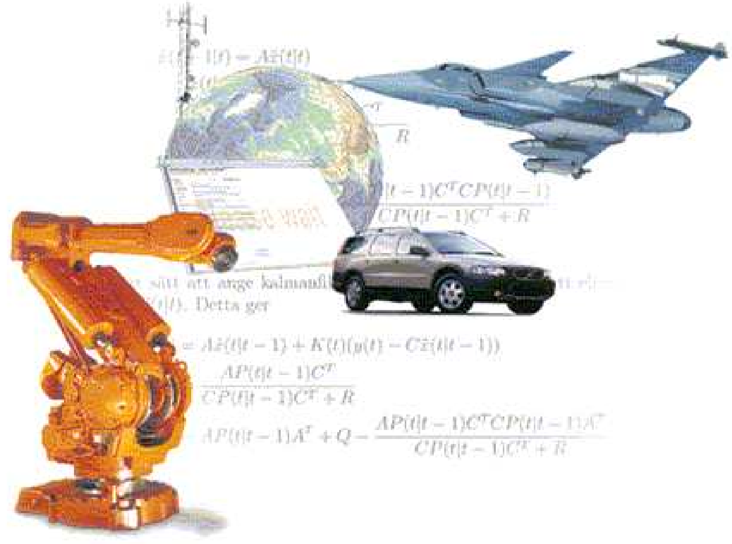
A (Simplistic) Perspective on Nonlinear System Identification



Lennart Ljung

Division of Automatic Control
Linköping University
Sweden

Outline: Two Themes

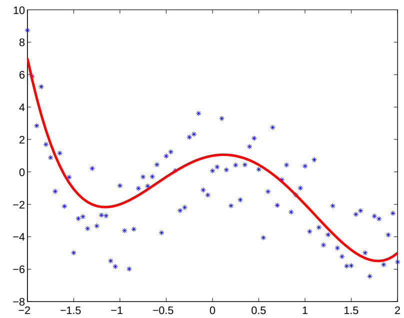
Abstract: Nonlinear System Identification is really curve fitting

Outline: Two Themes

Abstract: Nonlinear System Identification is really curve fitting

1. The basic questions and (statistical) tools illustrated for a simple curve fitting problem.
2. Nonlinear dynamical models: Parameterizations, problems and techniques.

Most basic ideas from system identification, choice of model structures and model sizes are brought out by considering the basic curve fitting problem from elementary statistics.



Unknown function $g_0(x)$. For a sequence of x-values (regressors) $\{x_1, x_2, \ldots, x_N\}$ (that may or may not be chosen by the user) observe the corresponding function values with some noise:

$$y(k) = g_0(x_k) + e(k)$$

Construct an estimate $\hat{g}_N(x)$ from $\{y(1), x_1, y(2), x_2, \ldots, y(N), x_N\}$

.

The Curve-fitting problem

$$y(k) = g_0(x_k) + e(k)$$

Construct an estimate $\hat{g}_N(x)$ from $\{y(1), x_1, y(2), x_2, \ldots, y(N), x_N\}$
The error $\hat{g}_N(x) - g_0(x)$ should be “as small as possible”
Approaches:

- Parametric: Construct $\hat{g}_N(x)$ by searching over a parameterized set of candidate functions.
- Non-parametric: Construct $\hat{g}_N(x)$ by smoothing over (carefully chosen subsets of) $y(k)$

Parametric Approach

Search for the function g_0 in a parameterized family of functions:

$$g(x,\theta) = \sum_{k=1}^{n} \alpha_k f_k(x,\tilde{\theta}_k), \quad \theta = \{\alpha_k, \tilde{\theta}_k,\ k = 1,\dots,n\}$$

Parametric Approach

Search for the function g_0 in a parameterized family of functions:

$$g(x,\theta) = \sum_{k=1}^{n} \alpha_k f_k(x, \tilde{\theta}_k), \quad \theta = \{\alpha_k, \tilde{\theta}_k,\ k = 1, \ldots, n\}$$

Examples:

Polynomial: $g(x,\theta) = \theta_1 + \theta_2 x + \ldots + \theta_n x^{n-1}$

Piecewise constant: $g(x,\theta) = \sum_{k=1}^{n} \alpha_k U(\beta_k(x - \gamma_k)),$

$U(x)$ is the unit pulse.

Parametric Approach

Search for the function g_0 in a parameterized family of functions:

$$g(x,\theta) = \sum_{k=1}^{n} \alpha_k f_k(x, \tilde{\theta}_k), \quad \theta = \{\alpha_k, \tilde{\theta}_k, \ k = 1, \dots, n\}$$

Examples:

Polynomial: $g(x,\theta) = \theta_1 + \theta_2 x + \ldots + \theta_n x^{n-1}$

Piecewise constant: $g(x,\theta) = \sum_{k=1}^{n} \alpha_k U(\beta_k(x - \gamma_k)),$

$U(x)$ is the unit pulse.

Parametric NL Black Box: Choice of g

The basic form is

$$g(x,\theta) = \sum_{k=1}^{N} \alpha_k \kappa(\beta_k (x - \gamma_k))$$

Parametric NL Black Box: Choice of g

The basic form is

$$g(x, \theta) = \sum_{k=1}^{N} \alpha_k \kappa(\beta_k (x - \gamma_k))$$

- Archetypical case:
  $\kappa(x) = U(x)$, (pulse or step) or $\kappa(x) = e^{-x^2/2}$, $\kappa(x) = \frac{1}{1+e^{-x}}$

Parametric NL Black Box: Choice of g

The basic form is

$$g(x,\theta) = \sum_{k=1}^{N} \alpha_k \kappa(\beta_k(x-\gamma_k))$$

- Archetypical case:
  $\kappa(x) = U(x),$ (pulse or step) or $\kappa(x) = e^{-x^2/2}, \quad \kappa(x) = \frac{1}{1+e^{-x}}$
- α coordinates, β scale or dilation, γ location

Parametric NL Black Box: Choice of g

The basic form is

$$g(x,\theta) = \sum_{k=1}^{N} \alpha_k \kappa(\beta_k(x-\gamma_k))$$

- Archetypical case:
  $\kappa(x) = U(x),$ (pulse or step) or $\kappa(x) = e^{-x^2/2}, \quad \kappa(x) = \frac{1}{1+e^{-x}}$
- α coordinates, β scale or dilation, γ location
- ANN: Radial basis, sigmoidal, etc
- LS Support Vector Machines
- Wavenets
- Neuro-Fuzzy

Parametric NL Black Box: Choice of g

The basic form is

$$g(x,\theta) = \sum_{k=1}^{N} \alpha_k \kappa(\beta_k(x-\gamma_k))$$

- Archetypical case:
  $\kappa(x) = U(x),$ (pulse or step) or $\kappa(x) = e^{-x^2/2}, \quad \kappa(x) = \frac{1}{1+e^{-x}}$
- α coordinates, β scale or dilation, γ location
- ANN: Radial basis, sigmoidal, etc
- LS Support Vector Machines
- Wavenets
- Neuro-Fuzzy

$$y(t) = g_0(x_t) + e(t)$$

Least Squares:

$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta)$$

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} |y(t) - g(x_t, \theta)|^2$$

.

$$y(t) = g_0(x_t) + e(t)$$

Weighted Least Squares:

$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta)$$

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} |y(t) - g(x_t, \theta)|^2 / \lambda_t$$

λ_t Proportional to 'reliability' of t:th measurement $\sim Ee^2(t)$

$$y(t) = g_0(x_t) + e(t)$$

Weighted Least Squares:

$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta)$$

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} L(x_t) |y(t) - g(x_t, \theta)|^2 / \lambda_t$$

λ_t Proportional to 'reliability' of t:th measurement $\sim Ee^2(t)$

A extra weighting $L(x_t)$ could also reflect the 'relevance' of the point x_t.

('Focus in fit')

$$y(t) = g_0(x_t) + e(t)$$

(Regularized) Least squares:

$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta) + \delta|\theta|^2$$

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} |y(t) - g(x_t, \theta)|^2$$

$\delta|\theta|^2$ penalizes excessive model flexibility. Could come in various forms.

Other Criteria and Regularization Terms

$$\hat{\theta}_N = \arg\min_{\theta} \frac{1}{N} \sum_{t=1}^{N} \ell(y(t) - g(x_t, \theta), t)$$

- Maximum likelihood $\ell(z) = -\log p(z)$
- “unknown-but-bounded”: $\min_\theta \max_t |y(t) - g(x_t, \theta)|$
- ‘Support vector machines”: $\min \sum |y(t) - g(x_t, \theta)|_\epsilon$ (ϵ-insensitive L_1 norm)

Regularization by

$$V_N(\theta) + \delta|\theta| \quad \text{or} \quad \min V_N(\theta), |\theta| < C$$

LARS, LASSO, nn-garotte ...

Choices in Parametric Methods

So, the choice of parameters within a parameterized model is not that difficult: Fit to the observed data, by one criterion or another.
The choice of model size and model parameterization is a more interesting issue.

Asymptotic Analysis: Basic Facts

Except for very simple parameterizations $g(x, \theta)$, the distribution of $\hat{\theta}_N$ cannot be calculated (mainly due to "arg min").

Asymptotic Analysis: Basic Facts

Except for very simple parameterizations $g(x, \theta)$, the distribution of $\hat{\theta}_N$ cannot be calculated (mainly due to "arg min"). However its asymptotic distribution as $N \to \infty$ can be established: (Straightforwad applications of the law of large numbers and the central limit theorem)

Except for very simple parameterizations $g(x, \theta)$, the distribution of $\hat{\theta}_N$ cannot be calculated (mainly due to “arg min”). However its asymptotic distribution as $N \to \infty$ can be established: (Straightforwad applications of the law of large numbers and the central limit theorem)

- $H(\theta) = \lim_{N\to\infty} H_N(\theta) = EL(x_t)|g_0(x_t) - g(x_t, \theta)|^2/\lambda_t$
- Main Result: $\lim_{N\to\infty} \hat{\theta}_N = \theta^* = \arg\min H(\theta)$

Asymptotic Analysis: Basic Facts

Except for very simple parameterizations $g(x, \theta)$, the distribution of $\hat{\theta}_N$ cannot be calculated (mainly due to "arg min"). However its asymptotic distribution as $N \to \infty$ can be established: (Straightforwad applications of the law of large numbers and the central limit theorem)

- $H(\theta) = \lim_{N\to\infty} H_N(\theta) = EL(x_t)|g_0(x_t) - g(x_t, \theta)|^2/\lambda_t$
- Main Result: $\lim_{N\to\infty} \hat{\theta}_N = \theta^* = \arg\min H(\theta)$
- The asymptotic distribution of $\sqrt{N}(\hat{\theta}_N - \theta^*)$ is normal with zero mean and covariance matrix $P = \lambda[E\psi(t)\psi^T(t)]^{-1}, \quad \psi(t) = \frac{d}{d\theta} g(x_t, \theta^*)$
- "Cov $\hat{\theta}_N \sim \frac{\lambda}{N}[E\psi(t)\psi^T(t)]^{-1}$" (Decreases with more regularization)

Consequences for Model Size and Parameterization

- Effective number of parameters (depending on parameter dimension and regularization) is a trade-off between bias and variance
- This trade-off is favored by grey-box models and by adaptive choices of basis functions for the parameterization

Non-Parametric Methods

A simple idea is to locally smooth the noisy observations of the function values:

$$\hat{g}_N(x) = \sum_{k=1}^{N} C(x, x_k) y(k)$$

$$\sum_{k=1}^{N} C(x, x_k) = 1 \; \forall x$$

Non-Parametric Methods

A simple idea is to locally smooth the noisy observations of the function values:

$$\hat{g}_N(x) = \sum_{k=1}^{N} C(x, x_k) y(k)$$

$$\sum_{k=1}^{N} C(x, x_k) = 1 \ \forall x$$

Often $C(x, x_k) = \tilde{c}(x - x_k)/\lambda_k$ and $\tilde{c}(r) = 0$ for $|r| > \beta, \ \beta =$ the “bandwidth”
These are known as “kernel methods” in statistics.

Non-Parametric Methods

A simple idea is to locally smooth the noisy observations of the function values:

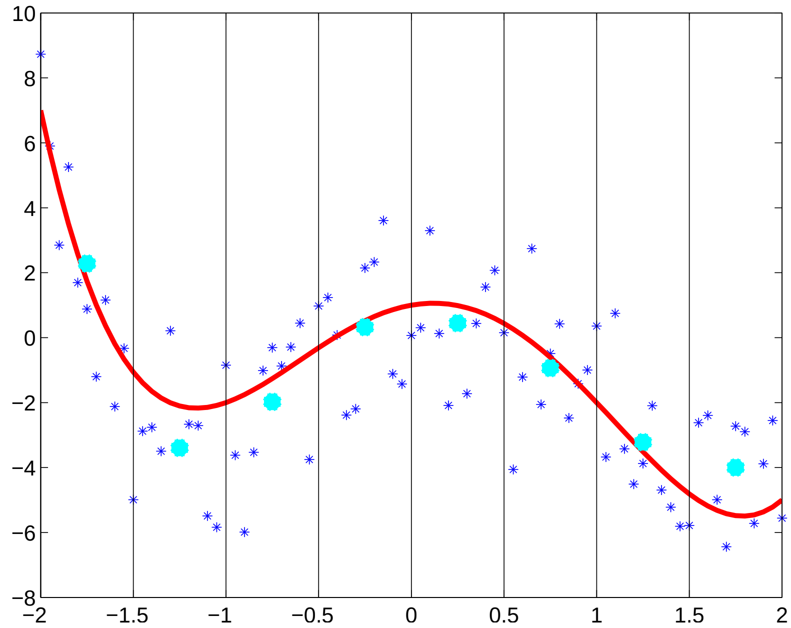
$$\hat{g}_N(x) = \sum_{k=1}^{N} C(x, x_k) y(k)$$

$$\sum_{k=1}^{N} C(x, x_k) = 1 \ \forall x$$

Often $C(x, x_k) = \tilde{c}(x - x_k)/\lambda_k$ and $\tilde{c}(r) = 0$ for $|r| > \beta, \ \beta =$ the "bandwidth"
These are known as "kernel methods" in statistics.

If $C(x, x_t)$ is chosen so that it is non-zero ($= 1/k$) only for k observed values x_t around x, this is the k-nearest neighbor method.

.



$C(x, x_k) = U((x - x_k)/\beta)$; $U(\cdot)$ the unit pulse. β = 0.25.
Cyan dots: Computed for $x = -1.75 : 0.5 : 1.75$

Bias-Variance Trade-off: ...

Some Variants of Non-Parametric Methods

- Local polynomial models
  - Adjust polynomials in local neighborhoods around x, Evaluate them in x.
- Direct weight optimization

$$\hat{g}_N(x) = \sum w_k(x) y(k), \quad \text{Choose } \{w_k\} \text{ for each } x$$

- Typically "Model-on-Demand" rather than "Off-the-Shelf"

Data: outputs and inputs

$$\{y(1), u(1), \ldots, y(N), u(N)\} = Z^N$$

- General aspects
- Black-box models
- Light-Grey-box models
- Dark-Grey-box models

General Aspects

A mathematical model for the system is a function from the past input-output data to the space where the output at time t, $y(t)$ lives, in general

$$\hat{y}(t|t-1) = g(Z^{t-1}, t)$$

The function can be thought of as a predictor of the next output.

A mathematical model for the system is a function from the past input-output data to the space where the output at time t, $y(t)$ lives, in general

$$\hat{y}(t|t-1) = g(Z^{t-1}, t)$$

The function can be thought of as a predictor of the next output.
Let us split it into one mapping from Z^{t-1} to a regression vector $\varphi(t)$ of fixed dimension d and a mapping g from R^d to R:

$$g(Z^{t-1}, t) = g(\varphi(t))$$

$$\varphi(t) = \varphi(Z^{t-1}, t)$$

Finding $\varphi(t)$ could itself be an estimation problem

General Aspects

A mathematical model for the system is a function from the past input-output data to the space where the output at time t, $y(t)$ lives, in general

$$\hat{y}(t|t-1) = g(Z^{t-1}, t)$$

The function can be thought of as a predictor of the next output.
Let us split it into one mapping from Z^{t-1} to a regression vector $\varphi(t)$ of fixed dimension d and a mapping g from R^d to R:

$$g(Z^{t-1}, t) = g(\varphi(t))$$

$$\varphi(t) = \varphi(Z^{t-1}, t)$$ Finding $\varphi(t)$ could itself be an estimation problem

Leaves two problems:

1. Choose the mapping $g(\varphi)$ – Same as in curvefitting
2. Choose the regression vector $\varphi(t)$ – “State”

Suppose $\varphi(t) = [y(t-1), u(t-1)]^T$
The (one-step ahead) predicted output at time for a given model θ is then

$$\hat{y}_p(t|\theta) = g([y(t-1), u(t-1)]^T, \theta)$$

It uses the previous measurement $y(t-1)$.

Simulation and Prediction

Suppose $\varphi(t) = [y(t-1), u(t-1)]^T$
The (one-step ahead) predicted output at time for a given model θ is then

$$\hat{y}_p(t|\theta) = g([y(t-1), u(t-1)]^T, \theta)$$

It uses the previous measurement $y(t-1)$.

A tougher test is to check how the model would behave in simulation, i.e. when only the input sequence u is used. The simulated output is obtained as above, by replacing the measured output by the simulated output from the previous step:

$$\hat{y}_s(t, \theta) = g([\hat{y}_s(t-1, \theta), u(t-1)]^T, \theta)$$

Suppose $\varphi(t) = [y(t-1), u(t-1)]^T$
The (one-step ahead) predicted output at time for a given model θ is then

$$\hat{y}_p(t|\theta) = g([y(t-1), u(t-1)]^T, \theta)$$

It uses the previous measurement $y(t-1)$.

A tougher test is to check how the model would behave in simulation, i.e. when only the input sequence u is used. The simulated output is obtained as above, by replacing the measured output by the simulated output from the previous step:

$$\hat{y}_s(t, \theta) = g([\hat{y}_s(t-1, \theta), u(t-1)]^T, \theta)$$

Notice a possible stability problem!

Color Coding: Shades of Grey

- Black
  - Parametric – Non-Parametric: see Curve Fitting
- Light-Grey
  - Physical modeling
- Dark-Grey
  - Semi-physical modeling
  - Block-oriented models
  - Local linear models and their cousins

AUTOMATIC CONTROL
COMMUNICATION SYSTEMS
LINKÖPINGS UNIVERSITET

Light Grey: Physical Modeling

Perform physical modeling (e.g. in MODELICA) and denote unknown physical parameters by θ. Collect the model equations as

$$\dot{x}(t) = f(x(t), u(t), \theta)$$
$$y(t) = h(x(t), u(t), \theta)$$

Light Grey: Physical Modeling

Perform physical modeling (e.g. in MODELICA) and denote unknown physical parameters by θ. Collect the model equations as

$$\dot{x}(t) = f(x(t), u(t), \theta)$$
$$y(t) = h(x(t), u(t), \theta)$$

(or in DAE, Differential Algebraic Equations, form.)

Light Grey: Physical Modeling

Perform physical modeling (e.g. in MODELICA) and denote unknown physical parameters by θ. Collect the model equations as

$$\dot{x}(t) = f(x(t), u(t), \theta)$$
$$y(t) = h(x(t), u(t), \theta)$$

(or in DAE, Differential Algebraic Equations, form.) For each parameter θ this defines a simulated (predicted) output $\hat{y}(t|\theta)$ which is the parameterized function

$$\hat{y}(t|\theta) = g(Z^{t-1}, \theta)$$

in somewhat implicit form.

Light Grey: Physical Modeling

Perform physical modeling (e.g. in MODELICA) and denote unknown physical parameters by θ. Collect the model equations as

$$\dot{x}(t) = f(x(t), u(t), \theta)$$
$$y(t) = h(x(t), u(t), \theta)$$

(or in DAE, Differential Algebraic Equations, form.) For each parameter θ this defines a simulated (predicted) output $\hat{y}(t|\theta)$ which is the parameterized function

$$\hat{y}(t|\theta) = g(Z^{t-1}, \theta)$$

in somewhat implicit form. To be a correct predictor this really assumes white measurement noise. Some more sophistical noise modeling is possible, usually involving *ad hoc* non-linear observers.

Light Grey: Physical Modeling

Perform physical modeling (e.g. in MODELICA) and denote unknown physical parameters by θ. Collect the model equations as

$$\dot{x}(t) = f(x(t), u(t), \theta)$$
$$y(t) = h(x(t), u(t), \theta)$$

(or in DAE, Differential Algebraic Equations, form.) For each parameter θ this defines a simulated (predicted) output $\hat{y}(t|\theta)$ which is the parameterized function

$$\hat{y}(t|\theta) = g(Z^{t-1}, \theta)$$

in somewhat implicit form. To be a correct predictor this really assumes white measurement noise. Some more sophistical noise modeling is possible, usually involving *ad hoc* non-linear observers.

The approach is conceptually simple, but could be very demanding in practice.

Dark-Grey: Semi-physical Models

Apply non-linear transformations to the measured data, so that the transformed data stand a better chance to describe the system in a linear relationship.

Dark-Grey: Semi-physical Models

Apply non-linear transformations to the measured data, so that the transformed data stand a better chance to describe the system in a linear relationship.

"Rules: Only high-school physics and max 10 minutes"

Dark-Grey: Semi-physical Models

Apply non-linear transformations to the measured data, so that the transformed data stand a better chance to describe the system in a linear relationship.

“Rules: Only high-school physics and max 10 minutes”

Simple examples: . . . .

Dark-Grey: Semi-physical Models

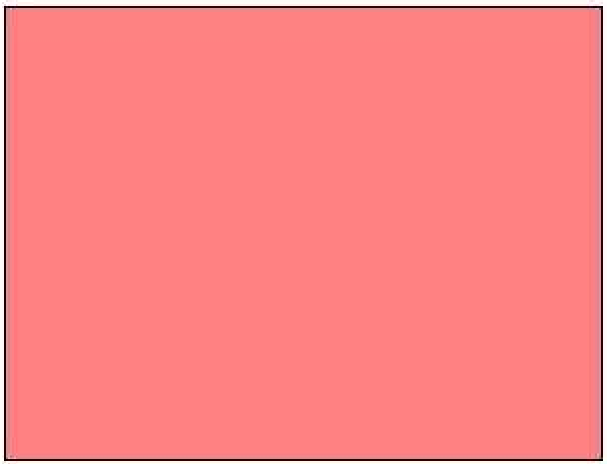
Apply non-linear transformations to the measured data, so that the transformed data stand a better chance to describe the system in a linear relationship.

"Rules: Only high-school physics and max 10 minutes"

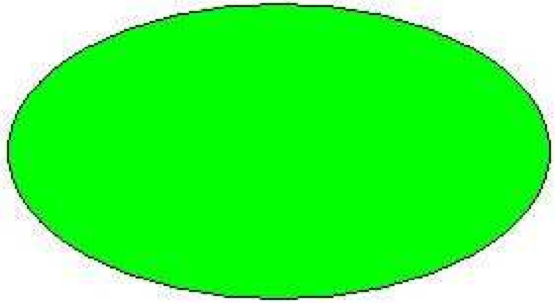
Simple examples: . . . .

Dark-Grey: Block-oriented Models

Building Blocks:

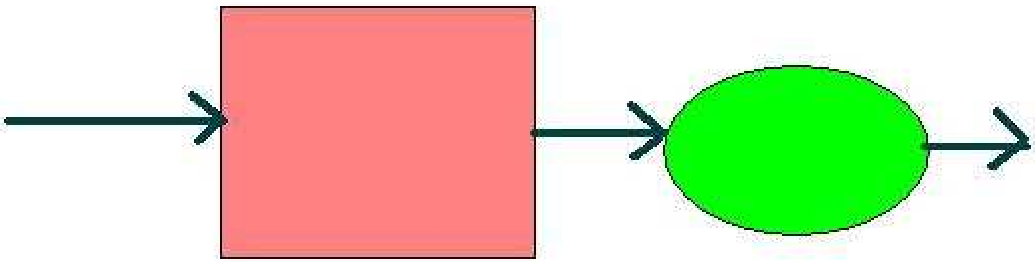


Linear Dynamic System $G(s)$

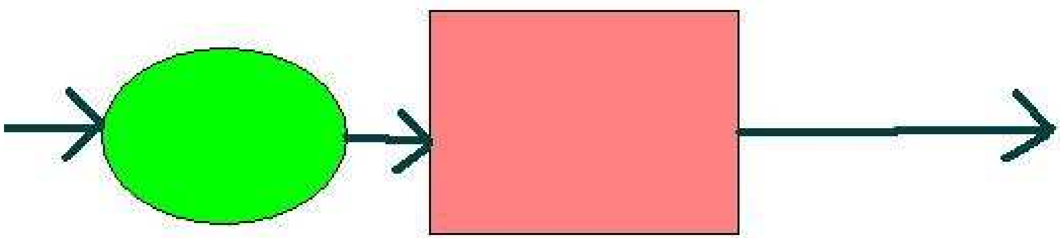


Nonlinear static function $f(u)$

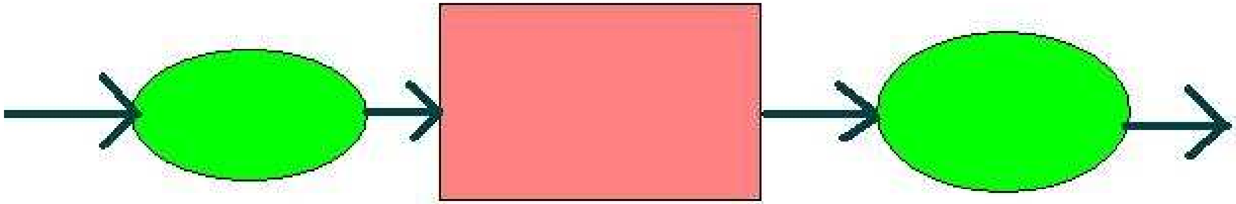
Common Models



Wiener

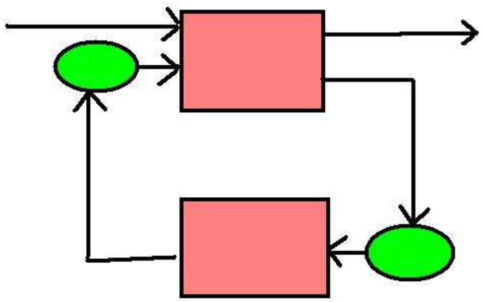


Hammerstein



Hammerstein-Wiener

Other Combinations



With the linear blocks parameterized as a linear dynamic system and the static blocks parameterized as a function (“curve”), this gives a parameterization of the output as

$$\hat{y}(t|\theta) = g(Z^{t-1}, \theta)$$

and the general approach of model fitting can be applied.

However, in this contexts many algorithmic variants have been suggested.

Dark Grey: Local Linear Models

Non-linear systems are often handled by linearization around a working point. The idea behind Local Linear Models is to deal with the nonlinearities by selecting or averaging over relevant linearized models.

Non-linear systems are often handled by linearization around a working point. The idea behind Local Linear Models is to deal with the nonlinearities by selecting or averaging over relevant linearized models.

Let the measured working point variable be denoted by $\rho(t)$ (sometimes called regime variable). If the regime variable is partitioned into d values ρ_k, the predicted output will be

$$\hat{y}(t) = \sum_{k=1}^{d} w_k(\rho(t), \rho_k, \eta)\hat{y}^{(k)}(t)$$

where η is a parameter that describes the partitioning

Dark Grey: Local Linear Models

Non-linear systems are often handled by linearization around a working point. The idea behind Local Linear Models is to deal with the nonlinearities by selecting or averaging over relevant linearized models.

Let the measured working point variable be denoted by $\rho(t)$ (sometimes called regime variable). If the regime variable is partitioned into d values ρ_k, the predicted output will be

$$\hat{y}(t) = \sum_{k=1}^{d} w_k(\rho(t), \rho_k, \eta)\hat{y}^{(k)}(t)$$

where η is a parameter that describes the partitioning   Choices of weights w_k : . . ..

Dark Grey: Local Linear Models

Non-linear systems are often handled by linearization around a working point. The idea behind Local Linear Models is to deal with the nonlinearities by selecting or averaging over relevant linearized models.
Let the measured working point variable be denoted by $\rho(t)$ (sometimes called regime variable). If the regime variable is partitioned into d values ρ_k, the predicted output will be

$$\hat{y}(t) = \sum_{k=1}^{d} w_k(\rho(t), \rho_k, \eta)\hat{y}^{(k)}(t)$$

where η is a parameter that describes the partitioning Choices of weights w_k : . . . ..
If the prediction $\hat{y}^{(k)}(t)$ corresponding to ρ_k is linear in the parameters, $\hat{y}^{(k)}(t) = \varphi^T(t)\theta^{(k)}$ the whole model will be a linear regression for a fixed η.

.

Hybrid Models and LPV Models

The model structure

$$\hat{y}(t,\theta,\eta) = \sum_{k=1}^{d} w_k(\rho(t),\eta)\varphi^T(t)\theta^{(k)}$$

is also an example of a hybrid model (piecewise linear). If the partition is to be estimated too, the problem is considerably more difficult.

Hybrid Models and LPV Models

The model structure

$$\hat{y}(t, \theta, \eta) = \sum_{k=1}^{d} w_k(\rho(t), \eta)\varphi^T(t)\theta^{(k)}$$

is also an example of a hybrid model (piecewise linear). If the partition is to be estimated too, the problem is considerably more difficult.

Linear Parameter Varying (LPV) models are also closely related:

$$\dot{x} = A(\rho(t))x + B(\rho(t))u$$
$$y = C(\rho(t))x + D(\rho(t))u$$

The model structure

$$\hat{y}(t,\theta,\eta) = \sum_{k=1}^{d} w_k(\rho(t),\eta)\varphi^T(t)\theta^{(k)}$$

is also an example of a hybrid model (piecewise linear). If the partition is to be estimated too, the problem is considerably more difficult.

Linear Parameter Varying (LPV) models are also closely related:

$$\dot{x} = A(\rho(t))x + B(\rho(t))u$$
$$y = C(\rho(t))x + D(\rho(t))u$$

Notice the link to non-parametric Local Polynomial Models in statistics!

Summary: Nonlinear Models

- A nonlinear model can be seen as nonlinear mapping from past data to the space where the output lives: $\hat{y}(t|t-1) = g(Z^{t-1}, t)$. Observations are then $y(t) = \hat{y}(t|t-1) + e(t)$.

Summary: Nonlinear Models

- A nonlinear model can be seen as nonlinear mapping from past data to the space where the output lives: $\hat{y}(t|t-1) = g(Z^{t-1}, t)$. Observations are then $y(t) = \hat{y}(t|t-1) + e(t)$.
- Useful split of mapping: $g(Z^{t-1}) = g(\varphi(Z^{t-1}, t))$

Summary: Nonlinear Models

- A nonlinear model can be seen as nonlinear mapping from past data to the space where the output lives: $\hat{y}(t|t-1) = g(Z^{t-1}, t)$. Observations are then $y(t) = \hat{y}(t|t-1) + e(t)$.
- Useful split of mapping: $g(Z^{t-1}) = g(\varphi(Z^{t-1}, t))$
- Non-parametric and Parametric methods – Essentially Curve-fitting

Summary: Nonlinear Models

- A nonlinear model can be seen as nonlinear mapping from past data to the space where the output lives: $\hat{y}(t|t-1) = g(Z^{t-1}, t)$. Observations are then $y(t) = \hat{y}(t|t-1) + e(t)$.
- Useful split of mapping: $g(Z^{t-1}) = g(\varphi(Z^{t-1}, t))$
- Non-parametric and Parametric methods – Essentially Curve-fitting
- Black-box and Grey-box parameterizations $g(\varphi, \theta)$

Summary: Nonlinear Models

- A nonlinear model can be seen as nonlinear mapping from past data to the space where the output lives: $\hat{y}(t|t-1) = g(Z^{t-1}, t)$. Observations are then $y(t) = \hat{y}(t|t-1) + e(t)$.
- Useful split of mapping: $g(Z^{t-1}) = g(\varphi(Z^{t-1}, t))$
- Non-parametric and Parametric methods – Essentially Curve-fitting
- Black-box and Grey-box parameterizations $g(\varphi, \theta)$
- Black-box parameterizations usually employ one basic basis-function, that is scaled and located at different points

Summary: Nonlinear Models

- A nonlinear model can be seen as nonlinear mapping from past data to the space where the output lives: $\hat{y}(t|t-1) = g(Z^{t-1}, t)$. Observations are then $y(t) = \hat{y}(t|t-1) + e(t)$.
- Useful split of mapping: $g(Z^{t-1}) = g(\varphi(Z^{t-1}, t))$
- Non-parametric and Parametric methods – Essentially Curve-fitting
- Black-box and Grey-box parameterizations $g(\varphi, \theta)$
- Black-box parameterizations usually employ one basic basis-function, that is scaled and located at different points
- Grey-boxes can be based on (serious) physical modeling and on more leisurely semi-physical modeling.

Summary: Nonlinear Models

- A nonlinear model can be seen as nonlinear mapping from past data to the space where the output lives: $\hat{y}(t|t-1) = g(Z^{t-1}, t)$. Observations are then $y(t) = \hat{y}(t|t-1) + e(t)$.
- Useful split of mapping: $g(Z^{t-1}) = g(\varphi(Z^{t-1}, t))$
- Non-parametric and Parametric methods – Essentially Curve-fitting
- Black-box and Grey-box parameterizations $g(\varphi, \theta)$
- Black-box parameterizations usually employ one basic basis-function, that is scaled and located at different points
- Grey-boxes can be based on (serious) physical modeling and on more leisurely semi-physical modeling.
- Non-convexity of the optimization remains one of the more serious problems for most parametric methods.

Conclusions for the SYSID Community

AUTOMATIC CONTROL
COMMUNICATION SYSTEMS
LINKÖPINGS UNIVERSITET

Conclusions for the SYSID Community

- For Tomorrow’s Panel Discussion ...