Mathematical Modelling FK (FRTN45) F2

Anders Rantzer

Department of Automatic Control, LTH

Outline

Thursday lecture

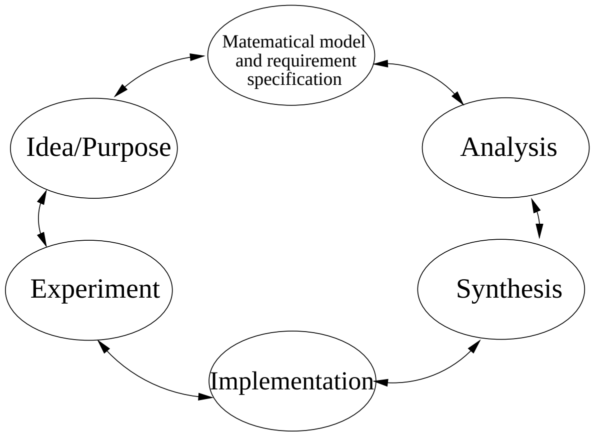
- Course introduction
- Models from physics (white boxes)

Friday lecture

- Models from data (black boxes)
  - Singular Value Decomposition (SVD)
  - Machine Learning
  - System Identification / Time Series Analysis
- Mixed models (grey boxes)

Modelling in three phases:

1. Problem structure
   - **Formulate purpose**, requirements for accuracy
   - Break up into subsystems — What is important?
2. Basic equations
   - Write down the relevant physical laws
   - Collect experimental data
   - Test hypotheses
   - Validate the model against fresh data
3. Model with desired features is formed
   - Put the model on suitable form. (Computer simulation or pedagogical insight?)
   - Document and illustrate the model
   - Evaluate the model: **Does it meet its purpose?**



Lecture 2

- Statistical modeling from data (statistical black boxes)
  - Singular Value Decomposition (SVD)
  - Principal Component Analysis (Factor Analysis)
- Dynamic experiments (dynamik black boxes)
  - Step response
  - Frequency response
  - Correlation analysis
- Gray boxes
  - Prediction error methods
  - Differential-Algebraic Equations revisited

Singular Value Decomposition (SVD)

A matrix M can always be factorized

$$M = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^*$$

with Σ diagonal and invertible and U, V unitary:

$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \qquad U^*U = I \qquad V^*V = I$$

Diagonal elements of Σ are called singular values of M and correspond to the square roots of the eigenvalues of M^*M.

Computation of SVD is *very numerically stable*.

Example of SVD

$$M \quad = \quad U \quad \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \quad V^*$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}_{U} \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \underbrace{\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \frac{1}{\sqrt{2}}}_{V^*}$$

What does it mean if a singular value is zero?

What does it mean if it is near zero?

Good children can have many names

Collect all the data into a large matrix. Then compute the **SVD**:

$$\begin{bmatrix} y_1(1) & y_1(2) & \dots & y_1(N) \\ y_2(1) & y_2(2) & \dots & y_2(N) \\ \vdots & & & \\ y_p(1) & y_p(2) & \dots & y_p(N) \end{bmatrix} = U \underbrace{\begin{bmatrix} \sigma & & 0 \\ & \ddots & \\ 0 & & \sigma_p \end{bmatrix}}_{\Sigma} V^*$$

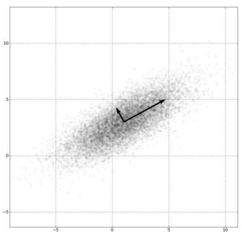
Singular values σ_i in decreasing order on the diagonal of Σ. The first columns of U give the direction of the main data area.

Principal Component Analysis: By replacing the small singular values σ_i with zeros focuses on the essential.

The name '**factor analysis**' is sometimes used as a synonymous, since large singular values σ_i highlight important factors.

Principal Component Analysis (PCA)

Data from a bi-dimensional Gaussian distribution centered in $(1, 3)$:



Principal component $(0.878, 0.478)$ has standard deviation 3.

Next component has standard deviation 1.

[Källa: Wikipedia]

Example: Image processing

What does this picture represent?

```
M =

     1     0     0     1     1     1     0     1     0     1
     1     0     0     0     1     0     0     1     0     1
     1     0     0     0     1     0     0     1     1     1
     1     0     0     0     1     0     0     1     0     1
     1     1     0     0     1     0     0     1     0     1
```

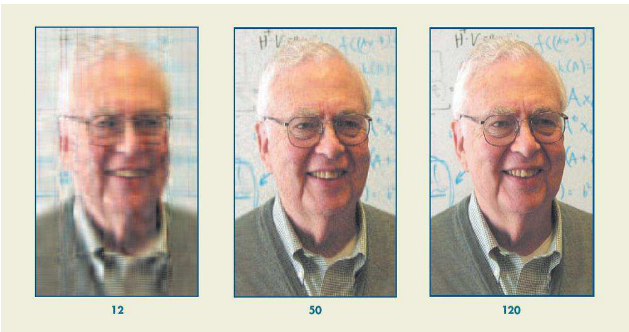
Example: Image processing with SVD

```
>> [U,S,V]=svd(M)

U =

   -0.4747    0.8662    0.0000   -0.1559    0.0000
   -0.4291   -0.1371   -0.0000    0.5450   -0.7071
   -0.4508   -0.3256   -0.7071   -0.4368   -0.0000
   -0.4291   -0.1371   -0.0000    0.5450    0.7071
   -0.4508   -0.3256    0.7071   -0.4368    0.0000

S =

4.5638 0      0      0      0      0      0      0      0      0
0      1.3141 0      0      0      0      0      0      0      0
0      0      1.0000 0      0      0      0      0      0      0
0      0      0      0.6670 0      0      0      0      0      0
0      0      0      0      0.0000 0      0      0      0      0
```

Example: Image processing with SVD

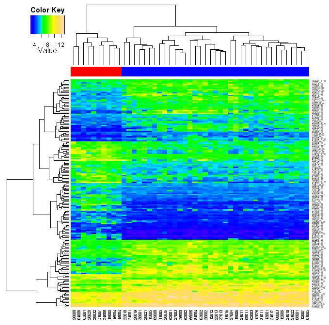
```
round(U*S1*V') =

     1     0     0     0     1     0     0     1     0     1
     1     0     0     0     1     0     0     1     0     1
     1     0     0     0     1     0     0     1     0     1
     1     0     0     0     1     0     0     1     0     1
     1     0     0     0     1     0     0     1     0     1

round(U*S2*V') =

     1     0     0     1     1     1     0     1     0     1
     1     0     0     0     1     0     0     1     0     1
     1     0     0     0     1     0     0     1     0     1
     1     0     0     0     1     0     0     1     0     1
     1     0     0     0     1     0     0     1     0     1
```

Example: Image processing



The original image has 897-by-598 pixels. Tacking red, green and blue vertically gives a 2691-by-598 matrix. Truncating all but 12 singular values gives the left picture. 120 gives the right.

Example: Correlations genes-proteines



Cancer research: microarrays (glass) with human genes are exposed to healthy cells, then to sick ones. Make a SVD of the data to find out which genes are important!

Outline

Thursday lecture

- Course introduction
- Models from physics (white boxes)

Friday lecture

- Models from data (black boxes)
  - Singular Value Decomposition (SVD)
  - **Machine Learning**
  - System Identification / Time Series Analysis
- Mixed models (grey boxes)

Fun stuff before we get started

3 / 24

Deep Dream version

4/24

Which one is Hemingway?

NO. 1:

Kilimanjaro is a snow-covered mountain 19,710 feet high, and is said to be the highest mountain in Africa. Its western summit is called the Masai "Ngaje Ngai," the House of God. Close to the western summit there is the dried and frozen carcass of a leopard. No one has explained what the leopard was seeking at that altitude.

NO. 2:

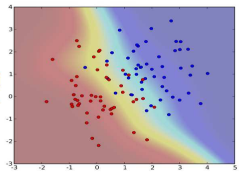
Kilimanjaro is a mountain of 19,710 feet covered with snow and is said to be the highest mountain in Africa. The summit of the west is called "Ngaje Ngai" in Masai, the house of God. Near the top of the west there is a dry and frozen dead body of leopard. No one has ever explained what leopard wanted at that altitude.

Before November 2016

Using language rule books:

Kilimanjaro is 19,710 feet of the mountain covered with snow, and it is said that the highest mountain in Africa. Top of the west, "Ngaje Ngai" in the Maasai language, has been referred to as the house of God. The top close to the west, there is a dry, frozen carcass of a leopard. Whether the leopard had what the demand at that altitude, there is no that nobody explained.

Components for deep lear



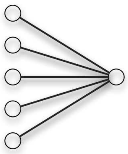
- One neuron
  - Example: Logistic regression
  - Classification model (x feature vector, (w,b) parameters, s smooth thresholding
    $x \in R^d, w \in R^d, b \in R, f(x) = s(w^T x + b)$
  - Logistic regression $s(z) = \dfrac{1}{1+e^{-x}}$
  - ML estimate of parameters (w,b) is a convex optimization problem

$$\min_{\boldsymbol{w}} \quad \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{l}\log(1+e^{-y_i\boldsymbol{w}^T\boldsymbol{x}_i}).$$

LUND UNIVERSITY

Single Layer Neural Networks One Neuron

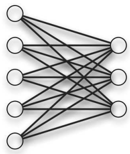
- One neuron
  $x \in R^d, w \in R^d, b \in R, f(x) = s(w^T x + b)$



LUND UNIVERSITY

Single Layer Neural Networks Several Neurons

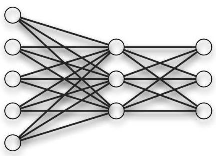
- Several parallell neurons
  $x \in R^d, y \in R^k, B \in R^d, W - k \times \text{dmatrix}$
  $y = s(Wx + B)$
- Elementwise smooth thresholding – s



LUND UNIVERSITY

Artificial Neural Networks One hidden layer

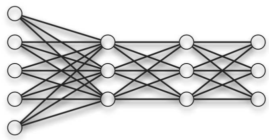
- Multi-class classification
- One hidden layer
- Trained by back-propagation
- Popular since the 1990ies



LUND UNIVERSITY

Deep Neural Networks Many layers

- However
- Naively implemented would give to many parameters
- Example
- 1M pixel image
- 1M hidden layers
- 10^{12} parameters between each pairs of layers



LUND UNIVERSITY

Outline

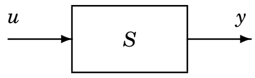
Thursday lecture

- Course introduction
- Models from physics (white boxes)

Friday lecture

- Models from data (black boxes)
  - Singular Value Decomposition (SVD)
  - Machine Learning
  - **System Identification / Time Series Analysis**
- Mixed models (grey boxes)

Basic idea of system identification



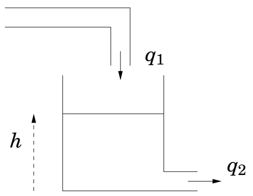
Measure U and y. Figure out a model of S, consistent with measured data.

Important aspects:

- We can only measure the u and y in discrete time points (sampling). Can be natural to use the discrete-time models.
- The system is affected by interference and measurement errors. We may need to signal models for dealing with this.

Example

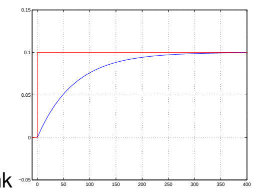
A tank which attenuates flow variations in q_1. Characterization of the tank system:



- Input: q_1
- Output: q_2 and/or h
- Internal variables / conditions: h

Step response

Step response for the tank



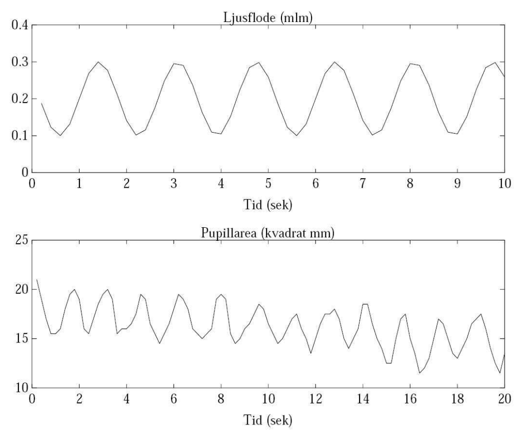
Can give idea of the dominant time constant, static reinforcement, character (overshoot or not)

Frequency response

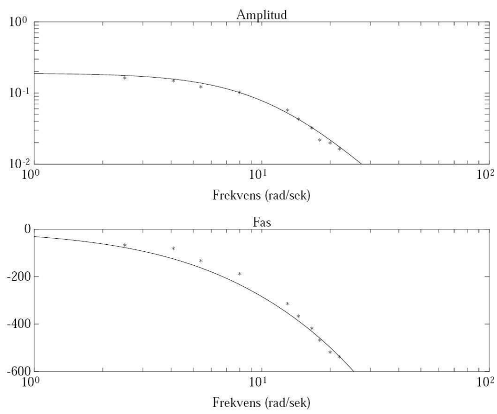
For good signal-to-noise ratio, an estimate of $G(i\omega)$ is obtained directly from the amplitudes and phase positions of u, y

$$
\begin{aligned}
u(t) &= A \sin \omega t \\
y(t) &= A|G(i\omega)| \sin(\omega t + \arg G(i\omega))
\end{aligned}
$$

How light affects pupil area



Bode-diagram for pupil



Correlation analysis

Can we estimate the impulse response with other inputs?

- Impulse response formula in discrete time ($T = 1$, $v =$ noise):

$$
y(t) = \sum_{k=1}^{\infty} g_k u(t-k) + v(t)
$$

- If v white noise with $\mathbf{E}v^2 = 1$, then

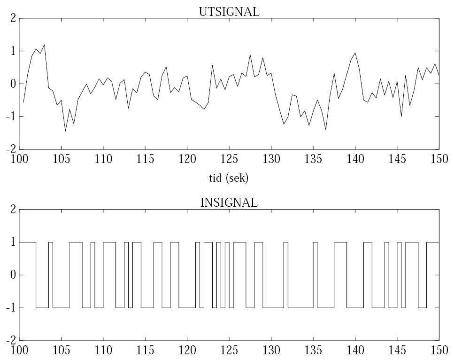
$$
R_{yu}(k) = \mathbf{E}y(t)u(t-k) = g_k
$$

- Covariance R_{yu} estimated by N data points with

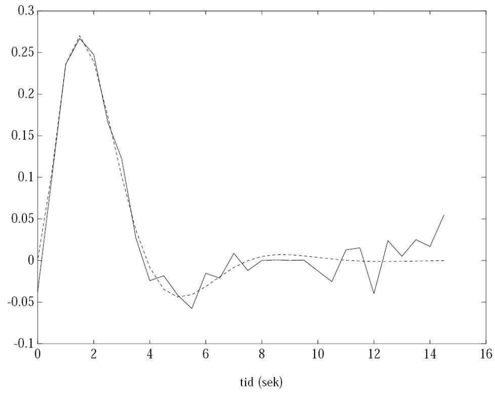
$$
\widehat{R}_{yu}^N(k) = \frac{1}{N} \sum_{t=1}^{N} y(t)u(t-k)
$$

Example

Correlation analysis for $\frac{1}{s^2+2s+1}$ (in- and out-put data)



Estimated and actual impulse responses



Basic rules

Make experiments with conditions similar to the conditions in which the model is to be used!

(Models from step response can be expected to work best on the stage.)

Save some data for model validation, i.e. check the model with data set different from the one that generated the model!

Outline

Thursday lecture

- Course introduction
- Models from physics (white boxes)

Friday lecture

- Models from data (black boxes)
  - Singular Value Decomposition (SVD)
  - Machine Learning
  - System Identification / Time Series Analysis
- **Mixed models (grey boxes)**

Prediction Error Methods

Find the unknown parameters θ by optimization:

$$\min_{\theta} \|\widehat{y}(t,\theta) - y(t)\|$$

Here $y(t)$ is the measured output at time t and $\widehat{y}(t,\theta)$ is the predicted output based on past measurements using a model with parameter values θ.

Prediction Error Method with Repeated Simulation

For a nonlinear grey-box model

$$0 = F(\dot{x}, x, t, \theta)$$
$$y(t) = h(x, t, \theta)$$

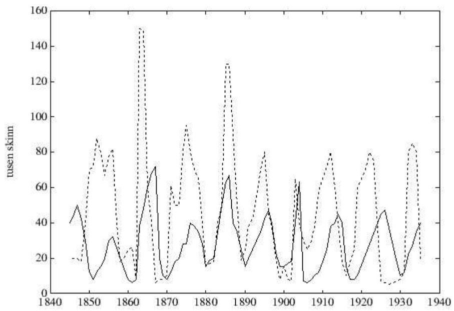
the unknown parameters θ could be determined by the prediction error method

$$\min_{\theta} \|\widehat{y}(t,\theta) - y(t)\|$$

where the output prediction $\widehat{y}(t,\theta)$ is computed by simulation.

(Repeated simulation for different values of θ could however be very time-consuming.)

Population dynamics / Ecology



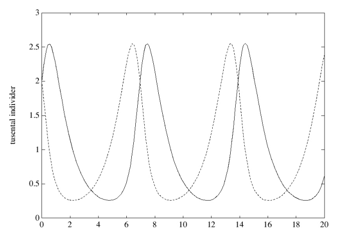
Variations in the number of lynx (solid) and hares (dashed) in Canada. Can you predict the periodic variations?

Population dynamics

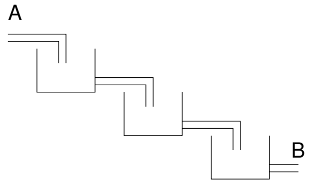
N_1 number of lynx, N_2 number of hares

$$\frac{d}{dt}N_1(t) = (\lambda_1 - \gamma_1)N_1(t) + \alpha_1 N_1(t)N_2(t)$$
$$\frac{d}{dt}N_2(t) = (\lambda_2 - \gamma_2)N_1(t) - \alpha_2 N_1(t)N_2(t)$$

Simulation:



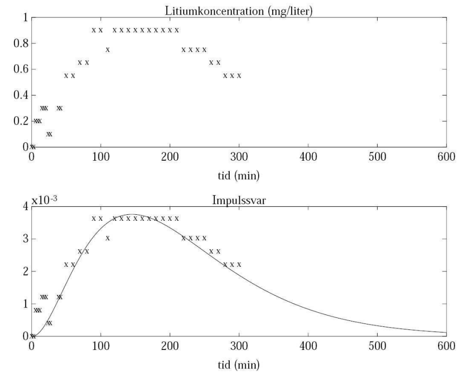
Mixing tanks in Skärblacka paper factory



A linear transfer function of three series-connected mixing tanks has the form $\frac{1}{(s\theta+1)^3}$.

To determine θ, radioactive lithium is added in **A**. Radioactivity was then measured by **B** as a function of time.

Impulse response



In the lower picture, θ has been chosen to adapt to the impulse response of $\frac{1}{(s\theta+1)^3}$

Grey Models — the best of both worlds

- White boxes: Physical laws provide some insight
- Black boxes: Statistics estimates complex relationships
- Gray boxes: Combine simplicity with insight

Nonlinear differential-algebraic equations (DAE)

Differential-algebraic equations, DAE

$$F(\dot{z}, z, u) = 0, \quad y = H(z, u)$$

u: input, y: output, z: "internal variable"

Special case: state model

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

u: input, y: output, x: state

Example: Pendulum

A pendulum with length L and position coordinates (x, y) moves according to the equations

$$\begin{aligned} &\dot{x} = u && \dot{y} = v && \\ &\dot{u} = \lambda x && \dot{v} = \lambda y && L^2 = x^2 + y^2 \end{aligned}$$

Differentiating the fifth equation gives

$$0 = \dot{x}x + \dot{y}y = ux + vy$$

Differentiating a second time gives

$$\begin{aligned} 0 &= \dot{u}x + u\dot{x} + \dot{v}y + v\dot{y} \\ &= \lambda(x^2 + y^2) - gy + u^2 + v^2 \\ &= \lambda L^2 - gy + u^2 + v^2 \end{aligned}$$

and a third time

$$0 = L^2\dot{\lambda} - 3gv$$

Finally, we have derivative expressions for all variables!

Mathematical modelling — Lectures

- Why modelling?
  - Natural sciences: Models for analysis (understanding)
  - Engineering sciences: Models for synthesis (design)
  - Specification: Model of a good technical solution
- Physical modeling (white boxes, **Tuesday's lecture**)
  Model derived from fundamental physical laws
- Statistical methods (black boxes, **today**)
  Model derived from measurement data
  - Singular Value Decomposition (SVD)
  - Machine Learning
  - System Identification / Time Series Analysis
- Combination of the two (gray boxes **today**)