Identification of Wiener-Hammerstein Models and Other Nonlinear Block Models

Jonas Sjöberg
Dept of Signals & Systems, Chalmers, Göteborg, Sweden

in colaboration with Giuseppe Giordano, Chalmers
and Johan Schoukens, ELEC, Vrije Universiteit Brussels, Belgium

November 8, 2014

Outline

- PEM and Maximum likelihood
- Wiener-Hammerstein model structure
- Algorithms
- Alternative Algorithms
- Example

The Maximum Likelihood Method

Given input and output data $Z^N = \{y(t), u(t)\}_{t=1}^N$ from a SISO system.

- Estimate a predictor (a model) $\hat{y}(t|t-1, \theta)$ for $y(t)$, where θ is a vector of unknown parameters.
- Let a probabilistic model be given as

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

where $\varepsilon(t, \theta)$ has PDF $f_e(x, t; \theta)$ (independent).

The *likelihood function* is given by

$$f(\theta, Z^N) = \prod_{t=1}^{N} f_e(\varepsilon(t,\theta), t; \theta)$$

This gives the maximum likelihood estimate

$$\hat{\theta}_{ML} = \arg\max_{\theta} \log f(\theta, Z^N) = \arg\min_{\theta} \frac{1}{N} \sum_{t=1}^{N} -\log f_e(\varepsilon, t; \theta)$$

Remark: This holds strictly only if the predictor is stable.

Hence, ML can be expressed as using prediction errors

$$\varepsilon(t,\theta) = y(t) - \hat{y}(t|t-1;\theta), \quad t = 1,\dots,N, \tag{1}$$

a *loss function* $V_N(\theta)$:

$$V_N(\theta) = \text{most cases} = \frac{1}{N}\sum_{t=1}^{N} \varepsilon^2(t,\theta) \tag{2}$$

where the last equality holds if f_e is Gaussian, which is a common assumption.
We thus have:

$$\hat{\theta}_N = \arg\min_{\theta\in\mathcal{D}_\mathcal{M}} V_N(\theta) \tag{3}$$

Search for the minimum of $V_N(\theta)$

This is optimization. Standard methods are available.
Given an initial estimate θ^0, iterate

$$\theta^{(i+1)} = \theta^{(i)} - \mu_i R_i \frac{dV_N(\theta)}{d\theta}$$

until the minimum is reached.
μ_i is a step length to assure downhill steps, and R_i is a matrix to modify the search direction.

To successfully compute the ML estimate

- To have good data $\{y(t), u(t)\}_{t=1}^{N}$
- To find a good model structure.
- To obtain a *good initial parameter estimate* for the numerical search!

A huge amount of research has been done on this, but often described in partly different way. Too many references to be listed here.

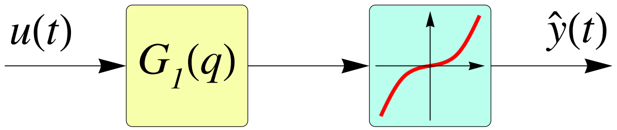
Linear SI: Instrumental variable methods (older) and Subspace identification methods (newer) are often used as initial estimates.

- P. Stoica, T. Söderström and B. Friedlander, Optimal instrumental variable estimates of the AR parameters of an ARMA process, IEEE Trans. Automatic Control, Vol. AC-30, No. 11, November 1985.
- Vector ARMA estimation: A reliable subspace approach J Mari, P Stoica, T McKelvey - IEEE Transactions on Signal Processing, 2000.

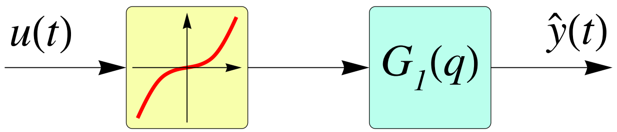
Initializing Wiener and Hammerstein Models

A lot of work has been published on methods to initialize block-based nonlinear models:

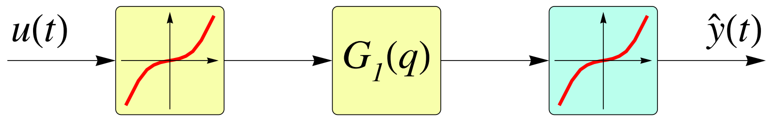
Wiener model:



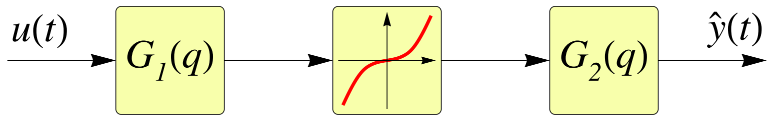
Hammerstein model:



Hammerstein-Wiener model:



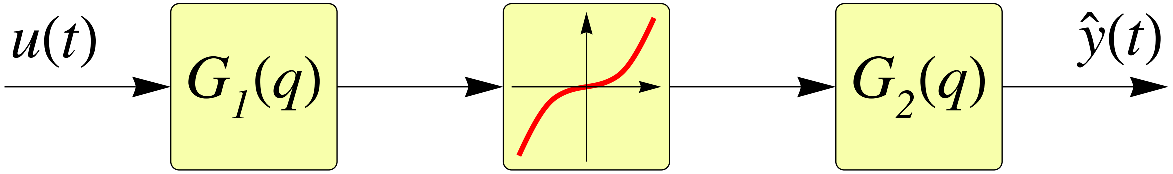
Wiener-Hammerstein model:



In the following, algorithms for Wiener-Hammerstein models are considered.

Wiener-Hammerstein model on state space form

These can all be represented as state-space models , eg, the Wiener-Hammerstein model:



$$
\begin{aligned}
x(t+1) &= \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} x(t) + \begin{pmatrix} B_1 \\ 0 \end{pmatrix} u(t) + \begin{pmatrix} 0 \\ B_2 \end{pmatrix} f(\theta, [C_1 \;\; 0] x(t)) \\
\hat{y}(t) &= \begin{pmatrix} 0 & C_2 \end{pmatrix} x(t)
\end{aligned}
$$

When is this a linear regression problem?

- If G_1 and G_2 are given and if the nonlinearity is linearly parameterized.

Block description

$$
\begin{aligned}
z(t) &= G_1(q^{-1}, \alpha)\, u(t) \\
x(t) &= f(\beta, z(t)) \\
\hat{y}(t) &= G_2(q^{-1}, \gamma)\, x(t) + \nu(t)
\end{aligned} \tag{4}
$$

$\hat{y}(t)$ is the prediction, and $G_1(q^{-1}, \alpha)$ and $G_2(q^{-1}, \gamma)$ are LTI transfer functions in q^{-1}, and parameterized with α and γ, respectively.
All parameters of the model structure are stored in a common parameter vector

$$\theta = [\alpha,\ \beta,\ \gamma]. \tag{5}$$

Start with a linear model

Lemma 1 *Suppose the input data $u(t)$ is a stationary normal distributed sequence and the output $y(t)$ is obtained by filtering $u(t)$ through a system of form (4) with linear parts G_1^0 and G_2^0 being stable, single input, single output finite order transfer functions, and the nonlinear part f^0 is a continuous function $\Re \to \Re$. Then best linear approximation (BLA), converge to*

$$\kappa G_1^0(q^{-1})G_2^0(q^{-1}) \tag{6}$$

where κ is a constant which value depends on $u(t)$, f^0, G_1^0 and G_2^0.

Proof: See Pintelon, R. and Schoukens, J. (2001). *System Identification: A Frequency Domain Approach*. IEEE-press, Piscataway.

Algorithm

- Start with the best linear model $G(z)$.
- Split the linear model into $G_1(z)$ and $G_2(z)$ so that $G(z) = G_1(z)G_2(z)$
- LS fitting of linear parameters.
- Fit all parameters of the models.

How should the dynamic be split?

Algorithm

- Split the linear model into $G_1(z)$ and $G_2(z)$ in *all possible* ways.

Sjöberg, J. och Schoukens, J. *Initializing Wiener-Hammerstein models based on partitioning of the best linear approximation.* Automatica, Vol. 48 (2012), 2, p. 353-359.

- *Brute force approach.*
- Correct division of poles and zeros is within the set of models.
- Computational aspect: feasible up to model order approximately 10.
- Better than all other method published up to then (which are all more complicated).

Alternative 1: Offer all poles and zeros on both sides

- Internal signals z and x

$$
\begin{aligned}
z(t) &= G_1\, u(t) \\
x(t) &= G_2^{-1}\, y(t).
\end{aligned} \tag{7}
$$

- Static nonlinearity as a concatenation of two functions f_1 and f_2

$$
f(x) = f_2^{-1}(f_1(x)). \tag{8}
$$

- Then, neglecting the influence of disturbances one obtains

$$f_1(G_1\,u(t)) = f_2(G_2^{-1}\,y(t)) \tag{9}$$

  solved approximately as a linear-in-the-parameters TLS problem.
- $\hat{G}_1$ and $\hat{G}_2^{-1}$ as linear combination of basis functions containing the poles and the zeros of G_{BLA}, respectively.

J. Sjöberg, L. Lauwers, J. Schoukens, *Identification of Wiener-Hammerstein models: Two algorithms based on the best split of a linear model applied to the SYSID'09 benchmark problem.*, Control Engineering Practice, Volume 20, Issue 11, November 2012, Pages 1119-1125.

Alternative 2: Offer all poles and zeros on both sides

Parameterize poles and zeros on both sides: β_k and $(1-\beta_k)$, respectively. Same thing with α_k.

$$G_1 = \frac{\prod_{k=1}^{n_B}(1 - z_k\, z^{-1})^{\beta_k}}{\prod_{k=1}^{n_A}(1 - p_k\, z^{-1})^{\alpha_k}} \qquad G_2 = \frac{\prod_{k=1}^{n_B}(1 - z_k\, z^{-1})^{(1-\beta_k)}}{\prod_{k=1}^{n_A}(1 - p_k\, z^{-1})^{(1-\alpha_k)}}$$

- Transformations between time- and frequency domain.
- The fractional power is calculated in frequency domain.
- The nonlinearity is calculated in time domain
- If nonlinear function is linear parameterized then a nonlinear minimization in β_k and α_k.

L. Vanbeylen, *A fractional approach to identify Wiener-Hammerstein systems*. Automatica, vol.50, n. 3, 2014, pp.903-909.

Alternative 3: Offer all poles and zeros on both sides

In time domain: Taylor expand poles and zeros of $G(z)$.
5th order expansion of one pole

$$\frac{1}{(1-\frac{0.8}{z})^{\alpha}} \approx 1 + \frac{0.8\alpha}{z} + \left(\frac{\alpha^2}{2}+\frac{\alpha}{2}\right)\left(\frac{0.8}{z}\right)^2 + \left(\frac{\alpha^3}{6}+\frac{\alpha^2}{2}+\frac{\alpha}{3}\right)\left(\frac{0.8}{z}\right)^3 + \left(\frac{\alpha^4}{24}+\frac{\alpha^3}{4}+\frac{11\alpha^2}{24}+\frac{\alpha}{4}\right)\left(\frac{0.8}{z}\right)^4 + \left(\frac{\alpha^5}{120}+\frac{\alpha^4}{12}+\frac{7\alpha^3}{24}+\frac{5\alpha^2}{12}+\frac{\alpha}{5}\right)\left(\frac{0.8}{z}\right)^5 + O\left(\left(\frac{0.8}{z}\right)^6\right)$$

Parameters indicating zero-pole position can, hence, be estimated.

Problem: Although number of parameter is not high, we have computational problems due to huge symbolic expressions when several poles and zeros are considered.

Alternative 3: Offer all poles and zeros on both sides

Consider 2 poles

$$\frac{1}{\left(1-\frac{0.8}{z}\right)^{\alpha_1}}\frac{1}{\left(1-\frac{0.4}{z}\right)^{\alpha_2}}$$

The coefficient of z^{-3} becomes:

$$0.085\alpha_1^3+0.13\alpha_2\alpha_1^2+0.26\alpha_1^2+0.064\alpha_2^2\alpha_1 \\ +0.19\alpha_2\alpha_1+0.17\alpha_1+0.011\alpha_2^3+0.032\alpha_2^2+0.021\alpha_2$$

Alternative 3: Offer all poles and zeros on both sides

Possibilities:

- Set $\alpha^k \to \alpha$, this reduce complexity but it not enough.
- Re-parameterize: replace complex expressions with new parameters

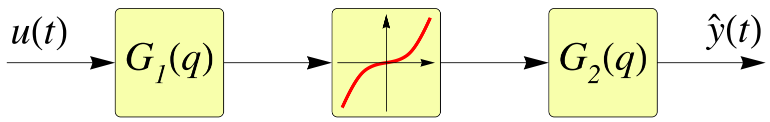
$$0.085\alpha_1^3 + 0.13\alpha_2\alpha_1^2 + 0.26\alpha_1^2 + 0.064\alpha_2^2\alpha_1$$
$$+ 0.19\alpha_2\alpha_1 + 0.17\alpha_1 + 0.011\alpha_2^3 + 0.032\alpha_2^2 + 0.021\alpha_2 \to \alpha_{new}$$

for some, or all, coefficients. After in a second step the regional parameters indicating zero-pole position can be estimated from α_{new}.

...and on this we are working.

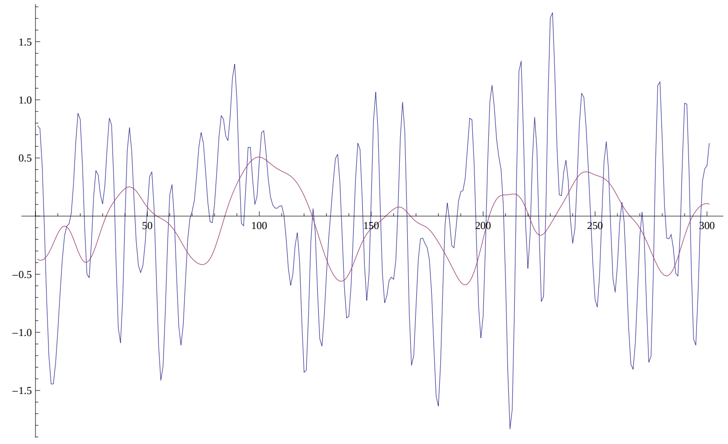
Example based on the benchmark data from SYSID 2009. Joint work with Johan Schoukens.

Data generated with an electronic nonlinear circuit with Wiener-Hammerstein structure. The two linear subsystems are of third order.

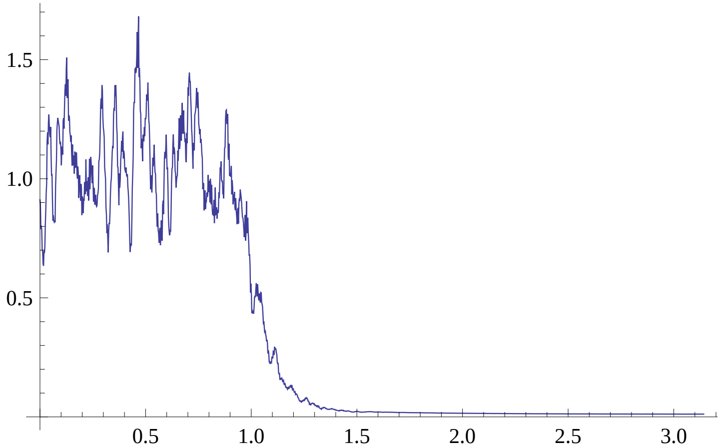


- General structure is known.
- Nonlinearity unknown.
- 100 000 estimation data
- 80 000 validation data.

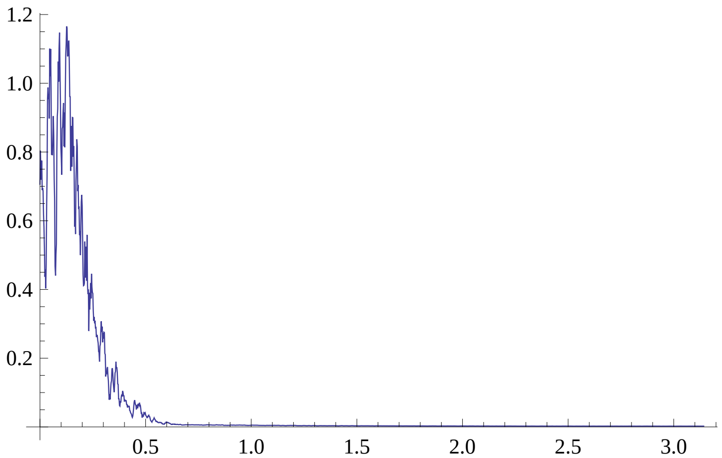
Part of Time domain data:



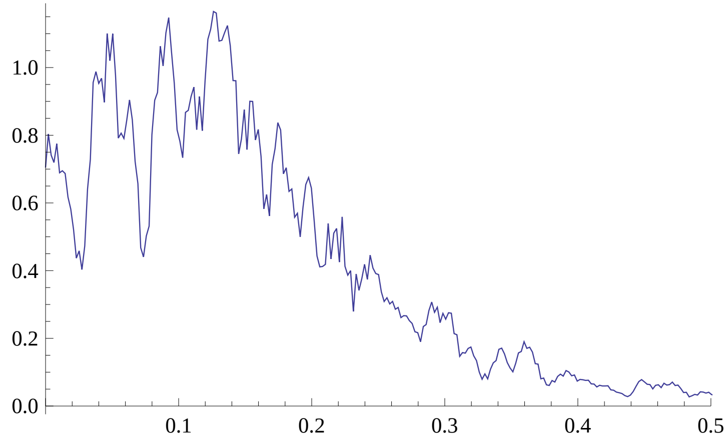
Input spectra:



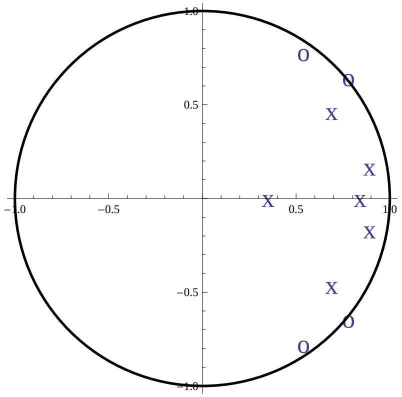
Output spectra:



Output spectra, close up:



6th order linear model: RMS error 43 mV. Poles and zeros:



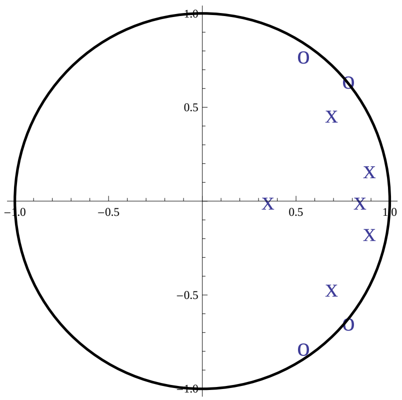
one zero at 3.5

Algorithm

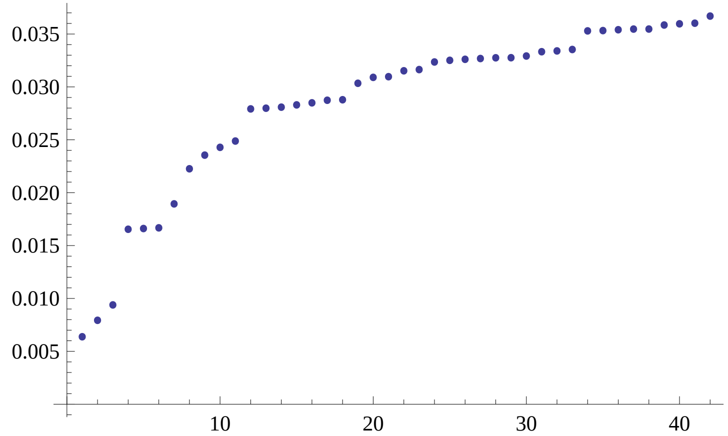
- Start with the best linear model $G(z)$.
- Split the linear model into all possible $G_1(z)$ and $G_2(z)$ so that $G(z) = G_1(z)G_2(z)$
- LS fitting of linear parameters in the nonlinearity as initialization.
- Order the initialized models with respect to their initial fit.
- Fit all parameters of the best, or some of the best models.

Initialization

- Constraint: linear subsystem order ≥ 1 gives 42 possible splittings.
- Nonlinearity: first order spline with 8 knots. Knot position initialized giving equally many data points in each interval. The linear spline parameters fitted with LS (=fast).



Sorted RMS error of the 42 initializations.

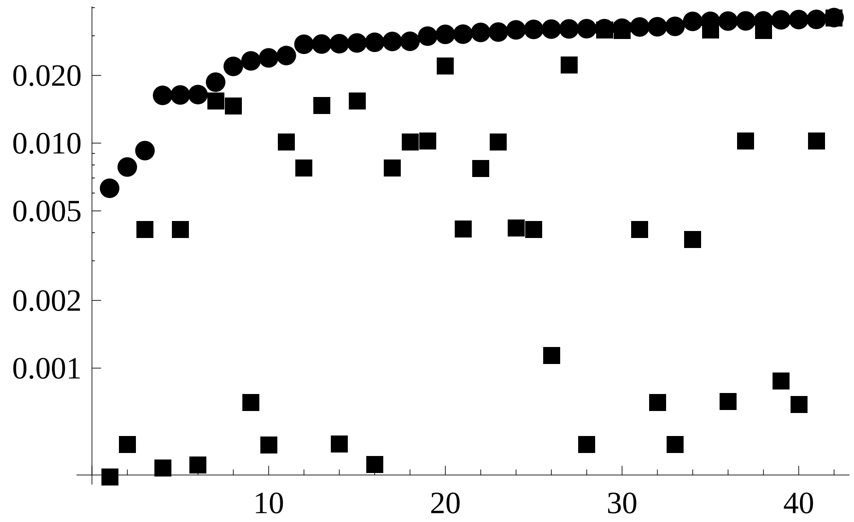


Best initialized Wiener Hammerstein model 6.4 mV

Minimization

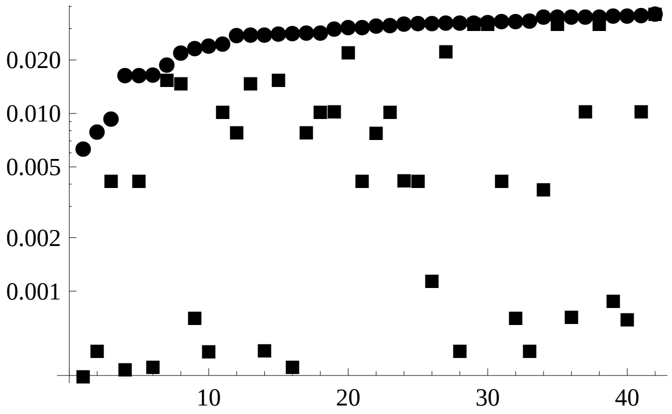
Minimize MSE for all 42 models, ie, apply the iterative minimization $\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} - \mu R_i \frac{dV}{d\theta}$.

Result

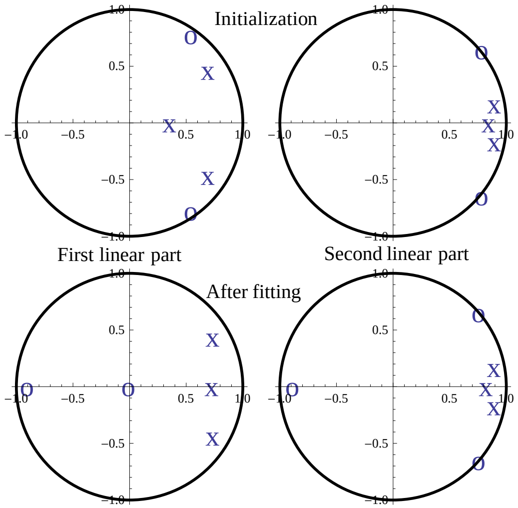


- Best initialized model also best after minimization.
- Some good initializations converge to bad minima.???
- Some bad initializations converge to good minima.???

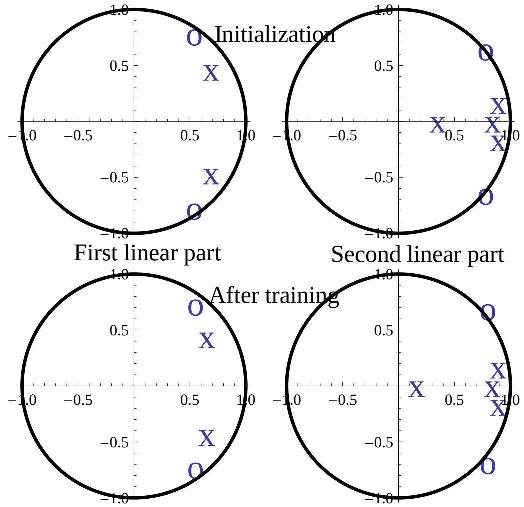
Look at poles and zeros for the two linear parts before and after minimization.



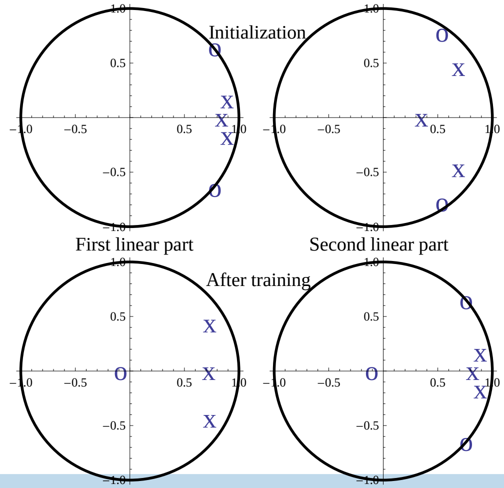
Best initialization



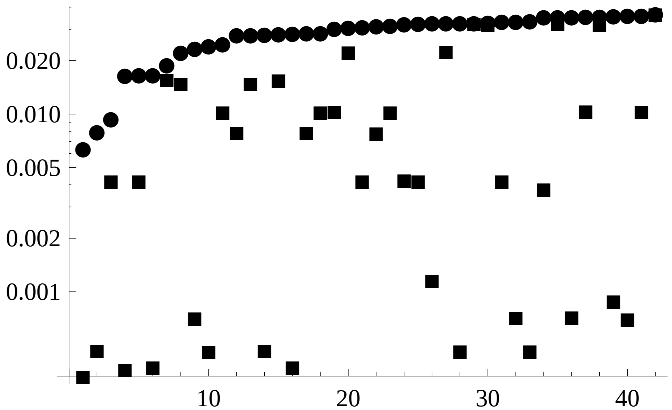
Third best initialization



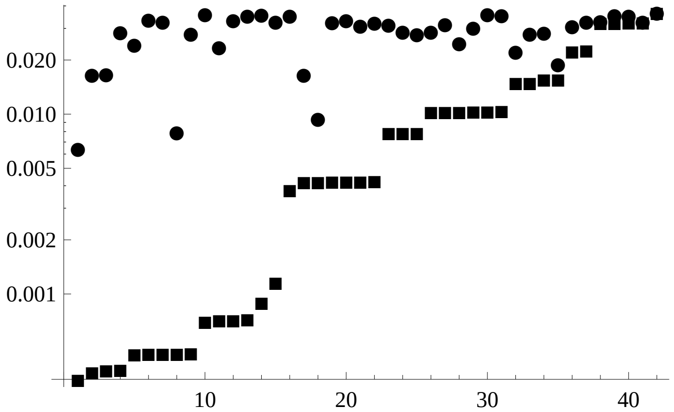
40th best initialization (third worst)



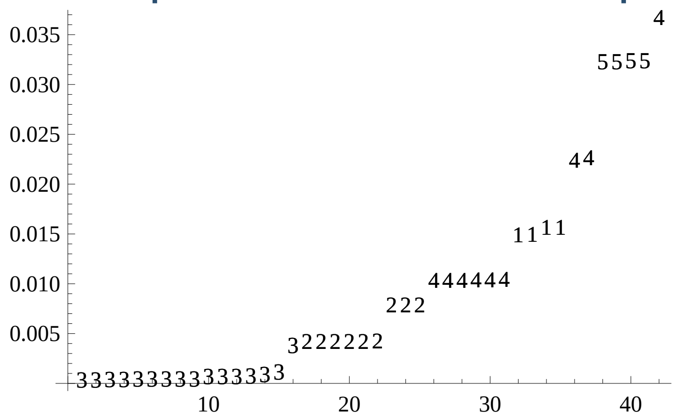
Look at the RMS errors again



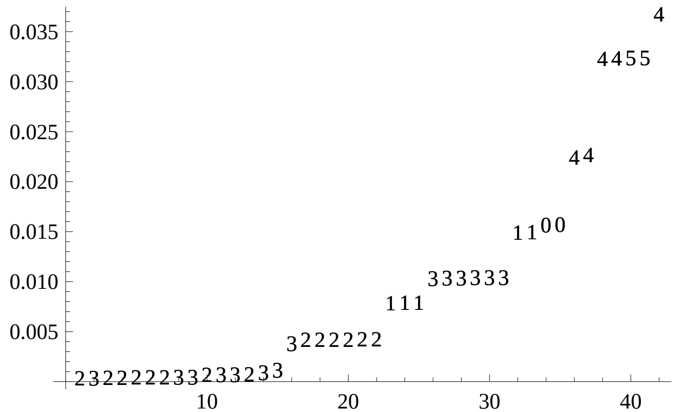
Sort the models with respect to RMS error after minimization



Number of poles in first linear part:

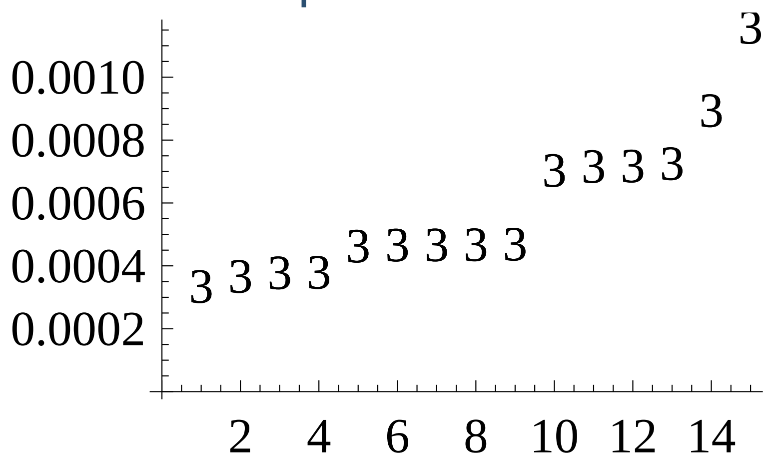


Number of zeros in first linear part:

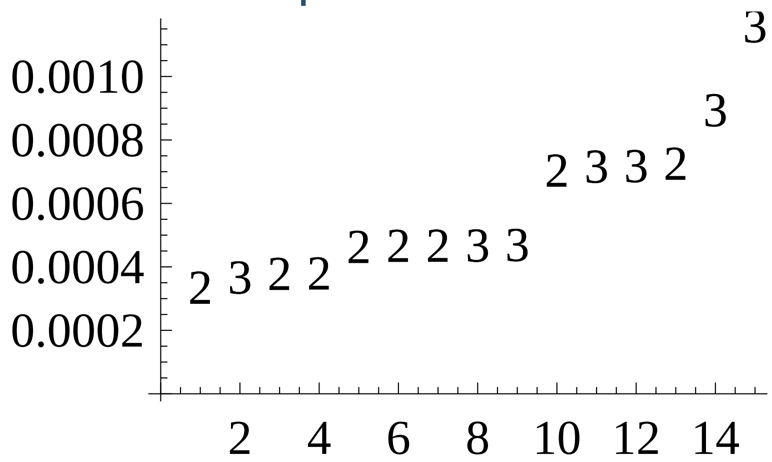


Close up on the 15 best models:

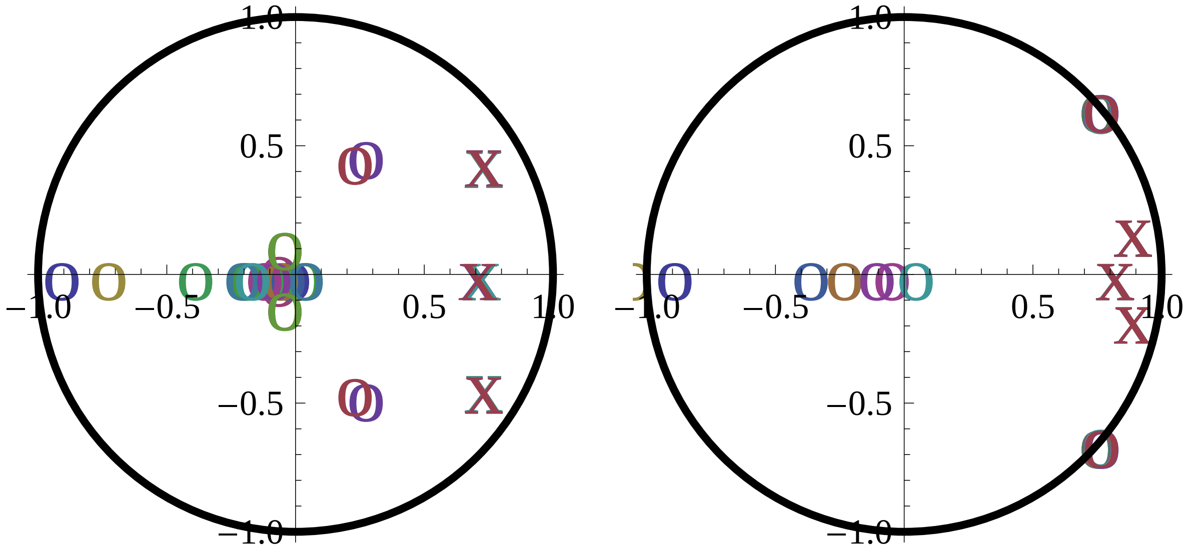
Number of poles in first linear part:



Number of zeros in first linear part:



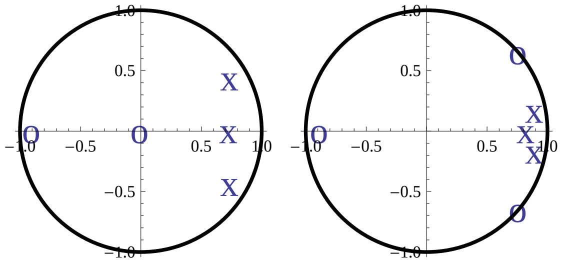
Zero and pole distribution for 15 best models



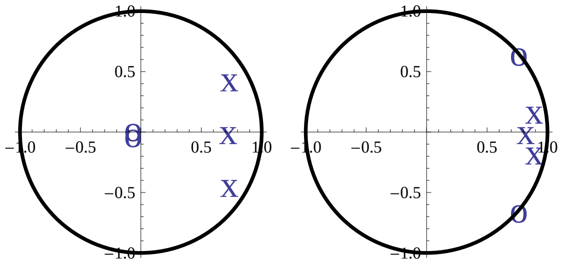
Best model

Outline
The Maximum Likelihood Method
Success with ML
Est W and H Models
Algorithm
Alternative 1: Offer all poles and zeros on both sides
Alternative 2: Offer all poles and zeros on both sides
Alternative 3: Offer all poles and zeros on both sides
Example

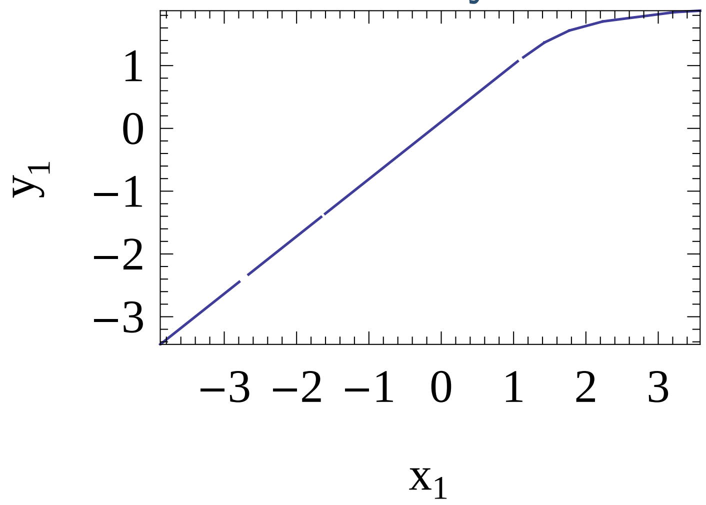
Best model:



Second best model:



Estimated nonlinearity:



Best model

- 30 parameters, 17 for the linear spline
- RMS on validation data: 0.31 mV

Can be compared to the best results at SYSID'09:

- Johan Paduart *et. al*, 0.42mv (polynomial state space model, 797 parameters)
- Wills & Ninnes , 0.49 mV (Same structure as presented here, but not using linear model for initialization).

4 of the models obtained from the linear model are better than J. Paduart's, 9 are better than Wills & Ninnes.
Additional step: add more flexibility to the nonlinear part, 8 knots to 24 knots, 0.27 mV RMS, estimated noise level 0.19 mV.

Conclusions

- Depending on initialization the solution ends up in different local minima.
- The number of poles and zeros in the linear subsystems most important for a good model.
- Correct division of the poles and zeros positions gives better initialization.
- With a consistent linear model, the best division of poles and zeros can be obtained from the fit of the initialized Wiener-Hammerstein models.