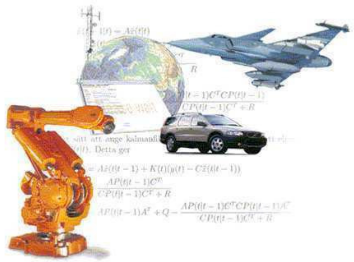
Sysid Course VT1 2018
Chapters 8 - 9



Lennart Ljung

Automatic Control, ISY, Linköpings Universitet

Basic Tools for Analysis

For a stationary stochastic process $e(\cdot)$ under mild conditions

- Law of large numbers (LLN)
  - $\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} e(t) = Ee(t)$
- Central limit theorem (CLT)
  If $e(t)$ has zero mean:
  - $\frac{1}{\sqrt{N}}\sum_{t=1}^{N} e(t)$ converges in distribution to the normal (Gaussian) distribution with zero mean and variance $\overline{\lambda} = \lim \frac{1}{N}\sum_{t,s=1}^{N} Ee(t)e(s)$.
    "$\frac{1}{\sqrt{N}}\sum_{t=1}^{N} e(t) \to N(0,\overline{\lambda})$"

Asymptotic Analysis: Probabilistic Setup

"Analytical Monte-Carlo Experiment": Curve fitting problem: Find an estimate of $g_0(x)$ in a paratererization $g(x,\theta)$. For a given $g_0(\cdot)$ and a given sequence x_t collect the data

$$y(t) = g_0(x_t) + e(t), \quad Ee(t)^2 = \lambda$$

where the stochastic process $e(\cdot)$ obeys the LLN and CLT and has variance λ. Form the estimate

$$\hat{\theta}_N = \arg\min \frac{1}{N}\sum_{t=1}^{N}(y(t) - g(x_t,\theta))^2$$

$$\hat{g}_N(x) = g(x,\hat{\theta}_N)$$

Then $\hat{\theta}_N$ and $\hat{g}_N(x)$ are random variables with properties inherited from e. What can be said about their distributions?

.

Asymptotic Analysis: Basic Facts – BIAS

Except for very simple parameterizations $g(x,\theta)$, the distribution of $\hat{\theta}_N$ cannot be calculated (mainly due to "arg min").
However its asymptotic distribution as $N \to \infty$ can be established:

- $\overline{E}$ = averaging over x_k: $\overline{E}f(x) = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} f(x_k)$
- $H(\theta) = \lim_{N\to\infty} H_N(\theta) = \overline{E}|g_0(x_t) - g(x_t,\theta)|^2$
- With relevance weighting: $H(\theta) = \overline{E}L(x_t)|g_0(x_t) - g(x_t,\theta)|^2$
- Best possible model in parameterization: $\theta^* = \arg\min H(\theta)$
- If $H(\theta^*) = 0$ we have a perfect curve fit, otherwise there be some bias in the curve fit.
- Main Result: $\lim_{N\to\infty} \hat{\theta}_N = \theta^*$

Proof: Formal Calculations

$$V_N(\theta) = \frac{1}{N}\sum(g_0(x_t) + e(t) - g(x_t,\theta))^2$$

$$= \frac{1}{N}\sum(g_0(x_t) - g(x_t,\theta))^2 + \frac{1}{N}\sum e^2(t) + \frac{2}{N}\sum(g_0(x_t) - g(x_t,\theta))e(t)$$

LLN: $\frac{1}{N}\sum(g_0(x_t) - g(x_t,\theta))e(t) \to 0$ (uniformly in θ!)

so $V_N(\theta) \to H(\theta)$ as $N \to \infty$

Asymptotic Analysis: Basic Facts – VARIANCE

Suppose the limit model is correct: $g(x,\theta^*) \approx g_0(x)$ and e white noise with variance λ:

- The asymptotic distribution of $\sqrt{N}(\hat{\theta}_N - \theta^*)$ is normal with zero mean and covariance matrix
  $P = \lambda[\bar{E}\psi(t)\psi^T(t)]^{-1}, \quad \psi(t) = \frac{d}{d\theta}g(x_t,\theta^*)$
- "Cov $\hat{\theta}_N \sim \frac{\lambda}{N}[\bar{E}\psi(t)\psi^T(t)]^{-1}$"

Proof: Formal Calculations

$$0 = V_N'(\hat{\theta}_N) = V_N'(\theta^*) + V_N''(\theta^*)(\hat{\theta}_N - \theta^*)$$

$$(\hat{\theta}_N - \theta^*) = -[V_N''(\theta^*)]^{-1}V_N'(\theta^*)$$

$$V_N'(\theta) = \frac{2}{N}\sum(y(t) - g(x_t,\theta))g'(x_t,\theta)$$

$$V_N'(\theta^*) = \frac{2}{N}\sum e(t)\psi(t)$$

LLN: $V_N''(\theta^*) = \frac{2}{N}\sum\psi(t)\psi^T(t) + \frac{2}{N}\sum e(t)g''(x_t,\theta^*) \to 2\bar{E}\psi\psi^T$

CLT: $\frac{1}{\sqrt{N}}\sum e(t)\psi(t) \to N(0,\lambda\bar{E}\psi\psi^T)$

$$\sqrt{N}(\hat{\theta}_N - \theta^*) \to N(0,\lambda[\bar{E}\psi\psi^T]^{-1})$$

A Remarkable Observation

Recall the curve fit $H(x,\theta) = |g_0(x) - g(x,\theta)|^2$,
$H(\theta) = \lim_{N\to\infty}\frac{1}{N}\sum H(x_t,\theta)$ (For the x-sequence of the estimation data.)
$H(\hat{\theta}_N)$ is a random variable, since the estimate depends on the e-sequence, and

$$EH(\hat{\theta}_N) = H(\theta^*) + \lambda\frac{d}{N}$$

where d is the number of estimated parameters independently of the parameterization!
(Proof: ....)

Proof: Formal Calculations

$$g_0(x) = g(x,\theta^*) \quad (\text{ assumption })$$
$$H(\hat\theta_N) = H(\theta^*) + H'(\theta^*)(\hat\theta_N - \theta^*) + \frac{1}{2}(\hat\theta_N - \theta^*)^T H''(\theta^*)(\hat\theta_N - \theta^*)$$
$$H'(\theta^*) = 0 \quad (\theta^* \text{ minimizes } H(\theta))$$
$$H'(\theta) = \frac{2}{N}\sum (g_0(x_t) - g(x_t,\theta))g'(x_t,\theta)^T$$
$$H''(\theta^*) = \frac{2}{N}\sum g'(x_t,\theta^*)g'(x_t,\theta^*)^T = 2\bar E\psi(t)\psi^T(t)$$
$$EH(\hat\theta_N) = H(\theta^*) + E\frac{1}{2}(\hat\theta_N - \theta^*)^T H''(\theta^*)(\hat\theta_N - \theta^*)$$
$$E\text{trace}[\frac{1}{2}(\hat\theta_N - \theta^*)^T H''(\theta^*)(\hat\theta_N - \theta^*)] = E\text{trace}[\frac{1}{2}H''(\theta^*)(\hat\theta_N - \theta^*)(\hat\theta_N - \theta^*)^T]$$
$$= \text{trace}\,\frac{1}{2}H''(\theta^*)\text{Cov}\,\hat\theta_N = \frac{\lambda}{N}\text{trace}\left[(\bar E\psi\psi^T)(\bar E\psi\psi^T)^{-1}\right] = d\frac{\lambda}{N}$$

Asymptotic Properties of Estimated Models

- Assume that data have been generated by $y(t) = G_0(q)u(t) + H_0(q)e(t)$, where e is white noise with variance λ.
-

$$\hat\theta_N = \arg\min \sum |L(q)\varepsilon(t,\theta)|^2$$
$$\varepsilon(t,\theta) = H^{-1}(q,\theta)[y(t) - G(q,\theta)u(t)]$$

- Properties of the estimate $\hat\theta_N$
- A parameter-free assessment of model quality for linear systems and models
- How to affect the model quality?

Properties of the Estimate $\hat\theta_N$ as $N \to \infty$

- 1. $\hat\theta_N \to \theta^* = \arg\min \mathrm{E}[L(q)\varepsilon(t,\theta)]^2$
  "The estimate converges to the best possible approximation of the true system"
  "Best possible": ....

  If $\varepsilon(t,\hat\theta_N) \approx$ white noise then

  $\mathrm{Cov}\ \hat\theta_N \approx \frac{\lambda}{N}[\mathrm{E}\psi(t)\psi^T(t)]^{-1}$
  $\lambda = \mathrm{E}\varepsilon^2(t,\hat\theta_N)$
  $\psi(t) = \frac{d}{d\theta}\hat y(t|\theta)$ d-dimensional vector

  NB! This covariance can be estimated!
  Also, asymptotic normality of $\sqrt{N}(\hat\theta_N - \theta^*)$

Quality in The Frequency Domain: Bias

True system: $y(t) = G_0(q)u(t) + v(t)$
Model: $\hat G_N(e^{i\omega}) = G(e^{i\omega},\hat\theta_N), \quad \hat H_N(e^{i\omega}) = H(e^{i\omega},\hat\theta_N)$
Translate properties of $\hat\theta_N$ to $\hat G_N$ & $\hat H_N$

$$\hat G_N(e^{i\omega}) \to G^*(e^{i\omega}), \quad \hat\theta_N \to \theta^* \qquad \text{as } N \to \infty$$

$$\theta^* \approx \arg\min \int_{-\pi}^{\pi} |\ G_0(e^{i\omega}) - G(e^{i\omega},\theta)\ |^2 \cdot \frac{\Phi_u(\omega)\cdot |\ L(e^{i\omega})\ |^2}{|\ H(e^{i\omega},\theta^*)\ |^2} d\omega$$

$G^*(e^{i\omega})$ is closest to $G_0(e^{i\omega})$ in the norm $Q(\omega) = \dfrac{\Phi_u(\omega)|L(e^{i\omega})|^2}{|H(e^{i\omega},\theta^*)|^2}$

Φ_u : Input spectrum, $\qquad L$: Pre-filter

A Paradox

A high order system is approximated in the ARX-structure

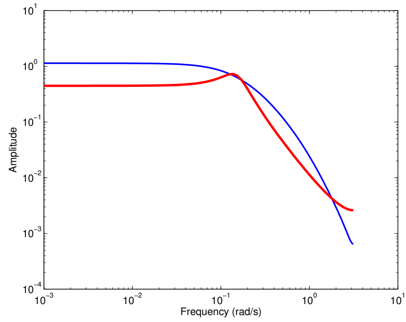
$$(1 + a_1 q^{-1} + a_2 q^{-2}) y(t) = (b_1 q^{-1} + b_2 q^{-2}) u(t) + e(t)$$

and in the OE-structure

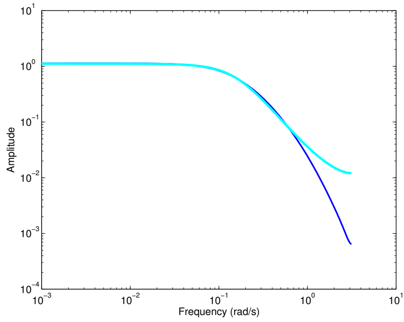
$$y(t) = \frac{b_1 q^{-1} + b_2 q^{-2}}{1 + f_1 q^{-1} + f_2 q^{-2}} u(t) + e(t)$$

Can you explain the difference?

True (thin) and Estimated (thick)

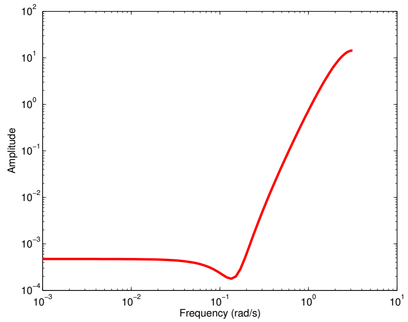


ARX

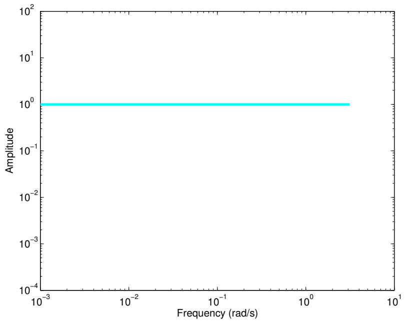


OE

The Weighting Functions



ARX

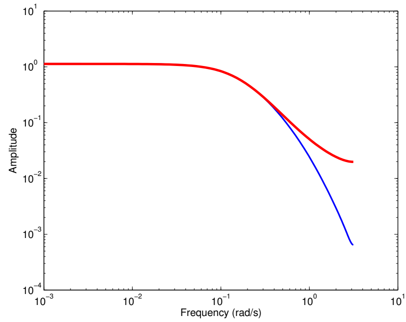


OE

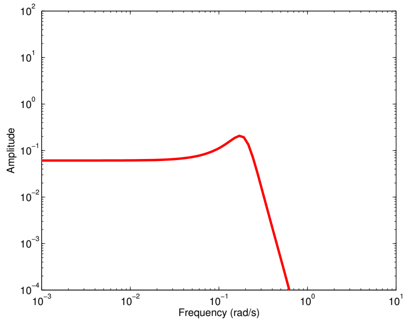
The Use of Pre-filtering

L is chosen as a low-pass filter

```
m=arx(data,[2 2 1],'focus',[0 0.2])
```



Estimated model



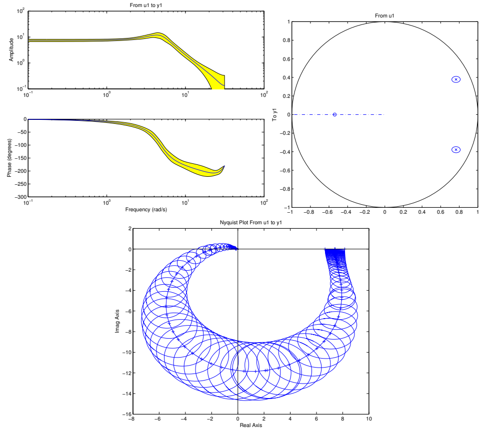
The weighting function

Model Quality in The Frequency Domain: Variance 17(20)

The asymptotic distribution and variance of $\hat{\theta}_N$ can be translated to those of $G(e^{i\omega}, \hat{\theta}_N) = \hat{G}_N(e^{i\omega})$ by
Gauss' Approximation Formula: $\text{Cov}\, f(\hat{\theta}_N) \approx f' \text{Cov}\, \hat{\theta}_N (f')^T$

Example: Model Properties with Uncertainty 18(20)

2.



Model Quality in The Frequency Domain: Variance 19(20)

The asymptotic distribution and variance of $\hat{\theta}_N$ can be translated to those of $G(e^{i\omega}, \hat{\theta}_N) = \hat{G}_N(e^{i\omega})$ by
Gauss' Approximation Formula: $\text{Cov}\, f(\hat{\theta}_N) \approx f' \text{Cov}\, \hat{\theta}_N (f')^T$
Approximate expressions for high order models:
Open Loop, Asymptotically in N and n

$$\text{Cov}\, \hat{G}_N(e^{i\omega}) \approx \frac{n}{N} \cdot \frac{\Phi_v(\omega)}{\Phi_u(\omega)}$$

n=model order, N=# of data

Φ_v: noise spectrum, Φ_u: input spectrum.

In General:

$$\text{Cov} \begin{bmatrix} \hat{G}_N(e^{i\omega}) \\ \hat{H}_N(e^{i\omega}) \end{bmatrix} \approx \frac{n}{N} \Phi_v(\omega) \begin{bmatrix} \Phi_u(\omega) & \Phi_{ue}(\omega) \\ \Phi_{ue}(-\omega) & \lambda \end{bmatrix}^{-1}$$

How To Affect the Model Quality 20(20)

$$\text{Cov}\, \hat{\theta}_N \approx \frac{\lambda}{N} \cdot [\text{E}\psi(t)\psi^T(t)]^{-1}$$

$$\text{Cov}\, \hat{G}_N \approx \frac{n}{N} \cdot \frac{\Phi_v(\omega)}{\Phi_u(\omega)}$$

$$\hat{\theta}_N \to \theta^* = \arg\min \text{E}[L(q)\varepsilon(t,\theta)]^2)$$

$$\hat{G}_N \to G^* = \text{closest to } G_0 \text{ in the } \frac{\Phi_u(\omega) \mid L(e^{i\omega}) \mid^2}{\mid H^*(e^{i\omega}) \mid^2} \text{ norm}$$