ESS101
Modelling and Simulation, 2025

Lecturer and examiner: Yasemin Bekiroğlu
Course assistant: Ahmet Tekden

Systems & Control Division
Department of Electrical Engineering
Chalmers University of Technology

September, 2025

Lecture 5 – System Identification

- Building models from data
- System identification
- Linear regression and least-squares

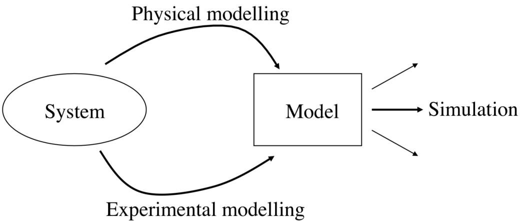
Learning objective:

- Use methods and tools to develop mathematical models of dynamical systems from measurement data.

How to build a model?

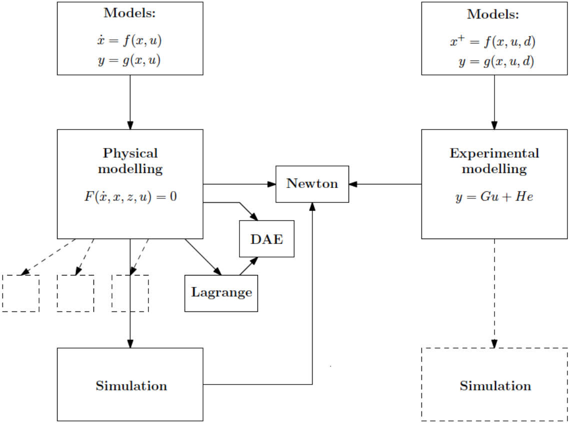
There are two main approaches to build a model:

1. **Physical (first principles) modelling:**
   Use the laws of physics (Newton, Kirchhoff, . . .).
2. **Experimental (data-driven) modelling, system identification:**
   Perform experiments on the system, analyze data to deduce a model.
   Connections with Machine learning.



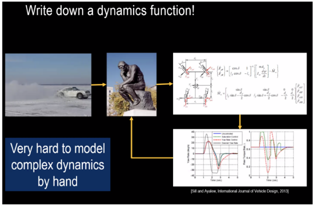
NB: In practice, often a combination of the two techniques is used.

How to build a model and simulate?



Physical vs Data-driven modelling

- Too complex for physical modelling

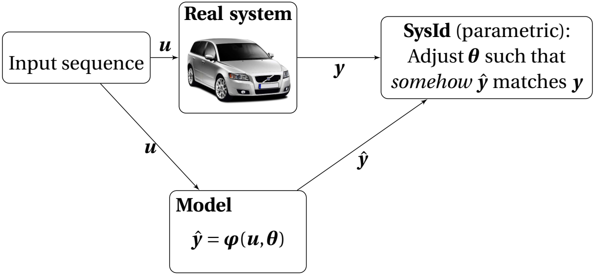


Data-driven modelling - System identification

- Linear regression and least-squares
- Prediction error methods
- Black-box models
- System identification workflow

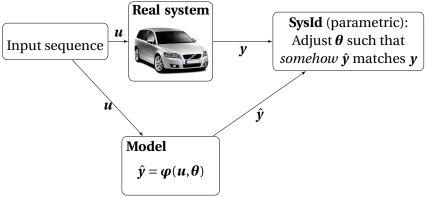
The system identification problem

SysId: Adjust a model/set of models (with adjustable parameters) to data.



The system identification problem

SysId: Adjust a model/set of models (with adjustable parameters) to data.

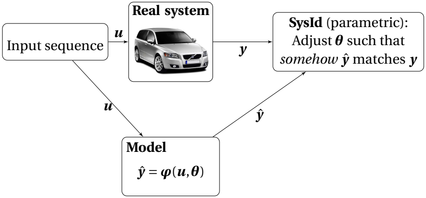


Some of the key issues:

- Experiment design: **selection of inputs and outputs** to be used and **construction of the input sequence** $\mathbf{u}$ to be applied to the system.

The system identification problem

SysId: Adjust the model (with adjustable parameters) to data.

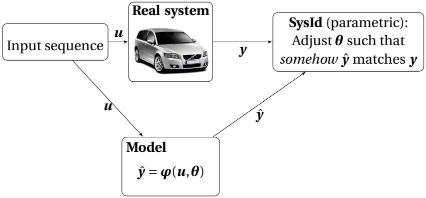


Some of the key issues:

- Experiment design: selection of inputs and outputs to be used and construction of the input sequence $\mathbf{u}$ to be applied to the system.
- **Selection of model structure**: the model $\hat{\mathbf{y}}(\mathbf{u}, \boldsymbol{\theta})$ can take various forms, e.g. both linear and nonlinear dynamics, different parametrizations etc.

The system identification problem

SysId: Adjust the model (with adjustable parameters) to data.

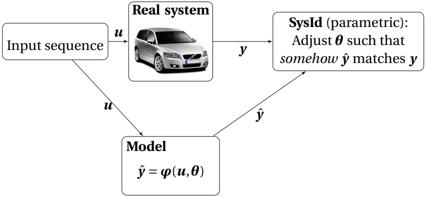


Some of the key issues:

- Experiment design: selection of inputs and outputs to be used and construction of the input sequence $\mathbf{u}$ to be applied to the system.
- Selection of model structure: the model $\hat{\mathbf{y}}(\mathbf{u}, \boldsymbol{\theta})$ can take various forms, allowing e.g. both linear and nonlinear dynamics, different parametrizations etc.
- Algorithm design: define **what is a good fit of the model** to data, and **how to find the best model parameter** vector $\boldsymbol{\theta}$.

The system identification problem

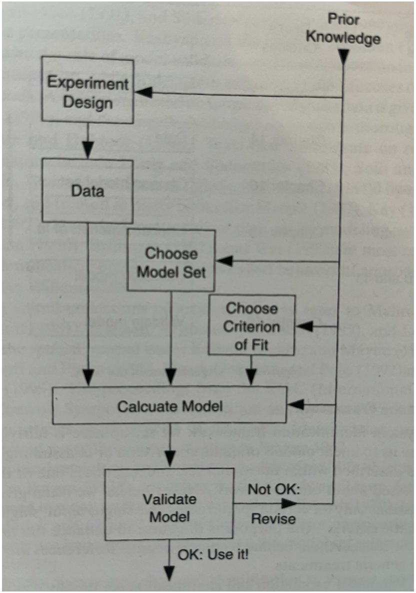
SysId: Adjust the model (with adjustable parameters) to data.



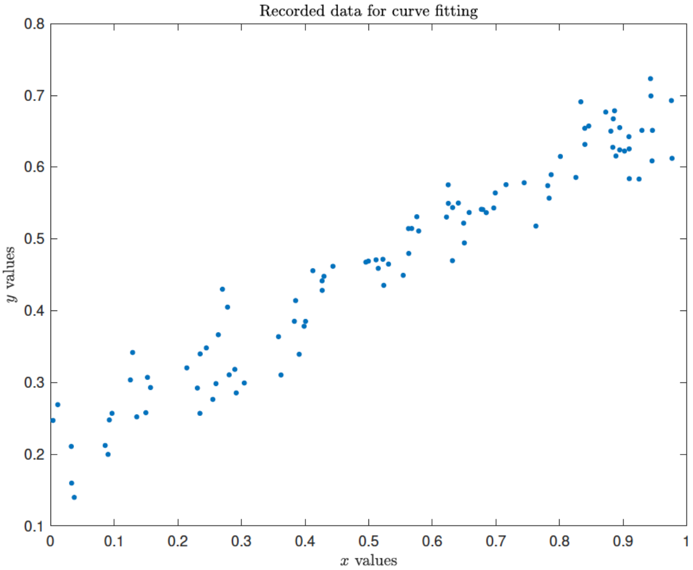
Some of the key issues:

- Experiment design: selection of inputs and outputs to be used and construction of the input sequence $\mathbf{u}$ to be applied to the system.
- Selection of model structure: the model $\hat{\mathbf{y}}(\mathbf{u}, \boldsymbol{\theta})$ can take various forms, allowing e.g. both linear and nonlinear dynamics, different parametrizations etc.
- Algorithm design: define what is a good fit of the model to data, and how to find the best model parameter vector $\boldsymbol{\theta}$.
- Model validation: **assess the resulting model** and whether it fills its purpose? (simulation, statistical tests)

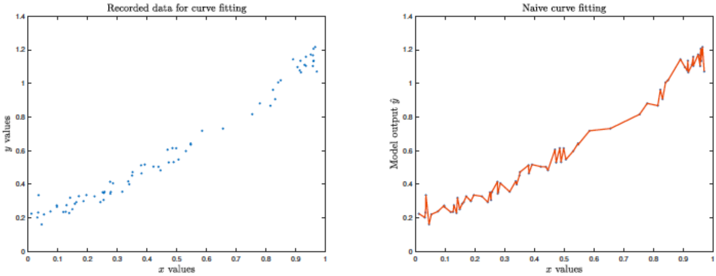
Data-driven modelling flow



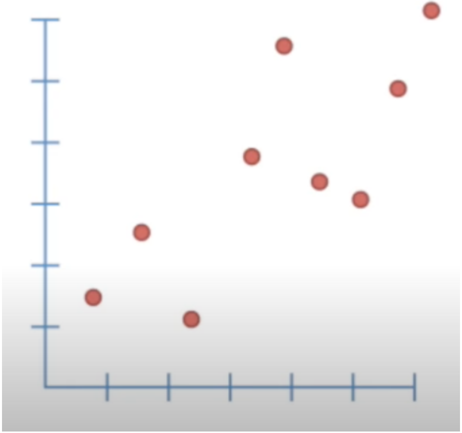
Example data for curve fitting: y = f(x)



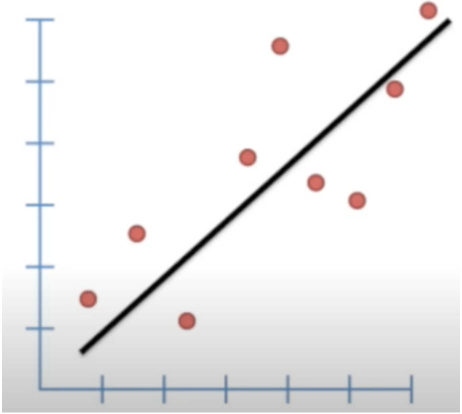
Overfitting



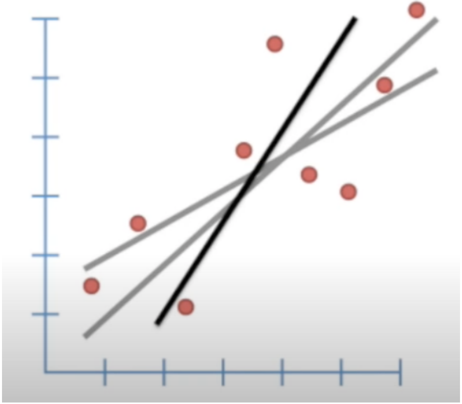
Curve fitting



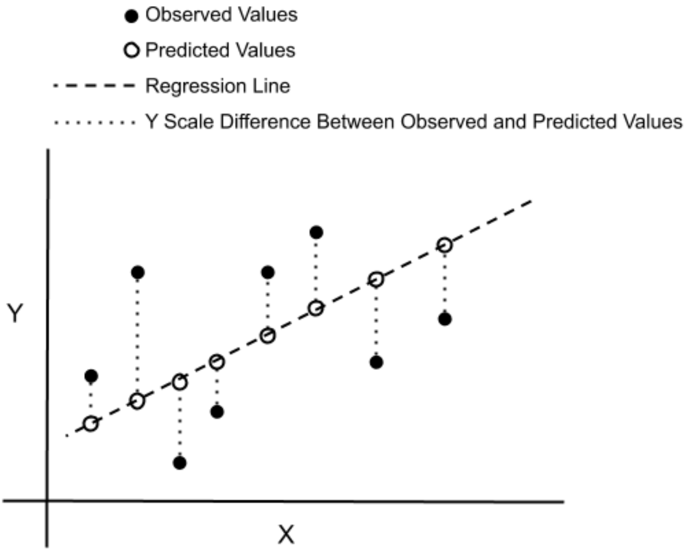
Curve fitting



Curve fitting



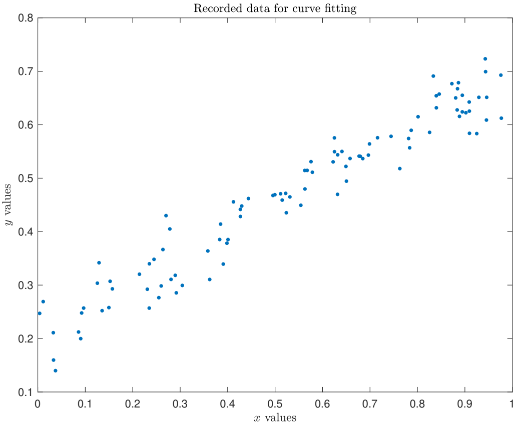
Curve fitting criterion



Example: Curve fitting using linear regression

Data: $x(i), y(i), \quad i = 1, \ldots, N$

Model: $y(i) = a + b \cdot x(i) = \theta^\top \varphi(i), \quad \theta = \begin{bmatrix} a \\ b \end{bmatrix}, \quad \varphi(i) = \begin{bmatrix} 1 \\ x(i) \end{bmatrix}$



Linear regression and least-squares

Consider the *linear-in-the-parameters* model

$$y(i) = \theta^\top \varphi(i), \qquad \theta = [\theta_1 \cdots \theta_d]^\top$$

where the *regression vector* $\varphi(i)$ contains known, deterministic signals.
Example: Polynomial trend.

The *least-squares (LS)* criterion is defined as

$$V_N(\theta) = \frac{1}{N} \sum_{i=1}^{N} \varepsilon^2(i,\theta),$$

where the *residual* ε expresses the discrepancy between data and model:

$$\varepsilon(i,\theta) = y(i) - \hat{y}(i|\theta) = y(i) - \theta^\top \varphi(i).$$

namely, the average squared difference between the estimated values and the actual values
The *least-squares estimate* minimizes the criterion, i.e.

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

Solution to the LS problem

The LS criterion can be written as:

$$\mathbf{y} = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}, \qquad \Phi = \begin{bmatrix} \varphi^\top(1) \\ \vdots \\ \varphi^\top(N) \end{bmatrix}, \tag{1}$$

$$V_N(\boldsymbol{\theta}) = \frac{1}{2}\|\mathbf{y} - \Phi\boldsymbol{\theta}\|^2 = \frac{1}{2}(\mathbf{y} - \Phi\boldsymbol{\theta})^\top(\mathbf{y} - \Phi\boldsymbol{\theta}) \tag{2}$$

The LS solution is found by:

$$\frac{dV_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}} = \boldsymbol{\theta}^\top\Phi^\top\Phi - \mathbf{y}^\top\Phi = 0, \tag{3}$$

giving

$$\hat{\boldsymbol{\theta}}_N = (\Phi^\top\Phi)^{-1}\Phi^\top\mathbf{y}, \tag{4}$$

$$\hat{\boldsymbol{\theta}}_N = R_N^{-1} f_N = \Big(\frac{1}{N}\sum_{i=1}^{N}\varphi(i)\varphi^\top(i)\Big)^{-1}\frac{1}{N}\sum_{i=1}^{N}\varphi(i)y(i) \tag{5}$$

Weighted least-squares

Define

$$\mathbf{y} = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} \qquad \Phi = \begin{bmatrix} \varphi^\top(1) \\ \vdots \\ \varphi^\top(N) \end{bmatrix}$$

The *weighted least-squares (WLS)* criterion can then be written as

$$V_N(\theta) = \frac{1}{2}\|\mathbf{y} - \Phi\theta\|_W^2 = \frac{1}{2}(\mathbf{y} - \Phi\theta)^\top W(\mathbf{y} - \Phi\theta)$$

The LS solution is found by differentiating w.r.t. θ:

$$\frac{dV_N(\theta)}{d\theta} = \theta^\top \Phi^\top W \Phi - \mathbf{y}^\top W \Phi = 0,$$

giving

$$\hat{\theta} = (\Phi^\top W \Phi)^{-1} \Phi^\top W \mathbf{y}$$

Note: the solution can be interpreted as an approximate solution of the overdetermined linear system of equations $\mathbf{y} = \Phi\theta$.

Solution to the LS problem

The *LS estimate* is

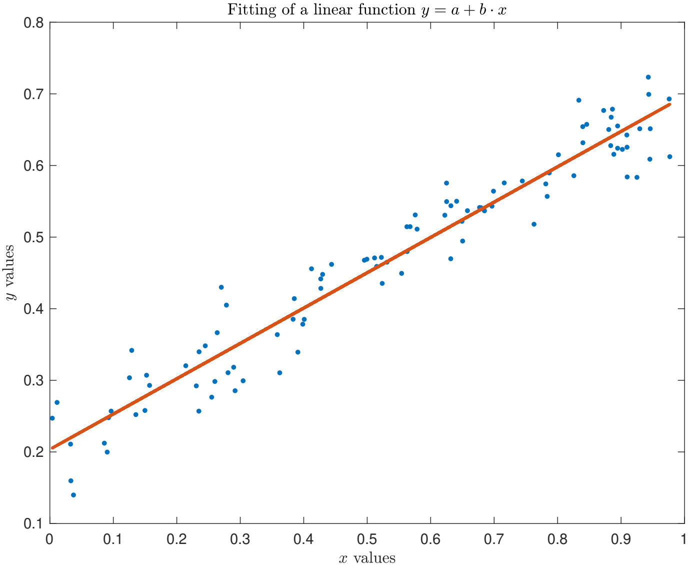
$$\hat{\theta}_N = R_N^{-1} f_N$$

$$\hat{\boldsymbol{\theta}}_N = \begin{bmatrix} \hat{a}_N \\ \hat{b}_N \end{bmatrix} = \Big(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\varphi}(i)\boldsymbol{\varphi}^\top(i)\Big)^{-1} \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\varphi}(i) y(i)$$

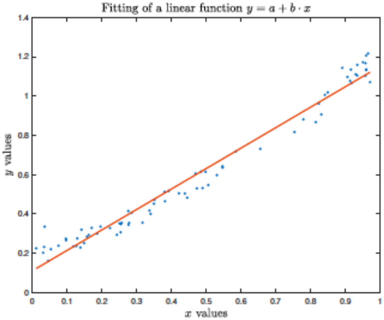
$$\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\varphi}(i)\boldsymbol{\varphi}^\top(i) = \frac{1}{N}\begin{bmatrix} N & \sum_{i=1}^{N} x(i) \\ \sum_{i=1}^{N} x(i) & \sum_{i=1}^{N} x^2(i) \end{bmatrix}$$

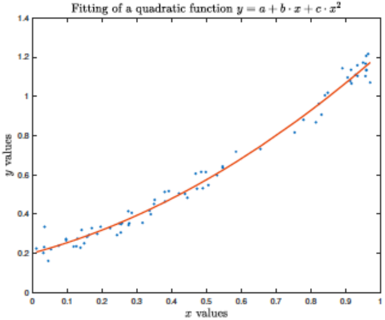
$$\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\varphi}(i) y(i) = \frac{1}{N}\begin{bmatrix} \sum_{i=1}^{N} y(i) \\ \sum_{i=1}^{N} x(i) y(i) \end{bmatrix}$$

Curve fitting, cont'd

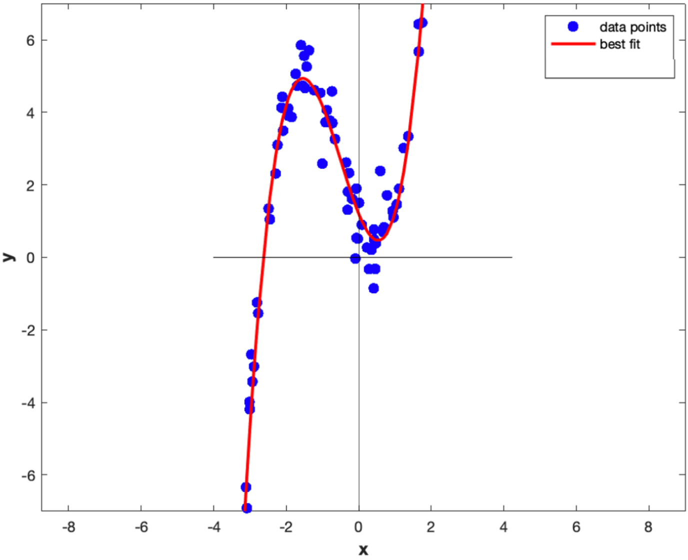


Curve fitting examples





The system identification problem



Linear regression and least-squares

Consider the *linear-in-the-parameters* model

$$y(i) = \theta^\top \varphi(i), \qquad \theta = [\theta_1 \cdots \theta_d]^\top$$

where the *regression vector* $\varphi(i)$ contains known, deterministic signals.
Example: Polynomial trend.

The *least-squares (LS)* criterion is defined as

$$V_N(\theta) = \frac{1}{N} \sum_{i=1}^{N} \varepsilon^2(i, \theta),$$

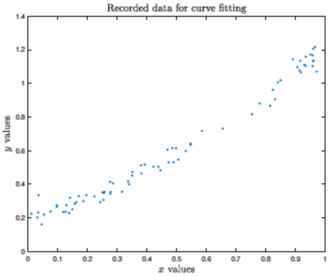
where the *residual* ε expresses the discrepancy between data and model:

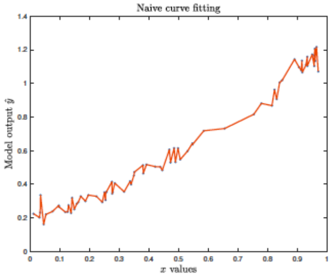
$$\varepsilon(i, \theta) = y(i) - \hat{y}(i|\theta) = y(i) - \theta^\top \varphi(i).$$

The *least-squares estimate* minimizes the criterion, i.e.

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

Overfitting





How to avoid overfitting?