Model-based Estimation Methods

Parameter Estimation,

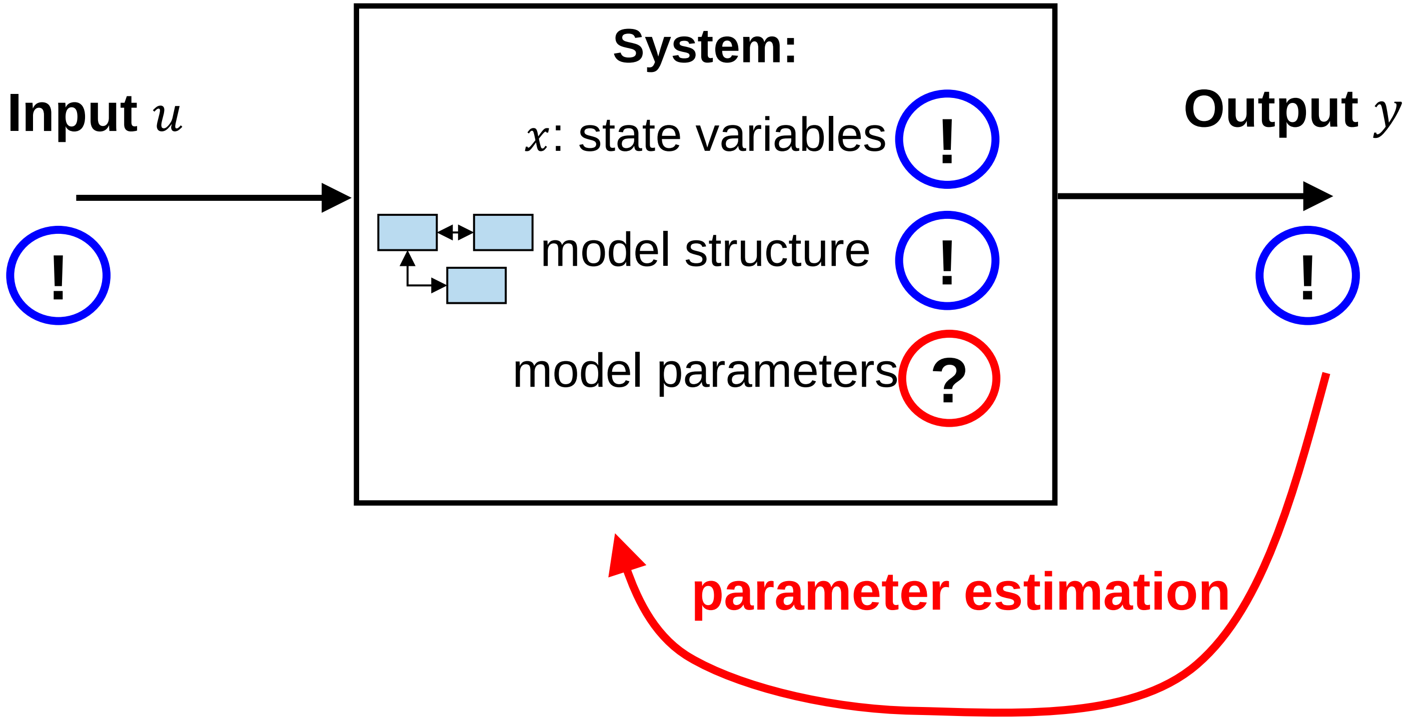
Error-in-Variables Estimation & Confidence Region

Dr.-Ing. Adel Mhamdi
AVT – Systemverfahrenstechnik

Lecture Outline

- What is a Parameter Estimation Problem?
- Linear Parameter Estimation
  - Least-Squares
  - Weighted Least-Squares
- Nonlinear Parameter Estimation
- Error-in-Variables Estimation
- Confidence Region
- Case study: Challenges in Spectral Analysis
- Data Reduction Using Principal Component Analysis

Parameter Estimation



Assume that **states, model structure, outputs and inputs** are known.

Find the **parameters** of this model, so that the model predicts as best as possible the measurement data (outputs).

Example: Reaction Kinetics

- Reaction processes

$$\text{reactant(s)} \xrightarrow{\color{red}{r}} \text{products,}$$

Application:
- chemical reactor design
- drug delivery

- Examples : - radioactive decay ${}^{235}U \rightarrow {}^{231}Th$
  - dimerization of butadiene $2C_4H_6 \rightarrow C_8H_{12}$
- Reaction rate model: reaction rate is function of concentration, temperature, …

$$r = f(x_A, k)\,,\;\; x_A \text{ is the concentration of reactant } A$$

  - First order reaction $r = kx_A$
  - Second order reaction $r = kx_A^2$

- **Parameter estimation problem:**

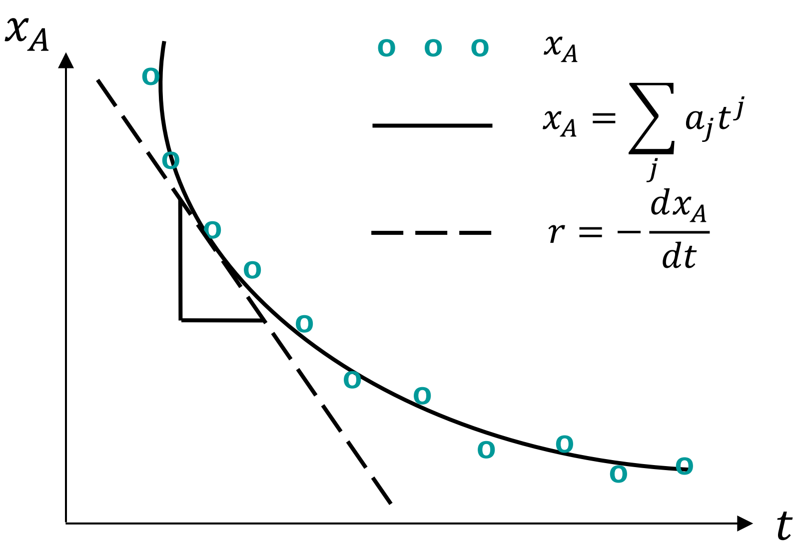
Estimate reaction rate constant k from concentration measurements !

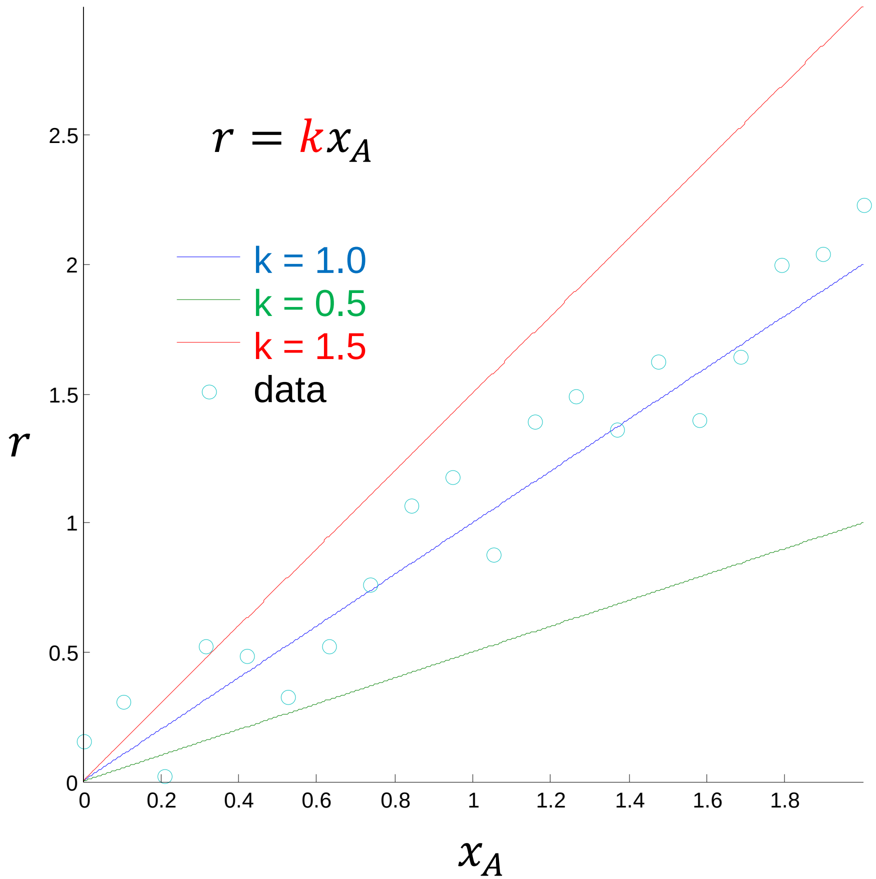
Example: Reaction Kinetics

Assuming measurements for x_A are available, how to estimate parameter k?

Use mass balance to compute r

$$x_A \Rightarrow r, \qquad since \ \frac{dx_A}{dt} = -r$$





Lecture Outline

- What is a Parameter Estimation Problem?
- Linear Parameter Estimation
  - Least-Squares
  - Weighted Least-Squares
- Nonlinear Parameter Estimation
- Error-in-Variables Estimation
- Confidence Region
- Case Study: Challenges in Spectral Analysis
- Data Reduction Using Principal Component Analysis

Linear Parameter Estimation

Consider a static model $y := f(u, \Theta)$ (a single algebraic equation),

linear in parameters Θ ,

but may be **nonlinear** in the input variables u,

i.e.

$$y = \sum_{j=1}^{n} x_j \Theta_j \text{ with } x_j := f_{u,j}(u)$$

Assume, we have $\boldsymbol{m}$ experiments (from m combinations of the inputs u),

thus we get m measurements

$$\tilde{y}_i = y_i + \epsilon_i, i = 1 \ldots m,$$

with ϵ_i an error term and

$$y_i = \sum_{j=1}^{n} x_{ij} \Theta_j \text{ , with } x_{ij} := f_{u,j}(u_i)$$

Linear Parameter Estimation - Matrix Notation

…we have m measurements $\tilde{y}_i = y_i + \epsilon_i, i = 1 \dots m$ and

$$y_i = \sum_{j=1}^{n} x_{ij}\Theta_j \text{ , with } x_{ij} := f_{u,j}(u_i)$$

Let $Y := [y_1 \quad \cdots \quad y_m]^T \in \mathbb{R}^m$, $\tilde{Y} := [\tilde{y}_1 \quad \cdots \quad \tilde{y}_m]^T \in \mathbb{R}^m$,

$\Theta := [\Theta_1 \quad \cdots \quad \Theta_n]^T \in \mathbb{R}^n$

and
$$X := \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n}$$

We get in matrix notation:

$$Y = X\Theta$$

Linear Parameter Estimation - Least-Squares (LS) Formulation

$$\widehat{\Theta} = \arg\min_{\Theta} \left\| Y - \tilde{Y} \right\|_2^2 = \arg\min_{\Theta} \left\| X\Theta - \tilde{Y} \right\|_2^2$$

(cf. Reusken, 2.3)

Necessary conditions:

$$\nabla_{\Theta} \left\| X\widehat{\Theta} - \tilde{Y} \right\|_2^2 = 0$$
$$\Leftrightarrow X^T X\widehat{\Theta} - X^T\tilde{Y} = 0$$
$$\Leftrightarrow \boxed{\widehat{\Theta} = X^{\dagger}\tilde{Y}}$$

where $X^{\dagger}$ is the pseudoinverse of X

Sufficient conditions:

all eigenvalues of $\nabla_{\Theta}^2 \left\| X\widehat{\Theta} - \tilde{Y} \right\|_2^2 = X^T X$ are positive and real

$$\Leftrightarrow \boxed{\operatorname{rank}(X) = n}$$

Remember: (Reusken, section 2.3)

Pseudoinverse of X

- $m = n,\ X^{\dagger} = \mathrm{X}^{-1}$
- $m > n,\ X^{\dagger} = (X^T X)^{-1} X^T$

Using SVD

$$X^{\dagger} = V\Sigma^{\dagger}U^T$$

$\Sigma^{\dagger} = diag\left(\sigma_1^{-1}, \dots, \sigma_{\min(m,n)}^{-1}, 0, \dots, 0\right)$ $\in \mathbb{R}^{n\times \mathrm{m}}$, where U, V are left and right singular vector

Linear Parameter Estimation - Illustrative Example Using LS

- **Model**

$$y = f(u,\Theta) = \Theta_0 + \Theta_1 u + \Theta_2 u^2 = [1\ u\ u^2][\Theta_0 \quad \Theta_1 \quad \Theta_2]^T$$

with true parameter values
$[\Theta_0 \quad \Theta_1 \quad \Theta_2]^T = [2,3,2]^T$

- **Measurements**
$\tilde{y} = y + \epsilon \ \ , \epsilon \sim N(0, 0.25^2)$

exp.	1	2	3	4	5	6
u	0	1	2	3	4	5
$\tilde{y}$	1.98	7.16	16.1	28.9	46.4	66.7

- **Parameter estimation by LS**

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix}, \tilde{Y} = \begin{bmatrix} 1.98 \\ 7.16 \\ 16.1 \\ 28.9 \\ 46.4 \\ 66.7 \end{bmatrix}$$

- Obviously $\text{rank}(X) = 3$

$$\Rightarrow \quad \hat{\Theta} = (X^T X)^{-1} X^T \tilde{Y} = \begin{bmatrix} 1.98 \\ 3.16 \\ 1.96 \end{bmatrix} \approx \Theta$$

Note: Linear parameter estimation is the same as linear regression!

Model-based Estimation Methods, Lecture 9
Dr.-Ing. Adel Mhamdi
SS2016

Linear Parameter Estimation – Weighted Least-Squares

Assumption in least-squares estimation:

normally distributed, Independent measurement errors with identical standard deviation

$$\widehat{\Theta} = \arg\min_{\Theta} \left\| Y - \tilde{Y} \right\|_2 = \arg\min_{\Theta} \left\| X\Theta - \tilde{Y} \right\|_2^2$$

Generalization… W weight matrix $\Rightarrow$ $X_W = WX,\ \tilde{Y}_W = W\tilde{Y}$

$$\Rightarrow \quad \widehat{\Theta} = \arg\min_{\Theta} \left\| X_W\Theta - \tilde{Y}_W \right\|_2 = \arg\min_{\Theta} \left\| X_W\Theta - \tilde{Y}_W \right\|_2^2$$

Choice of weights?

$W = \mathrm{diag}(\tilde{\sigma}_1^{-1} \quad \cdots \quad \tilde{\sigma}_n^{-1}),\ \tilde{\sigma}_i$ is standard deviation of measurement i

i.e. measurements with **large** $\tilde{\sigma}_i$ are **less important**, than those with **low** $\tilde{\sigma}_i$

Weighted least-squares (WLS)

$$\widehat{\Theta} = \arg\min_{\Theta} \|X_W\Theta - Y_W\|_2 = \arg\min_{\Theta} \|X_W\Theta - Y_W\|_2^2$$

Solution ?

necessary condition:

$$\nabla_{\Theta} \left\| X_W\widehat{\Theta} - Y_W \right\|_2^2 = 0$$
$$\Leftrightarrow X_W^T X\widehat{\Theta} - X_W^T Y_W = 0$$
$$\Leftrightarrow \boxed{\widehat{\Theta} = X_W^{\dagger} Y_W}$$

where $X_W^{\dagger}$ is the pseudoinverse of X_W

sufficient condition:

all eigenvalues of $\nabla_{\Theta}^2 \left\| X_W\widehat{\Theta} - Y_W \right\|_2^2 = X_W^T X_W$ are positive and real

$$\Leftrightarrow \boxed{\mathrm{rank}(X_W) = n}$$

Lecture Outline

- What is a Parameter Estimation Problem?
- Linear Parameter Estimation
  - Least-Squares
  - Weighted Least-Squares
- Nonlinear Parameter Estimation
- Error-in-Variables Estimation
- Confidence Region
- Case Study: Challenges in Spectral Analysis
- Data Reduction Using Principal Component Analysis

Nonlinear Parameter Estimation

The model is **nonlinear in parameters Θ** , e.g. $y = f(u, \Theta)$:

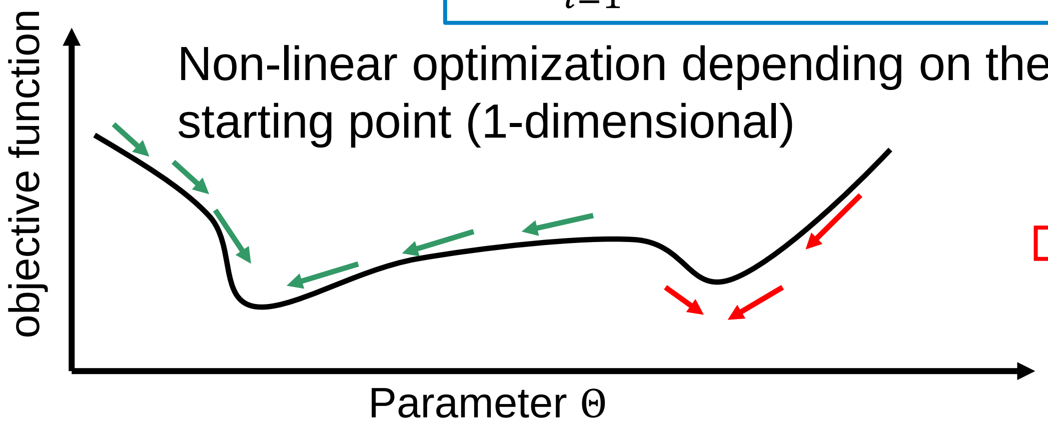
Least squares estimation

$$\min_{\Theta} \sum_{i=1}^{m} \left(f(u_i, \Theta) - \tilde{f}(u_i) \right)^2$$

is an **optimization problem**, that must be (in general) solved **iteratively** with appropriate numerical methods

NCO:

$$\boxed{\Rightarrow \sum_{i=1}^{m} \left(f(u_i, \Theta) - \tilde{f}(u_i) \right) \frac{\partial f(u_i, \Theta)}{\partial \Theta} = 0}$$



It can provide local solutions!

Lecture Outline

- What is a Parameter Estimation Problem?
- Linear Parameter Estimation
  - Least-Squares
  - Weighted Least-Squares
- Nonlinear Parameter Estimation
- Error-in-Variables Estimation
- Confidence Region
- Case Study: Challenges in Spectral Analysis
- Data Reduction Using Principal Component Analysis

Error-in-Variables Estimation (1)

- Not only the dependent output data y, but also the independent input data u are error-prone!
- Taking into account error in the inputs, the problem becomes implicit

  $\Rightarrow$ We cannot solve for corrupted variables in one step with WLS, only iteratively.

Error-prone output data	$\tilde{y} = y + \epsilon_y$
Error-prone input data	$\tilde{u} = u + \epsilon_u$
Unknown parameters	Θ_0, Θ_1

$$\tilde{y} = \Theta_0 + \Theta_1 \tilde{u} + \epsilon_y = \Theta_0 + \Theta_1 u + \Theta_1 \epsilon_u + \epsilon_y$$

$\Rightarrow$ We need a **more general method** that WLS, as it can cope only with error-prone outputs. The new method should work also in case of **nonlinear parameter estimation**!

Britt and Luecke, „The Estimation of Parameters in Nonlinear, Implicit Models“, Technometrics, 1973

Error-in-Variables Estimation (2)

We combine the error-prone inputs u and outputs y to a single variable z

General setup:

$$g(z, \Theta) = 0$$

where

- $z \in \mathbb{R}^q$ are the measurable variables
- $\Theta \in \mathbb{R}^r$ are the unknown parameter
- $g: \mathbb{R}^{q+r} \rightarrow \mathbb{R}^s$ is the vector of well behaved system functions
- and $q \geq s > r$.

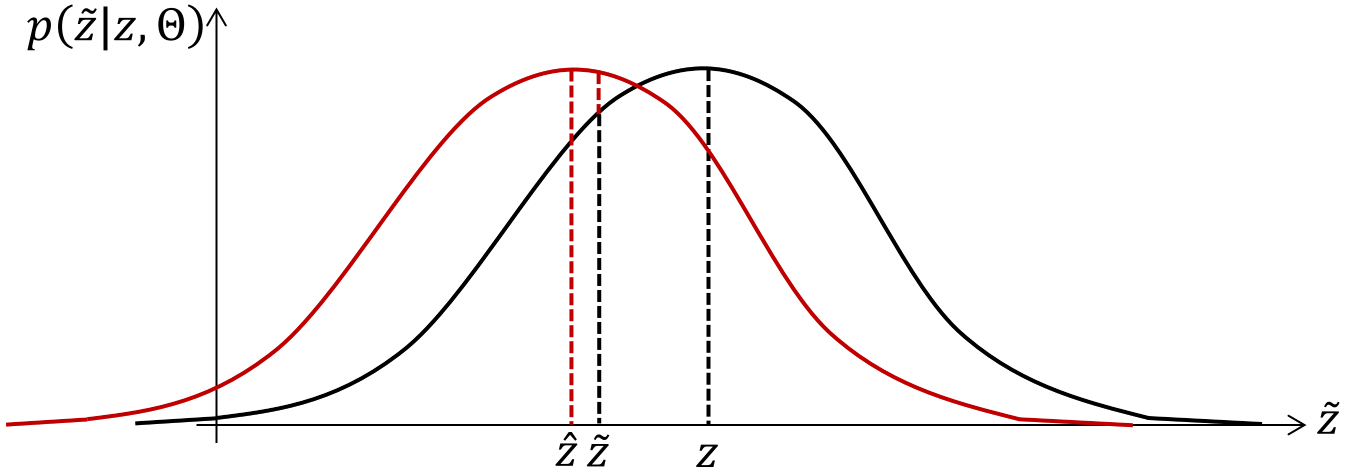
Well behaved: $g_i, \frac{\partial g_i}{\partial z_j}, \frac{\partial g_i}{\partial \Theta_j}$ are continuous, $\frac{\partial^2 g_i}{\partial z_j \partial \Theta_k}$ exist and are bounded

rank of Jacobian matrices $\mathrm{rank}(\nabla_z g(z, \Theta)) = s$ and $\mathrm{rank}(\nabla_\Theta g(z, \Theta)) = r$

The measurements $\tilde{z}$ of z contain random experimental errors $\epsilon_z \sim N(0, \Sigma^2)$:

$$\tilde{z} = z + \epsilon_z$$

Error-in-Variables Estimation (3)



Since $\epsilon_z \sim N(0,R)$, the probability density function $p(\tilde{z}|z,\Theta)$ for the measurement vector becomes:

$$\boxed{p(\tilde{z}|z,\Theta) = (2\pi)^{-\frac{q}{2}}|\Sigma|^{-1}\, e^{-\frac{1}{2}(\tilde{z}-z)^T\Sigma^{-2}(\tilde{z}-z)}}$$

Goal: Find the "best" estimate $(\hat{z},\widehat{\Theta})$ of true (z,Θ) for which holds that $g(\hat{z},\widehat{\Theta}) = 0$.

The measurement $\tilde{z}$ is the only information we have.

⇒ Idea: "for the "best" estimate $(\hat{z},\widehat{\Theta})$ it must hold that **the probability to get $\tilde{z}$ as measurement is maximal**."

⇒ maximum likelihood estimate

Maximum Likelihood Estimate (1)

Probability density function

$$p(\tilde{z}|z,\Theta) = (2\pi)^{-\frac{q}{2}}|\Sigma|^{-1}\, e^{-\frac{1}{2}(\tilde{z}-z)^T\Sigma^{-2}(\tilde{z}-z)}$$

The **likelihood function** for the measurement $\tilde{z}$ is defined as

$$L(\bar{z},\bar{\Theta}) = p(\tilde{z}|\bar{z},\bar{\Theta}) = (2\pi)^{-\frac{q}{2}}|\Sigma|^{-1}\, e^{-\frac{1}{2}(\tilde{z}-\bar{z})^T\Sigma^{-2}(\tilde{z}-\bar{z})}$$

where $(\bar{z},\bar{\Theta})$ have to fulfill the equations $g(\bar{z},\bar{\Theta}) = 0$.

The **maximum likelihood estimate**

$$\begin{aligned}(\hat{z},\hat{\Theta})_{ML} &= \arg\max_{(\bar{z},\bar{\Theta})} L(\bar{z},\bar{\Theta}) \\ &= \arg\min_{(\bar{z},\bar{\Theta})}\left((\tilde{z}-\bar{z})^T\Sigma^{-2}(\tilde{z}-\bar{z})\right) \\ s.t.\quad & g(\bar{z},\bar{\Theta}) = 0\end{aligned}$$

⇒This is the “constrained” WLS!

Aachener Verfahrenstechnik
RWTH AACHEN UNIVERSITY

Maximum Likelihood Estimate (2)

- If the equation vector $g(\bar{z}, \overline{\Theta}) = 0$ has a **unique solution** $\bar{z}$ for each $\overline{\Theta}$, the constrained minimization problem reduces to an **unconstrained** one with respect to $\overline{\Theta}$:

$$\left(\hat{z}, \widehat{\Theta}\right)_{ML} = \arg\min_{\overline{\Theta}}\left((\tilde{z} - \bar{z})^T \Sigma^{-2} (\tilde{z} - \bar{z})\right)$$

Note: Solving the minimization problem is an iterative process, since $\bar{z}$ has to be calculated from the implicit relation $g(\bar{z}, \overline{\Theta}) = 0$!

- If in addition $\bar{z}$ can be written as a explicit function of $\overline{\Theta}$, i.e. $\bar{z} = \bar{z}(\overline{\Theta})$, the estimation problem reduces to the standard nonlinear parameter estimation problem:

$$\widehat{\Theta} = \arg\min_{\overline{\Theta}}\left(\left(\tilde{z} - \bar{z}(\overline{\Theta})\right)^T \Sigma^{-2} \left(\tilde{z} - \bar{z}(\overline{\Theta})\right)\right)$$

Lecture Outline

- What is a Parameter Estimation Problem?
- Linear Parameter Estimation
  - Least-Squares
  - Weighted Least-Squares
- Nonlinear Parameter Estimation
- Error-in-Variables Estimation
- Confidence Region
- Case Study: Challenges in Spectral Analysis
- Data Reduction Using Principal Component Analysis

Quality of Parameter Estimation

model structure + model parameters = **model**

Can we evaluate the „quality" of the model?

Some quality assessment criteria (… check more than one):

Parameter covariance matrix: propagation of measurement noise into parameter uncertainty data on estimates

⇒ no output errors ...

Confidence regions: the region in parameter space containing true parameters with a certain probability

⇒ visualization of parametric uncertainty ...

Goodness-of-fit test: prediction capabilities of the model with respect to measured outputs

⇒ difficult to verify test assumptions...

Confidence Region (1)

Consider a model

$$y_i = f(u_i, \Theta),$$

where the model parameter Θ are unknown, but the model structure f is known.

For parameter estimation there are m noisy measurement available:

$$\tilde{y}_i = y_i + \epsilon_i, \qquad i = 1, \dots, m, \qquad \epsilon \sim N(0, \Sigma^2)$$

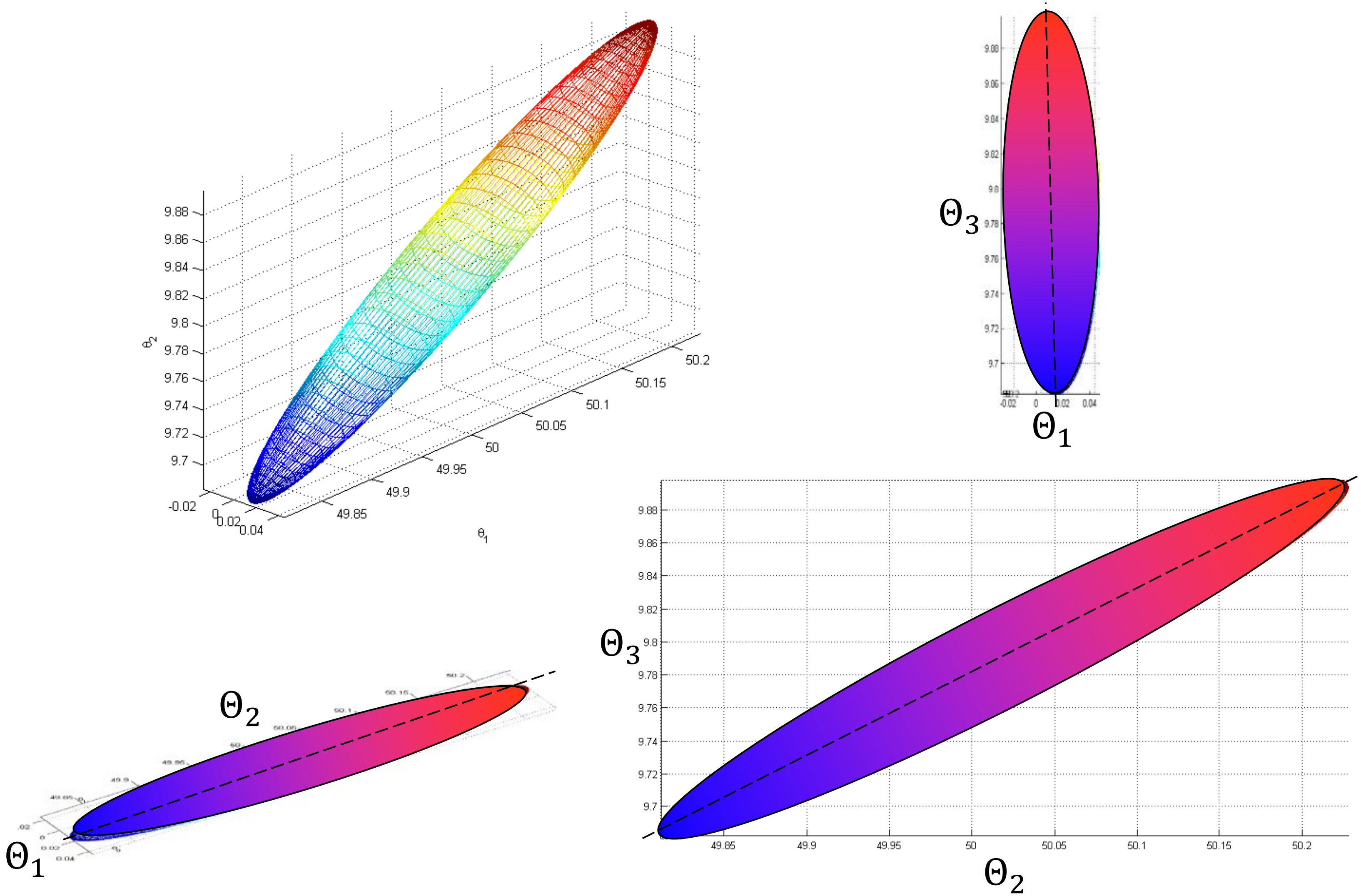
Goal: Instead of just determine a point estimate $\widehat{\Theta}$, we also want to determine a region, where the real parameter is situated with a given probability $(1-\alpha)$.

$\Rightarrow$ these regions are called **confidence regions Ω**

Let $p(\Theta|\tilde{y})$ be the probability density function of Θ, based on the measurement vector $\tilde{y}$, then for Ω it must hold:

$$\int_\Omega p(\Theta|\tilde{y}) d\Theta = 1 - \alpha$$

Confidence Region (2)



Approximation of Confidence Region (1)

In the parameter estimation case, the probability density function is

$$p(\Theta|\tilde{y}) = p(\tilde{y}|\Theta) = (2\pi)^{-\frac{m}{2}}|\Sigma|^{-1}\, e^{-\Phi(u,\Theta)},$$

where $\Phi(u,\Theta) = \frac{1}{2}(\tilde{y}-y)^T\Sigma^{-2}(\tilde{y}-y)$ and $y_i = f(u_i,\Theta)$.

The ML estimate is: $\widehat{\Theta}_{ML} = \arg\max_{\Theta} p(\tilde{y}|\Theta) = \arg\min_{\Theta} \Phi(u,\Theta)$

$J_\Theta(u,\Theta) := \nabla_\Theta \Phi(u,\Theta)$
$H_\Theta(u,\Theta) := \nabla_\Theta^2 \Phi(u,\Theta)$

Taylor series expansion of $\Phi(u,\Theta)$ at fixed inputs u and $\widehat{\Theta}_{ML}$:

$$\Phi(u,\Theta) = \Phi\left(u,\widehat{\Theta}_{ML}\right) + J_\Theta\left(u,\widehat{\Theta}_{ML}\right)\left(\Theta-\widehat{\Theta}_{ML}\right) + \frac{\left(\Theta-\widehat{\Theta}_{ML}\right)^T H_\Theta\left(u,\widehat{\Theta}_{ML}\right)\left(\Theta-\widehat{\Theta}_{ML}\right)}{2} + \cdots$$

since $J_\Theta\left(u,\widehat{\Theta}_{ML}\right) = 0$ according to necessary conditions of optimality and we assume to analyze a **neighborhood of $\widehat{\Theta}_{ML}$**, where 3^{rd} **and higher order derivatives are negligible**.

$$\boxed{\Phi(u,\Theta) \approx \Phi\left(u,\widehat{\Theta}_{ML}\right) + \frac{\left(\Theta-\widehat{\Theta}_{ML}\right)^T H_\Theta\left(u,\widehat{\Theta}_{ML}\right)\left(\Theta-\widehat{\Theta}_{ML}\right)}{2}}$$

Approximation of Confidence Region (2)

Then the probability density function can be approximated by

$$p(\Theta|\tilde{y}) \approx (2\pi)^{-\frac{m}{2}}|\Sigma|^{-1}\, e^{-\left(\Phi(u,\widehat{\Theta}_{ML})+\frac{\Delta\Theta^T H_\Theta(u,\widehat{\Theta}_{ML})\Delta\Theta}{2}\right)}$$

where $\Delta\Theta \coloneqq \left(\Theta - \widehat{\Theta}_{ML}\right)$.

The **smallest region Ω**, which fulfills

$$\int_\Omega p(\Theta|\tilde{y})d\Theta = 1 - \alpha$$

is now an **ellipsoid with center in $\widehat{\Theta}_{ML}$**:

$$\Omega = \left\{\Theta \in \mathbb{R}^r \,|\, \Delta\Theta^T H_\Theta\left(u, \widehat{\Theta}_{ML}\right)\Delta\Theta \leq \beta\right\}$$

where β **is a function dependent on α and the covariance Σ^2**.

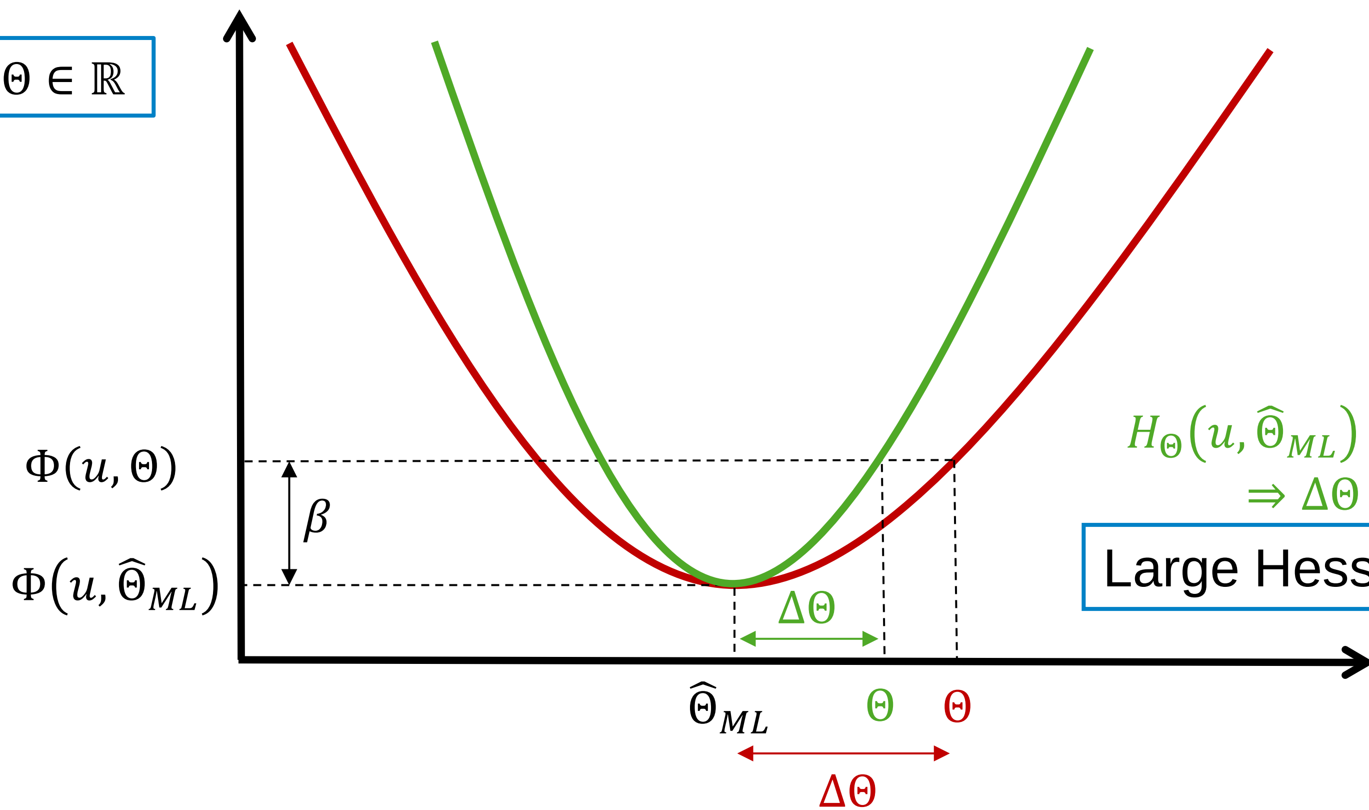
Parameter Precision (1)

The Θ on the border of the confidence ellipse fulfills the equation

$$\boxed{\Delta\Theta^T H_\Theta(u, \widehat{\Theta}_{ML})\Delta\Theta = \Phi(u, \Theta) - \Phi(u, \widehat{\Theta}_{ML}) = \beta}$$

$H_\Theta(u, \widehat{\Theta}_{ML})$ determines the curvature of $\Phi(u, \widehat{\Theta}_{ML})$ at $\widehat{\Theta}_{ML}$:

$\boxed{\Theta \in \mathbb{R}}$



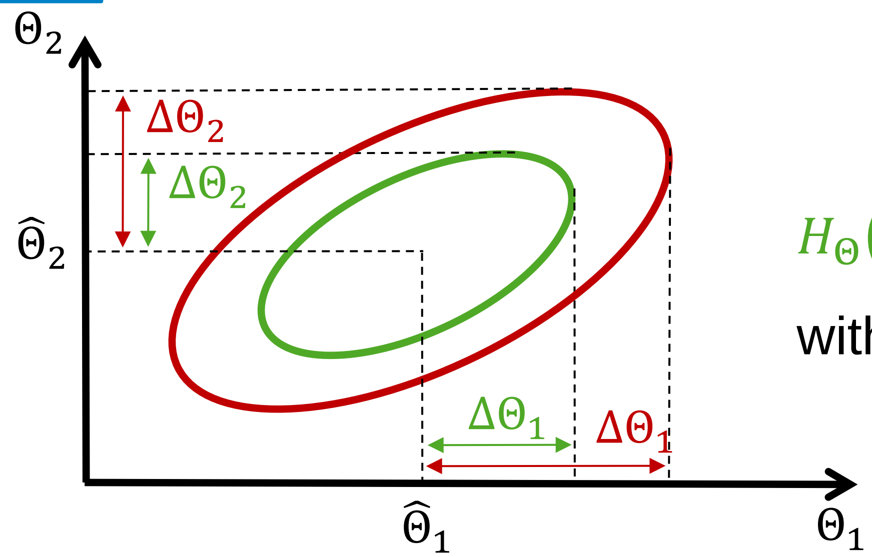
$H_\Theta(u, \widehat{\Theta}_{ML}) > H_\Theta(u, \widehat{\Theta}_{ML})$

$\Rightarrow \Delta\Theta < \Delta\Theta$

Large Hessian is desirable

Parameter Precision (2)

$\Theta \in \mathbb{R}^2$



$$\Delta\Theta^T H_\Theta(u, \widehat{\Theta}_{ML})\Delta\Theta = \beta$$

$$H_\Theta(u, \widehat{\Theta}_{ML}) = \gamma H_\Theta(u, \widehat{\Theta}_{ML})$$

with $\gamma > 1$

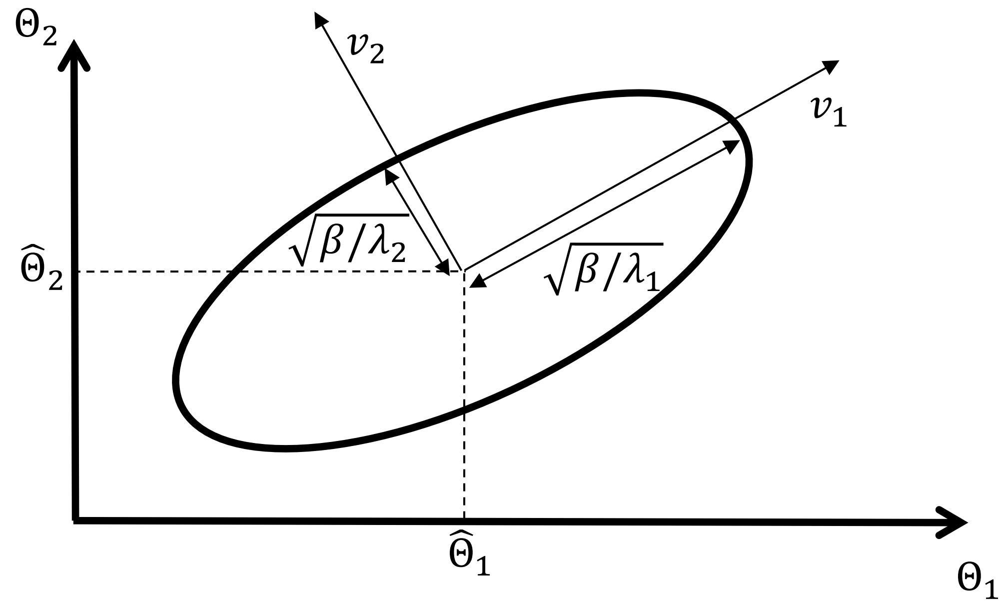
$$\Rightarrow \Delta\Theta_{1,2} < \Delta\Theta_{1,2}$$

- $H_\Theta(u, \widehat{\Theta}_{ML})$ and β determine $\Delta\Theta$: for given β, a larger Hessian results in smaller $\Delta\Theta$
- High parameter precision requires a small $\Delta\Theta$ and thus a small confidence region.

Confidence Region, Eigenvalues and Eigenvectors

Eigenvalue decomposition

$H_\Theta(u, \hat{\Theta}_{ML}) = V \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_r \end{bmatrix} V^{-1}$, where the i^{th} column of V is the eigenvector v_i



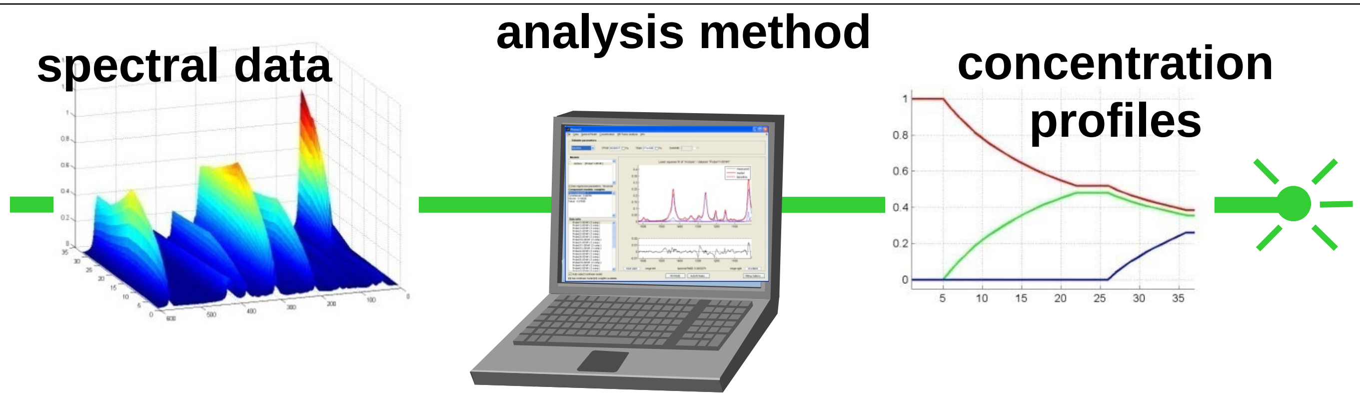
- Axis i of the confidence ellipsoid is characterized by
  - orientation → eigenvector v_i
  - length → eigenvalue λ_i and β

- For a given β, the extension of the confidence region can be measured by the eigenvalues λ_i of $H_\Theta(u, \hat{\Theta}_{ML})$.

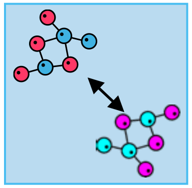
Lecture Outline

- What is a Parameter Estimation Problem?
- Linear Parameter Estimation
  - Least-Squares
  - Weighted Least-Squares
- Nonlinear Parameter Estimation
- Error-in-Variables Estimation
- Confidence Region
- Case Study: Challenges in Spectral Analysis
- Data Reduction Using Principal Component Analysis

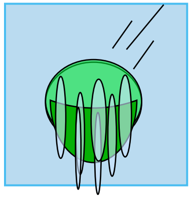
Case Study - Challenges in Spectral Analysis



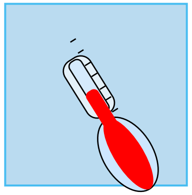
Challenges for **calibration** models:



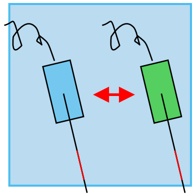
Molecular interaction



Reactive Mixtures



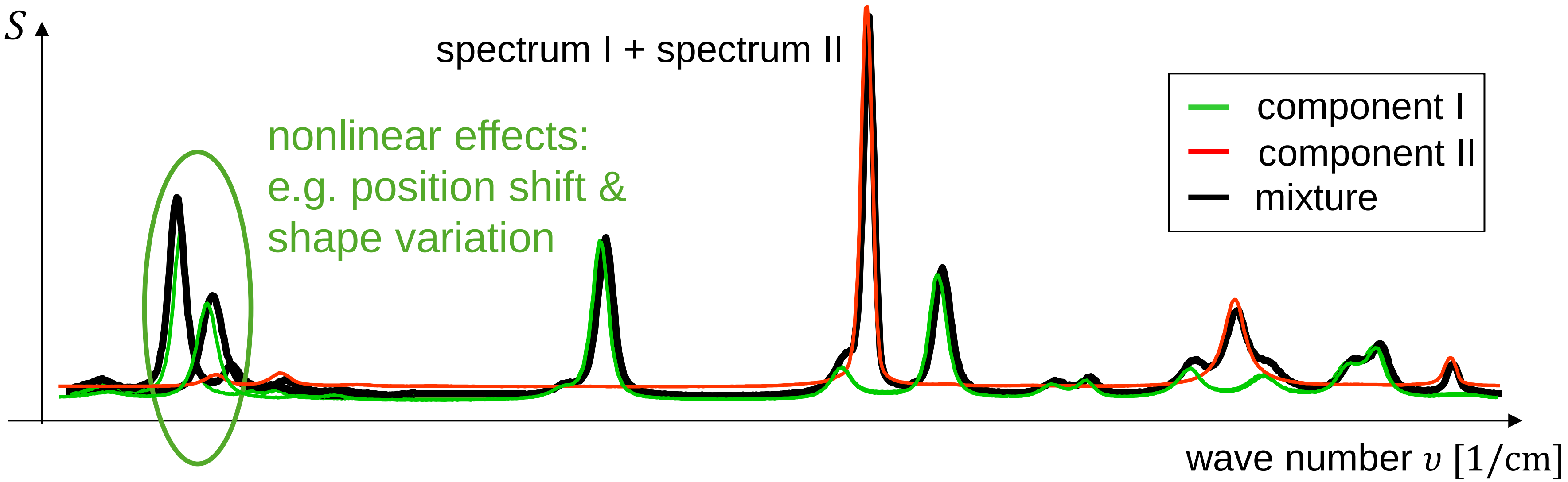
Temperature changes



Variable properties of measurement devices

Nonlinear spectral effects

Single- and Multi-Component Mixture Spectra



Raman spectrum – the Raman light scattering intensity in arbitrary units as a function of the wave number.

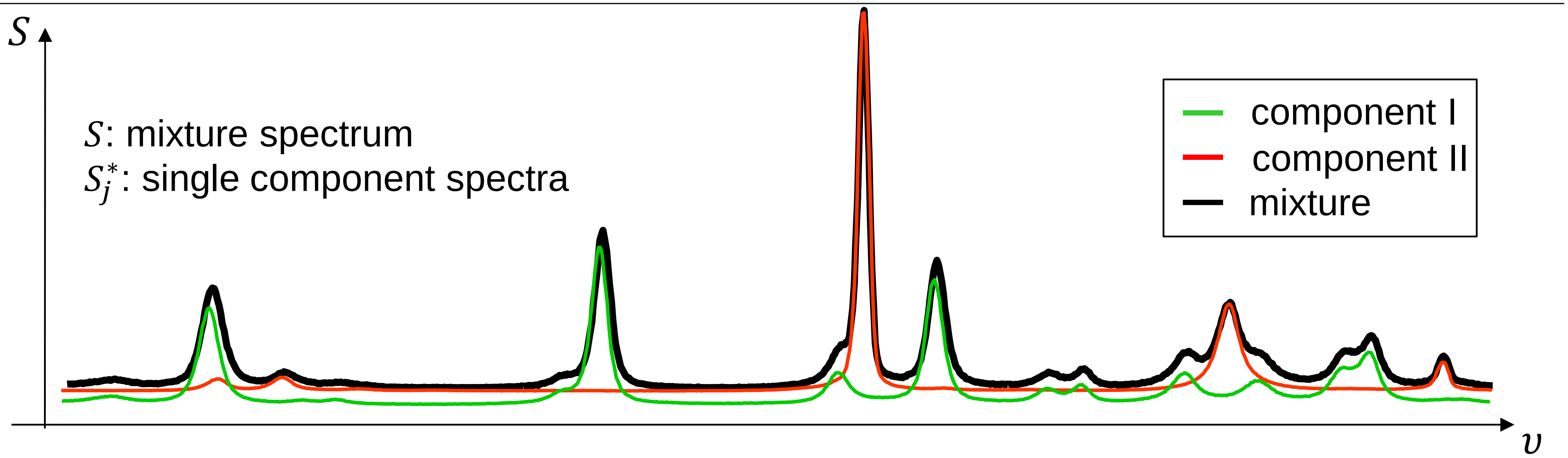
Pure component spectrum j:

$\Rightarrow S_j^*(\upsilon, \Theta)$ is nonlinear in parameters Θ

***N*-component mixture spectrum:**

$\Rightarrow S(\upsilon) = \sum_{j=1}^{n} \alpha_i S_j^*(\upsilon, \Theta)$ is linear combination of nonlinear function

Case Study: Spectral Analysis



Goal: Be able to infer from measurements of the mixture spectrum S the concentrations x_j of the components j in the mixture.

⇒ **Calibrate the model:** Use reference measurements of S, where the concentrations x_j are known, to estimate the model parameter

Case Study: Univariate Calibration – First Method

- Measure n pure component spectra S_j^* and one mixture spectrum S.
- Assume linear superposition (ideal mixing according to Lambert-Beer law)
- Weights α_j could be estimated from spectra S_j^* and S by WLS:

$$\min_{\alpha}\left(\left\|S(v)-\sum_{j=1}^{n}\alpha_j S_j^*(v)\right\|_2\right)$$

Objective: Determine the relationship between the fitted weights α_j and mixture concentrations $x_j,\ j=1,\dots,n$

Assumption (Lambert-Beer):

All $\frac{x_j}{x_i}$ depend linearly on $\frac{\alpha_j}{\alpha_i}$ i.e. $\frac{\alpha_j}{\alpha_i}K_{j,i}^{cali}=\frac{x_j}{x_i},\ j,i\in\{1,\dots,n\}$

The parameter $K_{j,i}^{cali}$ can now be used to determine the concentrations x_j of an unknown mixtures out of a spectrum $S(v)$:

1. Solve the optimization problem for the measured spectrum to get the weights α_j
2. Use the equations $\frac{\alpha_j}{\alpha_i}K_{j,i}^{cali}=\frac{x_j}{x_i}$ and $\sum_{j=1}^{n}x_j=1$ to get the concentrations x_j

Case Study: Multivariate Calibration – Second Method

- In reality, it is not possible to measure continuous spectra $S(v)$

$\Rightarrow$ measure the intensity of a mixture spectrum $S(v_i)$ at m discrete wave numbers v_i

Assumption: The concentrations x_j can be written as a linear combination of the intensities of the spectrum at the m discrete wave numbers, i.e.

$$x_j = \sum_{k=1}^{m} b_{k,j} S(v_k)$$

- For calibration l mixture spectra $S_k, k = 1, \dots, l$ are measured, where the concentration $x_{j,k}$ are known for all components $j = 1, \dots, n$

wavenumbers → components →

measurements ↓

$$\underbrace{\begin{bmatrix} S_1(v_1) & \cdots & S_1(v_m) \\ \vdots & & \vdots \\ S_l(v_1) & \cdots & S_l(v_m) \end{bmatrix}}_{=:X} \underbrace{\begin{bmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & & \vdots \\ b_{m,1} & \cdots & b_{m,n} \end{bmatrix}}_{=:\Theta} = \underbrace{\begin{bmatrix} x_{1,1} & \cdots & x_{n,1} \\ \vdots & & \vdots \\ x_{1,l} & \cdots & x_{n,l} \end{bmatrix}}_{=:Y}$$

measurements ↓

Solve the minimization problem $\min_{\Theta}(\|Y - X\Theta\|_2)$

Case Study: Multivariate Calibration – Second Method

Solve

$$\min_{\Theta}(\|Y - X\Theta\|_2)$$

$$\Rightarrow \boxed{\Theta = \left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1} X^T Y}$$

Problems:

- X is a large matrix with $l \geq m$ (l number of measurements, m number of wavenumbers)
- Strong collinearity in X, because only n components influence the spectra

$\Rightarrow$ **Condition number** of X is **very large**

$\Rightarrow$ **ill-posed problem**

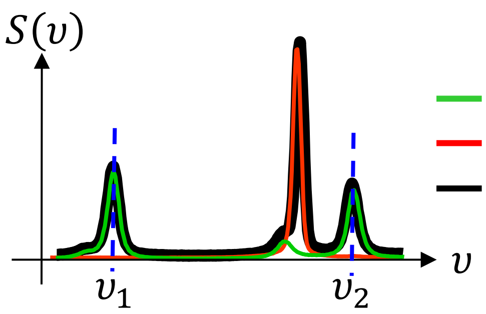
Idea: Reduce the dimension of the matrix by a skillful selection of a new coordinate system.

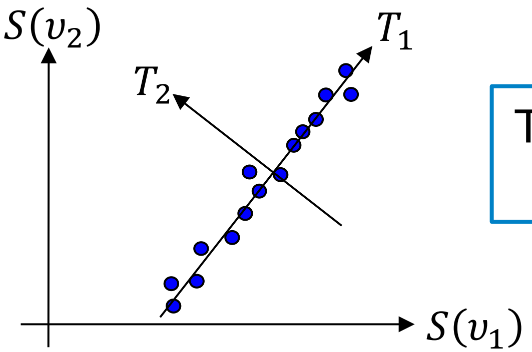
Lecture Outline

- What is an Parameter Estimation Problem?
- Linear Parameter Estimation
  - Least-Squares
  - Weighted Least-Squares
- Nonlinear Parameter Estimation
- Error-in-Variables Estimation
- Confidence Region
- Case Study: Challenges in Spectral Analysis
- Data Reduction Using Principal Component Analysis

Data Reduction with Principle Component Analysis

Consider two component mixture spectrum with $\mathrm{rank}(\Theta) = 2$ and measurements $S_k(v_i), i = 1{,}2$ and $k = 1, \dots, l$:





The data indicates a strong correlation

Identify a new coordinate system alongside the largest variants in the data

$\Rightarrow$ **Principle Component Analysis (PCA)**

Principle component (PC) are **right singular vectors** of the matrix X.

SVD of X:

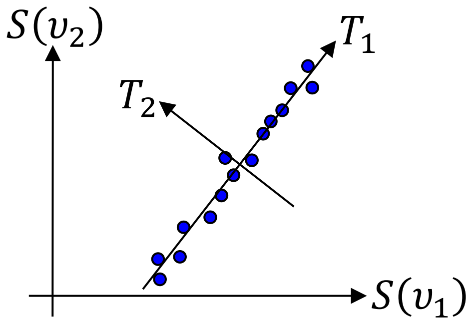
$$X = \underbrace{U\mathrm{diag}(\sigma)}_{=:T} V^T = TV^T$$

where $T^T T = diag(\sigma) U^T U diag(\sigma) = \mathrm{diag}(\sigma^2)$ and $V^T V = I$

Note: The singular values tell us, how much information is contained in a PC

Illustration of Principle Component Analysis

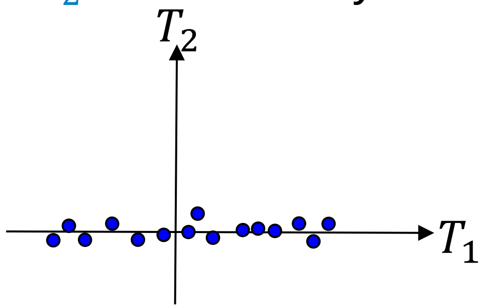
- Select those PC which correspond to large singular values, cf. truncated SVD (Reusken, section 3.2): $X_{red} = T_{red} V_{red}^T$



$$X = TV^T = \begin{bmatrix} T_{11} & T_{12} \\ \vdots & \vdots \\ T_{l1} & T_{l2} \end{bmatrix} \begin{bmatrix} V_{11} & V_{21} \\ V_{12} & V_{22} \end{bmatrix} \begin{array}{l} \text{PC1} \\ \text{PC2} \end{array}$$

T_1 contains the main information

T_2 contains only measurement signal noise and can be neglected



$$\mathrm{X} \approx X_{red} = T_{red} V_{red}^T = \begin{bmatrix} T_{11} \\ \vdots \\ T_{l1} \end{bmatrix} \begin{bmatrix} V_{11} & V_{21} \end{bmatrix}$$

Binary spectral data with two correlated intensities is reduced to **one dimension**. In general, spectral matrices can be represented in $(n-1)$-**dimensional spaces**!

Calibration by Principle Component Regression (PCR)

Idea: Identify the model parameters Θ in the reduced space X_{red}

$$\min_{\Theta}(\|Y - X_{red}\Theta_{red}\|_2)$$

$$\Rightarrow \quad \Theta_{red} = \left(X_{red}^T X_{red}\right)^{-1} X_{red}^T Y$$

$$= V_{red}\left(T_{red}^T T_{red}\right)^{-1} T_{red}^T Y$$

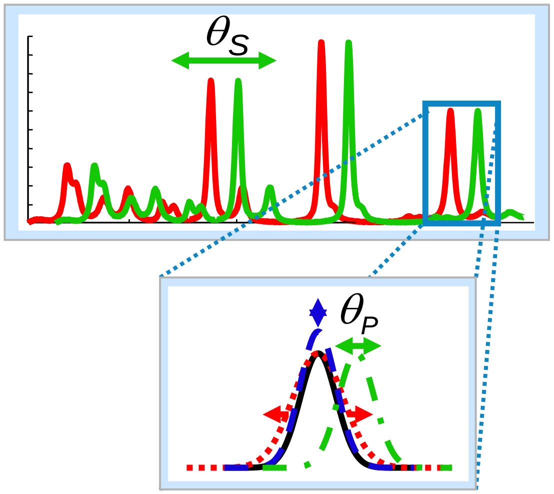
Since $X_{red} \approx X$, the identified model parameters θ_{red} are a good approximation of Θ.

Advantage:

- $T_{red} \in \mathbb{R}^{l\times(n-1)}$, where X and $X_{red} \in \mathbb{R}^{l\times m}$ with $m \geq n$
- $\mathrm{cond}(T_{red}) = \mathrm{cond}(X_{red}) < \mathrm{cond}(X)$, because the smallest singular values of X are not taken into account

Capturing Nonlinearity – Indirect Hard Modeling

PCR requires a linear spectral model, but **nature is nonlinear**

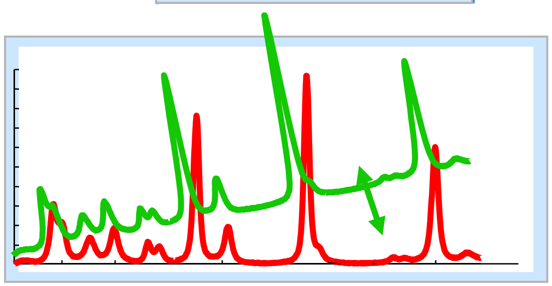


Spectral shifts

parameters $\Theta_{S,j}$

Peak variations

parameters $\Theta_{P,j}$
(position, intensity, width, shape)



Baseline effects

parameters Θ_B

Full spectral model: $S(\upsilon,\alpha,\Theta) = B(\upsilon,\Theta_B) + \sum_{j=1}^{n} \alpha_j S_j^*(\upsilon,\Theta_{S,j},\Theta_{P,j})$

Spectral model is **nonlinear in Θ,**
where $\Theta := [\Theta_B^T, \Theta_S^T, \Theta_P^T]^T$

$\Rightarrow$ **Nonlinear parameter estimation**

Review questions

- Why is the parameter estimation an inverse problem?
- Explain Least Squares Method (LS) using an example!
- What is the main difference between LS and Weighted LS?
- In what sense and under which conditions is the WLS-estimate optimal?
- Explain the method of data reduction by PCA!
- Explain the „Constrained“ WLS!
- What is confidence region?