System Identification
Lecture 9

Identification in the Frequency Domain

Contents

Frequency domain identification

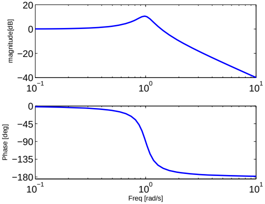
In several situations the "measurement data" is actually represented in the frequency domain:

- Periodic excitation of the system provides point estimates of the system's frequency response

$$\breve{G}(e^{i\omega_k}) \quad \text{for } k = 1, \cdots N_p$$

and you would like to estimate a parametric model.

E.g. frequency analyzers provide a direct frequency response "measurement".



Frequency domain identification

In several situations the "measurement data" is actually represented in the frequency domain:

- Periodic excitation of the system provides hugely sized data records that can be effectively represented by its Fourier transforms $(Y_N(\omega_k), U_N(\omega_k))$.
- In particular in mechanical engineering (vibration/modal analysis) and electrical engineering (circuits), a frequency-domain treatment of signals is widespread and insightful.

As a result it can be attractive to formulate the identification problem (i.c. the identification criterion) in the frequency domain also.

Attractive properties of the frequency domain treatment will typically appear in conjunction with periodic excitation signals. (The basis functions of the Fourier analysis).

The particular advantages of periodic excitation will be discussed later.

One-to-one relation between time and frequency domain

Through (discrete) Fourier Transform[1]:

$$U_N(\omega) = \sum_{t=0}^{N-1} u(t) e^{-i\omega t}$$

$$u(t) = \frac{1}{N} \sum_{\ell=0}^{N-1} U_N\left(\frac{2\pi\ell}{N}\right) e^{i\frac{2\pi\ell}{N}t}$$

This provides a unique relationship in two directions:
Mapping N (real-valued) points in t-domain, to $N/2$ (complex-valued) points in f-domain, and back.

[1] Note that in some literature the factor $1/N$ in $u(t)$ is replaced by a factor $1/\sqrt{N}$ in both expressions for $u(t)$ and $U_N(\omega)$.

Identification criteria

Starting from the time-domain (prediction error) approach:

$$\begin{aligned}\varepsilon(t,\theta) &= H(q,\theta)^{-1}\left[y(t) - G(q,\theta)u(t)\right] \\ \hat{\theta}_N &= \arg\min_{\theta} \frac{1}{N}\sum_{t=0}^{N-1}\varepsilon^2(t,\theta)\end{aligned}$$

Through F-transform:

$$\mathcal{E}_N(\omega,\theta) = H(e^{i\omega},\theta)^{-1}\left[Y_N(\omega) - G(e^{i\omega},\theta)U_N(\omega)\right] + R_N(\omega,\theta)$$

with R_N reflecting the transients of both H^{-1} and $H^{-1}G$, and

$$\frac{1}{N}\sum_{t=0}^{N-1}\varepsilon^2(t,\theta) = \frac{1}{N}\sum_{k=0}^{N-1}\left|\mathcal{E}_N\left(\frac{2\pi k}{N},\theta\right)\right|^2$$

Two options for “data”:

- F-transformed i/o data: $Y_N(\omega), U_N(\omega)$
- (Measured) frequency response (FRF) $\breve{G}(e^{i\omega_k})$
  e.g. through frequency analyzer

Identification criterion for f-domain data $Y_N(\omega_k), U_N(\omega_k)$

The typical identification criterion for given f-domain data $Y_N(\omega), U_N(\omega)$[2]:

$$\hat{\theta}_N := \arg\min_{\theta} \frac{1}{N} \sum_{k=0}^{N/2} \frac{|Y_N(\omega_k) - G(e^{i\omega_k}, \theta) U_N(\omega_k)|^2}{|H(e^{i\omega_k}, \theta)|^2}$$

with ω_k on the Fourier grid: $\omega_k = \frac{2\pi k}{N}$.

[2] Neglecting the effect of initial conditions in the term R_N

Identification criterion for FRF data $\breve{G}(e^{i\omega_k})$

If the frequency response function (FRF) $\breve{G}(e^{i\omega_k})$ is measured or estimated a priori, e.g. through ETFE, i.e.

$$\breve{G}(e^{i\omega_k}) = \frac{Y_N(\omega_k)}{U_N(\omega_k)}$$

then the previous criterion transforms to

Identification criterion for FRF data

$$\hat{\theta}_N = \arg\min_{\theta} \frac{1}{N} \sum_{k=0}^{N/2} \left| \breve{G}(e^{i\omega_k}) - G(e^{i\omega_k}, \theta) \right|^2 \cdot \frac{|U_N(\omega_k)|^2}{|H(e^{i\omega_k}, \theta)|^2}$$

Identification criterion for FRF data $\breve{G}(e^{i\omega_k})$

$$\hat{\theta}_N = \arg\min_{\theta} \frac{1}{N} \sum_{k=0}^{N/2} \left| \breve{G}(e^{i\omega_k}) - G(e^{i\omega_k}, \theta) \right|^2 \cdot \frac{|U_N(\omega_k)|^2}{|H(e^{i\omega_k}, \theta)|^2}$$

Since $U_N(\omega_k)$ is typically not available when $\breve{G}$ is "measured" this criterion is often replaced by

Simplified identification criterion for FRF data

$$\hat{\theta}_N = \arg\min_{\theta} \frac{1}{N} \sum_{k=0}^{N-1} \left| \breve{G}(e^{i\omega_k}) - G(e^{i\omega_k}, \theta) \right|^2 \cdot W_k$$

with W_k some (fixed) positive weighting function.

The set of frequencies ω_k over which the criterion is calculated,
can be chosen by the user (special choice of W_k)
to highlight particular frequency ranges

Role of the noise model

- In t-domain: $H(q, \theta)$ needs to be available as parametric transfer function for simulating the prediction error $\varepsilon(t, \theta)$.
- In f-domain: $|H(e^{i\omega_k}, \theta)|^2$ only serves purpose of frequency dependent weighting in a limited number of f-points

Modelling noise in the f-domain:

White noise signal $e(t)$ with $e \sim \mathcal{N}(0, \sigma_e^2)$

- $E_N(\omega_k)$ = zero mean & independent over $\omega_k = 2\pi\frac{k}{N}$.
- $E_N(\omega_k)$ is *circular complex normally distributed,*
- $E_N(\omega_k) \in \mathbb{C}$ with pdf $p(z) = \frac{1}{2\pi} e^{-z^*\Gamma^{-1}z}$

  $\mathbb{E}\left[E_N(\omega_k)\right] = 0, \quad \mathbb{E}\left[E_N^2(\omega_k)\right] = 0, \quad \mathbb{E}\left[|E_N(\omega_k)|^2\right] = \sigma_E^2$

For colored noise $v(t) = H(q)e(t)$

- $V_N(\omega_k)$ = zero mean & independent over $\omega_k = 2\pi\frac{k}{N}$ for $(N \to \infty)$
- $V_N(\omega_k)$ is *circular complex normally distributed,*
- $V_N(\omega_k) \to H(\omega_k)E_N(\omega_k)$ for $(N \to \infty)$
- $\mathbb{E}\left[V_N(\omega_k)\right] = 0, \quad \mathbb{E}\left[V_N^2(\omega_k)\right] = 0,$

$$\mathbb{E}\left[|V_N(\omega_k)|^2\right] := \sigma_V^2(\omega_k) = |H(\omega_k)|^2\sigma_E^2$$

The identification criterion for given f-domain data $Y_N(\omega), U_N(\omega)$ can then be written as:

$$\hat{\theta}_N := \arg\min_{\theta} \frac{1}{N} \sum_{k=0}^{N/2} \frac{|Y_N(\omega_k) - G(e^{i\omega_k}, \theta) U_N(\omega_k)|^2}{\sigma_V^2(\omega_k)}$$

with ω_k on the Fourier grid: $\omega_k = \frac{2\pi k}{N}$,

and this will be the maximum likelihood estimator for

- periodic input u and e Gaussian, or
- $N \rightarrow \infty$.

In general, $\sigma_V^2(\omega_k)$ will need to be estimated but not necessarily through a finite-dimensional LTI filter, but just as a sequence of points (see slides 25/26), resulting in the sample maximum likelihood estimator.

Model structures

Model parametrizations for $G(e^{i\omega_k}, \theta)$ and $H(e^{i\omega_k}, \theta)$ are completely similar to the time domain model structures, by replacing q by $e^{i\omega_k}$:

- Output error model structure

$$\begin{aligned} G(e^{i\omega_k}, \theta) &= \frac{B(e^{-i\omega_k}, \theta)}{F(e^{-i\omega_k}, \theta)} = \frac{b_0 + b_1 e^{-i\omega_k} + b_2 e^{-2i\omega_k}}{1 + f_1 e^{-i\omega_k} + f_2 e^{-2i\omega_k}} \\ H(e^{i\omega_k}, \theta) &= 1 \end{aligned}$$

- ARX model structure

$$\begin{aligned} G(e^{i\omega_k}, \theta) &= \frac{B(e^{-i\omega_k}, \theta)}{A(e^{-i\omega_k}, \theta)} = \frac{b_0 + b_1 e^{-i\omega_k} + b_2 e^{-2i\omega_k}}{1 + a_1 e^{-i\omega_k} + a_2 e^{-2i\omega_k}} \\ H(e^{i\omega_k}, \theta) &= \frac{1}{A(e^{-i\omega_k}, \theta)}. \end{aligned}$$

All present model structures lead to non-convex optimization schemes, except for the two structures that are linear-in-the-parameters:

- FIR (OE with $F = 1$; ARX with $A = 1$)
- ARX

they lead to convex optimization problems / analytical solutions for $\hat{\theta}_N$.
When a nonlinear-in-the-parameters optimization needs to be solved other methods are available:

- Sanathanan-Koerner,
- Gauss-Newton based algorithms.

Linear regression in the f-domain

An ARX model applied to FRF data leads to a criterion

$$V_N(\theta) = \frac{1}{N}\sum_{k=1}^{N}\left|\breve{G}(e^{i\omega_k})A(e^{-i\omega_k},\theta) - B(e^{-i\omega_k},\theta)\right|^2$$

With

$$\begin{aligned} A(e^{-i\omega_k},\theta) &= 1 + a_1 e^{-i\omega_k} + \cdots a_{n_a} e^{-n_a i\omega_k} \\ B(e^{-i\omega_k},\theta) &= b_1 e^{-i\omega_k} + \cdots b_{n_b} e^{-n_b i\omega_k} \\ \theta^T &= [a_1\ a_2 \cdots a_{n_a}\ b_1 \cdots b_{n_b}] \end{aligned}$$

this can be written as in a linear regression form:

With

$$\breve{G}(e^{i\omega_k})A(e^{-i\omega_k},\theta) - B(e^{-i\omega_k},\theta) = \\ \breve{G}(e^{i\omega_k})[1 + a_1 e^{-i\omega_k} + \cdots a_{n_a} e^{-n_a i\omega_k}] - [b_1 e^{-i\omega_k} + \cdots b_{n_b} e^{-n_b i\omega_k}]$$

it follows that

$$V_N(\theta) = \frac{1}{N}\sum_{k=1}^{N} \left| \breve{G}(e^{i\omega_k}) - \Omega^T(e^{i\omega_k})\theta \right|^2$$

where

$$\Omega^T(e^{i\omega_k}) := [-e^{-i\omega_k}\breve{G}(e^{i\omega_k}) \cdots - e^{-n_a i\omega_k}\breve{G}(e^{i\omega_k}) \; e^{-i\omega_k} \cdots e^{-n_b i\omega_k}]$$

With

$$\bar{G} := \begin{bmatrix} \text{Re } \breve{G}(e^{i\omega_1}) \\ \text{Im } \breve{G}(e^{i\omega_1}) \\ \vdots \\ \text{Re } \breve{G}(e^{i\omega_N}) \\ \text{Im } \breve{G}(e^{i\omega_N}) \end{bmatrix} \qquad \bar{\Omega} := \begin{bmatrix} \text{Re } \Omega^T(e^{i\omega_1}) \\ \text{Im } \Omega^T(e^{i\omega_1}) \\ \vdots \\ \text{Re } \Omega^T(e^{i\omega_N}) \\ \text{Im } \Omega^T(e^{i\omega_N}) \end{bmatrix}$$

it follows that

$$V_N(\theta) = \frac{1}{N}(\bar{G} - \bar{\Omega}\theta)^T(\bar{G} - \bar{\Omega}\theta)$$

where all terms are real-valued now, leading to

$$\hat{\theta}_N = (\bar{\Omega}^T\bar{\Omega})^{-1}\bar{\Omega}^T\bar{G}.$$

(compare similarity with time-domain approach).

Note however that the criterion actually has a high-frequency (ARX) weighting:

$$
\begin{aligned}
V_N(\theta) &= \frac{1}{N}\sum_{k=1}^{N}\left|\breve{G}(e^{i\omega_k})A(e^{-i\omega_k},\theta) - B(e^{-i\omega_k},\theta)\right|^2 \\
&= \frac{1}{N}\sum_{k=1}^{N}\left|\breve{G}(e^{i\omega_k}) - \frac{B(e^{-i\omega_k},\theta)}{A(e^{-i\omega_k},\theta)}\right|^2 \cdot |A(e^{-i\omega_k},\theta)|^2
\end{aligned}
$$

Sanathanan & Koerner (1963) iteration:

In order to counteract the negative effect of the weighting, apply a criterion with a fixed weight:

$$V_N(\theta) = \frac{1}{N}\sum_{k=1}^{N}\left|\breve{G}(e^{i\omega_k})A(e^{-i\omega_k},\theta) - B(e^{-i\omega_k},\theta)\right|^2 W_k$$

with W_k determined by the denominator polynomial from a previous estimate:

$$W_k = \frac{1}{|A(e^{-i\omega_k},\hat{\theta}_{prev})|^2}$$

leading to an iterative algorithm and consecutive linear regression estimation steps.

Gauss-Newton based methods:

The cost function of the general problem is nonlinear in the parameters:

$$\hat{\theta}_N := \arg\min_{\theta} \frac{1}{N} \sum_{k=0}^{N/2} \frac{|Y_N(\omega_k) - G(e^{i\omega_k}, \theta) U_N(\omega_k)|^2}{\sigma_V^2(\omega_k)}$$

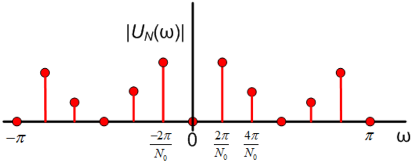
This requires nonlinear optimization techniques such as Guass-Newton based algorithms (see basic course on optimization).

1. Advantages of periodic excitation signals in identification

Periodic excitation (and taking a full number of periods into account) implies that

$$U_N(\omega)$$

takes values $\neq 0$ only in a finite frequency grid, being a subset of $\omega = 2\pi k/N_0$, $k = 1, .., N_0$, and N_0 the period length.



- Any non-periodic component in the measured y (after transients have disappeared) is due to noise
- Any component of $Y_N(\omega)$ outside of the grid with excited frequencies is due to noise (simple noise filtering)

- Information only in limited number of f-points, (no info in between these f-points),
- but in these points natural experiment repetition (noise reduction)
- If signals are non-periodic, transient effects (non-stationary initial conditions) have an influence on model accuracy and can/should be estimated too.

2. Non-parametric noise models from periodic inputs

Consider the data generating system:

$$y(t) = y_u + v(t), \quad \text{with } y_u(t) = G_0(q)u(t)$$

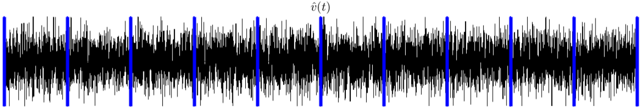
If u is periodic with period N_0, then y_u should be too (after a transient), so an improved estimate of y_u is obtained by averaging over K periods:

$$\bar{y}(t) = \frac{1}{K} \sum_{k=0}^{K-1} y(t + kN_0) \qquad t = 0, \cdots N_0 - 1$$

leading to an improved estimate of the noise

$$\hat{v}(t) = y(t) - \bar{y}(t) \quad \text{with } \bar{y}(t) \text{ periodically extended.}$$

Given noise $\hat{v}(t)$



DFT: $\hat{V}_{N_0}^{[1]}(\omega_k)$ $\hat{V}_{N_0}^{[2]}(\omega_k)$ ... $\hat{V}_{N_0}^{[K]}(\omega_k)$

The unbiased estimator of $\sigma_V^2(k)$ is

$\hat{\sigma}_V^2(k) := \frac{1}{K-1}\sum_{i=1}^{K}\left|\hat{V}_{N_0}^{[i]}(\omega_k)\right|^2$

This we can substitute as an appropriate weighting in the (maximum-likelihood) estimator:

$$\hat{\theta}_N := \arg\min_{\theta} \frac{1}{N}\sum_{k=0}^{N/2} \frac{|Y_N(\omega_k) - G(e^{i\omega_k},\theta)U_N(\omega_k)|^2}{\hat{\sigma}_V^2(\omega_k)}$$

thus avoiding the identification of a parametric noise model.

3. Prefiltering

Prefiltering in order to influence the approximation properties of estimated models becomes very simple:
a simple multiplication of the error term in the identification criterion:

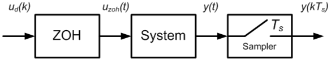
$$\hat{\theta}_N = \arg\min_{\theta} \frac{1}{N} \sum_{k=0}^{N-1} \left| \breve{G}(e^{i\omega_k}) - G(e^{i\omega_k}, \theta) \right|^2 \cdot W_k$$

(compare time-domain approach)

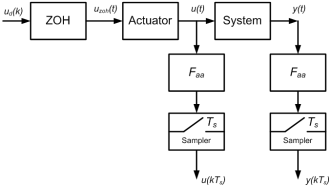
4. Continuous-time experimental setups

Experimental setups for identification typically are either:

- ZOH: The input to the system is a discrete-time signal, processed through a zero-order-hold (ZOH), to deliver an input to the (continuous-time) physical setup;



- BL: Input and output signals of the system are sampled versions of band-limited signals (e.g. after anti-aliasing).



- In the ZOH-setup: mapping from identified discrete-time (DT) model to continuous-time (CT) model through (inverse) Z- and Laplace transforms;

$$G(z) = (1 - z^{-1})\mathcal{Z}\left\{\frac{G(s)}{s}\right\}$$

or equivalently

$$\frac{G(s)}{s} = \mathcal{L}\left\{\frac{G(z)}{1 - z^{-1}}\right\}$$

(see control course 5ESD0)

- In the BL-setup: continuous-time frequency response can be obtained directly from DTFT (obey nyquist theorem!), i.e. CT frequency response = DT frequency response, for frequencies up to $\omega_s/2$.

$$G(i\omega) = G(e^{i\omega}), \qquad |\omega| < \omega_s/2.$$

Consequence in BL setup:

We can estimate a continuous-time (CT) model, directly from $\breve{G}(e^{i\omega})$.

The only thing that needs to be changed is the model parametrization:

$$G(i\omega_k, \theta) = \frac{B(i\omega_k, \theta)}{A(i\omega_k, \theta)} = \frac{b_0(i\omega_k)^n + b_1(i\omega_k)^{n-1} + \cdots + b_n}{(i\omega_k)^n + a_1(i\omega_k)^{n-1} + a_n}$$

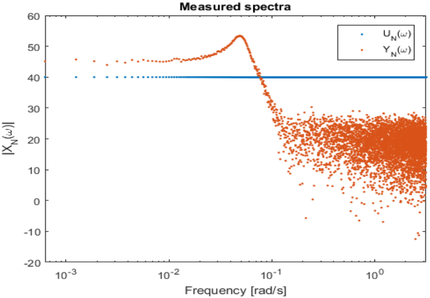
Estimating a transfer function model is a curve fitting problem in the frequency domain. It does not matter whether rational functions in $z = e^{i\omega_k}$, $s = i\omega_k$ or $\sqrt(s)$ (e.g. for distributed systems) are used.

5. Additional properties of f-domain identification

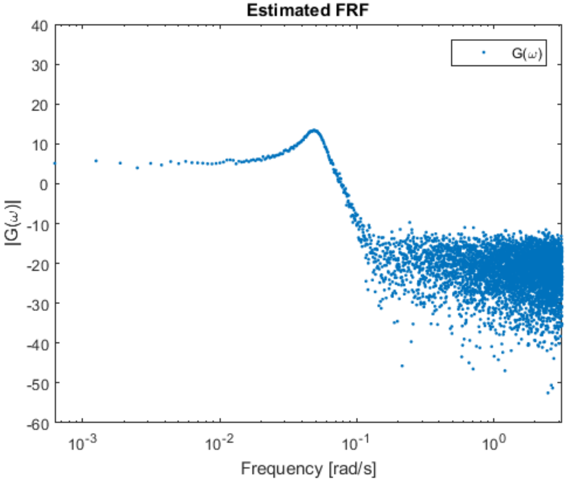
- Condensing large (time domain) data sets, to a limited number of points in the f-domain
- Combining data from different experiments (low frequent vs high frequent)
- Flexible choice of frequency region of interest

Example

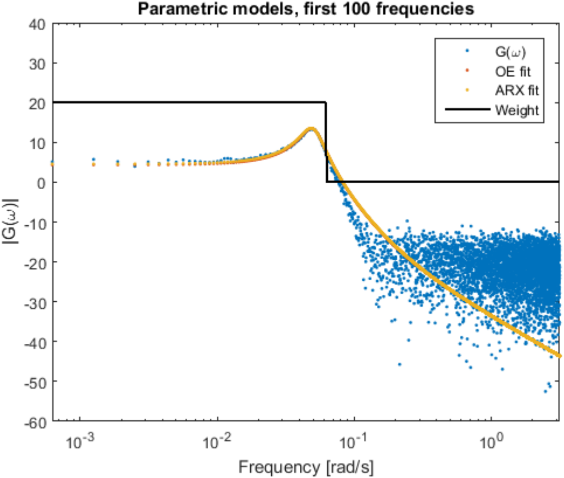
- System G_0 with 4 poles and 2 zeros
- 2 resonance peaks; 1 observed, 1 buried in noise
- Input: 2 repetitions of a white noise sequence, of which the second is used for identification; $N = 10,000$

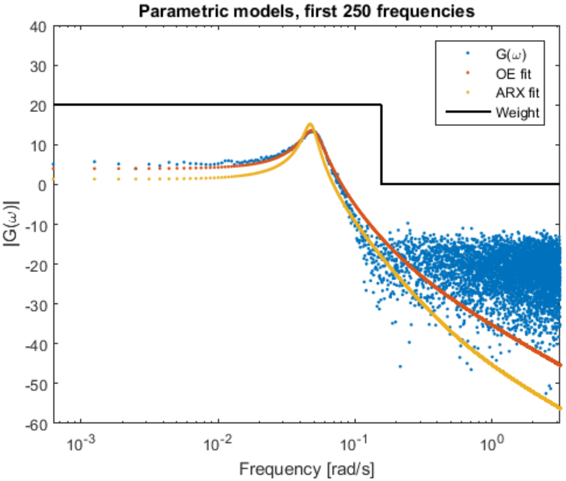


ETFE

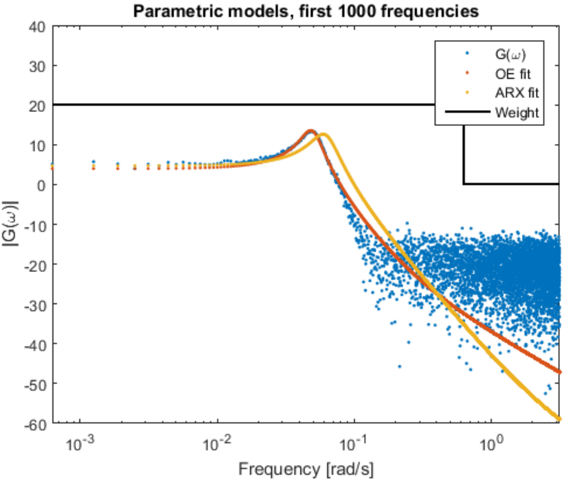

Estimated FRF
G(ω)
|G(ω)|
40
30
20
10
0
-10
-20
-30
-40
-50
-60
10^{-3}
10^{-2}
10^{-1}
10^{0}
Frequency [rad/s]

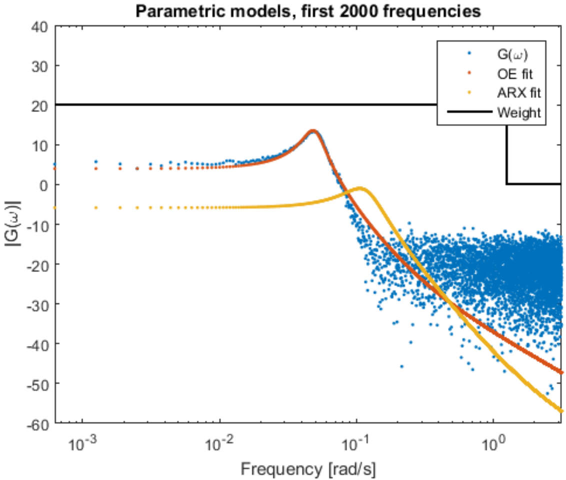
Parametric models: 2 poles, 2 zeros





Parametric models, first 250 frequencies
G(ω)
OE fit
ARX fit
Weight
|G(ω)|
Frequency [rad/s]
40
30
20
10
0
-10
-20
-30
-40
-50
-60
10^-3
10^-2
10^-1
10^0


Parametric models, first 1000 frequencies
40
30
20
10
0
-10
-20
-30
-40
-50
-60
|G(ω)|
10^{-3}
10^{-2}
10^{-1}
10^{0}
Frequency [rad/s]
G(ω)
OE fit
ARX fit
Weight


Parametric models, first 2000 frequencies
40
30
20
10
0
-10
-20
-30
-40
-50
-60
|G(ω)|
G(ω)
OE fit
ARX fit
Weight
10^{-3}
10^{-2}
10^{-1}
10^{0}
Frequency [rad/s]

Summary

- f-domain identification is to a large extent dual/similar to t-domain ID
- It allows two approaches, based on DFT of i/o data, or on measured/estimated FRF
- Periodic excitation is naturally connected to the f-domain identification, and has several advantages
- For periodic excitation there is less need for parametric noise models (selection of frequency grid, nonparam. noise model)
- Bandlimited setup allows direct identification of continuous-time models
- Theoretical principles and results are largely similar to the t-domain results

References

- R. Pintelon, P.Guillaume, Y. Rolain, J. Schoukens and H. Van hamme (1994). Parametric identification of transfer functions in the frequency domain - a survey. *IEEE Trans. Automatic Control, Vol. 39*, no. 11, pp. 2245-2260.
- R. Pintelon and J. Schoukens (2012). *System Identification - A Frequency Domain Approach*. John Wiley & Sons, NJ, 2nd edition.