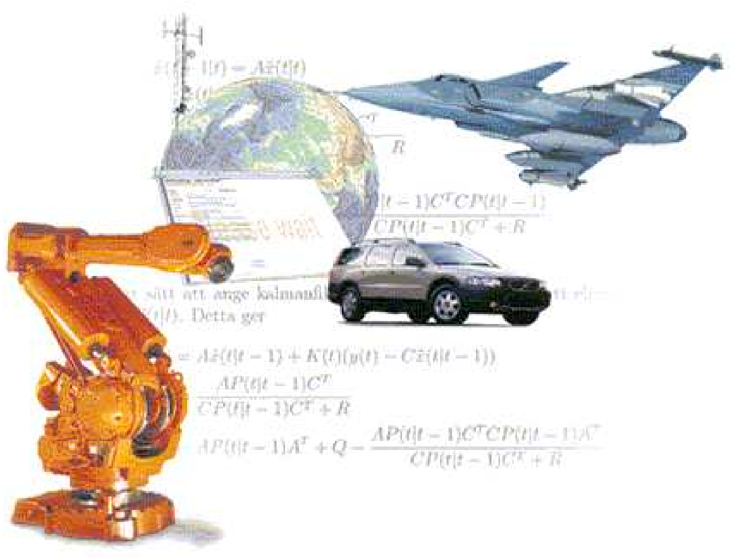
System Identification: From Data to Model

With Applications to Aircraft Modeling



Lennart Ljung

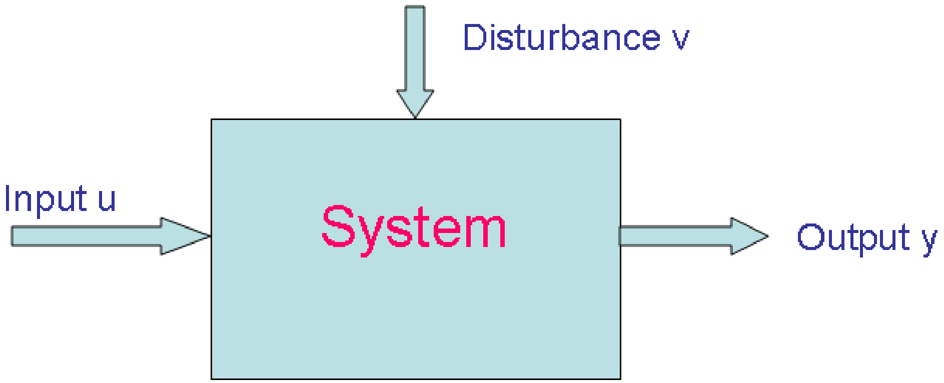
Division of Automatic Control
Linköping University
Sweden

Dynamic Systems

A Dynamic system has an output response y that depends on (all) previous values of an input signal u. It is also typically affected by a disturbance signal v. So the output at time t can be written as

$$y(t) = g(u^t, v^t)$$

where superscript denotes the signal's values from the remote past up to the indicated time.
The input signal u is known (measured), while the disturbance v is unmeasured.



The Problem

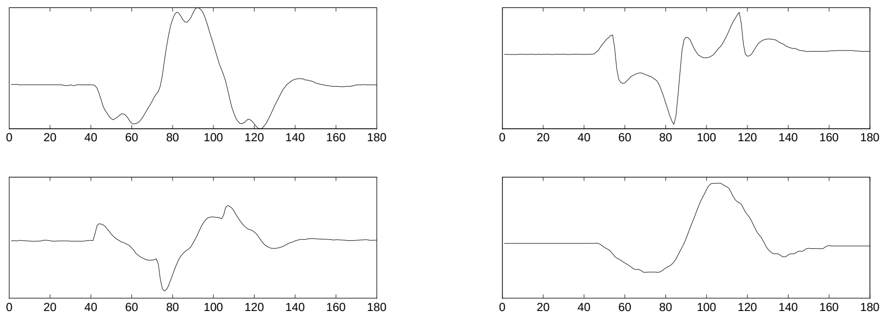
Data from Gripen

Pitch Rate

Elevator

Canard

Edge flap



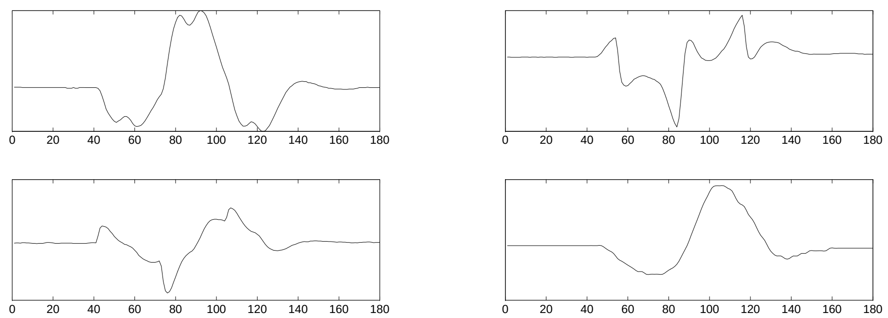
Data from Gripen

Pitch Rate

Canard

Elevator

Edge flap



- How do the control surface angles affect the pitch rate?
- Aerodynamic derivatives?
- How to use the information in flight data?

More Formalized Questions

Think discrete time data sequences:
$u^t, y^t = [u(1), u(2), ..., u(t), y(1), y(2), ..., y(t)]$
We need to get hold of a “simulation function”

$$y(t) = g(u^t)$$

and/or a prediction function

$$\hat{y}(t|t-1) = \tilde{f}(u^{t-1}, y^{t-1})$$

in order to

- be able to simulate and/or predict the input-output behavior of the system — “black-box”
- find parameters associated with a physical description (like the aerodynamic derivatives) of the system — “grey-box”

More concretely, assume that the relevant past can be condensed into a finite dimensional state, $x(t)$

$$x(t) = h(u^{t-1}, y^{t-1})$$

that is sufficient for the prediction

$$\hat{y}(t|t-1) = \tilde{f}(u^{t-1}, y^{t-1}) = f(x(t))$$

We thus need to find the function f (and h).

State or Regression Vector

More concretely, assume that the relevant past can be condensed into a finite dimensional state, $x(t)$

$$x(t) = h(u^{t-1}, y^{t-1})$$

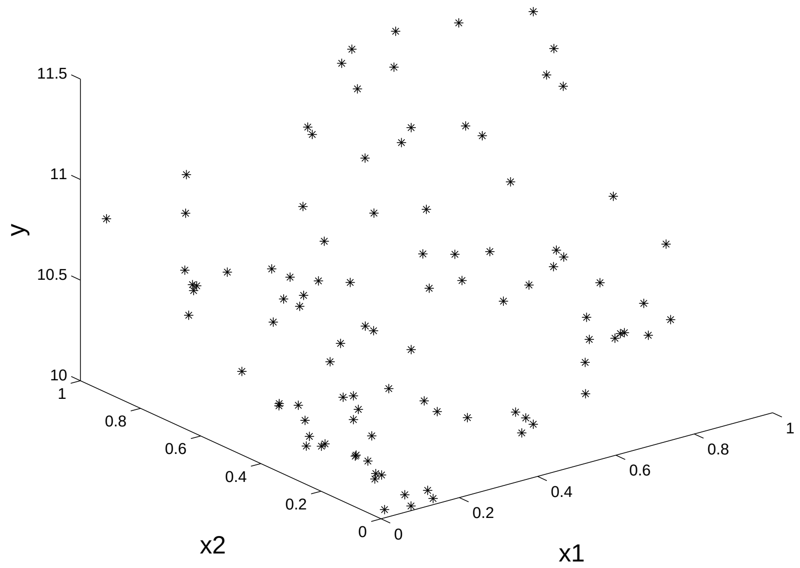
that is sufficient for the prediction

$$\hat{y}(t|t-1) = \tilde{f}(u^{t-1}, y^{t-1}) = f(x(t))$$

We thus need to find the function f (and h). For simplicity, in this talk think of

$$x(t) = \big[y(t-1), \ldots, y(t-n), u(t-1), \ldots u(t-m)\big]^T$$

So, the estimation problem is, given $Z^N = [y^N, u^N]$ find the mapping f.



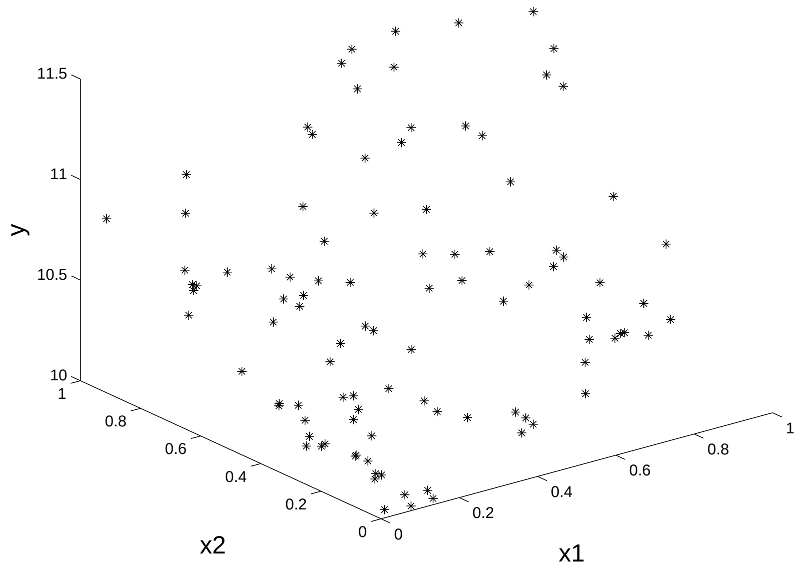
The Estimation Problem

So, the estimation problem is, given $Z^N = [y^N, u^N]$ find the mapping f.

Suppose that the mapping h is known.

Then the observations are equivalent to $[y(t), x(t)], t = 1, \dots N$ and we seek a mapping so that

$$y(t) \approx \hat{y}(t|t-1) = f(x(t))$$



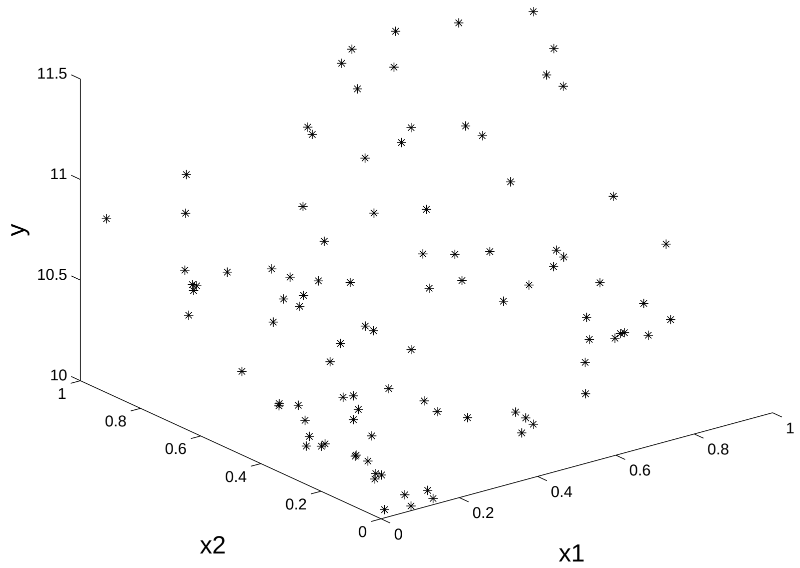
So, the estimation problem is, given $Z^N = [y^N, u^N]$ find the mapping f.

Suppose that the mapping h is known.

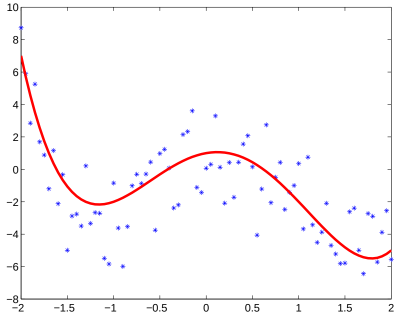
Then the observations are equivalent to $[y(t), x(t)], t = 1, \ldots N$ and we seek a mapping so that

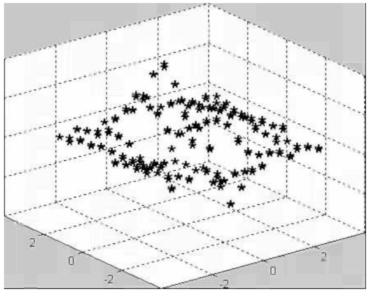
$$y(t) \approx \hat{y}(t|t-1) = f(x(t))$$

Statistically, this is a classical curve fitting problem

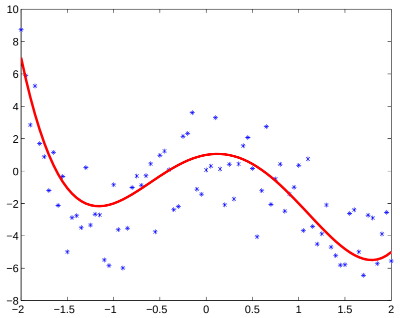


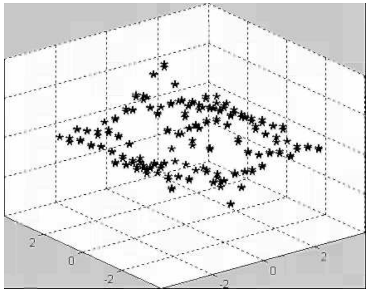
How are Models Adjusted to Data?





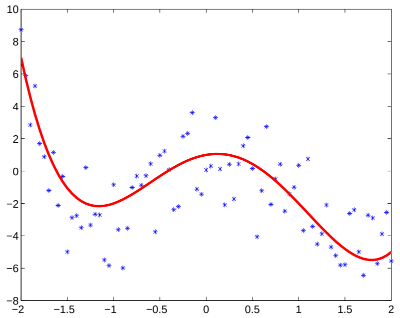
How are Models Adjusted to Data?

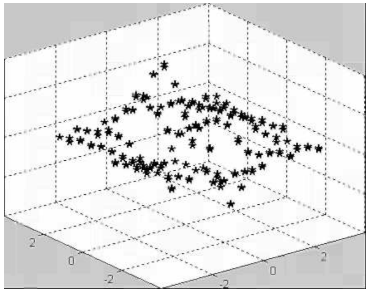




- Non-parametric: Smooth observed data over suitable neighborhoods

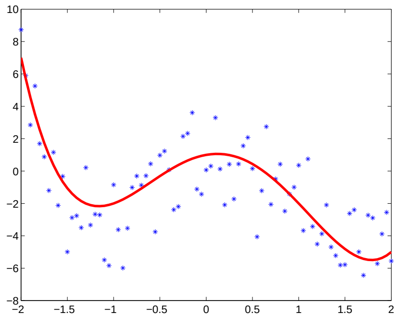
How are Models Adjusted to Data?

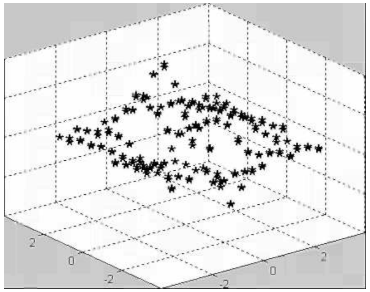




- Non-parametric: Smooth observed data over suitable neighborhoods
- Parametric: Postulate a parametrized surface and adjust its parameters to the data

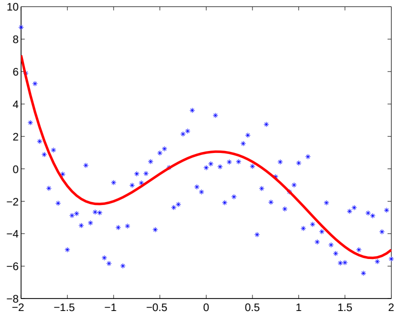
How are Models Adjusted to Data?

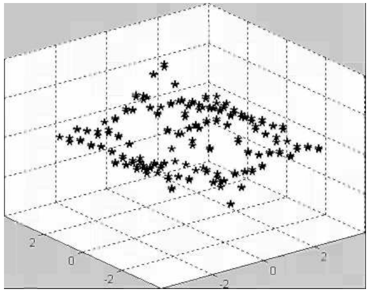




- Non-parametric: Smooth observed data over suitable neighborhoods
- Parametric: Postulate a parametrized surface and adjust its parameters to the data
- Semi-parametric: Adjust local, parameterized surface patches.

How are Models Adjusted to Data?

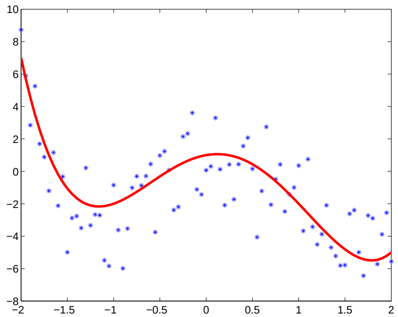


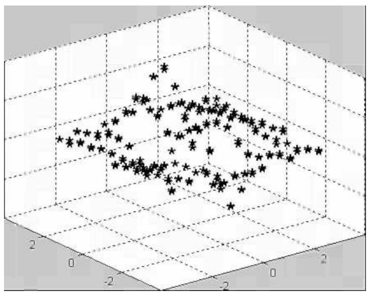


- Non-parametric: Smooth observed data over suitable neighborhoods
- Parametric: Postulate a parametrized surface and adjust its parameters to the data
- Semi-parametric: Adjust local, parameterized surface patches.

- Always a “knob” that controls bias/variance trade-off

How are Models Adjusted to Data?





- Non-parametric: Smooth observed data over suitable neighborhoods
- Parametric: Postulate a parametrized surface and adjust its parameters to the data
- Semi-parametric: Adjust local, parameterized surface patches.

- Always a “knob” that controls bias/variance trade-off
- Basic headache: Curse of dimensionality

1. The estimated function value $\hat{f}(x)$ is a random variable, inheriting its pdf from the disturbances v in a rather complex manner. One must therefore be content with finding its asymptotic distribution as $N \to \infty$. This can be calculated using the LLN (limit value) and CLT (Gaussian distribution around this limit). The standard calculations carry over to the case of dynamic systems with more or less effort.

Some Statistical Complements

1. The estimated function value $\hat{f}(x)$ is a random variable, inheriting its pdf from the disturbances v in a rather complex manner. One must therefore be content with finding its asymptotic distribution as $N \to \infty$. This can be calculated using the LLN (limit value) and CLT (Gaussian distribution around this limit). The standard calculations carry over to the case of dynamic systems with more or less effort.

2. The basic Bias/Variance trade-off knob for dynamic models is the order of the model, i.e. the number of states required to describe its prediction function. It can be complemented by regularization and all its modern variants (lasso, lars, nn-garrote ...)

$$\sum |y(t) - f(x(t), \theta)|^2 + \delta |\theta|^2$$

in the same way as for pure curve-fitting problems.

Some Issues Specific for System Identification

- Parameterization of the predictor mapping $\hat{y}(t|t-1) = f(u^{t-1}, y^{t-1})$

Some Issues Specific for System Identification

- Parameterization of the predictor mapping $\hat{y}(t|t-1) = f(u^{t-1}, y^{t-1})$
- Adopting and adapting the statistical theory to analyze the asymptotic properties in this context.

Some Issues Specific for System Identification

- Parameterization of the predictor mapping $\hat{y}(t|t-1) = f(u^{t-1}, y^{t-1})$
- Adopting and adapting the statistical theory to analyze the asymptotic properties in this context.
- Special techniques for minimizing the function of fit for these structures.

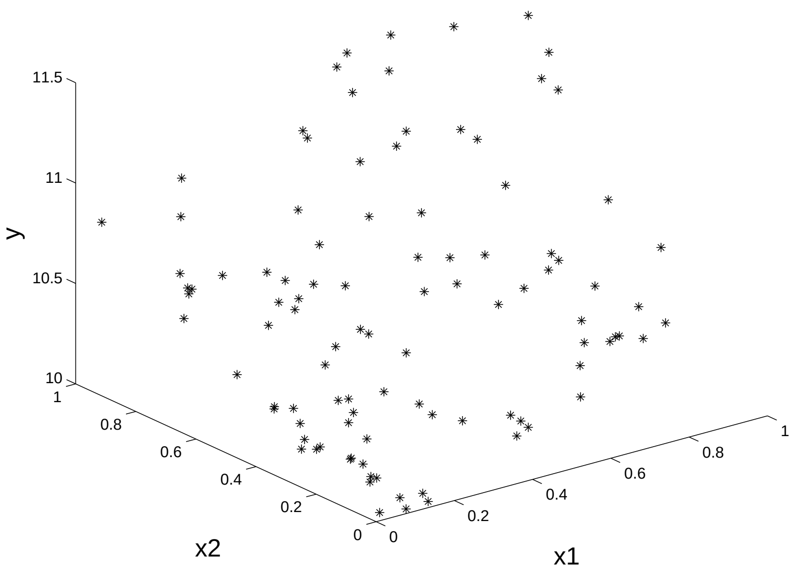
Some Issues Specific for System Identification

- Parameterization of the predictor mapping $\hat{y}(t|t-1) = f(u^{t-1}, y^{t-1})$
- Adopting and adapting the statistical theory to analyze the asymptotic properties in this context.
- Special techniques for minimizing the function of fit for these structures.
- Choice of the properties of the input signal.

Some Issues Specific for System Identification

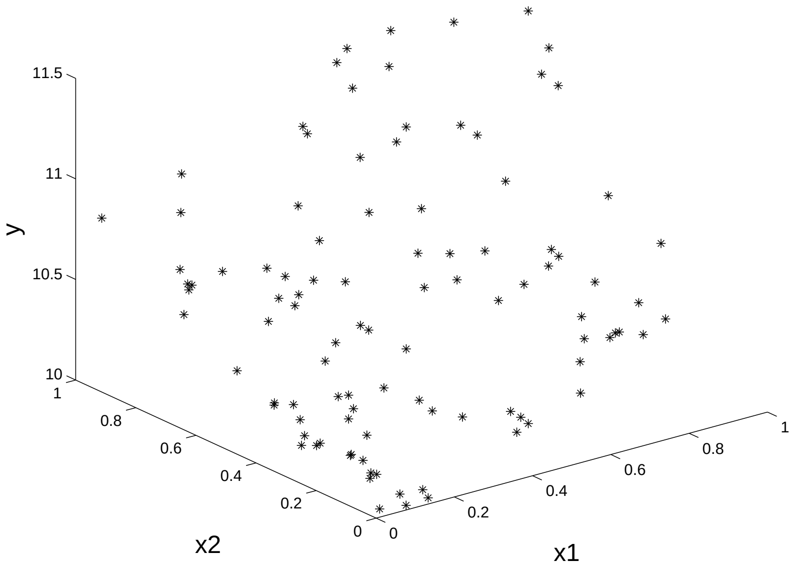
- Parameterization of the predictor mapping $\hat{y}(t|t-1) = f(u^{t-1}, y^{t-1})$
- Adopting and adapting the statistical theory to analyze the asymptotic properties in this context.
- Special techniques for minimizing the function of fit for these structures.
- Choice of the properties of the input signal.
- Selection of which signals to measure and when to measure them.

Linearity and nonlinearity concerns $\hat{y}(t|t-1) = f(y^{t-1}, u^{t-1})$ as a function of past data.



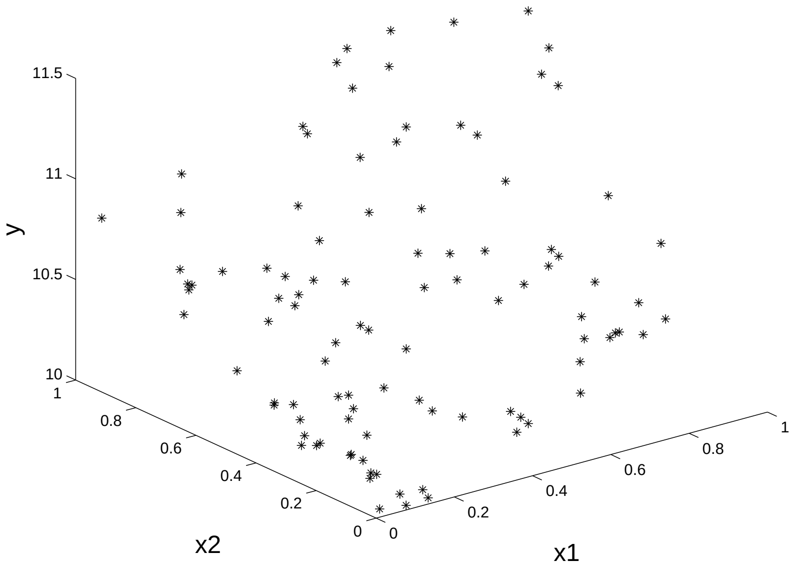
Linear and Non-Linear Models

Linearity and nonlinearity concerns $\hat{y}(t|t-1) = f(y^{t-1}, u^{t-1})$ as a function of past data. If this function is linear in the data we have a linear model, and are seeking a linear hyperplane to fit the observations in the data space.



Linear and Non-Linear Models

Linearity and nonlinearity concerns $\hat{y}(t|t-1) = f(y^{t-1}, u^{t-1})$ as a function of past data. If this function is linear in the data we have a linear model, and are seeking a linear hyperplane to fit the observations in the data space.
For a nonlinear model, an arbitrary, suitably fitting surface is sought.



Parameterization of Linear Dynamic Models

A linear dynamic model is written in transfer function form

$y(t) = G(q)u(t) + H(q)e(t)$ G and H functions of the delay operator q

e.g. $y(t) = g_1 u(t-1) + g_2 u(t-2) + e(t) + h_1 e(t-1)$ e white noise

A linear dynamic model is written in transfer function form

$y(t) = G(q)u(t) + H(q)e(t)$ G and H functions of the delay operator q

e.g. $y(t) = g_1 u(t-1) + g_2 u(t-2) + e(t) + h_1 e(t-1)$ e white noise

Typical parameterizations: rational functions in q (Black-Box)

$$G(q,\theta) = \frac{b_1 q^{-1} + \ldots + b_n q^{-n}}{1 + a_1 q^{-1} + \ldots + a_n q^{-n}}$$

A linear dynamic model is written in transfer function form

$y(t) = G(q)u(t) + H(q)e(t)$ G and H functions of the delay operator q

e.g. $y(t) = g_1 u(t-1) + g_2 u(t-2) + e(t) + h_1 e(t-1)$ e white noise

Typical parameterizations: rational functions in q (Black-Box)

$$G(q,\theta) = \frac{b_1 q^{-1} + \ldots + b_n q^{-n}}{1 + a_1 q^{-1} + \ldots + a_n q^{-n}}$$

State Space (Grey-Box, originating from a system of first order ODEs)

$$G(q,\theta) = C(\theta)(qI - A(\theta))^{-1}B(\theta)$$

A Quick Classifiation of Non-Linear Models: B/W

1. Basis-function expansion models (Black)

$$\hat{y}(t|t-1) = f(x(t), \theta)$$

$$f(x, \theta) = \sum \alpha_k \kappa(\beta_k(x - \gamma_k)) \quad \theta = \{\alpha_k, \beta_k, \gamma_k\}$$

- The whole neuro-fuzzy, SVM etc business

A Quick Classifiation of Non-Linear Models: B/W

1. Basis-function expansion models (Black)

$$\hat{y}(t|t-1) = f(x(t), \theta)$$

$$f(x, \theta) = \sum \alpha_k \kappa(\beta_k(x - \gamma_k)) \quad \theta = \{\alpha_k, \beta_k, \gamma_k\}$$

- The whole neuro-fuzzy, SVM etc business

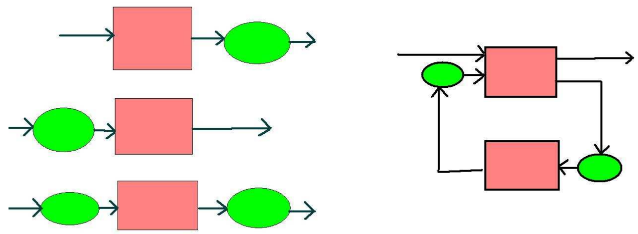
2. Physically parameterized DAE models (Light-Grey)

$$\sin(\theta_1 \phi(t) + \theta_2) + \theta_3 u_1(t)^2 = 0; \quad \theta_4 u_2(t) \dot{y}_1(t)^3 + 34 \ddot{y}_2(t) = 0$$

- Often obtained as modules in object oriented physical modeling (e.g. MODELICA)

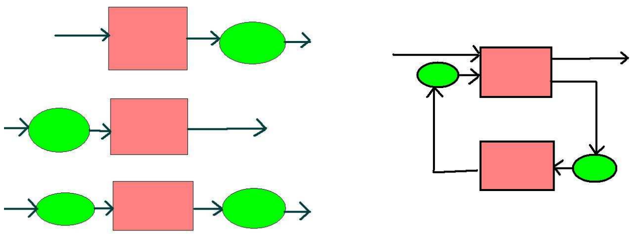
More nonlinear models (Darker Grey)

3. Block-oriented (nonlinear static blocks mixed with linear dynamic blocks)



More nonlinear models (Darker Grey)

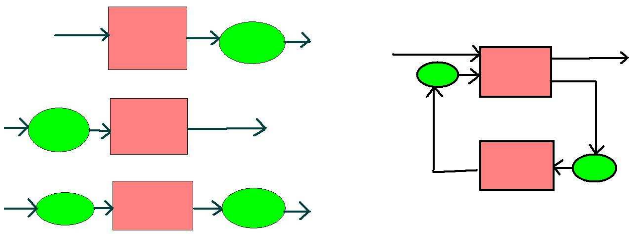
3. Block-oriented (nonlinear static blocks mixed with linear dynamic blocks)



4. Semi-physical models (non-linear transformations of measured data, based on simple insights)

More nonlinear models (Darker Grey)

3. Block-oriented (nonlinear static blocks mixed with linear dynamic blocks)



4. Semi-physical models (non-linear transformations of measured data, based on simple insights)
5. Composite Local models (local linear models)

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

Experiment Design Issues

Experiment Design is about making the experiment as informative as possible about important features in the models, given existing constraints. The main tool is the Fisher Information Matrix as a function of given options.

Experiment Design is about making the experiment as informative as possible about important features in the models, given existing constraints. The main tool is the Fisher Information Matrix as a function of given options.

A beacon for experiment (input) design for linear systems is the asymptotic expression for the variance of the estimated frequency function:

$$\text{Var}\hat{G}(\omega) \sim \frac{n}{N}\frac{\Phi_v(\omega)}{\Phi_u(\omega)}$$

n: Model order, N: number of observations, Φ_v: disturbance spectrum, Φ_u input spectrum

A Special Feature for Experiment Design: Feedback

In control applications the input signal is often chosen as output feedback: $u(t) = k(y^t, r^t)$. This creates complications:

A Special Feature for Experiment Design: Feedback

In control applications the input signal is often chosen as output feedback: $u(t) = k(y^t, r^t)$. This creates complications:

1. Looking for connections between u and y gives two possibilities: The system describes how y depends on past u. The feedback regulator describes how u depends on past y. One may have to be careful not to mix these up.

A Special Feature for Experiment Design: Feedback

In control applications the input signal is often chosen as output feedback: $u(t) = k(y^t, r^t)$. This creates complications:

1. Looking for connections between u and y gives two possibilities: The system describes how y depends on past u. The feedback regulator describes how u depends on past y. One may have to be careful not to mix these up.
2. One can loose “identifiability”. Consider the simple case

   $$y(t) + ay(t-1) = bu(t-1) + e(t); \quad u(t) = ky(t) \quad \text{System with P-feedback}$$

   Clearly we have $y(t) + (a - bk)y(t-1) = e(t)$ and all values of a and b such that $a - bk$ has certain value give identical input-output signals, and hence cannot be distinguished.

Concluding Remarks

- Describing the dynamic system as a predictor of future outputs, and finding models that minimize the prediction error gives a good common denominator for identification of both linear and nonlinear models.

Concluding Remarks

- Describing the dynamic system as a predictor of future outputs, and finding models that minimize the prediction error gives a good common denominator for identification of both linear and nonlinear models.
- The process can be seen as a basic curve-fitting problem allowing both parametric and nonparametric techniques.

Concluding Remarks

- Describing the dynamic system as a predictor of future outputs, and finding models that minimize the prediction error gives a good common denominator for identification of both linear and nonlinear models.
- The process can be seen as a basic curve-fitting problem allowing both parametric and nonparametric techniques.
- Classic statistical tools for curve-fitting and model quality analysis can be readily transformed to the system identification problem.

Concluding Remarks

- Describing the dynamic system as a predictor of future outputs, and finding models that minimize the prediction error gives a good common denominator for identification of both linear and nonlinear models.
- The process can be seen as a basic curve-fitting problem allowing both parametric and nonparametric techniques.
- Classic statistical tools for curve-fitting and model quality analysis can be readily transformed to the system identification problem.
- For linear models, the time/frequency domain duality and the Parseval relationship allow useful complementary view-points.

Concluding Remarks

- Describing the dynamic system as a predictor of future outputs, and finding models that minimize the prediction error gives a good common denominator for identification of both linear and nonlinear models.
- The process can be seen as a basic curve-fitting problem allowing both parametric and nonparametric techniques.
- Classic statistical tools for curve-fitting and model quality analysis can be readily transformed to the system identification problem.
- For linear models, the time/frequency domain duality and the Parseval relationship allow useful complementary view-points.
- The special features of dynamic system identification concern primarily area-specific model parameterizations and experiment design issues involving feedback configurations.

Concluding Remarks

- Describing the dynamic system as a predictor of future outputs, and finding models that minimize the prediction error gives a good common denominator for identification of both linear and nonlinear models.
- The process can be seen as a basic curve-fitting problem allowing both parametric and nonparametric techniques.
- Classic statistical tools for curve-fitting and model quality analysis can be readily transformed to the system identification problem.
- For linear models, the time/frequency domain duality and the Parseval relationship allow useful complementary view-points.
- The special features of dynamic system identification concern primarily area-specific model parameterizations and experiment design issues involving feedback configurations.
- The intended model use plays a prominent role in experiment design.