The Decoupled Polynomial NARX Model: Parameter Reduction and Structural Insights for the Bouc-Wen Benchmark

David Westwick

Department of Electrical and Computer Engineering
Schulich School of Engineering, University of Calgary

Gabriel Hollander, Johan Schoukens
Department of Fundamental Electricity,
Vrije Universiteit Brussel

April 26, 2017

The Nonlinear Autoregressive Exogenous Input Model

$$
\begin{aligned}
y(t) &= f(\boldsymbol{z}(t)) + e(t) = \sum_{k=1}^{K} \gamma_k \phi_k(\boldsymbol{z}(t)) + e(t) \\
\boldsymbol{z}(t) &= \begin{bmatrix} u(t) & \dots & u(t-n_b) & y(t-1) & \dots & y(t-n_a) \end{bmatrix}^T
\end{aligned}
$$

Pros

- Linear in Parameters
- Generality

Cons

- Many, many parameters
- Difficult to Interpret
- Lack of a Noise Model

Applications of NARX Models

There have been many recent applications of NARX modelling, *with no direct connection to the identification community*.

- Analysis of EEG
- Model of biogas producing waste digestor
- Electric Power Load Forecasting
- Wind Speed and Power Forecasting
- Hysteresis Modelling in Passive Building Control Systems

Polynomial NARX Models

$$y(t) = \sum_{k=1}^{K} \gamma_k \phi_k(\boldsymbol{z}(t)) + e(t)$$

- Many choices possible for basis elements: $\phi_k(\boldsymbol{z})$
  - Polynomials, Sigmoids, Radial Basis Functions, etc....
- Polynomials are commonly used,

Cons

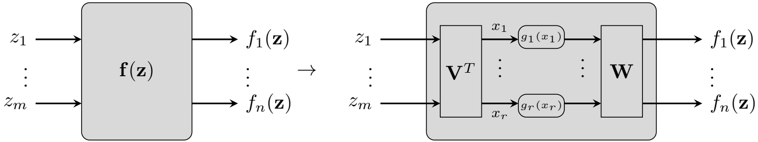
- Poor extrapolation
- Can lead to instability in simulation

Pros

- Connection to Volterra theory

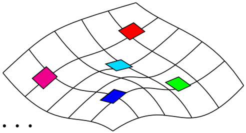
MIMO Polynomial Decoupling

- Decoupling a polynomial can reveal structure.
- Applications in Parallel Wiener-Hammerstein models, Polynomial State-Space models



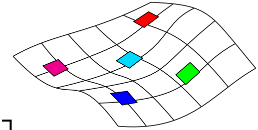
Jacobian-Based Polynomial Decoupling

- Probe $\mathbf{f}(\mathbf{z})$ at sampling points
- Collect Jacobian matrices



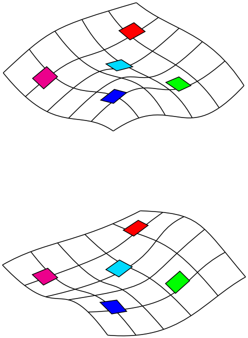
$$\mathbf{J_f}(\mathbf{z}^{(1)}), \mathbf{J_f}(\mathbf{z}^{(2)}), \mathbf{J_f}(\mathbf{z}^{(3)}), \mathbf{J_f}(\mathbf{z}^{(4)}), \mathbf{J_f}(\mathbf{z}^{(5)}), \dots$$

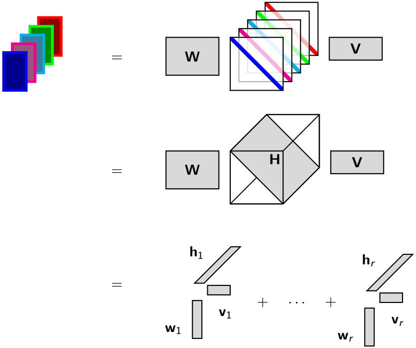
- Jacobian of $\mathbf{f}(\mathbf{z})$



$$\mathbf{f}(\mathbf{z}) = \mathbf{W}\left[g_i(\mathbf{v}_i^T\mathbf{z})\right]$$
$$\Downarrow$$
$$\mathbf{J_f}(\mathbf{z}) = \mathbf{W}\begin{bmatrix} g_1'(\mathbf{v}_1^T\mathbf{z}) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & g_r'(\mathbf{v}_r^T\mathbf{z}) \end{bmatrix}\mathbf{V}^T$$

Tensor of Stacked Jacobian Matrices





Single-Output Case

- Single output, each Jacobian is now a vector, not a matrix.
- Stacking operating points creates a matrix.

$$\boldsymbol{J} = \boldsymbol{V}\boldsymbol{H}^T$$

- Matrix factorization problem
- Tensor uniqueness results no longer apply!

Enforcing Polynomial Structure on $\boldsymbol{H}$

$$
\begin{aligned}
x_k(t) &= \boldsymbol{z}^T(t)\boldsymbol{v}_k \\
g_k'(x_k(t)) &= \sum_{\ell=0}^{n-1} c_{k,\ell}\, x_k^{\ell}(t) \\
\boldsymbol{X}_k &= \begin{bmatrix} 1 & x_k(1) & x_k^2(1) & \dots & x_k^{n-1}(1) \\ 1 & x_k(2) & x_k^2(2) & \dots & x_k^{n-1}(2) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_k(N) & x_k^2(N) & \dots & x_k^{n-1}(N) \end{bmatrix} \\
\boldsymbol{H}(\boldsymbol{V},\boldsymbol{C}) &= \begin{bmatrix} \boldsymbol{X}_1\boldsymbol{c}_1 & \boldsymbol{X}_2\boldsymbol{c}_2 & \dots & \boldsymbol{X}_r\boldsymbol{c}_r \end{bmatrix}
\end{aligned}
$$

SCHULICH School of Engineering

VUB VRIJE UNIVERSITEIT BRUSSEL

Decoupled NARX Algorithm

1. Fit a polynomial NARX model between $u(t)$ and $y(t)$

$$\hat{y}(t|t-1) = f(\mathbf{z}(t))$$

2. Compute the Jacobian

$$J(i,t) = \frac{\partial f(\mathbf{z}(t))}{\partial z_i(t)}$$

3. Factor the Jacobian using the optimization

$$(\hat{\mathbf{V}}, \hat{\mathbf{C}}) = \min_{\mathbf{V},\mathbf{C}} \left\| \mathbf{J} - \mathbf{V}\mathbf{H}(\mathbf{V},\mathbf{C}) \right\|_F$$

Final Optimization

- Optimization on Identification Data

$$\begin{aligned}\hat{\mathbf{y}}(\mathbf{V},\mathbf{C}) &= \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \dots & \mathbf{X}_r \end{bmatrix} \mathrm{vec}(\mathbf{C}) \\ (\hat{\mathbf{V}},\hat{\mathbf{C}}) &= \min_{\mathbf{V},\mathbf{C}} \|\mathbf{y} - \hat{\mathbf{y}}(\mathbf{V},\mathbf{C})\|_2\end{aligned}$$

- Linear in polynomial coeffiients $\mathbf{C}$.
- Nonlinear in mixing matrix, $\mathbf{V}$.
- Remove redundancies (constants, gains).
- Similarity to Wiener model identification.

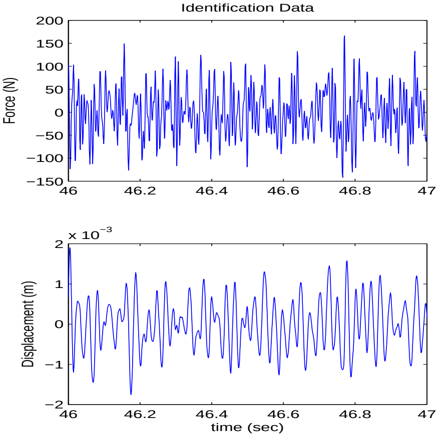
Bouc-Wen Benchmark

- Hysteresis model
- Continuous time state-space relization:

$$\begin{bmatrix} \dot{y}(t) \\ \ddot{y}(t) \\ \dot{z}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -\frac{K_L}{m_L} & -\frac{c_L}{m_L} & -\frac{1}{m_L} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y(t) \\ \dot{y}(t) \\ z(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{m_L} \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -\beta\gamma & -\beta\delta \end{bmatrix} \begin{bmatrix} |\dot{y}(t)|z(t) \\ \dot{y}(t)|z(t)| \end{bmatrix}$$

$$y(t) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} y(t) \\ \dot{y}(t) \\ z(t) \end{bmatrix}$$

Identification Data

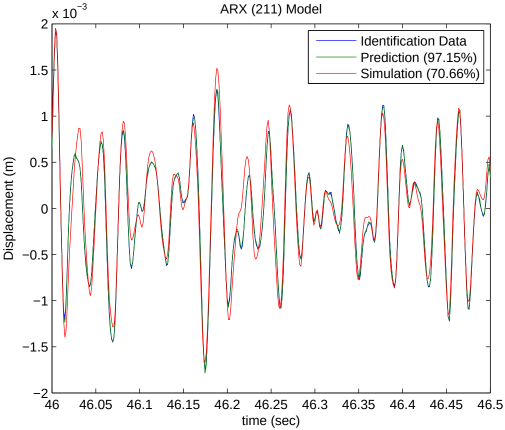


- Data generation using example provided with Benchmark
- 5 periods of 8192 samples of multisine, 50 N RMS.
- 750 Hz sampling, 54.6 seconds of data
- SNR 40dB

SCHULICH School of Engineering

VUB VRIJE UNIVERSITEIT BRUSSEL

Linear ARX models



- Minimizing AIC or MDL suggests $na = 2$, $nb = 1$, $nk = 1$.
- Prediction: 97.15%
- Simulation: 70.66%

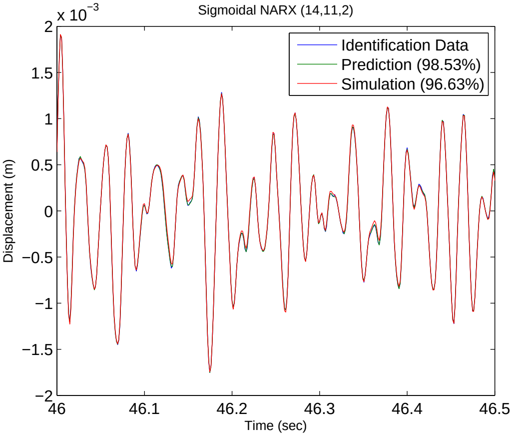
Standard NARX models

- Nonlinearities: Wavenet and Sigmoid
- Scan over
  - past outputs: $1 \leq n_a \leq 15$
  - past inputs: $1 \leq n_b \leq 15$
  - input delay: $0 \leq n_k \leq 5$
- Minimize AIC to select structure

$$n_a = 14, n_b = 11, n_k = 2$$

- Probably better to split data, and cross-validate.

Sigmoidal NARX model

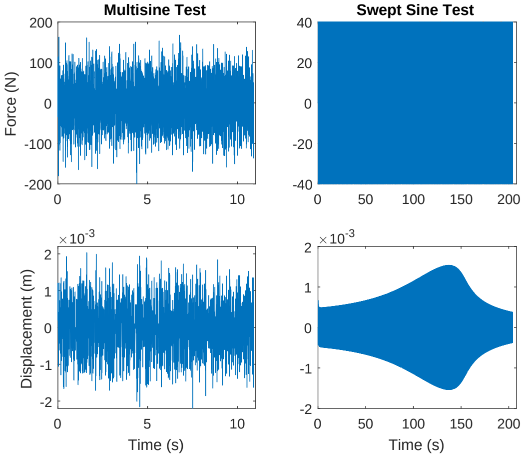


- Excellent performance in both Prediction and Simulation
- 1571 parameters!

NARX Results (estimation data)

NARX Model Type	Structure	# Param.	Pred.	Sim.
Linear	(2,1,1)	3	97.2%	70.7%
Sigmoidal 10 units	(14,11,2)	1571	98.5%	96.6%
Full Degree 3 Polynomial	(14,11,2)	3250	98.6%	97.0%
Pruned Degree 3 Polynomial	(14,11,2)	568	98.6%	96.7%
Decoupled 5 Branch Deg. 3	(14,11,2)	141	98.3%	94.6%
Decoupled 5 Branch Deg. 5	(14,11,2)	151	98.4%	95.2%
Decoupled 5 Branch Deg. 7	(14,11,2)	161	98.5%	96.1%

Validation Data



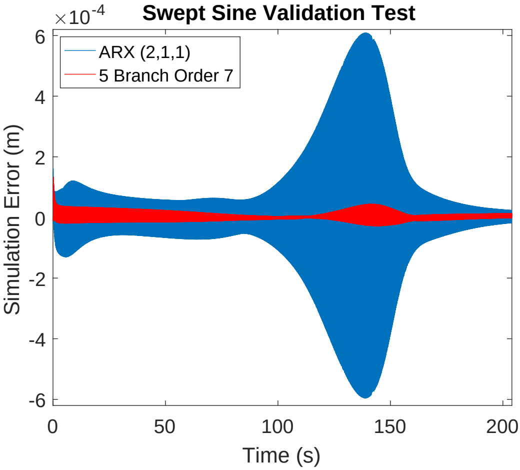
- Noise-Free Validation Signals
- Figure of Merit: Unnormalized RMS Error
- One-Step Ahead Prediction and Pure Simulation

NARX Results (validation data)

Simulation RMSE for two validation signals

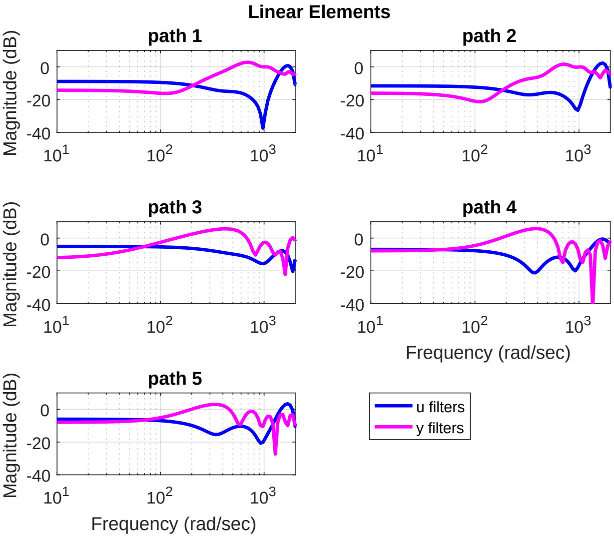
NARX Model Type	MultiSine $\times 10^{-4}$	Swept Sine $\times 10^{-4}$
Linear	1.71	1.61
Sigmoidal 10 units	0.53	0.15
Decoupled 5 Branch Deg. 3	0.59	0.21
Decoupled 5 Branch Deg. 5	0.57	0.19
Decoupled 5 Branch Deg. 7	∞	0.14

Swept Sine Test

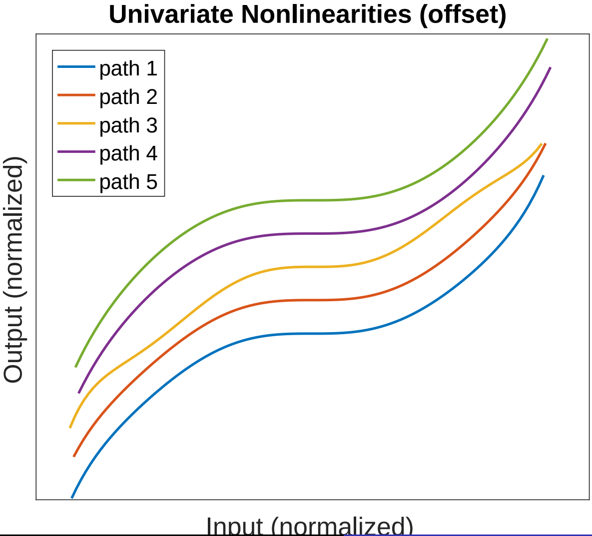


- Linear Model: 3 parameters, RMS error: -75.9dB
- Decoupled Model: 161 parameters, RMS error: -97.1 dB

Linear Elements



Nonlinearities



Conclusions

- Reduction in number of model parameters
- Insight from Linear Elements
- Simplicity of dealing with Univariate Nonlinearities.
- Poor extrapolation can lead to instability in simulation!