System Identification & Parameter Estimation

Wb2301: SIPE

Lecture 9: Physical Modeling, Model and Parameter Accuracy

Erwin de Vlugt, Dept. of Biomechanical Engineering (BMechE), Fac. 3mE

April 26 2010

Contents

- Parameter estimation in time domain
  - resume previous lecture(s)

- Overview of an experiment
  - Basic steps in an 'ideal' experiment

- Parameter estimation in frequency domain:
  - Non-parametric models: frequency response function (FRF)
    - Can be derived from 'non-parametric' time-domain models
  - Models with physical parameters
    - Transfer function of model (as function of Laplace operator **s**)
    - Model structure & model parameters: linear models
    - Optimization algorithms: adapt model parameters for best fit in frequency domain

Parameter estimation in time-domain

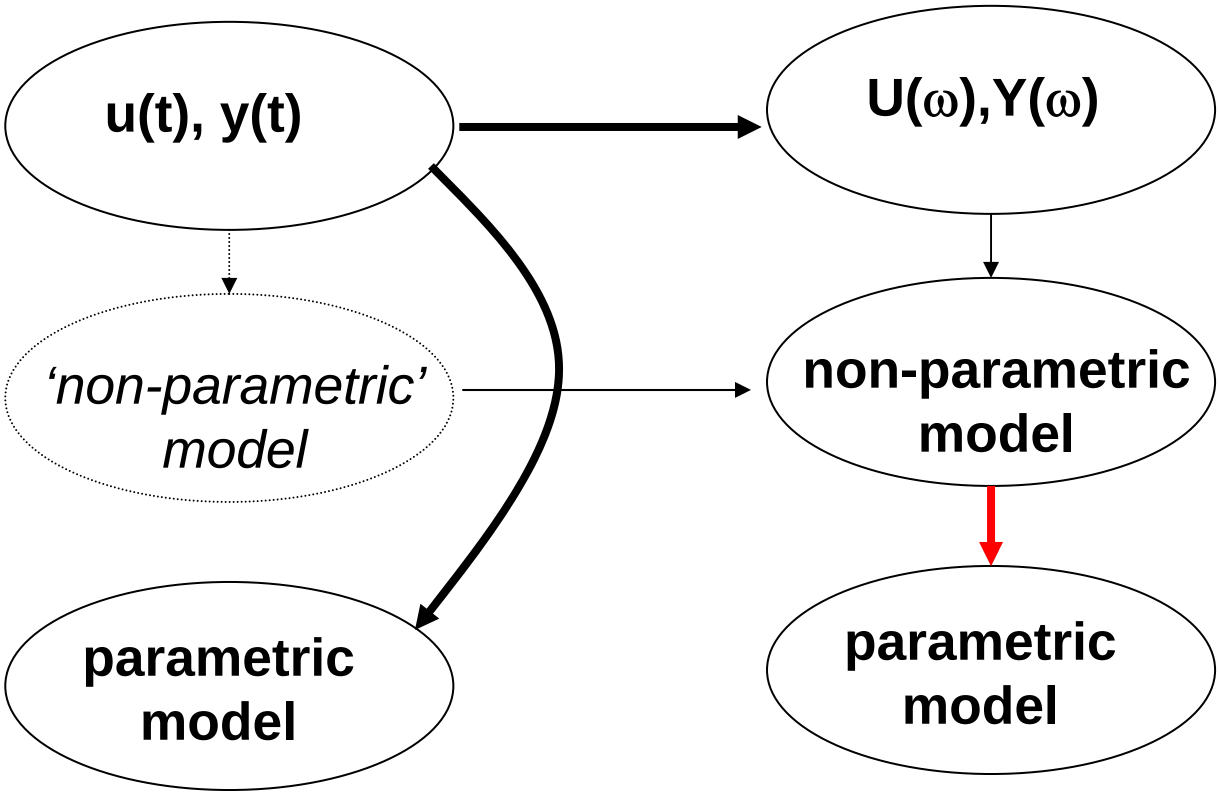
- ‘Non-parametric’ models:
  - ARMA, OE, Box-Jenkins, etc.

- Models with physical parameters
  - Input-output data, simulation of model (time domain)
  - Model structure & model parameters:
    - linear and non-linear models
  - Optimization algorithm:
    - adapt model parameters for best fit to simulation

- Note that ARX is a special case!
  - ARX is ‘linear in the parameters’: no simulation/optimization required!

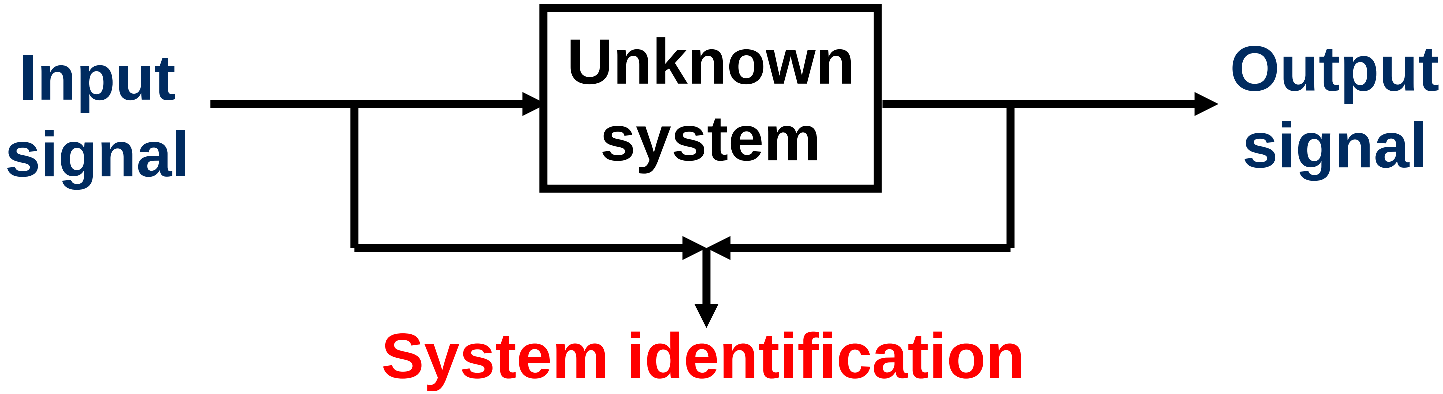
System identification & parameter estimation

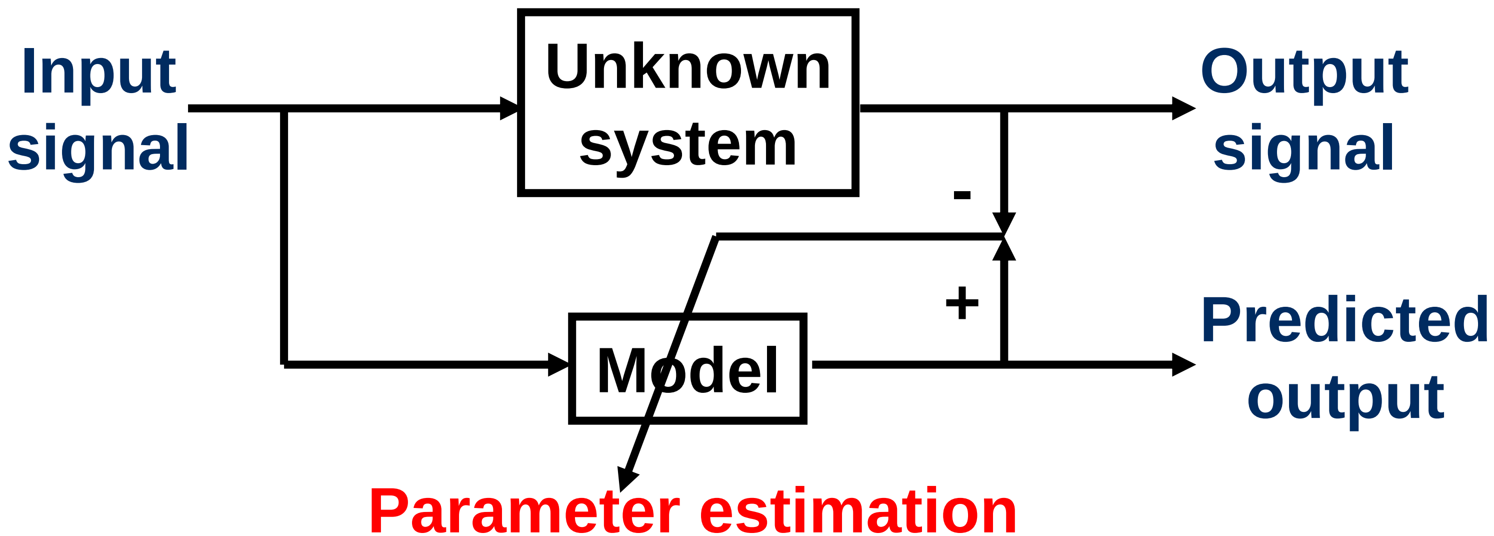
ARX
ARMA
Etc.



Frequency Response Function (FRF)

System identification & parameter estimation





Quantification of validity

- Variance-Accounted-For (VAF) values: How much of the variance in the data can be explained by the model?

$$VAF = 1 - \frac{\sum_{i=1}^{N} (y(t_i) - \hat{y}(t_i))^2}{\sum_{i=1}^{N} y(t_i)^2}$$

$$\hat{y}(t_i) = f(\theta, u; t)$$

$$\mathrm{y(t_i), u(t_i): recorded\ data}$$

Coherence and VAF

- High coherence, low VAF:
  - Linear system, good SNR, wrong model!
- High coherence, high VAF:
  - Linear system, good SNR, good model
- Low coherence, high VAF:
  - Non-linear system, good SNR, good <u>non-linear</u> model
- Low coherence, low VAF:
  - Non-linear system or poor SNR, poor model

Parameter estimation of static and dynamic systems

- Measured data
  - input signal $x(k)$
  - output signal $y(k)$
- Model (linear or non-linear)
  - Predicted output: $\hat{y}(k)=f(\theta,u(k))$
  - θ: parameter vector
  - => simulation of the model, depends on θ and $u(k)$

- error function
  - $e(k) = y(k) - f(\theta,u(k))$
- criterion function (least squares)
  - $J(x(k),y(k),\theta)= \Sigma e(k)^2$

- find θ which minimizes J
  - iterative search requires many simulations!

Accuracy of parameter fit

- Single parameter:
  - SEM: ‘Standard Error of the Mean’

- Multiple parameters:
  - Covariance matrix
  - Estimated from Jacobian and residual error

‘Standard error of the mean’ (SEM)

- How accurate can the <u>parameters</u> be estimated?
- Example:
  - Normal distribution of data x_N: μ_x, σ_x
  - Standard Error of the Mean:

$$\sigma_{\mu_x}^2 = \frac{1}{N}\sigma_x^2$$ variance of the mean

$$\sigma_{\mu_x} = \frac{\sigma_x}{\sqrt{N}}$$ standard error (deviation) of the mean

  - the more data samples to more accurate the estimation of the mean

Co-variance matrix

$$P_\theta \approx \hat{P}_N = \frac{1}{N} e^T e.[\frac{1}{N} J^T J]^{-1}$$

$$\text{Cov } \hat{\theta}_N \approx \frac{1}{N} P_\theta$$

- P_θ: Parameter co-variance matrix for parameter vector θ_N
  - Approximated by P_N (limited number of data samples N for estimation)

- Cov θ_N: Variance of P_θ
  - $\sigma_{\theta N} = \sqrt{(\text{diag}(\text{Cov } \theta_N))}$

Covariance Matrix P_N

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

- Z^N: data vector with input vector u and output vector y
- $V_N(\theta,Z^N)$: criterion value

$$V_N^{'}(\theta_N, Z^N) = \frac{\partial V_N(\theta_N, Z^N)}{\partial \theta} = 0$$

- θ^o: True, optimal parameter vector (unknown!)
- Expanding Taylor series (1st order) around θ^o:

$$0 = V_N^{'}(\theta^o, Z^N) + V_N^{''}(\theta^o, Z^N).(\hat{\theta}_N - \theta^o)$$

$$(\hat{\theta}_N - \theta^o) = -[V_N^{''}(\theta^o, Z^N)]^{-1}.V_N^{'}(\theta^o, Z^N)$$

$$\sigma^2_{\hat{\theta}_N} = \frac{1}{N}(\hat{\theta}_N - \theta^o)^2 = \frac{1}{N}.P_N$$

Derivation P_N

$$P_N = V_N^{''-1} . V_N^{'} . V_N^{'T} . V_N^{''-T}$$

$$V_N = e^2$$

$$V_N^{'} = \frac{\partial V_N}{\partial \theta} = J^T e$$

$$V_N^{''} = \frac{\partial V_N^{'}}{\partial \theta} = J^T J + \frac{\partial J^T}{\partial \theta} . e$$

- *e* is ‘white noise’ at θ^o, and hence $\partial J^T / \partial \theta = \partial^2 e / \partial \theta^2 \approx 0$

Derivation P_N

- P_N becomes

$$P_N = V_N''^{-1}.V_N'.V_N'^{T}.V_N''^{-T}$$
$$= (J^T J)^{-1}.J^T \left(\sum_i^N e_i.e_i \right).J.(J^T J)^{-1}$$
$$= \lambda_N.(J^T J)^{-1}.J^T J.(J^T J)^{-1}$$
$$= \lambda_N.(J^T J)^{-1}$$

- Where

$$\lambda_N = \sum_i^N e_i.e_i = .e^T.e$$

Co-variance matrix

$$\mathrm{Cov}\,\hat{\theta}_N = \begin{bmatrix} \sigma_{\theta_1}^2 & \sigma_{\theta_1}.\sigma_{\theta_2} & \cdots & \sigma_{\theta_1}.\sigma_{\theta_M} \\ \sigma_{\theta_2}.\sigma_{\theta_1} & \sigma_{\theta_2}^2 & \cdots & \sigma_{\theta_2}.\sigma_{\theta_M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\theta_M}.\sigma_{\theta_1} & \sigma_{\theta_M}.\sigma_{\theta_2} & \cdots & \sigma_{\theta_M}^2 \end{bmatrix}$$

- And $\sigma_{\theta 1} = \sqrt{\mathrm{cov}\ \theta_N(1,1)}$, etc.

Matlab demo: parameter accuracy

Basic ‘steps’ in identification scheme

1. Prepare experiment
   - Choose sample frequency, observation time, and number of repetitions
   - Choose/design input signal

2. Perform experiment
   - Perform experiments with care and prevent possible noise sources

3. Analyze results
   - Check linearity! (e.g. coherence)
   - Open-loop or closed-loop algorithms required?
   - Do nonparametric analysis (FRF or ARX/OE/ARMAX)
   - Fit (parametric) model onto data
   - Check residue (should be small and preferably white)
   - Check validity (VAF) and parameter uncertainty (e.g. SEM)

1. Prepare experiment

- Sample frequency
  - Should be high enough to ‘see’ all relevant dynamics
  - High sample frequency will give more data (storage!) but will not necessarily give more information!
  - Prevent aliasing
- Observation time
  - Determines resolution in frequency domain
  - In general longer is better (as long as system is time-invariant)
- Number of repetitions
  - Multiple observations => variations between observations
- Choose/design input signal
  - ‘persistently’ exciting => excite all relevant dynamics
  - Prevent leakage

2. Perform experiment

- Perform experiments with care and prevent possible noise sources
  - Electromagnetic interference?
  - Human subjects
    - Unpredictable, to prevent anticipation
    - clear instruction, no distractions, etc

- Often data can not be ‘fixed’ afterwards!

3. Analyze results

- Check linearity
  - Calculate coherence
- Open-loop or closed-loop algorithms required?
  - Try to make a block scheme
- Do nonparametric analysis (FRF or ARX/OE/ARMAX e.d.)
  - Bode diagram can give indication of system under investigation
- Fit (parametric) model onto data
  - Do a first check by inspecting the Bode diagram of data and model!

- Check residue (should be small and preferably white)
  - What is not captured with fitted model?
- Check validity of model (VAF) and parameter uncertainty (e.g. SEM)

Options in parameter estimation

Time-domain:

- direct fit using derivatives (HMC: inverse dynamics)
  - Noise is amplified by differentiation
- direct fit using simulation (previous lecture)
  - Requires multiple model simulations:
    a lot of CPU power
  - Can handle non-linear models!

Options in parameter estimation

- Frequency-domain:

1. FFT, estimation of FRF, estimation of parameters
   - No prior assumptions are needed!
   - Can easily cope with systems in closed-loop
   - Estimates can be biased if (very) much noise is present
   - E.g. depends on number frequency bands used for averaging

2. OE/ARMAX fit, estimation of parameters
   - In general reasonable fast and accurate
   - Order selection is needed (requires a choice!)
   - Frequency reconstruction desired to estimate the model structure (if system is unknown)

Error function in frequency domain

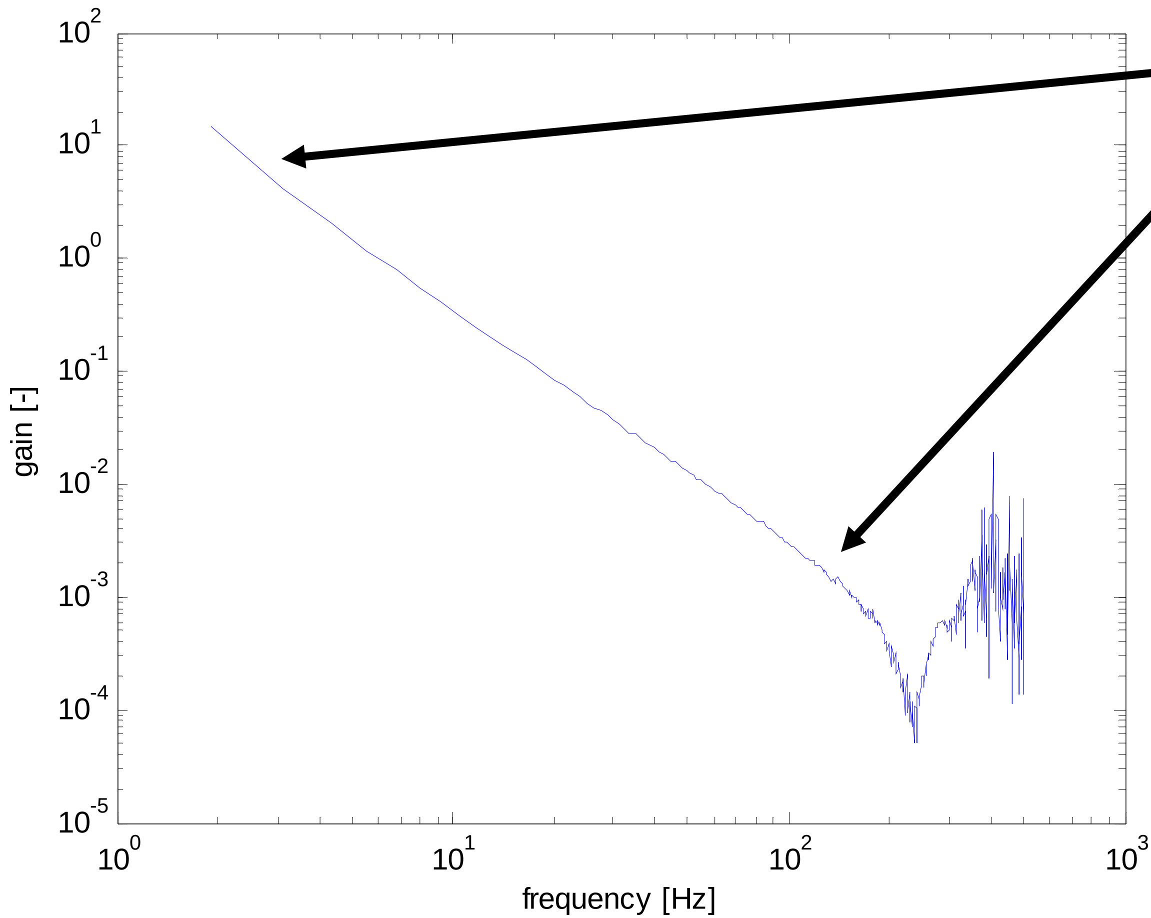
Simple approach:

$$J(f) = \sum_f e(f)^2$$

$$e(f) = H_{est}(f) - H_{\mathrm{mod}}(f)$$

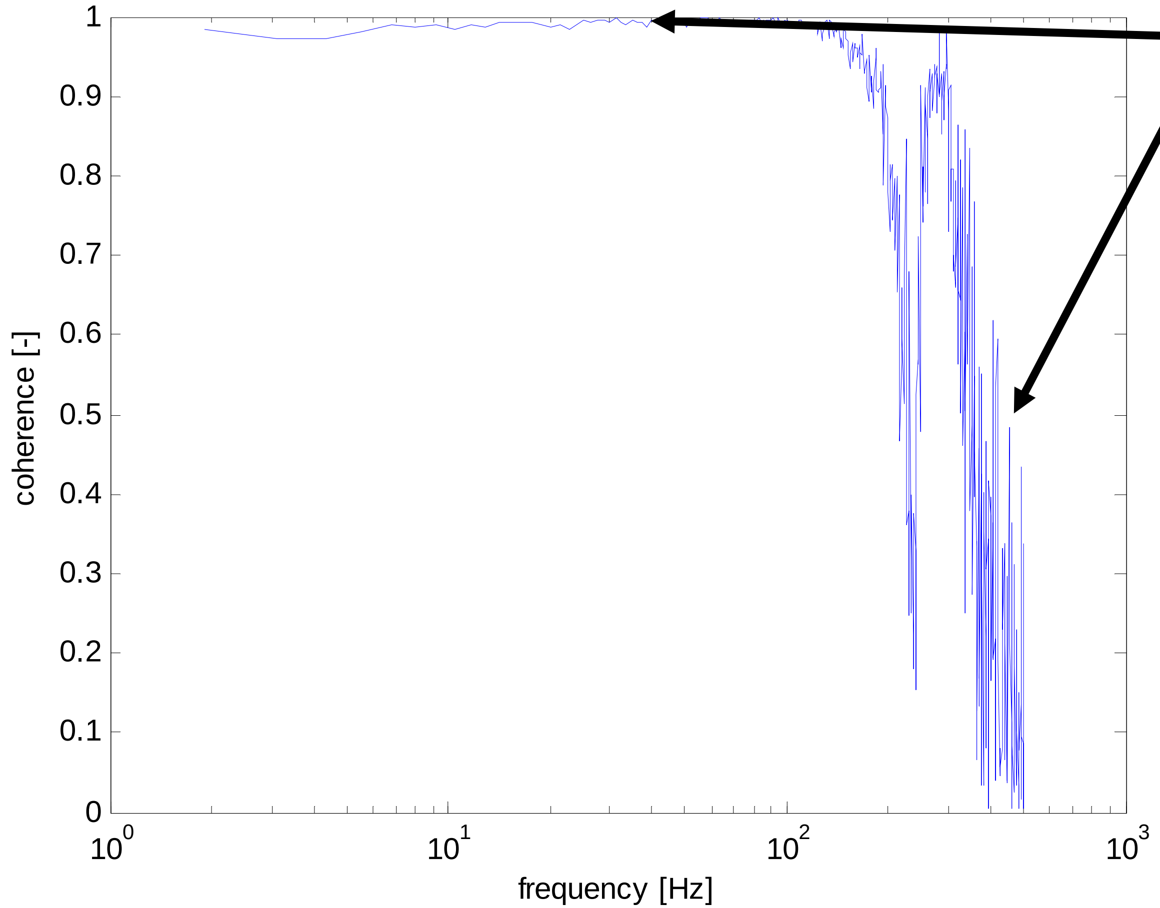
- Wrong approach
- Can give severely biased results

Model fit in frequency domain



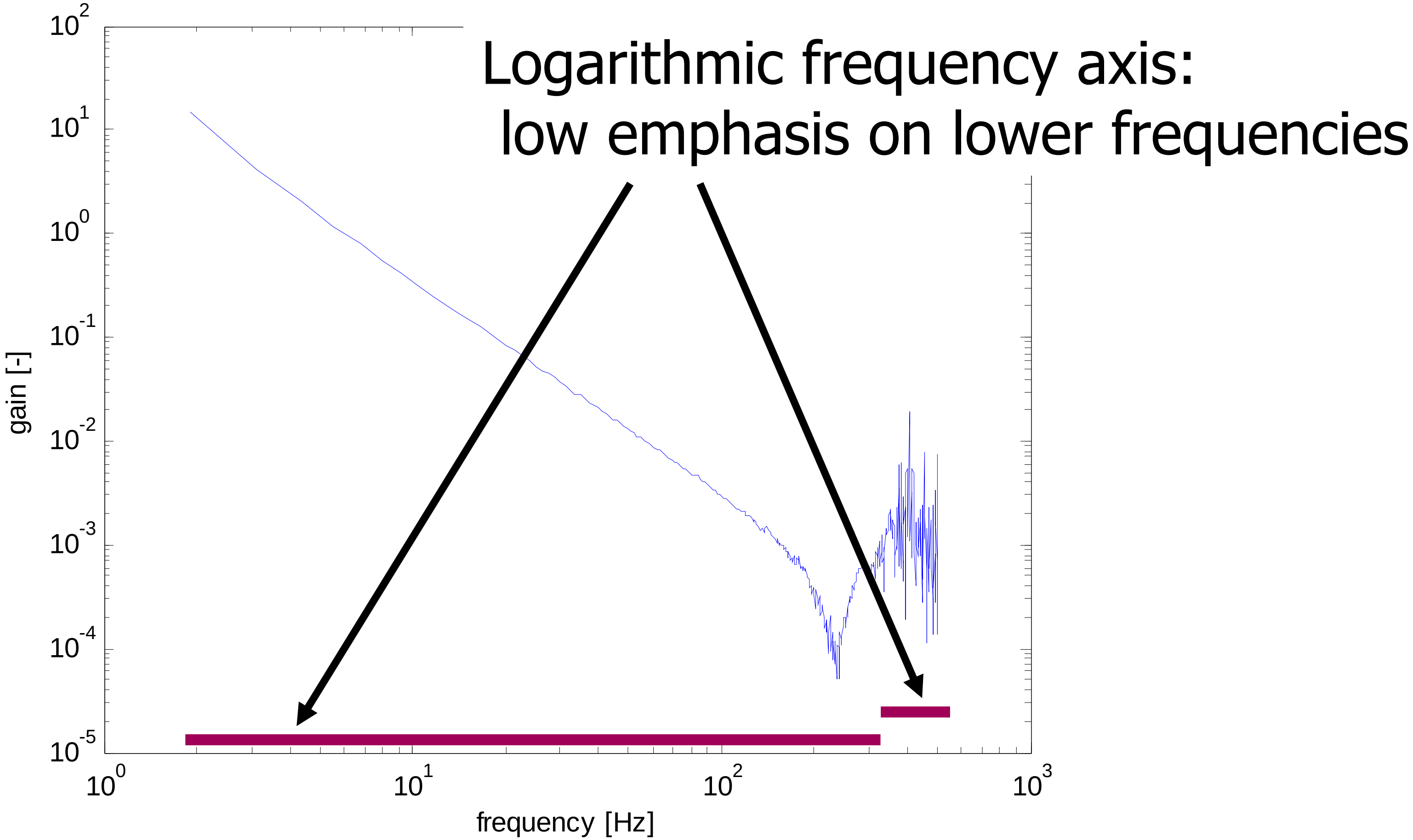
Dynamic range: absolute errors vary between 10^1 vs 10^{-4}

Model fit in frequency domain



Reliability: coherence varies with frequency

Model fit in frequency domain



LogN of Transfer function H

$$H(\omega) = a(\omega) + j.b(\omega)$$
$$= A(\omega).e^{j\varphi(\omega)}$$

$$\ln(H(\omega)) = \ln(A(\omega).e^{j\varphi(\omega)})$$
$$= \ln(A(\omega)) + \ln(e^{j\varphi(\omega)})$$
$$= \ln(|H(\omega)|) + j.\varphi(\omega)$$

- Logarithm effects the gain, not the phase!

Error function in frequency domain

$$J(f)=\sum_f e(f)^2$$

$$e(f)=\sqrt{\frac{1}{f}}.\gamma(f).\left|\ln(H_{est}(f))-\ln(H_{\mathrm{mod}}(f))\right|$$

$$=\sqrt{\frac{1}{f}}.\gamma(f).\left|\ln(\left|H_{est}(f)\right|)-\ln(\left|H_{\mathrm{mod}}(f)\right|+i.(\varphi(H_{est}(f))-\varphi(H_{\mathrm{mod}}(f))\right|$$

$$\sim\sqrt{\frac{1}{f}}*\gamma(f)*(\text{difference in log(gain)}+\text{difference in phase})$$

$$J(f)=\sum_f\frac{1}{f}.\gamma^2(f).\left|\ln(H_{est}(f))-\ln(H_{\mathrm{mod}}(f))\right|^2=\sum_f\frac{1}{f}.\gamma^2(f).\left|\ln\left(\frac{H_{est}(f)}{H_{\mathrm{mod}}(f)}\right)\right|^2$$

weighted by coherence to put more emphasis on reliable frequencies

weighted by 1/frequency to compensate for few data in low frequency region

Error function in frequency domain

Example:
measurements on a mass-spring-damper system

$$J(f) = \sum_f e(f)^2$$

$$e(f) = \sqrt{\frac{1}{f}}.\gamma(f).\left|\ln(H_{est}(f)) - \ln(H_{\mathrm{mod}}(f))\right| = \sqrt{\frac{1}{f}}.\gamma(f).\left|\ln\left(\frac{H_{est}(f)}{H_{\mathrm{mod}}(f)}\right)\right|$$

$$H_{est}(f) = \frac{S_{uy}(f)}{S_{uu}(f)}$$

$$H_{\mathrm{mod}}(f) = \frac{1}{M.s^2 + B.s + K} = \frac{1}{M.(2\pi j.f)^2 + B.2\pi j.f + K}$$

Assignment this week

- Estimate parameters in time and frequency domain
- Compare the results between both approaches

- Goal: Show the (dis-)advantages and peculiarities of estimation in both time domain and frequency domain (and compare the two approaches)