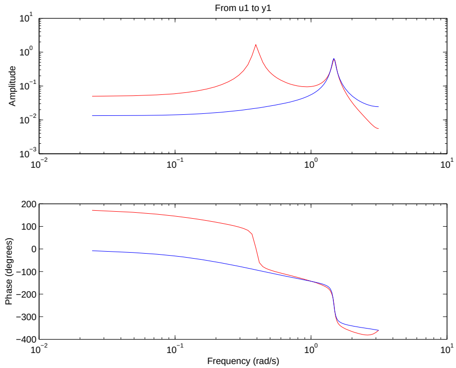
$G_0$ (red) and $G(\hat{\theta}_N)$ (blue) identified with the filtered data.




$\Longrightarrow G(\hat{\theta}_N)$ is OK.

What if we do not use a pre-filter $L$ ?

$G_0$ (red) and $G(\hat{\theta}_N)$ (blue) identified with the data in $Z^N$




$\Longrightarrow G(\hat{\theta}_N)$ is KO

## Summary

- The linearity-in-the-parameters property of ARX and FIR models (linear regression), leads to simple (convex) computational identification algorithms with an analytical solution.
- Consistency result for these model structures follow the earlier presented consistency results.
- If model sets can not capture reality, i.e. $\mathcal{S} \notin \mathcal{M}$, or $G_0 \notin \mathcal{G}$, the asymptotically estimated models can be characterized with an approximation criterion.
- Approximative properties of identified models can be influenced by input signal, prefilter and noise model.