Nonlinear System Identification
A User-Oriented Roadmap

Johan Schoukens

Extended presentation - May 25, 2020

VRIJE UNIVERSITEIT BRUSSEL

Technische Universiteit Eindhoven University of Technology

This presentation is an extended version of the plenary lecture

Nonlinear System Identification. A User-Oriented Roadmap.

18th IFAC Symposium on System Identification, SYSID 2018, Stockholm, Sweden, July 9-11, 2018.

The paper

Nonlinear System Identification: A User-Oriented Road Map

Johan Schoukens and Lennart Ljung

IEEE Control Systems Magazine, vol. 39 (6), pp. 28-99, 2019

▸▸ Schoukens and Ljung (2019) IEEE Control Systems Magazine

provides more information on the topic and references to the material used in this presentation.


TC57
A380
A380

Linear

Nonlinear

Time-Varying

Outline

Outline

- Why is nonlinear SI so involved?
  - From hyperplane to manifold
  - Model errors
  - Process noise

  Linear or nonlinear SI? A users decision

  The lead actors in SI

  Linear identification in the presence of nonlinear distortions

  Nonlinear SI: Extensive case study

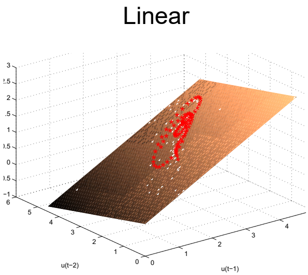
  Conclusions

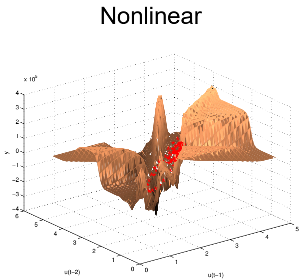
From hyperplane to manifold [1]

Linear models: a hyperplane

Nonlinear models: a manifold

- only known where domain is sampled
- extrapolation should be avoided





[1] Acknowledgement Ljung, Bode Lecture IEEE CDC 2003

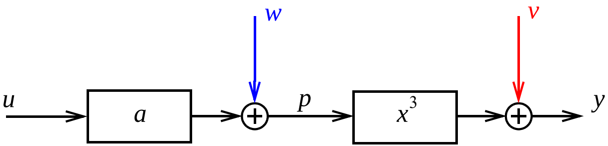
Model errors

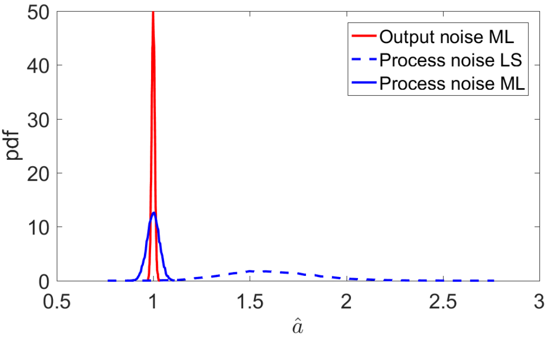
Impossible to avoid in many cases

Affect experiment design and choice cost function

Residuals no longer independent of input

Process noise





Cost functions

Outline

Linear or nonlinear SI? A users decision

Nonlinear SI much more 'expensive' than linear SI

Make a well informed decision

- Do we face a nonlinear identification problem?
- Safe to use a linear system identification approach?
- How much to gain with a nonlinear model?

Linear or nonlinear SI? A users decision

Nonlinear SI much more 'expensive' than linear SI

Make a well informed decision

- Do we face a nonlinear identification problem?
- Safe to use a linear system identification approach?
- How much to gain with a nonlinear model?

Detection, qualification, quantification NL distortions

- Characterize nonlinear behavior
- No increase of the measurement time
- Little user interaction

Linear or nonlinear SI? A users decision

Nonlinear SI much more 'expensive' than linear SI

Make a well informed decision

- Do we face a nonlinear identification problem?
- Safe to use a linear system identification approach?
- How much to gain with a nonlinear model?

Detection, qualification, quantification NL distortions

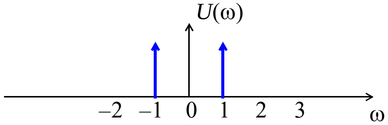
- Characterize nonlinear behavior
- No increase of the measurement time
- Little user interaction

Tool: well-designed periodic excitations

$$u_0(t) = \frac{2}{\sqrt{N}} \sum_{k=1}^{N/2-1} U_k cos\left(2\pi k f_0 t + \varphi_k\right)$$

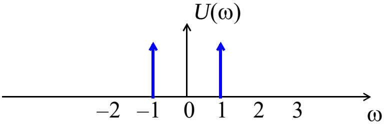
Understanding nonlinear systems: $y = u^3$

$$u(t) = 2\cos\omega t$$
$$= e^{j\omega t} + e^{-j\omega t} \text{ with } \omega = 1$$



Understanding nonlinear systems: $y = u^3$

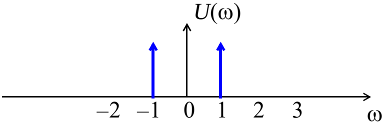
$$u(t) = 2\cos\omega t$$
$$= e^{j\omega t} + e^{-j\omega t} \text{ with } \omega = 1$$



$$y = u^3$$
$$= (e^{j\omega t} + e^{-j\omega t})(e^{j\omega t} + e^{-j\omega t})(e^{j\omega t} + e^{-j\omega t})$$

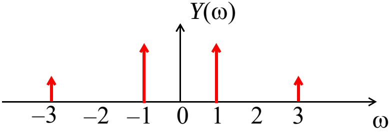
Understanding nonlinear systems: $y = u^3$

$$u(t) = 2\cos\omega t$$
$$= e^{j\omega t} + e^{-j\omega t} \text{ with } \omega = 1$$

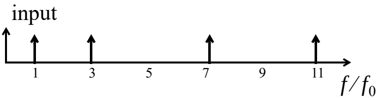


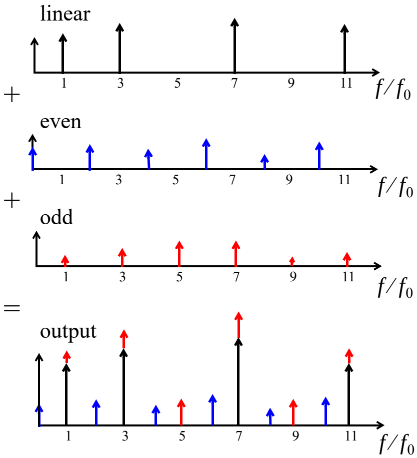
$$y = u^3$$
$$= (e^{j\omega t} + e^{-j\omega t})(e^{j\omega t} + e^{-j\omega t})(e^{j\omega t} + e^{-j\omega t})$$

1	1	1	3
1	1	-1	1
1	-1	1	1
1	-1	-1	-1
-1	1	1	1
-1	1	-1	-1
-1	-1	1	-1
-1	-1	-1	-3



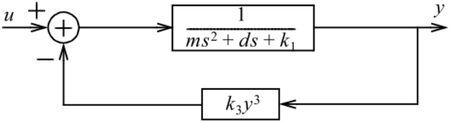
Detection and qualification of nonlinear distortions



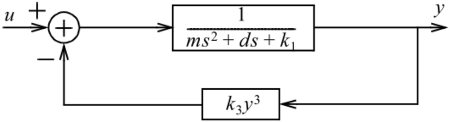


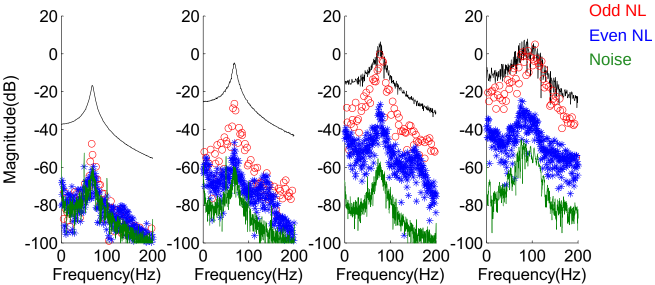
▸▸ A detailed example

Example: the forced Duffing oscillator



Example: the forced Duffing oscillator



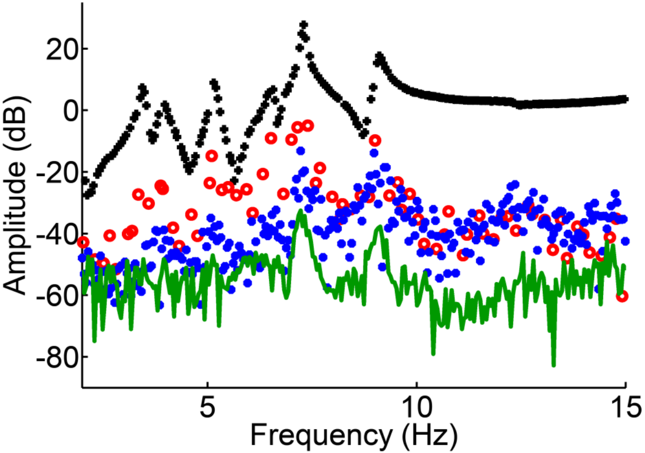


▸▸ Estimation nonparametric noise model

Example: an air fighter[2]

[2] Acknowledgement M. Vaes (VUB), B. Peeters, J. Debille (Siemens Industry Sofware), T. Dossogne, J.P. Noël, C. Grappasaonni, G. Kerschen (ULg)

Example: an air fighter[2]

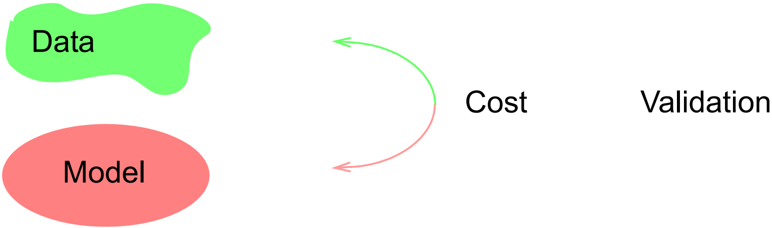


Output Odd NL Even NL Noise

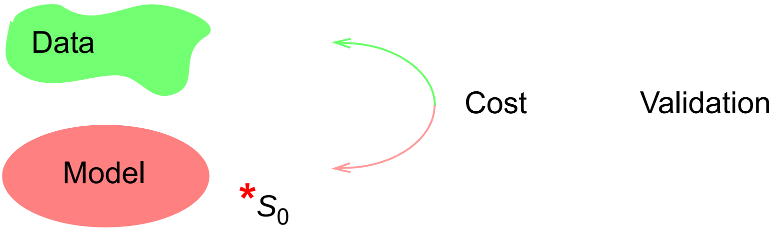
[2] Acknowledgement M. Vaes (VUB), B. Peeters, J. Debille (Siemens Industry Sofware), T. Dossogne, J.P. Noël, C. Grappasaonni, G. Kerschen (ULg)

Outline

Lead actors in SI from a nonlinear perspective

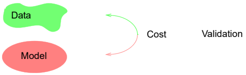


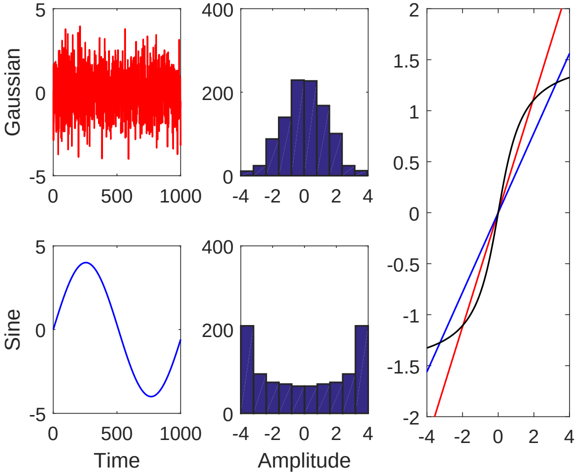
Lead actors in SI from a nonlinear perspective



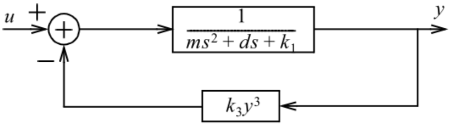
▸▸ User choices in the presence of model errors

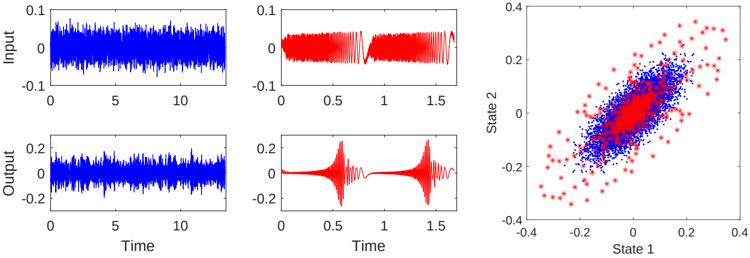
Data: amplitude distribution





Data: cover domain of interest





⏩ Error plot

Data: experiment design

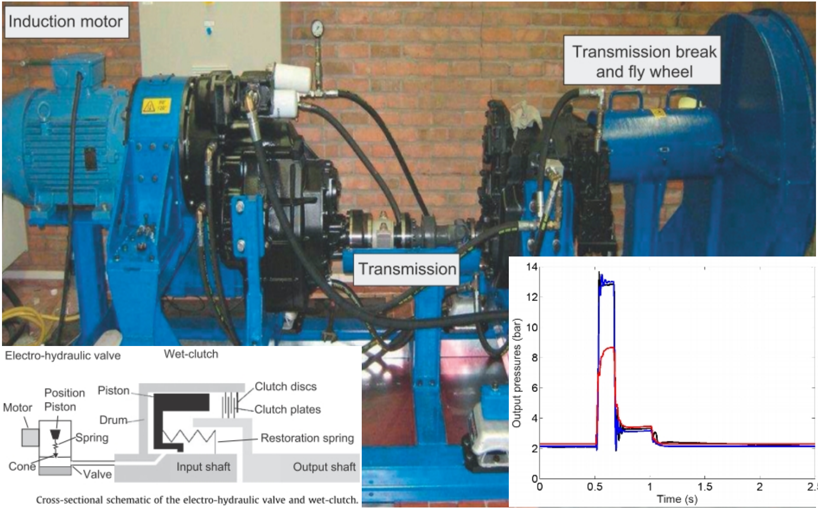
Optimal experiment design is still an open problem

User guidelines

- Use periodic excitations
  - Nonparametric distortion analysis
  - No user interaction
  - Separation of plant and noise model
  - No inference with model errors
- Cover amplitude and frequency range of interest
  - Necessary but not sufficient condition
  - Pay attention to the state domain
  - Strongly linked to application: use excitation with same nature
- Repeat experiment with new realization of random excitation
- Use linear SI insights to excite the dynamics

Design a periodic excitation

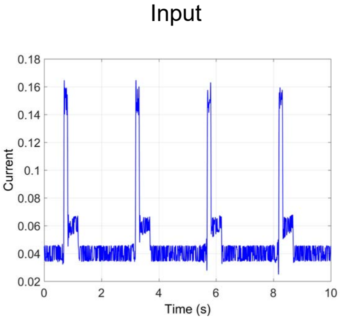
Experiment design: example[3]

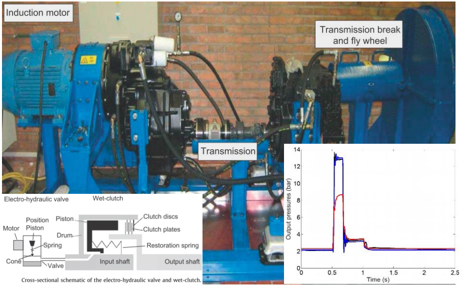


Cross-sectional schematic of the electro-hydraulic valve and wet-clutch.

[3] Acknowledgement W.D. Widanage, A. Van Mulders (VUB), J. Stoev, G. Pinte (FMTC)

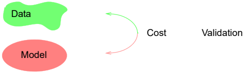
Experiment design: example[3]





[3] Acknowledgement W.D. Widanage, A. Van Mulders (VUB), J. Stoev, G. Pinte (FMTC)

Cost



Minimize the distance between the data and the model

Model errors dominate $\rightarrow$ move away from ML paradigm

Should reflect user's need how to shape model errors

Least squares cost functions in TD and FD

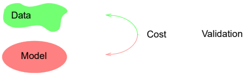
$$\hat{\theta}_N = \text{argmin}_\theta \sum_{t=1}^{N} \|y(t) - \hat{y}(t|\theta)\|^2$$

$$\hat{\theta}_N = \text{argmin}_\theta \sum_{f \in F} \|Y(f) - \hat{Y}(f|\theta)\|^2$$

Can be combined with regularization

Basic idea regularization

Model



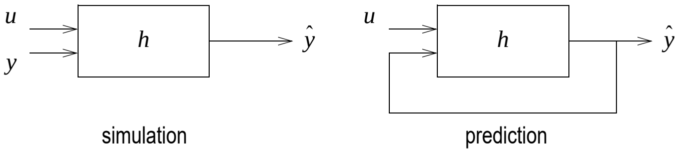
Estimates future outputs

$$\hat{y}(t|\theta, Z^{t-})$$

What data are used?

Model structure?

Model: what data are used?



Prediction model

uses past inputs and outputs: $Z^t = \{u^t, y^{t-1}\}$

Simulation model

uses only past inputs: $Z^t = u^t$

Basic idea prediction

Model: selection model structure

$$\hat{y}(t|\theta, Z^{t-})$$

System behavior

open loop - dynamic NL closed loop

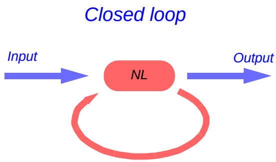
Users choice

white box - black box models

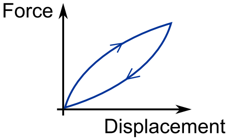
Nonlinear function

present in every nonlinear model
$q(t) = f(p(t))$ with p, q signals in the model
f a static multivariate function

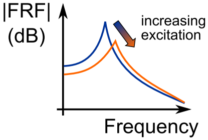
System behavior: open loop - dynamic NL closed loop



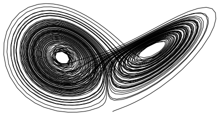
Hysteresis



Varying dynamics



Chaos



System behavior: open loop - dynamic NL closed loop

NL Open loop

- NL not captured in a dynamic feedback loop
- Fading memory
- Examples
  - NFIR
  - Volterra ⏩ Volterra theory in a nutshell
  - Block-oriented models: Wiener, Hammerstein, Wiener-Hammerstein, Hammerstein-Wiener
  - Nonlinear state space with lower triangular structure

Dynamic NL closed loop

- Covers complex non-fading memory behavior
  - shifting resonances, hysteresis, chaos,. . .
- Can become unstable
- Examples
  - NIIR and NARX
  - Closed loop block-oriented systems
  - Nonlinear state space

Users choice: white box - black box models

White models

- Dedicated: new model for new problem
- Expensive
- Compact
- Provide physical insight

Black box models

- Generic methodology
- Behavior modeling
- Exploding number of parameters
- Can provide intuitive insight

Users choice: white box - black box models

- White models: physical models
  - Estimate value physical parameters
- Smoke-grey models: semi-physical modeling
  - Natural selection of the variables
- Steel-grey models: linearization based models
  - Models depend on working point and nature of excitation
- Slate-grey models: block oriented models
  - Structural insight can be injected
- Black models: universal approximators
  - Volterra, NARX, Nonlinear state space
- Pit-black models: nonparametric smoothing

White Box Black Box Examples

Black box models complexity
Keep the exploding number of parameters under control

- Regularization
  - Sparse solutions
    - Force parameters to zero
    - Brute force
  - Smooth solutions
    - Impose smoothness
    - Number of parameters not changed
- Data driven structure retrieval
  - Decouple multivariate nonlinear function $p = f(q)$
  - Search for a natural basis
  - Reduction from combinatorial to linear grow complexity

Basic idea regularization

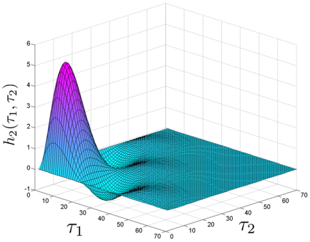
Regularization: impose smoothness on Volterra kernel

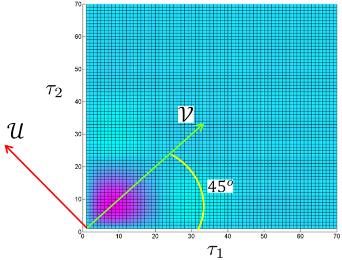
Model

$$y_0(t) = \sum_{0}^{n_1} g_1(\tau)u(t-\tau) + \sum_{0}^{n_2}\sum_{0}^{n_2} g_\alpha(\tau_1,\tau_2)u(t-\tau_1)u(t-\tau_2) + \dots$$

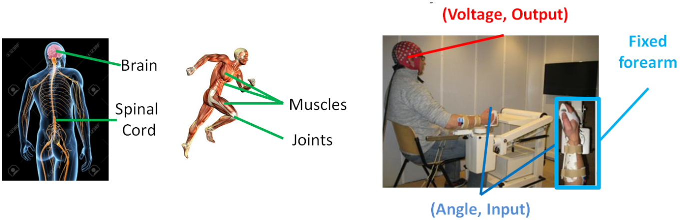
Cost

$$V = \frac{1}{N}\sum_{t=1}^{N}(y(t)-y_{mod}(t,\theta))^2 + [\theta_1^T \theta_2^T]\begin{bmatrix} P_1^{-1} & 0 \\ 0 & P_2^{-1}\end{bmatrix}\begin{bmatrix}\theta_1 \\ \theta_2\end{bmatrix}$$





Example: Volterra model wrist-brain sensorimotor system[4]



Experiment Design

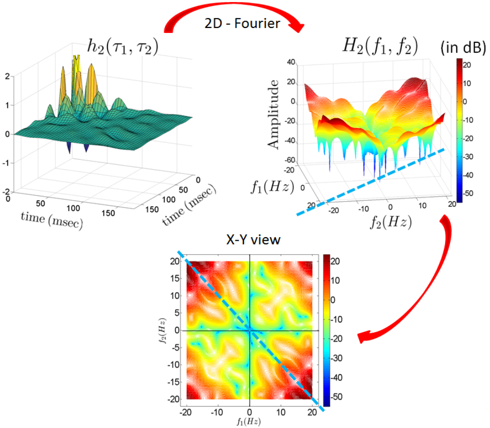
Random odd multisine $[1, 3, 5, 7, 9, 11, 13, 15, 19, 23]$Hz
Averaged over 210 periods
7 Realizations

Model

2^{nd} degree Volterra kernel
Fixed delay 20 ms
Memory length 130 ms (33 samples)

[4] Acknowledgement G. Birpoutsoukis (VUB), M. Vlaar, F. Van der Helm and A. Schouten (TU Delft)

Example: Volterra model wrist-brain sensorimotor system



Results

Linear kernel 10% VAF (Variance Acounted For)
2^{nd} degree Volterra model 45% VAF
2^{nd} degree Volterra model 60% VAF on improved experiment
high-pass system

Black box models complexity
Keep the exploding number of parameters under control

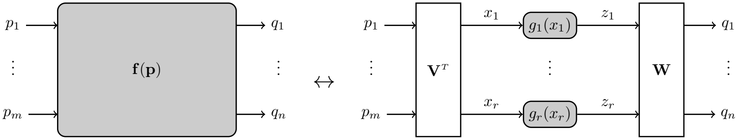
- Regularization
  - Sparse solutions
    - Force parameters to zero
    - Brute force
  - Smooth solutions
    - Impose smoothness
    - Number of parameters not changed
- Data driven structure retrieval
  - Decouple multivariate nonlinear function $p = f(q)$
  - Search for a natural basis
  - Reduction from combinatorial to linear grow complexity

Data driven structure retrieval: decoupling[5]

Multivariate nonlinear function $q = f(p)$

Decouple

$$q = W\mathbf{g}(V^\top p)$$
$$\mathbf{g}_i = g_i(x_i), i = 1, \ldots, r \text{ with } x = V^\top p$$

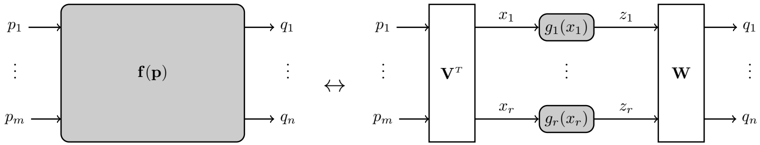


$O(m^d)$ $\qquad$ $O(rd)$

[5] Acknowledgement D. Philippe, M. Ishteva, K. Tiels (VUB)

Decoupling: example

$$
\begin{aligned}
q_1 &= f_1(p_1, p_2) \\
&= 54p_1^3 - 54p_1^2 p_2 + 8p_1^2 + 18p_1 p_2^2 + 16p_1 p_2 - 2p_2^3 + 8p_2^2 + 8p_2 + 1 \\
q_2 &= f_2(p_1, p_2) \\
&= -27p_1^3 + 27p_1^2 p_2 - 24p_1^2 - 9p_1 p_2^2 - 48p_1 p_2 - 15p_1 + p_2^3 - 24p_2^2 - 19p_2 - 3
\end{aligned}
$$



$$
\begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ -3 & -1 \end{bmatrix} \begin{bmatrix} 2x_1^2 - 3x_1 + 1 \\ x_2^3 - x_2 \end{bmatrix}, \quad \text{with} \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -2 & -2 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}
$$

Decoupling: from complexity control to model reduction

Complexity control

Number of parameters

Full model $O(n^d)$

Decoupled model $O(rd)$

Force all $g_i(x) = g(x), \forall i$

$$q = W\mathbf{g}(V^\top p)$$

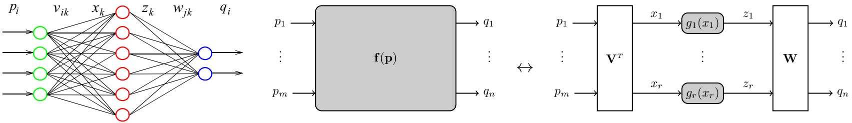
$$\mathbf{g}_i = g(x_i), i = 1, \ldots, r \text{ with } x = V^\top p$$

Model reduction

Reduce number of branches

Balance complexity versus model errors

Decoupling: Link with Neural Networks



Neural network

- Activation functions (red circles) prior user choice
  - Example: sigmoids, Gaussian bells, ReLU functions
- Parameters tuned on the data

Decoupled model

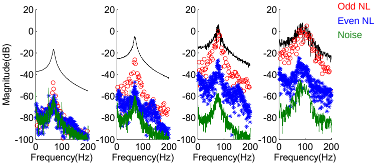
- Nonlinear functions set by the data
  - Nonparametric or parametric representation

Outline

Outline

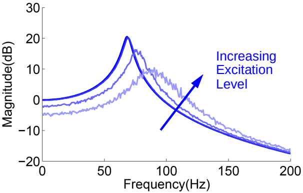
Forced Duffing Oscilator: Initial nonparametric analysis

Nonlinear distortion analysis



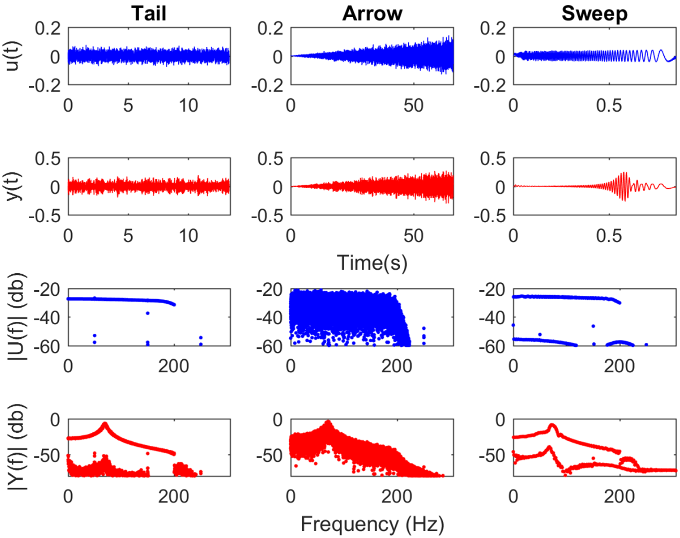
Conclusion: nonlinear distortions dominate noise

FRF measurement



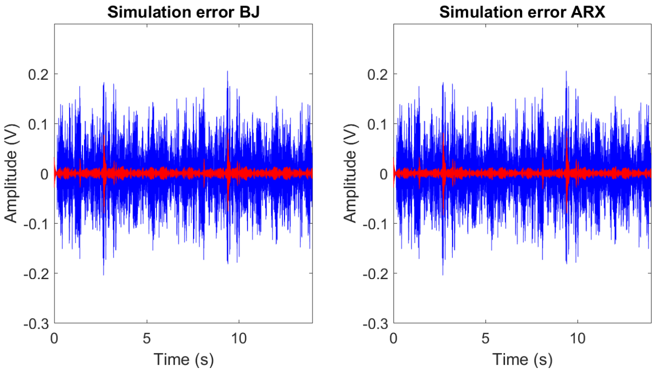
Conclusion: Nonlinear feedback model needed

Forced Duffing Oscilator: the data[6]

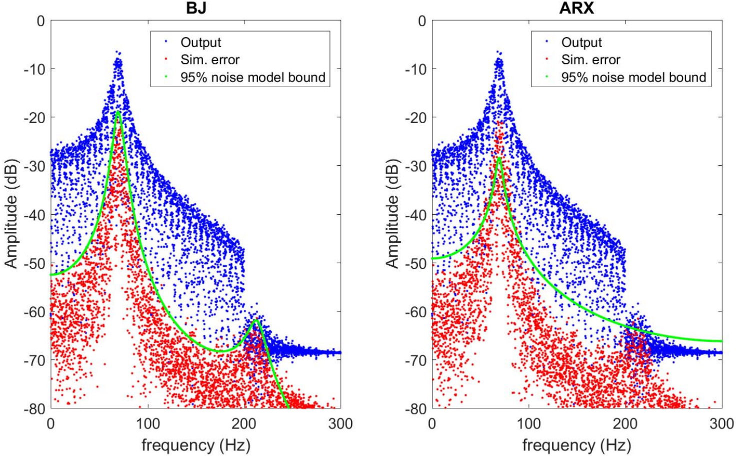


[6] Data available on http://www.nonlinearbenchmark.org (Silverbox Benchmark)

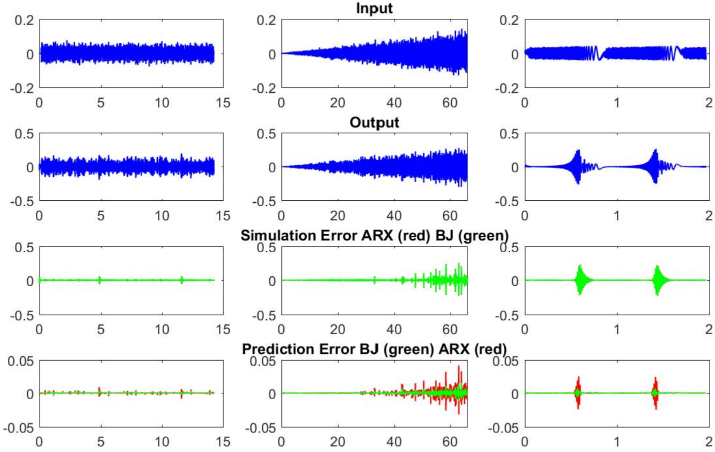
Forced Duffing Oscilator: linear models



Forced Duffing Oscilator: linear models



Forced Duffing Oscilator: linear models

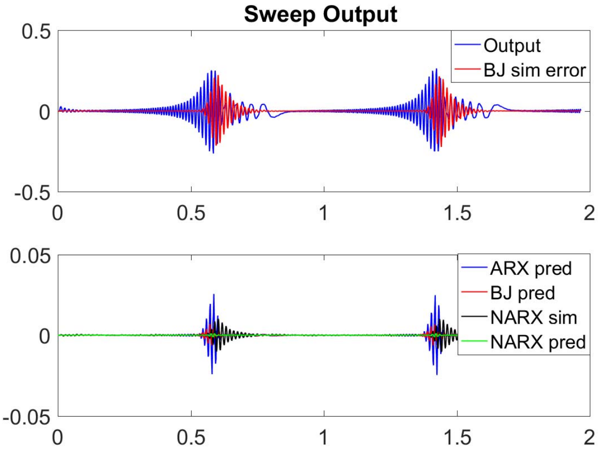


Forced Duffing Oscilator: from linear to nonlinear models

NARX

$$y(t) = P(u(t), u(t-1), u(t-2), y(t-1), y(t-2))$$

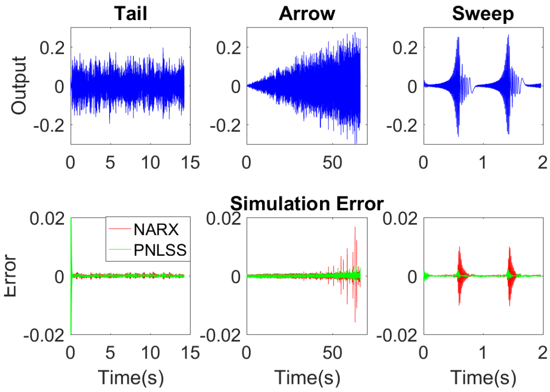
P Polynomial degree 3



Forced Duffing Oscilator: black box nonlinear state space model[7]

Nonlinear State space

- 2 states
- Polynomial degree 3



[7] Acknowledgement K. Tiels (University of Uppsala)

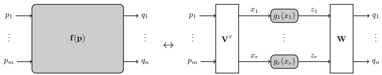
Forced Duffing Oscilator: from black box to highly structured models[8]

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = A \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + Bu(k) + \begin{bmatrix} f_1(x_1(k), x_2(k), u(k)) \\ f_2(x_1(k), x_2(k), u(k)) \end{bmatrix}$$

$$y(k) = C \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + Du(k)$$

f polynomial degree 3

Decouple $f(x_1(k), x_2(k), u(k))$



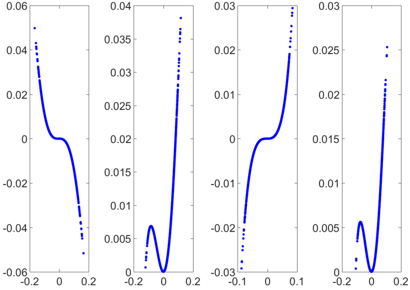
[8] Acknowledgement J. Decuyper (VUB)

Forced Duffing Oscilator: Decoupled Model

4 branches

Polynomial degree 3 $\rightarrow$ 5

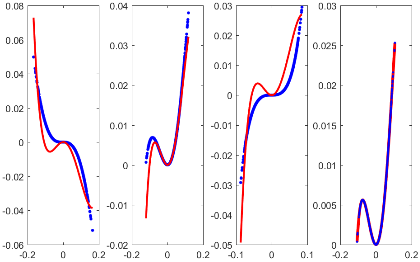
	BLA	NLSS	Decoupled
RMS error	12%	0.49%	0.40%
n_{θ_L}	5	5	5
$n_{\theta_{NL}}$	0	30	12



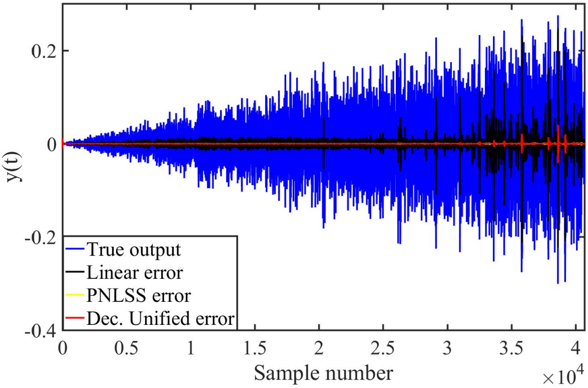
Forced Duffing Oscilator: Decoupled + Equal Branches

Impose all branches are equal

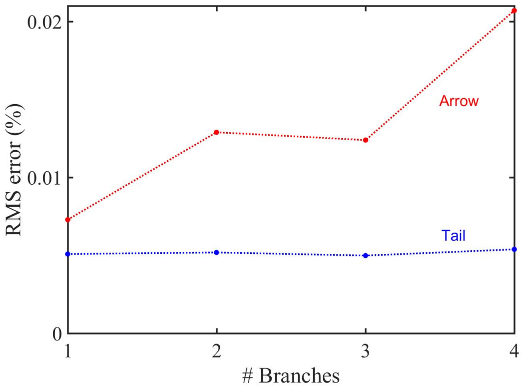
	BLA	NLSS	Decoupled	Equal Branches
RMS error	12%	0.49%	0.40%	0.40%
n_{θ_L}	5	5	5	5
$n_{\theta_{NL}}$	0	30	12	6



Forced Duffing Oscilator: Decoupled + Equal Branches



Forced Duffing Oscilator: Decoupled, Equal + Single Branch

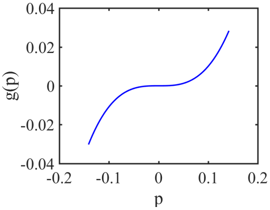


Forced Duffing Oscilator: Decoupled, Equal, Single Branch

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = A \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + Bu(k) + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} g(p)$$

$$y(k) = C \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + Du(k)$$

$$p = v_1 x_1(k) + v_2 x_2(k) + v_3 u(k)$$



Forced Duffing Oscilator: Final model

Black box

Data driven structure retrieval

Single branch

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = A \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + Bu(k) + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} g(p)$$

$$y(k) = C \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + Du(k)$$

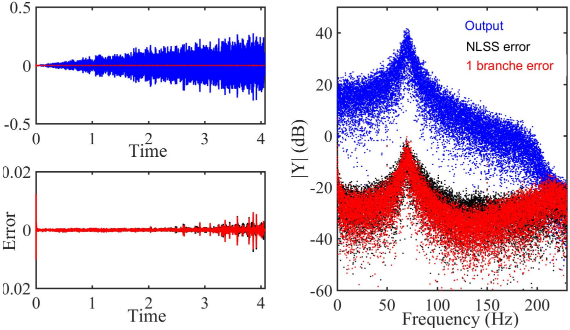
$$p = v_1 x_1(k) + v_2 x_2(k) + v_3 u(k)$$

Forced Duffing Oscilator: Final model

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = A \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + Bu(k) + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} g(p)$$

$$y(k) = C \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + Du(k)$$

$$p = v_1 x_1(k) + v_2 x_2(k) + v_3 u(k)$$

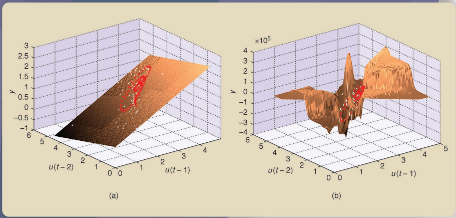


Conclusions

- Why is nonlinear SI so involved?
  - From hyperplane to manifold
  - Model errors
  - Process noise
- Linear or nonlinear SI? A users decision
  - Nonparametric distortion analysis
  - Guidance model structure selection
- The lead actors in SI
  - Impact model errors on Experiment, Model, Cost Function
- Data driven insight
  - Structure retrieval
  - Model reduction

Nonlinear System Identification

A USER-ORIENTED ROAD MAP



JOHAN SCHOUKENS and LENNART LJUNG

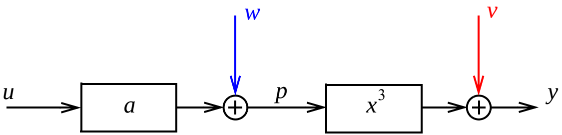
Nonlinear system identification is an extremely broad topic, since every system that is not linear is nonlinear. That makes it impossible to give a full overview of all aspects of the field. For this reason, the selection of topics and the organization of the discussion are strongly colored by the personal journey of the authors in this nonlinear universe.

The identification of linear dynamic systems started in the late 1950s. Zadeh [1] prioritized the need for a well-developed system identification framework at the very outset, followed by early overviews of the field [2]. A series of books established the field [3]–[7]. Linear system identification presented many successes, and data-driven modeling became an enabling factor in modern design methods. Nonlinear system identification [8]–[29] began when linear system identification [6], [7], [30] failed to address users' questions. The real world is nonlinear and time varying, and, in some applications, these aspects cannot be ignored (see Figure 1). Therefore, linear models are imprecise or do not reproduce essential aspects of system behavior. This article is focused on nonlinear system identification. Overviews of time-varying system identification are given in [31] and the references therein.

Nonlinear behavior appears in many engineering problems. In mechanical engineering, nonlinear stiffness, damping, and interconnections influence ground vibration tests of airplanes and satellites, resulting in resonance frequencies and dampings that vary with the excitation level (see Figure 2, [32], and [33]). In telecommunications, power amplifiers are pushed into a nonlinear operation regime to improve power efficiency. Distillation columns exhibit nonlinear dynamic behavior. Many biological systems (for example, the eyes, ears, and sense of touch) first apply a nonlinear compression (known as the *Weber–Fechner law*) to cover the very large dynamic range of the inputs.

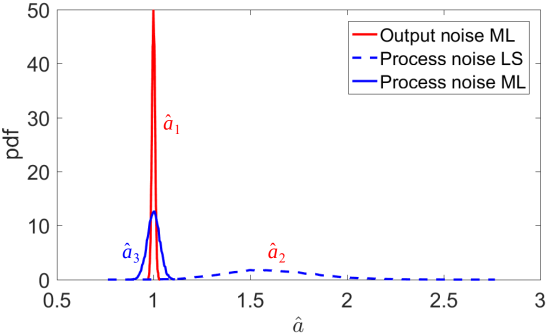
Digital Object Identifier 10.1109/MCS.2019.2938121
Date of current version: 13 November 2019

1066-033X/19©2019IEEE



$$\hat{a}_{1,2} = \text{argmin}_a \sum_{t=1}^{N} \| y(t) - [au(t)]^3 \|^2$$

$$\hat{a}_3 = \text{argmin}_a \sum_{f \in F} \| \sqrt[3]{y(t)} - au(t) \|^2$$

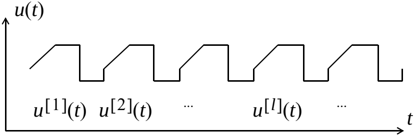


return

Noise model, prior analysis
periodic excitation

Estimate $\sigma_U^2(k)$, $\sigma_Y^2(k)$ and $\sigma_{YU}^2(k)$

Additional information: the signals are periodic



$$\hat{U}(k) = \frac{1}{M}\sum_{l=1}^{M} U^{[l]}(k), \quad \hat{Y}(k) = \frac{1}{M}\sum_{l=1}^{M} Y^{[l]}(k),$$

$$\hat{\sigma}_U^2(k) = \frac{1}{M-1}\sum_{l=1}^{M} \left|U^{[l]}(k) - \hat{U}(k)\right|^2 \text{ and } \hat{\sigma}_Y^2(k) = \frac{1}{M-1}\sum_{l=1}^{M} \left|Y^{[l]}(k) - \hat{Y}(k)\right|^2$$

$$\hat{\sigma}_{YU}^2(k) = \frac{1}{M-1}\sum_{l=1}^{M} (Y^{[l]}(k) - \hat{Y}(k))\overline{(U^{[l]}(k) - \hat{U}(k))}$$

⏩ return

Noise model, prior analysis
periodic excitation

Properties

consistency: $M = 4$ periods are enough

efficiency:

$M = 6$ periods are enough

'loss' in efficiency $C_{\theta SML} = \frac{M-2}{M-3} C_{\theta ML}$

normality: $M = 7$ is enough

Recent results

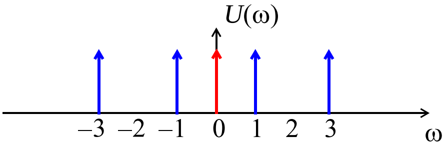
2 periods + overlapping windows are enough

See:

Welch Method Revisited: Nonparametric Power Spectrum Estimation Via Circular Overlap
Barbe, K.; Pintelon, R.; Schoukens, J.
IEEE TRANSACTIONS ON SIGNAL PROCESSING, Vol. 58, pp. 553-565, 2010

return

Frequency analysis nonlinear distortions
Impact DC component



	No DC	DC
ω	-3 -1 1 3	-3 -1 0 1 3
u	-3 -1 1 3	-3 -1 0 1 3
u^2	0 1-1 3-3 2 1+1 3-1 …	1 1+0 3 3+0 …
u^3	1 1+1-1 3-1-1 3 1+1+1 3+1-1 …	0 1+0-1 2+0-2 2 3+0-1 2+1+0 …

return

Snow-White Models

$$x(t+1) = f(x(t), u(t), w(t))$$
$$y(t) = h(x(t), u(t)) + v(t)$$

$v(t), w(t)$ sequences of independent random variables.

f, h obtained from physical insight

return

Off-White Models

Model depends on physical parameters θ with unknown value

$$x(t+1) = f(x(t), u(t), w(t), \theta)$$
$$\hat{y}(t|\theta) = h(x(t), u(t), \theta) + v(t)$$

f, h parameterized on θ

$\hat{y}(t|\theta)$ predicted output for parameter value θ

return

Steel-Grey Models: Semi-Physical Modelling

Using qualitative reasoning rather than formal equations

Example: electrical motor

- acceleration $\frac{d}{dt}\omega \sim T_e - T_f - T_L$
  - electrical torque $T_e \sim i$
  - friction torque $T_f \sim \omega$
  - load torque $T_L \sim \omega$
- current $i \sim u - u_{bef}$
  - u applied input voltage
  - back electromotive force $u_{bef} \sim \omega$

$$\theta_1 \frac{d}{dt}\omega(t) + \theta_2 \omega(t) = u(t)$$

return

Steel-Grey Models: Semi-Physical Modelling

Nonlinear transformations of the measured data

Example: heat production in resistor

$$P(t) = Ri(t)^2$$
$$i(t) \sim u(t)$$

Use $u^2(t)$ as input of linear model $P = f(u^2(t))$

return

Steel-Grey Models: Linearization-Based Models

Best Linear Approximation of a nonlinear system: BLA

$$G_{BLA} = \arg\min_{G} E\left\{\left|y_0(t) - G(q)\,u(t)\right|^2\right\}$$

Local linear models

$$\hat{y}(t|\theta, Z^{1-1}) = \sum_{i=1}^{d} \rho(p(t), p_i)\hat{y}_i(t|\theta, Z^{t-1})$$

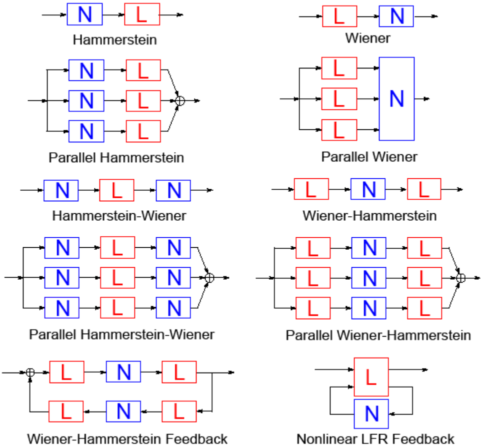
ρ is a weighting or *validity function*

p_i regime variable: the local working point

$\hat{y}_i(t|\theta, Z^{t-1})$ local linear model around p_i

return

Block-Oriented Models



N: static nonlinear block *L*: dynamic linear block

return

Black-Box Models: Universal Approximators

Nonlinear state space models

$$x(t+1) = f(x(t), u(t), \theta),$$
$$\hat{y}(t|\theta) = h(x(t), \theta).$$

Special case: NARX Nonlinear Autoregressive Exogenous models

states $x(t)$: finite number of past inputs and outputs

$$x(t) = [y(t-1), \ldots, y(t-n_a), u(t-1), \ldots, u(t-n_b)]^T$$

regressors $\varphi(t) = x(t)$

NARX model

$$\hat{y}(t|\theta) = h(\varphi(t), \theta)$$

return

Pit-Black Models: Nonparametric Smoothing

Model: $y(t) = g(\varphi(t)) + e(t)$

Regressors $\varphi(t)$: n-dimensional vector of past observations

Assumption: model surface $[y, \varphi(t)]$ is smooth

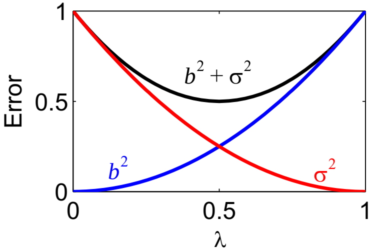
Model

$$\hat{g}(\varphi_*) = \sum_{i=1}^{N} y_i w(|\varphi_* - \varphi_i|)$$

Kernel w: weights the observations in neighborhood of φ_*

return

Basic idea regularization: pull parameters to zero $\tilde{\theta} = \lambda\hat{\theta}$



true parameter: $\theta_0 = 1$

unbiased estimate: $\hat{\theta}$

bias $b = 0$ and variance $\sigma^2 = 1$

scaled estimator: $\tilde{\theta} = \lambda\hat{\theta}$

bias $\tilde{\theta}$: $b = (1 - \lambda)$ and variance $\tilde{\theta}$: $\sigma^2_{\tilde{\theta}} = \lambda^2$

MSE: $b^2 + \sigma^2_{\tilde{\theta}} = (1 - \lambda)^2 + \lambda^2$

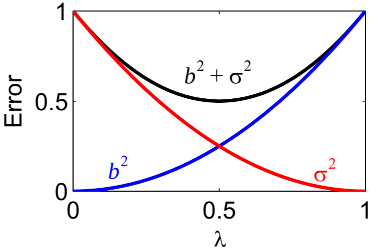
return

Extended cost function

$$\hat{\theta}_N = \text{argmin}_\theta \sum_{t=1}^{N} \|y(t) - \hat{y}(t|\theta)\|^2 + \theta^T P^{-1} \theta$$

return

Basic idea regularization: pull parameters to zero $\tilde{\theta} = \lambda\hat{\theta}$



true parameter: $\theta_0 = 1$

unbiased estimate: $\hat{\theta}$

  bias $b = 0$ and variance $\sigma^2 = 1$

scaled estimator: $\tilde{\theta} = \lambda\hat{\theta}$

  bias $\tilde{\theta}$: $b = (1 - \lambda)$ and variance $\tilde{\theta}$: $\sigma^2_{\tilde{\theta}} = \lambda^2$

  MSE: $b^2 + \sigma^2_{\tilde{\theta}} = (1 - \lambda)^2 + \lambda^2$

return

Extended cost function

$$\hat{\theta}_N = \text{argmin}_\theta \sum_{t=1}^{N} \|y(t) - \hat{y}(t|\theta)\|^2 + \theta^T P^{-1}\theta$$

return

Basic idea prediction methods
Model the errors using past output observations

Simulation model: $\hat{y}_s(t) = G(u, \theta)$

Simulation error: $v_s(t) = y(t) - \hat{y}_s(t)$

If simulation error $v_s(t)$ is correlated

$v_s(t)$ can be predicted from past values v_s^{t-1}

Correlation error model: $v_s(t) = H(q)e(t)$

$$\hat{v}_p(t|t-1) = (1 - H^{-1}(q))v_s(t)$$

Use $\hat{v}_p(t|t-1)$ to improve $\hat{y}_s(t)$

$$\hat{y}_p(t) = \hat{y}_s(t) + \hat{v}_p(t|t-1)$$

$$\hat{y}_p(t) = G(u, \theta) + (1 - H^{-1}(q))v_s(t)$$

return

Best choice: Prediction of Simulation?

- Simulation model
  - Goal: simulate the system output for new inputs
  - No output data available
  - Prediction model cannot be used
  - Use simulation model
- Prediction model
  - Goal: give the best estimate for the next output
  - Past output data available
  - Prediction error is smaller than simulation error
  - Use prediction model

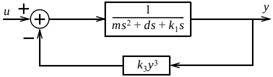
▸▸ return

Best choice: Prediction of Simulation?

- $v(t)$ dominated by measurement or sensor noise
  - $v(t)$ not related to process of interest
  - $v(t)$ should be eliminated
  - Use simulation model
- $v(t)$ dominated by process noise
  - Process noise affects the process of interest
  - $v(t)$ should be part of the model
  - Use prediction model
- $v(t)$ dominated by structural model errors
  - Structural model errors related to process of interest
  - $v(t)$ should be part of the model
  - Use prediction model

return

Example: Forced Duffing Oscilator

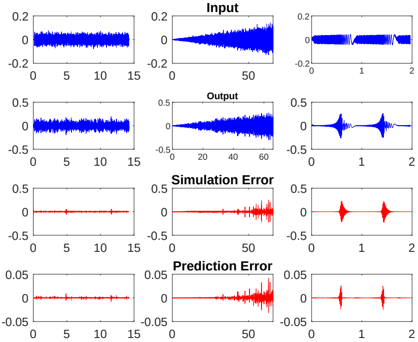


Simulation

$$\hat{y}_s(t) = b_0 u(t) + b_1 u(t-1) + b_2 u(t-2) - a_1 \hat{y}_s(t-1) - a_2 \hat{y}_s(t-2)$$

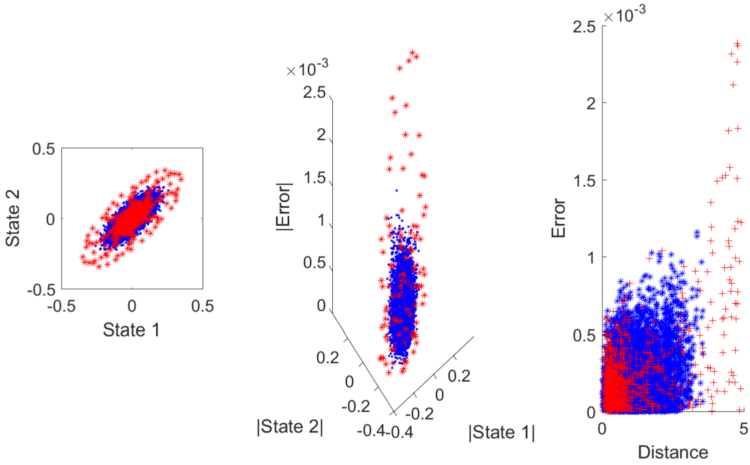
Prediction

$$\hat{y}_p(t) = b_0 u(t) + b_1 u(t-1) + b_2 u(t-2) - a_1 y(t-1) - a_2 y(t-2)$$



return

Error plot



return

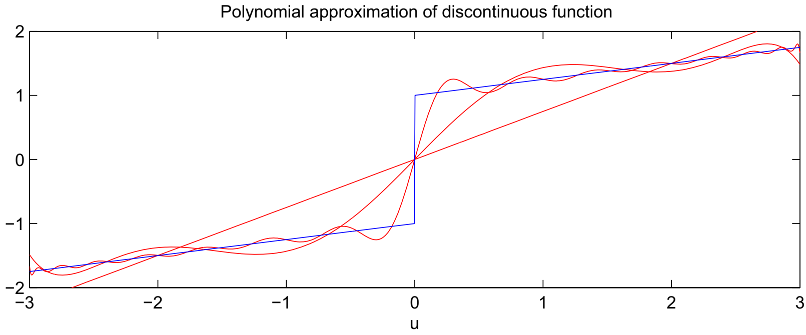
Identification in the Presence of Model Errors

User Choices

- convergence criterion
- approximation method
- excitation

return

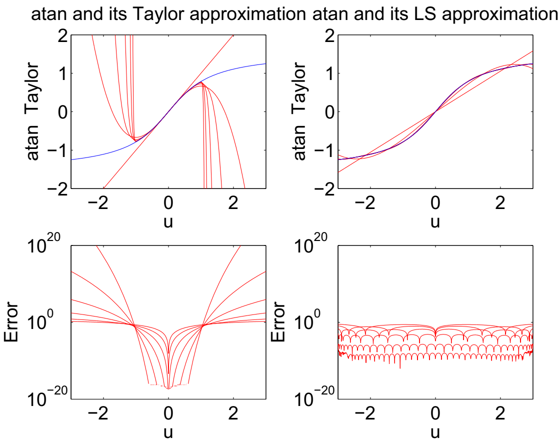
User choices: convergence criterion



uniform convergence >< point wise convergence

return

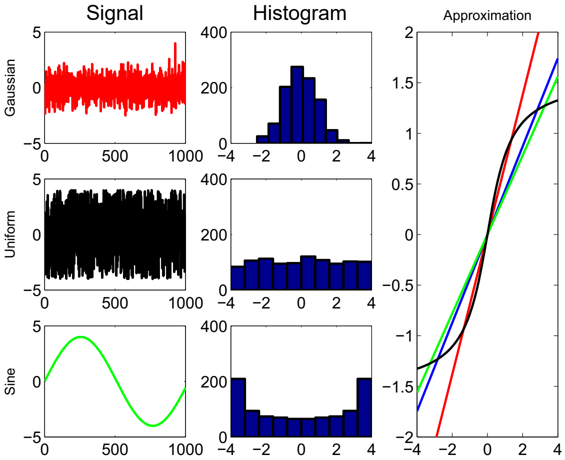
User choices: Approximation method



Taylor >< Least Squares

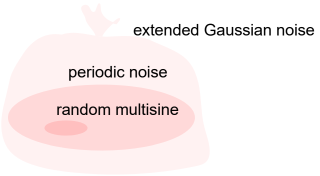
return

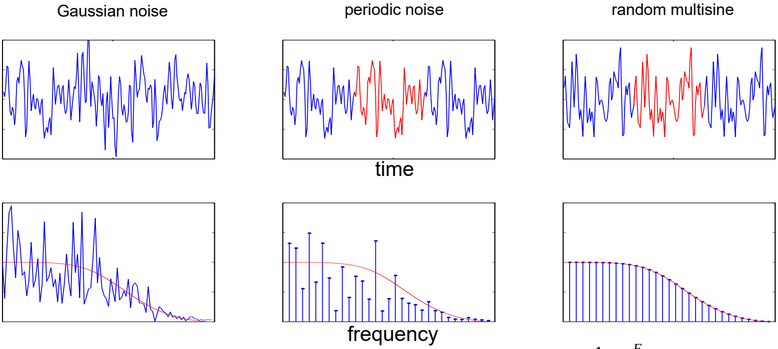
User choices: Excitation



▸▸ return

Class of excitation signals





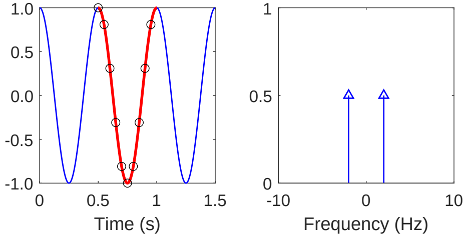
$$u(t) = \frac{1}{F}\sum_{k=1}^{F} A_k \cos(2\pi k f_0 t + \varphi_k)$$

return

Design of discrete time periodic excitation: a sine

$$u(t) = U_1 cos(2\pi f_0 t)$$
$$t = kT_s$$
$$k = 1, \ldots, N$$

f_s: sample frequency
$f_0 = f_s/N$: fundamental frequency
$T_s = 1/f_s$: sample period
$T = 1/f_0$: period signal
$T = NT_s$, N samples in one period



return

Design multisine

$$u(t) = \sum_{k=1}^{F} U_k cos\left(2\pi k f_0 t + \varphi_k\right)$$

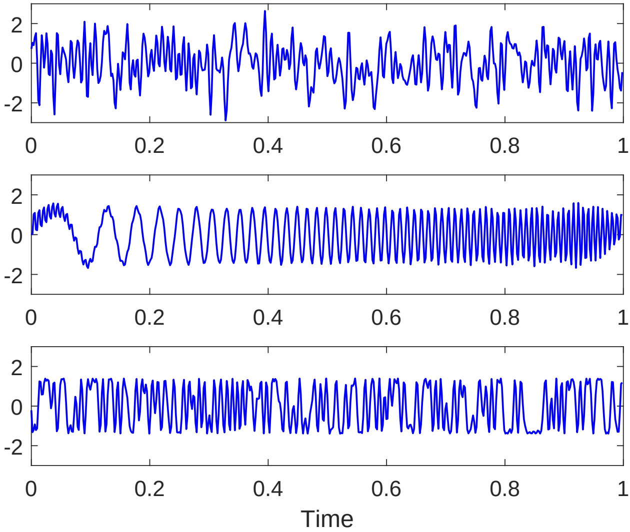
$T = 1/f_0$: period signal

$f_0 = f_s/N$: frequency resolution

N samples in one period

return

Multisine Examples



return

User guidelines to design a multisine

Spectral resolution $f_0 = f_s/N$

miss no sharp resonances

Period length $N = f_s/f_0$

higher frequency resolution $\rightarrow$ longer measurement time

Amplitude spectrum $U_k, k = 1, \ldots, F$

cover the frequency band of interest

Phases φ_k

use random phases in $[0, 2\pi[$

Signal amplitude

cover the input amplitude range of interest

Number of periods

measure 3 or more periods

return

Volterra theory in a nut shell
time domain

$$y(t) = \sum_{k=1}^{\infty} y^{[k]}(t)$$

with

$$y^{[1]}(t) = \int_{-\infty}^{\infty} u(\sigma_1) h_1(t - \sigma_1) d\sigma_1$$

$$y^{[2]}(t) = \iint_{-\infty}^{\infty} u(\sigma_1) u(\sigma_2) h_2(t - \sigma_1, t - \sigma_2) d\sigma_1 d\sigma_2$$

...

Volterra theory in a nut shell
multi dimensional frequency domain

$$y(t) = \sum_{k=1}^{\infty} y^{[k]}$$

Define

$$y^{[2]}(t_1, t_2) = \iint_{-\infty}^{\infty} u(\sigma_1)u(\sigma_2)h_2(t_1 - \sigma_1, t_2 - \sigma_2)d\sigma_1 d\sigma_2$$

Then

$$Y^{[2]}(\omega_1, \omega_2) = \iint_{-\infty}^{\infty} y^{[2]}(t_1, t_2)e^{-j\omega_1 t_1}e^{-j\omega_2 t_2}dt_1 dt_2$$

return

Volterra theory in a nut shell
collapsing the multi dimensional frequency domain

Inverse Fourier transform

$$y^{[2]}(t) = y^{[2]}(t_1, t_2) = \frac{1}{(2\pi)^2}\iint_{-\infty}^{\infty} Y^{[2]}(\omega_1, \omega_2)e^{j\omega_1 t_1}e^{j\omega_2 t_2}d\omega_1 d\omega_2 \text{ with } t = t_1 = t_2$$

or

$$y^{[2]}(t) = \frac{1}{(2\pi)^2}\iint_{-\infty}^{\infty} Y^{[2]}(\omega_1, \omega_2)e^{j(\omega_1+\omega_2)t}d\omega_1 d\omega_2$$

Put

$$\omega = \omega_1 + \omega_2 \text{ --> } \omega_2 = \omega - \omega_1 \text{, and } d\omega = d\omega_2$$

Then

$$y^{[2]}(t) = \frac{1}{(2\pi)^2}\iint_{-\infty}^{\infty} Y^{[2]}(\omega_1, \omega - \omega_1)d\omega_1 e^{j\omega t}d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} Y^{[2]}(\omega)e^{j\omega t}dt$$

with

$$Y^{[2]}(\omega) = \int_{-\infty}^{\infty} y^{[2]}(\omega, \omega - \omega_1)d\omega_1$$

return

Volterra theory in a nut shell
frequency domain relations

$$Y^{[n]}(\omega_1, \omega_2, \ldots, \omega_n) = H^{[n]}(\omega_1, \omega_2, \ldots, \omega_n) U(\omega_1) \ldots U(\omega_n)$$

with

$$H^{[n]}(\omega_1, \omega_2, \ldots, \omega_n) = \int \ldots \int_{-\infty}^{\infty} h_n(t_1, t_2, \ldots, t_n) e^{-j\omega_1 t_1} \ldots e^{-j\omega_n t_{nn}} dt_1 dt_2 \ldots dt_n$$

Corresponding one-dimensional frequency representation

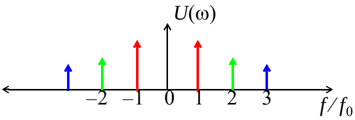
$$Y^{[n]}(\omega_1, \omega_2, \ldots, \omega_n) \rightarrow Y(\omega_1 + \omega_2 + \ldots + \omega_n)$$

$\omega_1 + \omega_2 + \ldots + \omega_n$ indicates that contribution results from n^{th} degree NL

return

Volterra theory in a nut shell
frequency domain relations for periodic signals

$$Y^{[n]}(\omega_1, \omega_2, \ldots, \omega_n) = H(\omega_1, \omega_2, \ldots, \omega_n)U(\omega_1)\ldots U(\omega_n)$$



with

$$Y^{[2]}(\omega) = \int_{-\infty}^{\infty} y^{[2]}(\omega, \omega - \omega_1)d\omega_1 \rightarrow Y^{[2]}(k) = \sum_{l} y^{[2]}(l, k-l) = \sum_{l} H^{[2]}(l, k-l)U(l)U(k-l)$$

similar

$$Y^{[3]}(k) = \sum_{l_1}\sum_{l_2}\ldots U(l_1)U(l_2)U(k-l_1-l_2)$$

Conclusion

$Y^{[3]}(k)$ sum over all combination $U(l_1)U(l_2)U(l_3)$ such that $l_1 + l_2 + l_3 = k$

return

Linear identification in the presence of nonlinear distortions
BLA: best linear approximation

User choices

- convergence criterion

- approximation method

- excitation

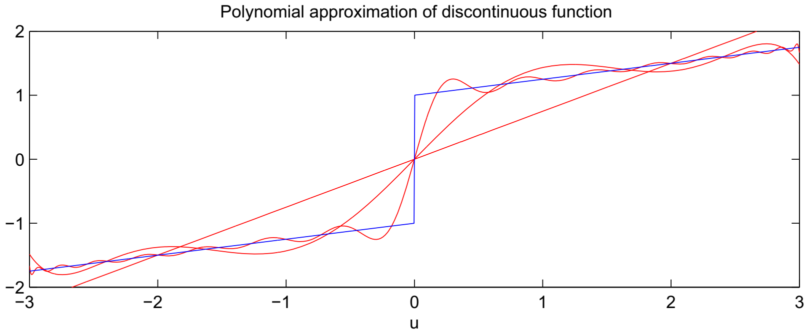
Class of nonlinear systems

Linear identification in the presence of nonlinear distortions

- Understanding the impact of nonlinear distortions
- Nonparametric identification: FRF measurements
- Parametric identification

return

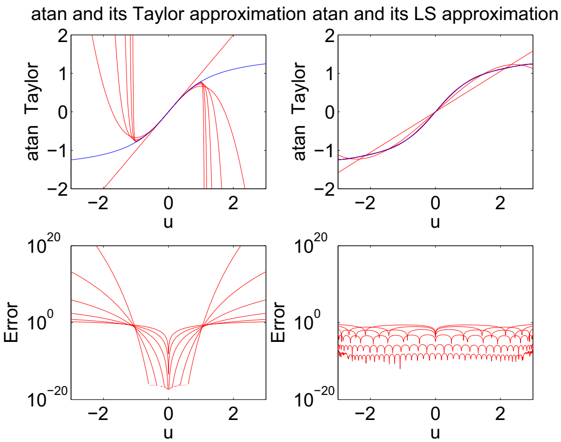
User choices: convergence criterion



uniform convergence >< point wise convergence

return

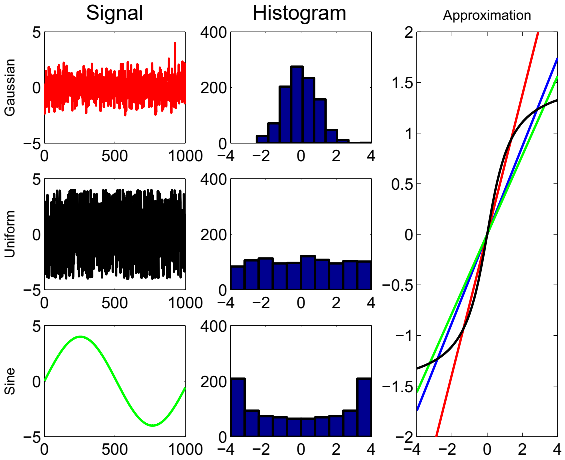
User choices: Approximation method



Taylor >< Least Squares

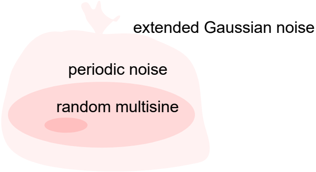
return

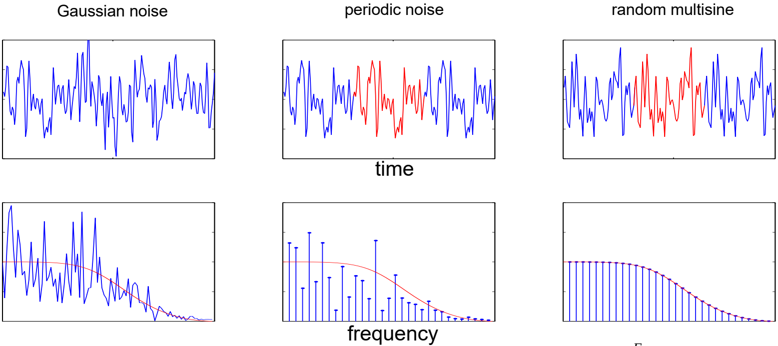
User choices: Excitation



return

Class of excitation signals





$$u(t) = \frac{1}{F}\sum_{k=1}^{F} A_k \cos(2\pi k f_0 t + \varphi_k)$$

return

Linear identification in the presence of nonlinear distortions
BLA: best linear approximation

User choices

- convergence criterion

- approximation method

- excitation

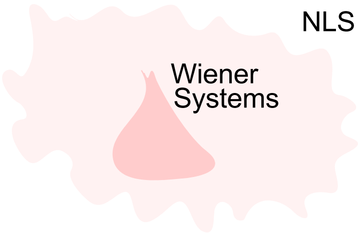
Class of nonlinear systems

Linear identification in the presence of nonlinear distortions

- Understanding the impact of nonlinear distortions
- Nonparametric identification: FRF measurements
- Parametric identification

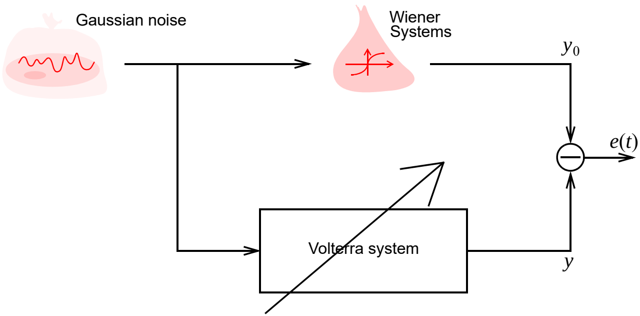
return

Class of nonlinear systems



return

Wiener systems?

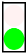
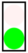
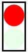


$$\mathcal{E}_u\{e^2_{RMS}\} \to 0$$

return

Major properties

- A periodic input $\rightarrow$ a periodic output with the same period

- Approximates the output in mean squares sense
  -  dynamic saturations
  -  discontinuities
  -  chaos

return

Linear identification in the presence of nonlinear distortions
BLA: best linear approximation

User choices

- convergence criterion
- approximation method
- excitation

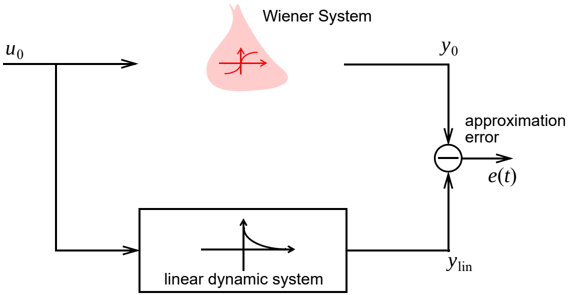
Class of nonlinear systems

Linear identification in the presence of nonlinear distortions

- Understanding the impact of nonlinear distortions
- Nonparametric identification: FRF measurements
- Parametric identification

return

Understanding the impact of nonlinear distortions on the linear framework

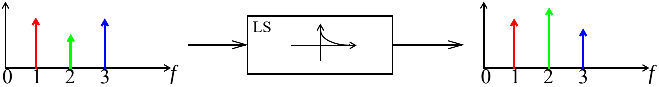


$$G_{BLA} = \arg\min_{G} E_U\{|Y - GU|^2\}$$

return

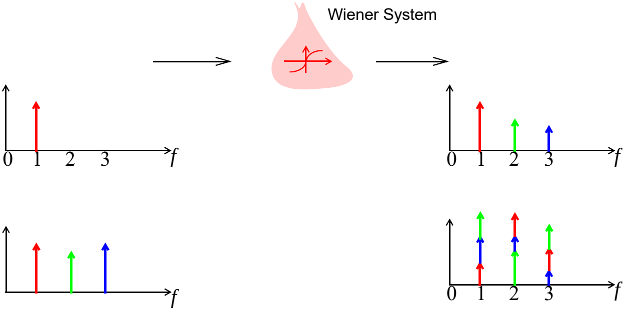
Behaviour of a nonlinear system

A linear system



return

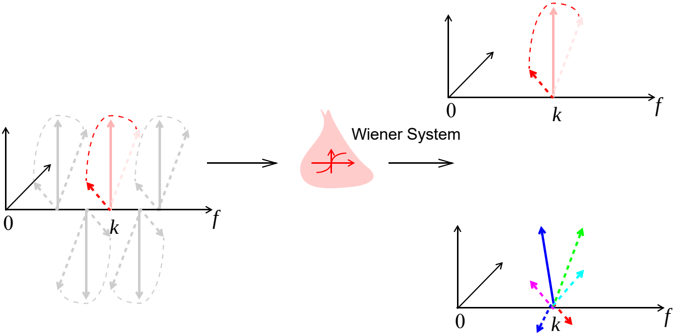
A Nonlinear system



return

Behaviour of a nonlinear system

coherent $\angle Y(k) = \angle U(k) + \varphi_0(k)$



non coherent

return

(non) Coherent contributions

Example: cubic contributions

$$Y^{[3]}(k) = \sum_{l_1}\sum_{l_2} H^{[3]}(l_1, l_2, k - l_1 - l_2) U(l_1) U(l_2) U(k - l_1 - l_2)$$

Frequency combinations s.t. $\angle U(l_1) U(l_2) U(k - l_1 - l_2) = \angle U(k)$?

Yes

$U(k)U(-l)U(l) = U(k)|U(l)|^2$ --> coherent contribution

NO

$U(k-2)U(1)U(1)$ --> non coherent contribution

return

(non) Coherent contributions (Cont'd)

Example: quadratic contributions

$$Y^{[2]}(k) = \sum_{l_1} H^{[2]}(l_1, k - l_1) U(l_1) U(k - l_1)$$

Frequency combinations s.t. $\angle U(l_1) U(k - l_1) = \angle U(k)$?

Yes
$U(k)U(0)$ --> coherent contribution requires DC

No
$U(k-1)U(1)$ --> non coherent contribution

return

(non) Coherent contributions

Conclusions

Put $U(0) = 0$

Even nonlinearities

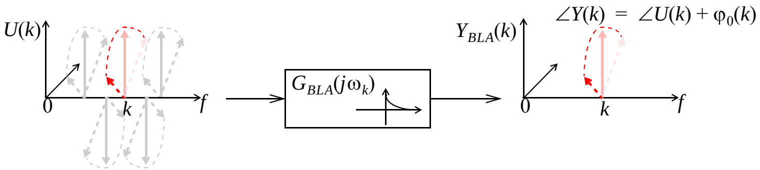
always non coherent

Odd nonlineairities

coherent
+
non coherent contributions

return

Coherent output



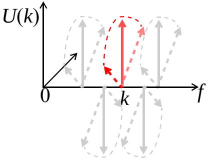
$$Y_{RBLA}(k) = G_{BLA}(j\omega_k)U(k)$$

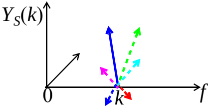
$G_{BLA}(j\omega_k)$ is the best linear approximation

$G_{BLA}(j\omega_k)$ is a function of S_{UU}

return

Non coherent output



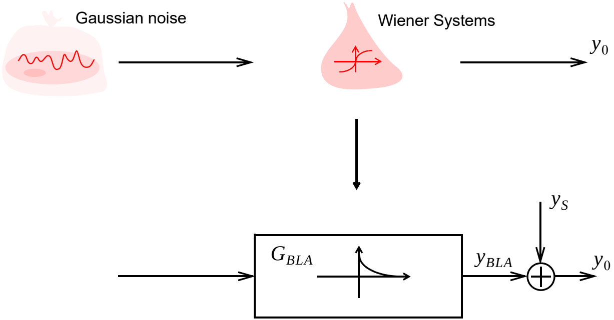


The phase of $Y_S(k)$ depends on $U(l)$ $l \neq k$

$Y_S(k)$ acts as a noise source

return

A new paradigm



$$Y(k) \; = \; G_{BLA}(j\omega_k)U(k) + Y_S(k)$$

return

A ‘new’ paradigm

Properties

$$Y(k) = G_{BLA}(j\omega_k)U(k) + Y_S(k)$$

$G_{BLA}(j\omega_k)$ is the ‘best linear approximation’

- smooth
- $O(N^0)$
- same for all excitations in the set (with same power spectrum)
- only odd nonlinearities contribute

$Y_S(k)$ is the ‘nonlinear noise source’

- smooth power spectrum
- zero mean
- circular complex normally distributed
- $O(N^0)$
- same power spectrum for all excitations in the set
- even and odd nonlinearities contribute

return

zero mean circular complex normally distributed

$$x = a + jb \in C$$

Zero mean circular complex: $E[x^2] \equiv 0$

$$E[(a + jb)^2] = E[a^2 - b^2 + 2jab] = 0 \quad \Rightarrow \quad E[a^2] = E[b^2] = \sigma^2 \text{ and } E[ab] = 0$$

Zero mean circular complex normally distributed: $E[x^n] \equiv 0$

x is a complex vector

- without a preferred direction
- no relation between amplitude and phase

return

Linear identification in the presence of nonlinear distortions
BLA: best linear approximation

User choices

- convergence criterion

- approximation method

- excitation

Class of nonlinear systems

Linear identification in the presence of nonlinear distortions

- Understanding the impact of nonlinear distortions
- Nonparametric identification: FRF measurements
- Parametric identification

return

Best Linear Approximation : Nonparametric measurement

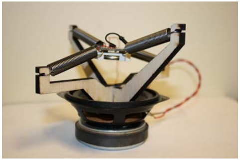
The classic linear equations still hold

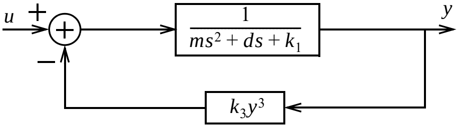
$$G_{BLA}(\omega) = \frac{S_{YU}(\omega)}{S_{UU}(\omega)}$$

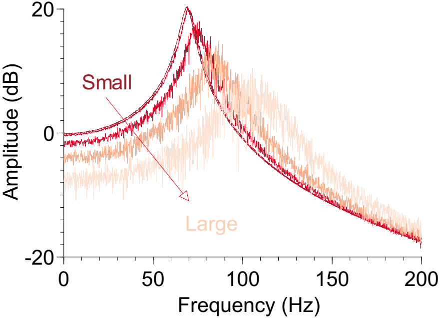
$$\sigma^2_{G_{BLA}}(\omega) = |G_{BLA}(\omega)|^2 \frac{1-\gamma^2(\omega)}{\gamma^2(\omega)}$$

return

Example : hardening spring

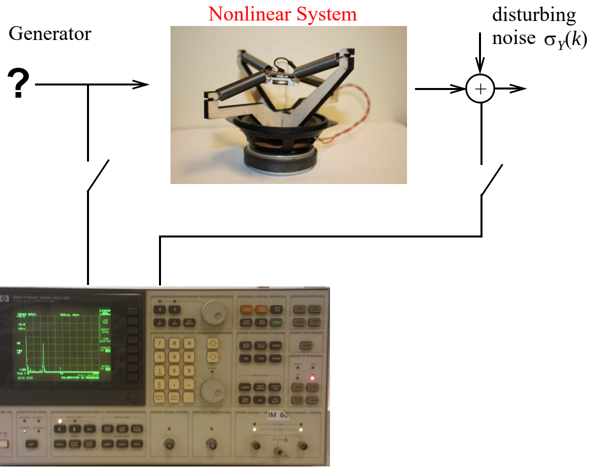






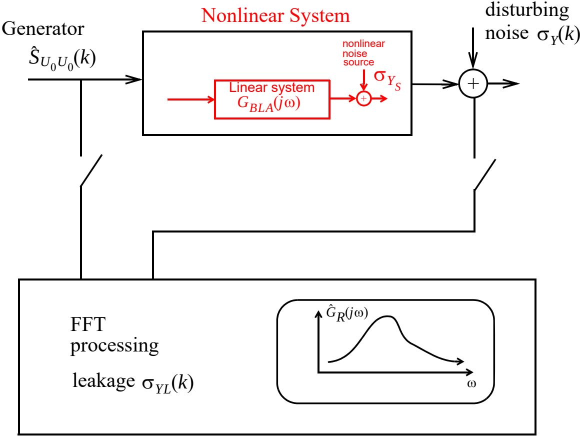
return

FRF-measurements in the presence of NL-distortions



return

FRF-measurements in the presence of NL-distortions



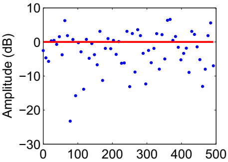
$$\sigma^2_{G_{BLA}}(k) = \frac{\sigma^2_{YL}(k) + \sigma^2_{Y_S}(k) + \sigma^2_Y(k)}{\hat{S}_{U_0U_0}(k)}$$

return

FRF-measurements in the presence of NL-distortions (Cont'd)

$$\sigma^2_{G_{BLA}}(k) = \frac{\sigma^2_{YL}(k) + \sigma^2_{Y_S}(k) + \sigma^2_{Y}(k)}{\hat{S}_{U_0U_0}(k)}$$

Avoid dips in $\hat{S}_{U_0U_0}(k)$



deterministic signals >< noise

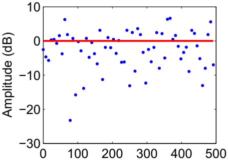
▸▸ return

FRF-measurements in the presence of NL-distortions (Cont'd)

$$\sigma^2_{\hat{G}_{BLA}}(k) = \frac{\sigma^2_{\hat{Y}L}(k) + \sigma^2_{\hat{Y}_S}(k) + \sigma^2_{\hat{Y}}(k)}{\hat{S}_{U_0U_0}(k)}$$

Avoid dips in $\hat{S}_{U_0U_0}(k)$

deterministic signals >< noise



Reduction of the leakage errors $\sigma^2_{\hat{Y}L}$

periodic signals

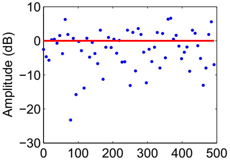
return

FRF-measurements in the presence of NL-distortions (Cont'd)

$$\sigma^2_{\hat{G}_{BLA}}(k) = \frac{\sigma^2_{\hat{Y}L}(k) + \sigma^2_{\hat{Y}_S}(k) + \sigma^2_{\hat{Y}}(k)}{\hat{S}_{U_0U_0}(k)}$$

Avoid dips in $\hat{S}_{U_0U_0}(k)$

deterministic signals >< noise

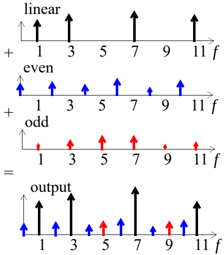


Reduction of the leakage errors $\sigma^2_{\hat{Y}L}$

periodic signals

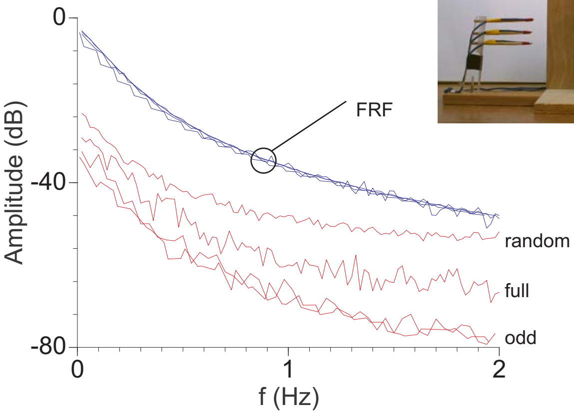
Reduction of the impact of nonlinear distortions $\sigma^2_{\hat{Y}_S}$

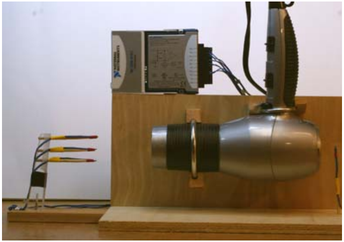
Odd excitations



return

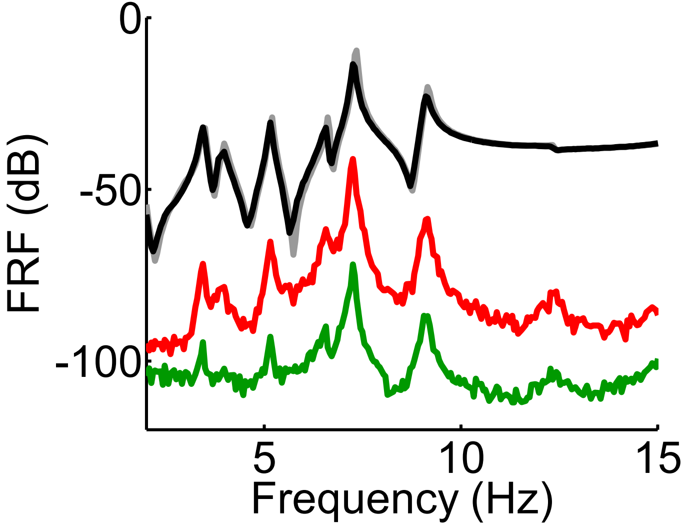
Hair dryer experiment





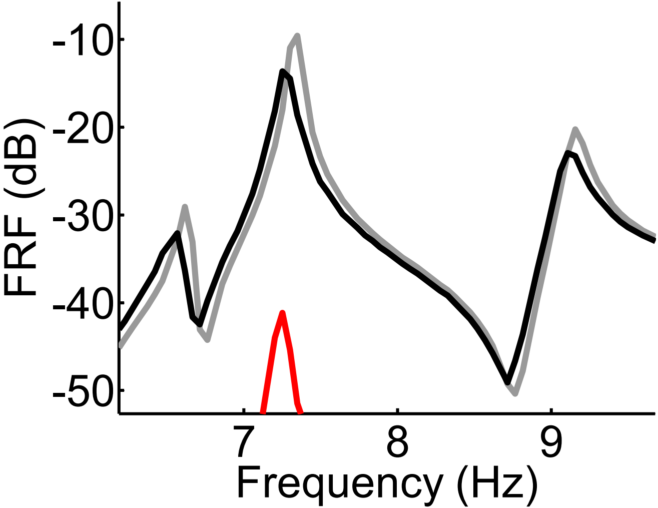
return

Example: F16-fighter measurements



return

Example: zoom F16-fighter measurements



return

Linear identification in the presence of nonlinear distortions
BLA: best linear approximation

User choices

- convergence criterion

- approximation method

- excitation

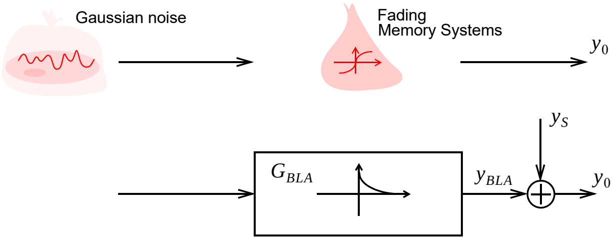
Class of nonlinear systems

Linear identification in the presence of nonlinear distortions

- Understanding the impact of nonlinear distortions
- Nonparametric identification: FRF measurements
- Parametric identification

return

Best Linear Approximation: Parametric modelling



$$Y(k) \; = \; G_{BLA}(j\omega_k)U(k) + Y_S(k)$$

Goal: find a parametric model $G_{BLA}(j\omega, \theta)$ and its uncertainty bound

return

Best Linear Approximation : Parametric modelling

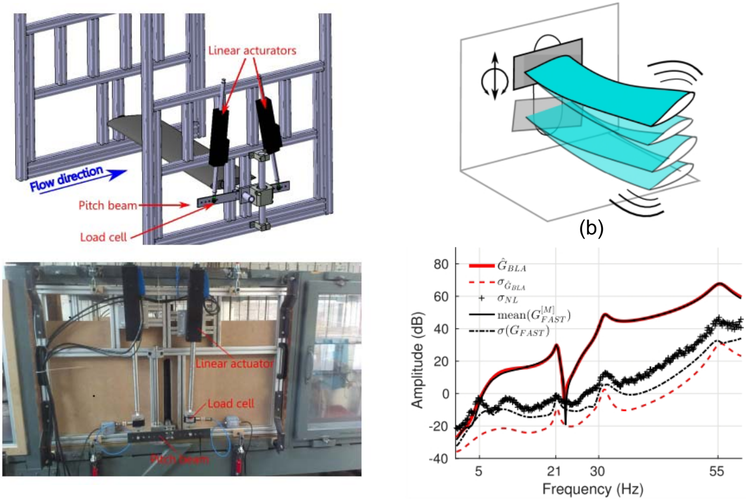
$$G_{BLA}(j\omega, \theta)$$

Linear identification framework

- Consistent estimate
  - True model retrieved for large data sets
- Uncertainty bounds are wrong
  - Nonlinear induced variance underestimated by factor 7 or more

return

Example



(a) (c)

return