Aalto University
School of Electrical
Engineering

ELEC-E8103 Modelling, Estimation and Dynamic Systems

6. Curve fitting

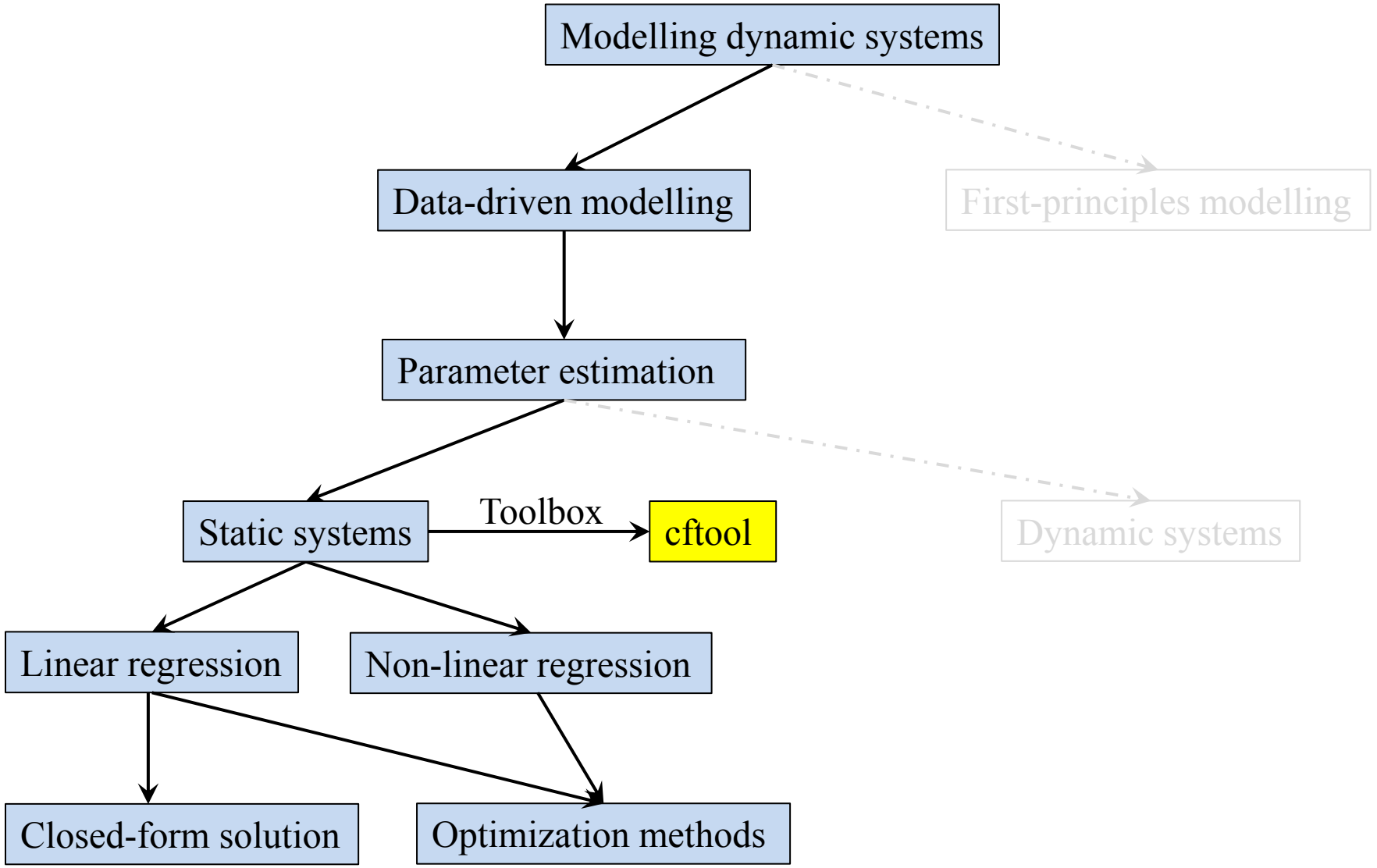
Artur Kopitca

Department of Electrical Engineering and Automation

Aalto University, School of Electrical Engineering

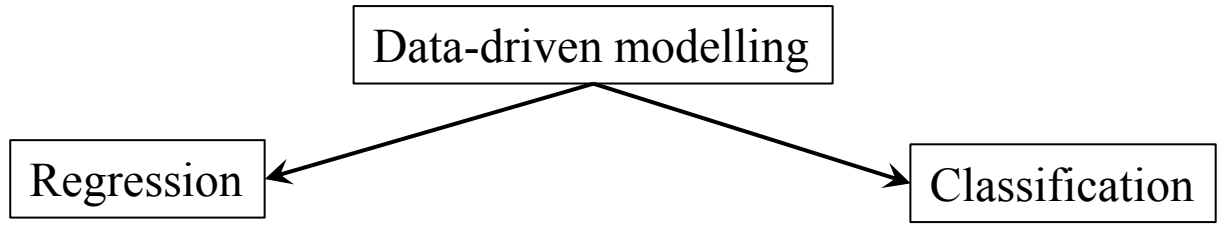
Email: artur.kopitca@aalto.fi

Recap


Modelling dynamic systems
Data-driven modelling
First-principles modelling
Parameter estimation
Static systems
Toolbox
cftool
Dynamic systems
Linear regression
Non-linear regression
Closed-form solution
Optimization methods

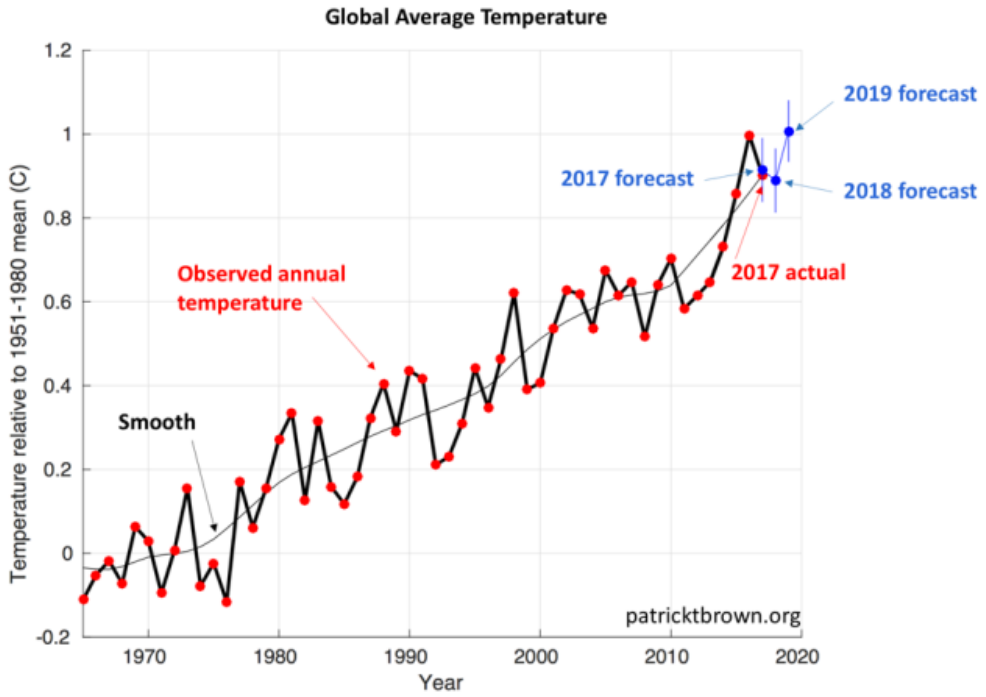
Data-driven modelling

Data-driven modelling Machine learning Grey-box modelling



Predicting a continuous-valued output
(e.g., temperature)

Predicting a discrete-valued output
(e.g., Day or Night)



Day or night?

Linear vs. nonlinear regression

Linear or nonlinear regression?

$y = k_1 x + k_2$	Linear regression, linear function
$y = k_1 x^2 + k_2 x + k_3$	Linear regression, nonlinear function
$z = k_1 x^2 + k_2 xy + k_3 y^2$	Linear regression, nonlinear function
$y = k_1 \sin(k_2 x + k_3) + k_4 \cos(k_5 x + k_6)$	Nonlinear regression, nonlinear function
$y = k_1 (1 - e^{-k_2 x})$	Nonlinear regression, nonlinear function
$y = k_1 x^{k_2}$	Nonlinear → Linear

Nonlinear → Linear regression problem

Nonlinear regression: $y = k_1 x^{k_2}$

Convert to a **linear** regression problem:

$$\ln y = \ln\left(k_1 x^{k_2}\right)$$

$$\ln y = \ln k_1 + k_2 \ln x$$

$$y_{\text{new}} = \alpha_1 + \alpha_2 x_{\text{new}}$$

Solve the linear regression problem and extract k_1 and k_2 from α_1 and α_2 thereafter. 👍

MATLAB Curve Fitting Toolbox

- Interactive curve-fitting procedure:

  - ✓ Generate data
  - ✓ Import training and validation data sets
  - ✓ Select the model structure
  - ✓ Parameter estimation → Confidence bounds
  - ✓ Fit post-processing
  - ✓ Analyze the goodness of fit
  - ✓ Code generation
  - ✓ Using generated code in MATLAB scripts

MATLAB Curve Fitting Toolbox

- Generate training and validation data with the following structure:

$$y = k_1 x^2 + k_2 x + k_3$$

- Add noise to x, y, xv, and yv data.
- Fit polynomials with orders 1, 2, 3, and 4 to the data.
- Compare the results of fitting on training and validation data sets.
- Generate a MATLAB code to fit data with your selected structure.
- Plot the resulted function in MATLAB for a range of x. (for instance x ϵ [0,20])

- Estimate the parameters of the following nonlinear function for the second dataset,

$$y = k_1 \sin(k_2 x + k_3) + k_4 \sin(k_5 x + k_6)$$

Closed-form solution

- Linear regression:

$$y = \theta_1 + \theta_2 \cdot x$$

- Use matrix notation:

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \qquad \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

$$\begin{cases} y_1 = \theta_1 + \theta_2 \cdot x_1 \\ y_2 = \theta_1 + \theta_2 \cdot x_2 \\ \quad\quad \dots \\ y_n = \theta_1 + \theta_2 \cdot x_n \end{cases} \longrightarrow \mathbf{Y} = \mathbf{X} \cdot \boldsymbol{\theta}$$

- Estimating $\boldsymbol{\theta}$:

$$\mathbf{Y} = \mathbf{X} \cdot \boldsymbol{\theta}$$

$$\mathbf{X}^{\mathrm{T}} \cdot \mathbf{Y} = \mathbf{X}^{\mathrm{T}} \cdot \mathbf{X} \cdot \boldsymbol{\theta}$$

$$(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \mathbf{Y} = (\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \mathbf{X} \cdot \boldsymbol{\theta}$$

$$\boxed{(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \mathbf{Y} = \boldsymbol{\theta}}$$

Closed-form solution

- Linear regression:

$$y = \theta_1 + \theta_2 \cdot x$$

Number of samples: n Number of parameters: k

- Closed-form solution

最小二乘法

$$\boldsymbol{\theta} = \left(\mathbf{X}^{\mathbf{T}} \cdot \mathbf{X}\right)^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \mathbf{Y}$$

Parameters matrix

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \rightarrow k \times 1$$

Input matrix

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \rightarrow n \times k$$

Output matrix

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \rightarrow n \times 1$$

Closed-form solution: Example

- Linear regression for **nonlinear functions**:

$$z = \theta_1 + \theta_2 x + \theta_3 xy + \theta_4 \sin(x)$$

Input matrix

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1 y_1 & \sin(x_1) \\ 1 & x_2 & x_2 y_2 & \sin(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n y_n & \sin(x_n) \end{bmatrix} \rightarrow n \times k$$

Output matrix

$$\mathbf{Y} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \rightarrow n \times 1$$

Parameters matrix

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} \rightarrow k \times 1$$

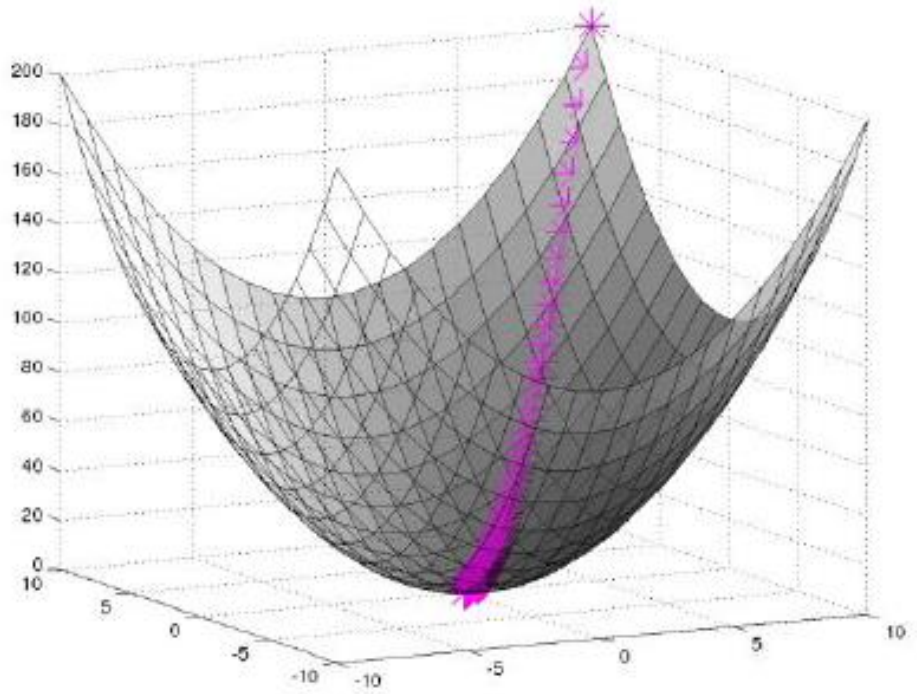
Nonlinear regression

- Cost function

$$J = \sum_{i=1}^{n} (y_i - f(x_i, k))^2 = \sum_{i=1}^{n} \left(y_i - \frac{k_1 x_i}{k_2 + x_i} \right)^2$$

Data

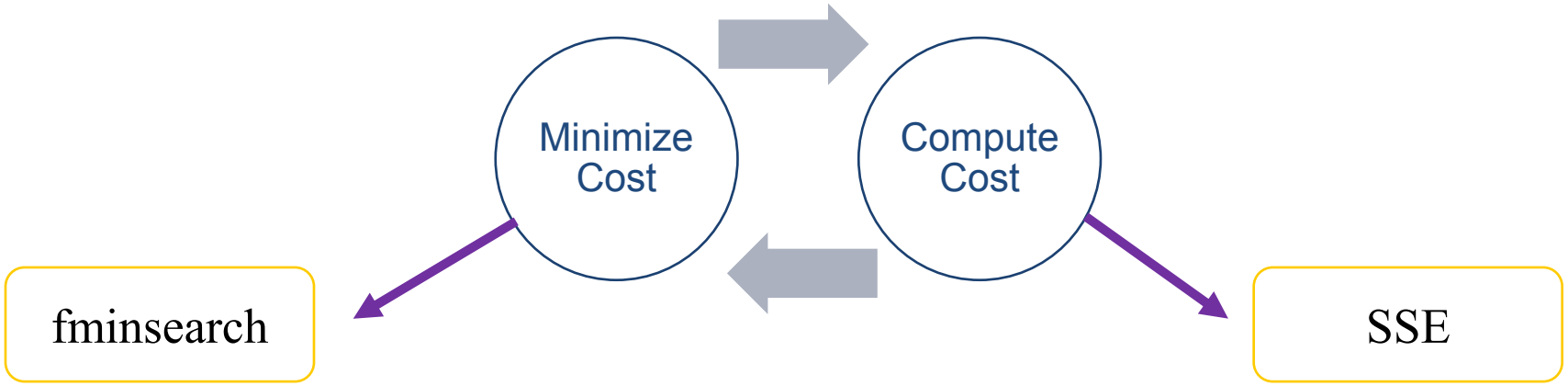
$$y = f(x) = \frac{k_1 x}{k_2 + x}$$



- Minimization of cost function using search algorithms in MATLAB
  - ✓ fminsearch
  - ✓ fminunc
  - ✓ Gradient descent
  - ✓ ...

Nonlinear regression

- How to solve the nonlinear regression problems using search algorithms systematically?



- MATLAB implementation

```
k = fminsearch('computeCost', k0);
```

Function to minimize

Initial point

Readings

- MATLAB Curve Fitting Toolbox guide:
  https://se.mathworks.com/help/curvefit/curvefitting-app.html

- Polynomial curve fitting in MATLAB:
  https://se.mathworks.com/help/matlab/ref/polyfit.html

- Nonlinear optimization in MATLAB:
  https://se.mathworks.com/help/matlab/ref/fminsearch.html

Questions