System Identification
Lecture 2

Prediction error method

Paul Van den Hof

Control Systems Group
Eindhoven University of Technology

Contents

Schedule

Apr 9	PVdH	Introduction; concepts
Apr 16	PVdH	Prediction error identification
Apr 23	PVdH	Asymptotic analysis
Apr 30	GB	Bayesian estimation; machine learning
May 7	JS	Frequency domain identification
May 14	JS	Nonlinear identification
May 28	XB	Experiment design
June 4	PVdH	Closed-loops and dynamic networks

One-step-ahead prediction

For a dynamical model $y(t) = G(q)u(t) + H(q)e(t)$ the one-step-ahead prediction $\hat{y}(t|t-1) := \mathbb{E}\{y(t)|y^{t-1}, u^t\}$ is given by

$$\hat{y}(t|t-1) = H^{-1}(q)G(q)u(t) + [1 - H^{-1}(q)]y(t)$$

At the same time:

$$y(t) = \hat{y}(t|t-1) + e(t)$$

The one-step ahead predictor of a dynamic model predicts the output of the model up to the white noise term $e(t)$, called the innovation process.

One-step-ahead prediction is going to be used as a way to measure the deviation/distance between a data set $\{(y(t), u(t))\}_{t=1,\cdots N}$ and a predictor model $(G(q), H(q))$.

If the data has been generated by (G, H) then

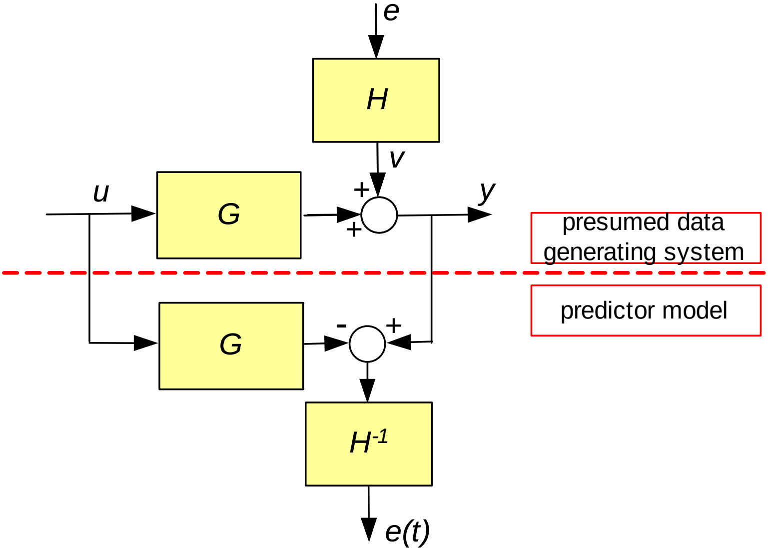
$$y(t) - \hat{y}(t|t-1) = e(t)$$

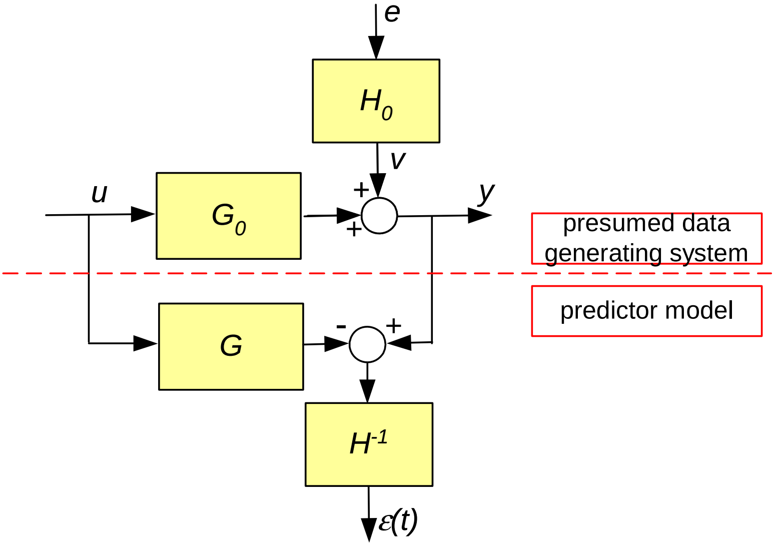
If the data comes from another system, but we calculate the prediction on the basis of (G, H), then

$$y(t) - \hat{y}(t|t-1) = \varepsilon(t)$$

the one-step-ahead prediction error.

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


e
H
v
+
u
G
+
y
presumed data generating system
predictor model
G
-
+
H-1
e(t)


e
H_0
v
u
G_0
+
+
y
presumed data generating system
predictor model
G
-
+
H^{-1}
$\varepsilon(t)$

$\varepsilon(t)$ is going to serve as a basis for the identification criterion.

Black box model structures and model sets

Predictor model: $(G(q), H(q))$

Model set:

$$\mathcal{M} = \{(G(q,\theta), H(q,\theta)), \theta \in \Theta \subset \mathbb{R}^d\}$$

A parametrization is used to represent models by real-valued coefficients

Example of parametrization:

- Coefficients in polynomial fractions $\Leftarrow \theta = (b_1 \ a_1 \ c_1 \ d_1)^T$

$$G(q,\theta) = \frac{b_1 q^{-1}}{1 + a_1 q^{-1}}; \quad H(q,\theta) = \frac{1 + c_1 q^{-1}}{1 + d_1 q^{-1}}$$

- Coefficients in series expansions
- Matrix coefficients in state space models

ARX Model structure

$$G(q,\theta) = \frac{B(q^{-1},\theta)}{A(q^{-1},\theta)}; \qquad H(q,\theta) = \frac{1}{A(q^{-1},\theta)}$$

with

$$\begin{aligned} B(q^{-1},\theta) &= q^{-n_k}\{b_0 + b_1 q^{-1} + \cdots + b_{n_b-1} q^{-n_b+1}\} \\ A(q^{-1},\theta) &= 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a} \\ \theta &= [a_1 \; a_2 \; \cdots a_{n_a} \; b_0 \; b_1 \; \cdots \; b_{n_b-1}]^T. \end{aligned}$$

n_a, n_b are the number of parameters in the A and B polynomial.
n_k number of time delays

Predictor:

$$\hat{y}(t|t-1;\theta) = B(q^{-1},\theta)u(t) + [1 - A(q^{-1},\theta)]y(t)$$

Model structures

Model structure	$G(q, \theta)$	$H(q, \theta)$
ARX	$\frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)}$	$\frac{1}{A(q^{-1}, \theta)}$
ARMAX	$\frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)}$	$\frac{C(q^{-1}, \theta)}{A(q^{-1}, \theta)}$
OE - Output Error	$\frac{B(q^{-1}, \theta)}{F(q^{-1}, \theta)}$	1
FIR	$B(q^{-1}, \theta)$	1
BJ - Box-Jenkins	$\frac{B(q^{-1}, \theta)}{F(q^{-1}, \theta)}$	$\frac{C(q^{-1}, \theta)}{D(q^{-1}, \theta)}$

Properties of model structures

- Linearity-in-the-parameters (ARX)

$$\begin{aligned}\hat{y}(t|t-1;\theta) &= B(q^{-1})u(t) + (1 - A(q^{-1}))y(t) \\ &= \phi^T(t)\theta\end{aligned}$$

  is a linear function in θ.
  $\Rightarrow$ Important computational advantages.

- Independent parametrization of $G(q,\theta)$ en $H(q,\theta)$
  There are no common parameters in G and H.
  $\Rightarrow$ Advantages for independent identification of G and H.

Both properties to be utilized later on.

System in the model set

Data-generating system $\mathcal{S}$: $[G_0, H_0]$
Model set: $\mathcal{M}$: $\{[G(q,\theta), H(q,\theta)], \theta \in \Theta \subset \mathbb{R}^d\}$.

$$\mathcal{S} \in \mathcal{M}$$

denotes that the data generating system can exactly be represented within $\mathcal{M}$, i.e. $\exists \theta_0 \in \Theta$ such that

$$\begin{aligned} G(q,\theta_0) &= G_0(q) \\ H(q,\theta_0) &= H_0(q) \end{aligned}$$

The notion $G_0 \in \mathcal{G}$ with $\mathcal{G} = \{G(q,\theta), \theta \in \Theta \subset \mathbb{R}^d\}$ denotes that only G_0 can exactly be represented, i.e.

$$G(q,\theta_0) = G_0(q)$$

Identification criterion

Identification criterion

Consider the data-generating system:

$$y(t) = G_0(q)u(t) + H_0(q)e(t)$$

and the parametrized model $G(q,\theta), H(q,\theta)$.
Denote: $\bar{V}(\theta) = \bar{\mathbb{E}}\varepsilon^2(t,\theta)$.
Then: $\bar{V}(\theta) \geq \sigma_e^2$, with equality for $\hat{\theta}$ if

$$\begin{aligned} G(q,\hat{\theta}) &= G_0(q) \\ H(q,\hat{\theta}) &= H_0(q) \end{aligned}$$

Uniqueness of this solution requires some more conditions, to be specified later.

Reasoning:

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

u and e uncorrelated;
First term becomes 0 for $G(q,\hat{\theta}) = G_0(q)$.

$$\frac{H_0(q)}{H(q,\theta)} e(t) = [1 + \gamma_1(\theta) q^{-1} + \gamma_2(\theta) q^{-2} + \cdots] e(t)$$

$$\bar{\mathbb{E}} \varepsilon^2(t,\theta) = [1 + \gamma_1(\theta)^2 + \gamma_2(\theta)^2 \cdots] \sigma_e^2$$

is minimal for $\gamma_1(\hat{\theta}) = \gamma_2(\hat{\theta}) = \cdots = 0$.

$\Rightarrow$ Cost function minimum σ_e^2 is achieved for

$$G(q,\hat{\theta}) = G_0 \quad \text{and} \quad H(q,\hat{\theta}) = H_0.$$

Identification criterion

Power of prediction error is estimated from a data sequence through the quadratic criterion:

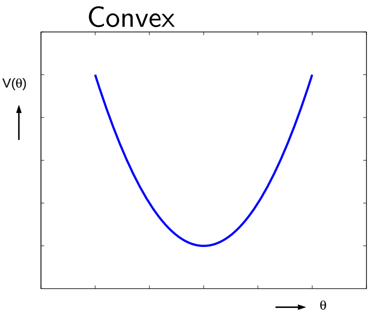
$$V_N(\theta, Z^N) = \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t, \theta)$$

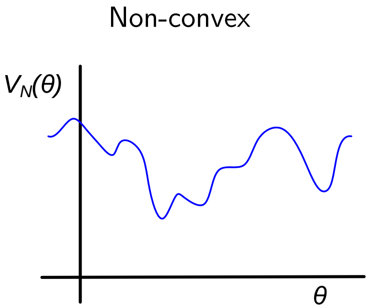
Parameter estimation through minimizing V_N:

$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta, Z^N)$$

Optimization problem that is convex or not....

Optimization:





Optimization is convex if the criterion is quadratic in θ.
In that situation, the solution can be found analytically.

Linear regression

If the predictor is linear-in-the-parameters (ARX,FIR):

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

then

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

and an analytical solution is obtained:

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

Assumptions for analysis:

- Data has been generated by a linear dynamical system, possibly operating in a (stabilized) feedback situation, with bounded excitation signals, and v having bounded moments of order 4;
- $\mathcal{M}$ is uniformly stable ($\rightarrow$ the predictor filters and their derivatives w.r.t. θ are uniformly stable).

Convergence result

Under the above assumptions, $\hat{\theta}_N = \arg\min_\theta V_N(\theta, Z^N)$, for $N \rightarrow \infty$ converges with probability 1 to the minimizing argument θ^* of

$$\bar{V}(\theta) := \bar{\mathbb{E}}\varepsilon^2(t, \theta)$$

Differently stated:

$$\hat{\theta}_N \rightarrow \theta^* := \arg\min_{\theta} \bar{\mathbb{E}}\varepsilon^2(t,\theta) \quad \text{w.p.} \quad 1 \text{ for } N \rightarrow \infty.$$

The estimate converges to the best possible predictor (in the sense of $\bar{\mathbb{E}}\varepsilon^2$) within the model set $\mathcal{M}$, and becomes independent of the particular noise sequence (variance goes to 0)

Persistence of excitation

Definition - persistently exciting input

A quasi-stationary signal u is persistently exciting of order n if the spectral density $\Phi_u(\omega)$ is unequal to 0 in n points in the interval $(-\pi, \pi]$.

Example

The signal

$$u(t) = sin(\omega_0 t)$$

is persistently exciting of order 2.
(Φ_u has a contribution in $\pm\omega_0$)
2 degrees of freedom (amplitude and phase).

Equivalent formulation

Proposition

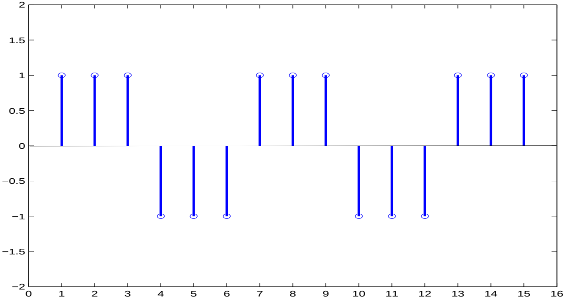
A quasi-stationary signal u is persistently exciting of order n if and only if the (Toeplitz) matrix

$$\bar{R}_n := \begin{bmatrix} R_u(0) & R_u(1) & \cdots & R_u(n-1) \\ R_u(1) & R_u(0) & \cdots & R_u(n-2) \\ \vdots & \ddots & \ddots & \vdots \\ R_u(n-1) & \cdots & R_u(1) & R_u(0) \end{bmatrix}$$

with $R_u(\tau) := \bar{\mathbb{E}}\{u(t)u(t-\tau)\}$, is non-singular.

Example:
A white noise process $(R_u(\tau) = \delta(\tau))$ is persistently exciting of **any** order.
Note: $\bar{R}_n = I_n$

Example block signal



$$R_u(0) = 1 \quad R_u(1) = \tfrac{1}{3} \quad R_u(2) = -\tfrac{1}{3} \quad R_u(3) = -1$$

$$R_u(4) = -\tfrac{1}{3} \quad R_u(5) = \tfrac{1}{3} \quad R_u(6) = 1 \quad \text{etcetera}$$

$$\bar{R}_4 = \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} & -1 \\ \frac{1}{3} & 1 & \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & 1 & \frac{1}{3} \\ -1 & -\frac{1}{3} & \frac{1}{3} & 1 \end{bmatrix}$$

$\bar{R}_3$ is regular, $\bar{R}_4$ is singular
so u is p.e. of order 3

A step signal:

$$\begin{aligned} u(t) &= 0 \quad t < 0 \\ &= c \quad t \geq 0 \end{aligned}$$

is actually a transient signal, and therefore does not fit well into the "persistent signals" framework.
Although a step response reveals all dynamics of an LTI system, a step signal is p.e. of order 1, only having a spectral contribution at $\omega = 0$.
So it only persistently excites frequency 0.

Consistency result for $\mathcal{S} \in \mathcal{M}$

Consistency for $\mathcal{S} \in \mathcal{M}$

Let data be generated by a system (G_0, H_0), and consider a model set $\mathcal{M}$ with $G(q, \theta)$ parametrized according to

$$G(q,\theta) = q^{-n_k} \cdot \frac{b_0 + b_1 q^{-1} + \cdots + b_{n_b-1} q^{-n_b+1}}{1 + f_1 q^{-1} + \cdots + f_{n_f} q^{-n_f}}.$$

If $\mathcal{S} \in \mathcal{M}$ and u is persistently exciting of order $\geq n_f + n_b$ then

$$\begin{aligned} G(e^{i\omega}, \theta^*) &= G_0(e^{i\omega}) \\ H(e^{i\omega}, \theta^*) &= H_0(e^{i\omega}) \qquad -\pi \leq \omega \leq \pi \end{aligned}$$

Note that n_b and n_f are number of parameters in numerator/denominator of $G(q, \theta)$.

Consistency result for $G_0 \in \mathcal{G}$

Consistency result for $G_0 \in \mathcal{G}$

Consider the same situation as on the previous slide.
If

- $G_0 \in \mathcal{G}$, and
- G and H are independently parametrized in $\mathcal{M}$, and
- u is persistently exciting of order $\geq n_f + n_b$

then

$$G(e^{i\omega}, \theta^*) \;=\; G_0(e^{i\omega}) \qquad -\pi \leq \omega \leq \pi$$

G_0 can be estimated consistently, irrespective of the question whether H_0 can be modeled exactly.

Independent parametrization

Model structure	$G(q,\theta)$	$H(q,\theta)$
OE	B/F	1
FIR	B	1
BJ	B/F	C/D

Not: ARX, ARMAX because of common denominator in G en H.

Asymptotic distribution

The asymptotic pdf of the estimate $\hat{\theta}_N$ is specified as follows:

Asymptotic distribution

Consider the data assumptions as present in the convergence result. Then for $N \to \infty$,

$$\sqrt{N}(\hat{\theta}_N - \theta^*) \to \mathcal{N}(0, P_\theta)$$

where

$$\begin{aligned} P_\theta &= [\bar{V}''(\theta^*)]^{-1} Q [\bar{V}''(\theta^*)]^{-1} \\ Q &= \lim_{N\to\infty} N \cdot \mathbb{E}\{[V_N'(\theta^*, Z^N)][V_N'(\theta^*, Z^N)]^T\} \end{aligned}$$

while $(\cdot)'$, $(\cdot)''$ respectively denote first and second derivative with respect to θ.

Justification of normal distribution

As example: linear regression structure, $\mathcal{S} \in \mathcal{M}$, leading to

$$\hat{\theta}_N - \theta_0 = \left[\frac{1}{N}\sum_t \varphi(t)\varphi^T(t)\right]^{-1} \cdot \frac{1}{N}\sum_t \varphi(t)e(t)$$

with $e(t)$ a random variable and e a white noise process.
Then, for $N \to \infty$, the above expression becomes a weighted sum of an infinite number of random variables with a fixed distribution.
According to the law of large numbers, this weighted sum will get a Gaussian pdf.

P_θ/N has the interpretation of covariance matrix of the parameter estimate:

$$P_\theta/N = \mathbb{E}[(\hat{\theta}_N - \theta_0)(\hat{\theta}_N - \theta_0)^T].$$

quantifying the variability of the estimate

Covariance matrix in case of $\mathcal{S} \in \mathcal{M}$

For the situation $\theta^* = \theta_0$ ($\mathcal{S} \in \mathcal{M}$) it holds that

$$P_\theta = \sigma_e^2 \cdot \left[\bar{\mathbb{E}}\psi(t,\theta_0)\psi^T(t,\theta_0)\right]^{-1}$$

$$\psi(t,\theta_0) = \left.\frac{\partial \hat{y}(t|t-1;\theta)}{\partial \theta}\right|_{\theta=\theta_0} = -\left.\frac{\partial \varepsilon(t,\theta)}{\partial \theta}\right|_{\theta=\theta_0}$$

Justification of the covariance matrix expression

In the linear regression situation, $\mathcal{S} \in \mathcal{M}$, we have

$$\hat{\theta}_N - \theta_0 = \left[\frac{1}{N}\sum_t \varphi(t)\varphi^T(t)\right]^{-1} \cdot \frac{1}{N}\sum_t \varphi(t)e(t)$$

so that $\bar{\mathbb{E}}\tilde{\theta}\tilde{\theta}^T$ can be written as

$$\bar{\mathbb{E}}\left\{\left[\frac{1}{N}\sum_t \varphi(t)\varphi^T(t)\right]^{-1} \frac{1}{N}\sum_t \varphi(t)e(t) \cdot \right.$$

$$\left. \cdot \frac{1}{N}\sum_t \varphi^T(t)e(t)\left[\frac{1}{N}\sum_t \varphi(t)\varphi^T(t)\right]^{-1}\right\}$$

With $\varphi(t)$ deterministic and $\mathbb{E}e^2(t) = \sigma_e^2$, then

$$P_\theta := \bar{\mathbb{E}}\tilde{\theta}\tilde{\theta}^T = \sigma_e^2 \cdot \left[\frac{1}{N}\sum_t \varphi(t)\varphi^T(t)\right]^{-1}$$

In linear regression: $\hat{y}(t|t-1;\theta) = \varphi^T(t)\theta$ so that

$$\varphi(t) = \frac{\partial \hat{y}(t|t-1;\theta)}{\partial \theta}.$$

In general structures, the role of $\varphi(t)$ is replaced by

$$\psi(t,\theta_0) = \left.\frac{\partial \hat{y}(t|t-1;\theta)}{\partial \theta}\right|_{\theta=\theta_0}.$$

end justification

$P_\theta/N = E(\hat{\theta}_N - \theta_0)(\hat{\theta}_N - \theta_0)^T$ can be estimated from the data and $\hat{\theta}_N$ as:

$$\hat{P}_\theta = \hat{\sigma}_e^2 [\frac{1}{N}\sum_{t=1}^{N}\psi(t,\hat{\theta}_N)\psi^T(t,\hat{\theta}_N)]^{-1}$$

$$\hat{\sigma}_e^2 = \frac{1}{N}\sum_{t=1}^{N}\varepsilon^2(t,\hat{\theta}_N)$$

with $\hat{\sigma}_e^2$ an estimate of σ_e^2.

Covariance matrix in case of $G_0 \in \mathcal{G}$

Let $\mathcal{M} = \{(G(q,\rho), H(q,\eta)),\ \theta = \begin{pmatrix} \rho \\ \eta \end{pmatrix} \in \Theta\}$ and

$\theta^* = \begin{pmatrix} \rho_0 \\ \eta^* \end{pmatrix}$. Then

$$\sqrt{N}(\hat{\rho}_N - \rho_0) \in As\,\mathcal{N}(0, P_\rho)$$

with

$$P_\rho = \sigma_e^2[\bar{\mathbb{E}}\psi_\rho(t)\psi_\rho^T(t)]^{-1}[\bar{\mathbb{E}}\tilde{\psi}(t)\tilde{\psi}^T(t)][\bar{\mathbb{E}}\psi_\rho(t)\psi_\rho^T(t)]^{-1}$$

$$\psi_\rho(t) = H^{-1}(q,\eta^*)\,\frac{\partial}{\partial\rho}G(q,\rho)\Big|_{\rho=\rho_0} u(t)$$

$$\tilde{\psi}(t) = \sum_{i=0}^{\infty} f_i\psi_\rho(t+i); \qquad F(z) := \frac{H_0(z)}{H(z,\eta^*)} = \sum_{i=0}^{\infty} f_i z^{-i}.$$

Parameter uncertainty regions

The asymptotical normal distribution can be used to quantify parameter uncertainty regions.
Start with:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \rightarrow \mathcal{N}(0, P_\theta)$$

Let P_θ^{-1} be decomposed as $P_\theta^{-1} = R_\theta^T R_\theta$, then

$$\sqrt{N} R_\theta(\hat{\theta}_N - \theta_0) \rightarrow \mathcal{N}(0, I_n)$$

i.e. a vector of n standard (independent) normally distributed random variables.

If $x \in \mathcal{N}(0, I_n)$ then for the sum of the quadratic variables:

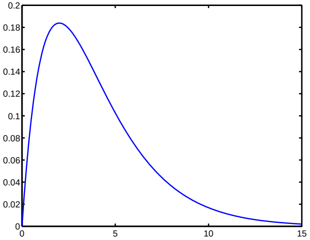
$$x^T x \in \chi^2(n)$$

i.e. follows a χ^2 distribution, scalar valued.

This implies that the sum of quadratic variables

$$N(\hat{\theta}_N - \theta_0)^T P_\theta^{-1}(\hat{\theta}_N - \theta_0) \to \chi^2(n).$$

Example χ^2-distribution



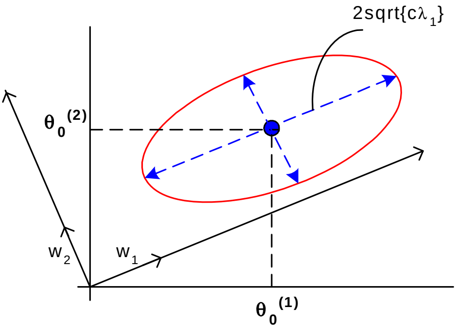
For $\mathbf{x} \in \chi^2(n)$:

$$\int_0^{c_\chi(\alpha,n)} x \, dx = \alpha$$

Note that contour lines:

$$N(\hat{\theta}_N - \theta_0)^T P_\theta^{-1}(\hat{\theta}_N - \theta_0) = c$$

are ellipsoidal contour lines of the multivariable Gaussian distribution:



(P_θ has eigenvalues λ_i and eigenvectors w_i)

while the level of probability related to the event

$$N(\theta - \theta_0)^T P_\theta^{-1}(\theta - \theta_0) < c$$

is determined by the $\chi^2(n)$-distribution.

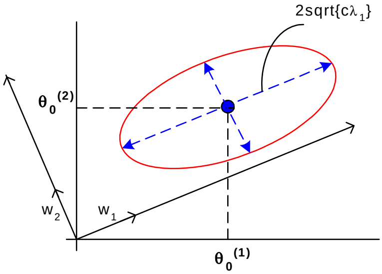
Denote with $c_\chi(\alpha, n)$ the value that satisfies

$$Pr(x \leq c_\chi(\alpha, n)) = \alpha$$

for $x \in \chi^2(n)$.
Then $\hat{\theta}_N \in \mathcal{D}_{\theta_0}$ with probability α, with

$$\mathcal{D}_{\theta_0} = \{\theta \mid (\theta - \theta_0)^T P_\theta^{-1}(\theta - \theta_0) < \frac{c_\chi(\alpha, n)}{N}\}$$

ellipsoid centered at θ_0.



Alternatively, denote:

$$\mathcal{D}_{\hat{\theta}_N} = \{\theta \mid (\theta - \hat{\theta}_N)^T P_\theta^{-1} (\theta - \hat{\theta}_N) < \frac{c_\chi(\alpha, n)}{N}\}$$

ellipsoid centered at $\hat{\theta}_N$
then it can simply be verified that

$$\hat{\theta}_N \in \mathcal{D}_{\theta_0} \Leftrightarrow \theta_0 \in \mathcal{D}_{\hat{\theta}_N}$$

so that $\theta_0 \in \mathcal{D}_{\hat{\theta}_N}$ with probability α.

Uncertainty regions for frequency responses of estimated models

The identified parameter vector $\hat{\theta}_N$ is a random variable distributed as $\hat{\theta}_N \sim As\mathcal{N}(\theta_0, P_\theta/N) \Longrightarrow$
the identified models (frequency responses) $G(e^{i\omega}, \hat{\theta}_N)$ (and $H(e^{i\omega}, \hat{\theta}_N)$) are also random variables:

- $G(e^{i\omega}, \hat{\theta}_N)$ is an (asymptotically) unbiased estimate of $G(e^{i\omega}, \theta_0)$
- the variance of $G(e^{i\omega}, \hat{\theta}_N)$ is defined in the frequency domain as:

$$cov(G(e^{i\omega}, \hat{\theta}_N)) := E\left(|G(e^{i\omega}, \hat{\theta}_N) - G(e^{i\omega}, \theta_0)|^2\right)$$

Covariance of the frequency response

$$cov\{G(e^{i\omega}, \hat{\theta}_N\} = \mathbb{E}\left(|G(e^{i\omega}, \hat{\theta}_N) - G(e^{i\omega}, \theta_0)|^2\right)$$

is obtained by

$$cov\{G(e^{i\omega}, \hat{\theta}_N\} \sim \left.\frac{\partial G(e^{i\omega}, \theta)}{\partial \theta}^*\right|_{\theta_0} \cdot \frac{P_\theta}{N} \cdot \left.\frac{\partial G(e^{i\omega}, \theta)}{\partial \theta}\right|_{\theta_0}$$

which is a first order Taylor approximation that is exact for models that are linear in the parameters.
($(\cdot)^*$ is complex conjugate transpose)
$cov\{G(e^{i\omega}, \hat{\theta}_N\}$ can be estimated by replacing P_θ by $\hat{P}_\theta$, and θ_0 by $\hat{\theta}_N$.

Proof of the expression of $cov(G(e^{i\omega}, \hat{\theta}_N))$

First order (Taylor) approximation:

$$G(e^{i\omega}, \hat{\theta}_N) \approx G(e^{i\omega}, \theta_0) + \left(\hat{\theta}_N - \theta_0\right)^T \Lambda_G(e^{i\omega}, \theta_0)$$

with $\Lambda(e^{i\omega}, \theta_0) = \left.\frac{\partial G(e^{i\omega}, \theta)}{\partial \theta}\right|_{\theta_0}$.

Consequently: $|G(e^{i\omega}, \hat{\theta}_N) - G(e^{i\omega}, \theta_0)|^2 = ..$

$$\begin{aligned} ... &= (G(e^{i\omega}, \hat{\theta}_N) - G(e^{i\omega}, \theta_0))^*(G(e^{i\omega}, \hat{\theta}_N) - G(e^{i\omega}, \theta_0)) \\ &\approx \Lambda_G(e^{i\omega}, \theta_0)^* \; \left(\hat{\theta}_N - \theta_0\right) \left(\hat{\theta}_N - \theta_0\right)^T \Lambda_G(e^{i\omega}, \theta_0) \end{aligned}$$

$$E(.) = \Lambda_G(e^{i\omega}, \theta_0)^* \cdot \frac{P_\theta}{N} \cdot \Lambda_G(e^{i\omega}, \theta_0)$$

Observation

The larger N and/or the larger the power of $u(t)$, the smaller $cov(G(e^{i\omega}, \hat{\theta}_N))$

direct consequence of the fact that P_θ / N has this property

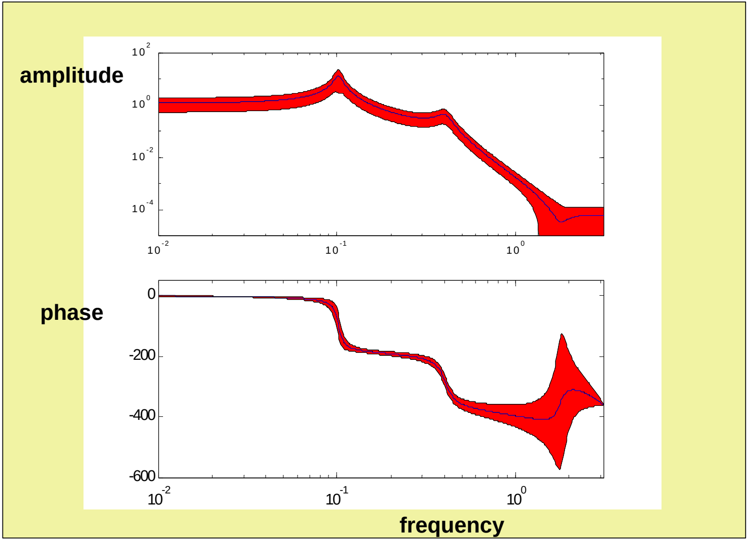
Separate bounds for amplitude and phase of $G(e^{i\omega}, \hat{\theta}_N)$

$$
\begin{aligned}
f_{a,\omega}(\theta) &= |G(e^{i\omega}, \theta)|, \\
f_{p,\omega}(\theta) &= arg\{G(e^{i\omega}, \theta)\}
\end{aligned}
$$

then covariance information on $f_{a,\omega}$ and $f_{p,\omega}$ is obtained from the first order approximations

$$
\begin{aligned}
Cov\{f_{a,\omega}(\hat{\theta}_N)\} &\approx \left.\frac{\partial f_{a,\omega}(\theta)}{\partial \theta}^T\right|_{\theta=\hat{\theta}_N} \frac{P_\theta}{N} \left.\frac{\partial f_{a,\omega}(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_N} \\
Cov\{f_{p,\omega}(\hat{\theta}_N)\} &\approx \left.\frac{\partial f_{p,\omega}(\theta)}{\partial \theta}^T\right|_{\theta=\hat{\theta}_N} \frac{P_\theta}{N} \left.\frac{\partial f_{p,\omega}(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_N}
\end{aligned}
$$

This is implemented in Matlab's System Identification Toolbox (3σ bounds).


amplitude
10^2
10^0
10^-2
10^-4
10^-2
10^-1
10^0
phase
0
-200
-400
-600
10^-2
10^-1
10^0
frequency

Remarks

- Estimated uncertainty bounds are reliable only if the models are correct (validated), i.e. $\mathcal{S} \in \mathcal{M}$, or $G_0 \in \mathcal{G}$.
- If the uncertainty bound for $G(e^{i\omega}, \hat{\theta}_N)$ is too large in a particular frequency range $\rightarrow$
  (a) redo experiment with increased $\Phi_u(\omega)$ at those frequencies, or
  (b) increase number of data (length of experiment)
- Related analysis can be developed for $H(e^{i\omega}, \hat{\theta}_N)$

Asymptotic-in-order covariance of frequency response

Consider the situation as in the convergence result.
Let $\mathcal{S} \in \mathcal{M}$, and let $G(q,\theta)$ and $H(q,\theta)$ have McMillan degree n.
Denote $T(q,\theta) := \begin{bmatrix} G(q,\theta) & H(q,\theta) \end{bmatrix}$.
If $n, N \to \infty$, and $n/N \to 0$, then

$$Cov\ (\hat{T}_N(e^{i\omega})) \sim \frac{n}{N}\Phi_v(\omega)[\Phi_\chi(\omega)]^{-1}$$

with $\chi(t) = [u(t)\ e(t)]^T$

If the system operates in open loop, u and e are uncorrelated, and the result simplifies to

$$
\begin{aligned}
Cov\ (\hat{G}_N(e^{i\omega})) &\sim \frac{n}{N}\frac{\Phi_v(\omega)}{\Phi_u(\omega)} \\
Cov\ (\hat{H}_N(e^{i\omega})) &\sim \frac{n}{N}\frac{\Phi_v(\omega)}{\sigma_e^2} = \frac{n}{N}|H_0(e^{i\omega})|^2
\end{aligned}
$$

But note the double limit, $n, N \to \infty$.

Summary - discussion

- Comprehensive PE theory
- Asymptotic results in N, convergence, consistency, covariance matrix
- Variance expressions leading to uncertainty sets
- Principle reasoning: find the "best" model in a given structured model set (of particular order).
- Relation with Maximum Likelihood estimation to be shown (next)
- as well as model order selection.