Introduction into Linear Regression

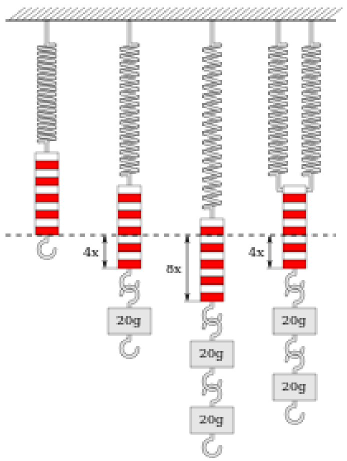
- Theory and experiment:
  How to confirm the congruence?

- Example:
  Elongation y [cm] of a spring subject to a force t [N]

  Hooke's law:

  $$y = y_0 + \frac{1}{D} t$$

  (D spring constant, describes stiffness of spring, y_0 pre-elongation of spring due to own weight)



Source: Wikipedia

- Are the free parameters D, y_0 constant over different experiments?

Introduction
Power series
Differentiation in Higher Dimensions
Limits and Continuity in $\mathbb{R}^n$
Differentiability in $\mathbb{R}^n$
Optimization in $\mathbb{R}^n$
Introduction into Linear Regression
Integration in Higher Dimensions
Further Topics in Calculus
Summary - Outlook and Review

Linear Regression - Intro

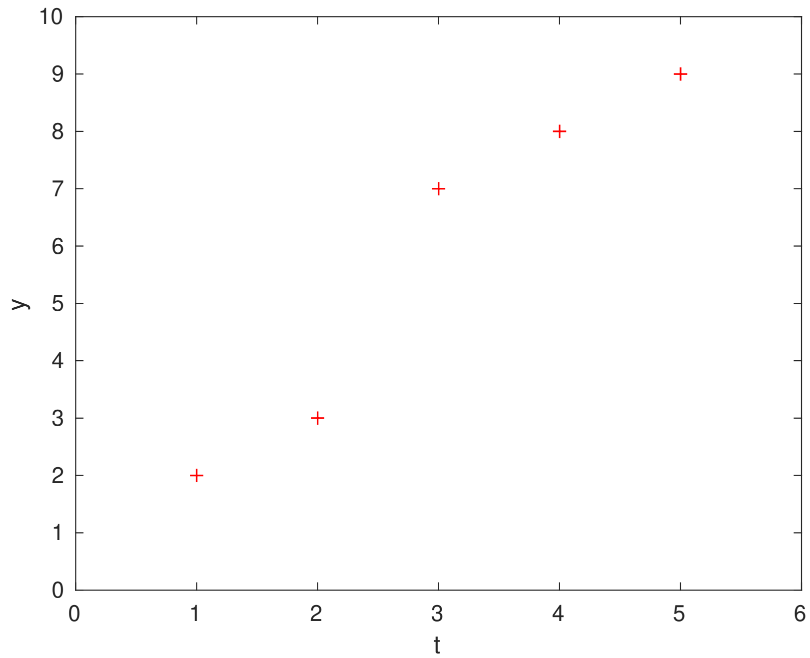
Introduction
Power series
Differentiation in Higher Dimensions
Limits and Continuity in $\mathbb{R}^n$
Differentiability in $\mathbb{R}^n$
Optimization in $\mathbb{R}^n$
Introduction into Linear Regression
Integration in Higher Dimensions
Further Topics in Calculus
Summary - Outlook and Review

Theory: $y(t) = x_1 + x_2\, t$

Data: (t_i, y_i)

Error (residuum): $r_i = y_i - y(t_i)$

i	t_i	y_i	$y(t_i)$	r_i	r_i^2
1	1	2			
2	2	3			
3	3	7			
4	4	8			
5	5	9			



Linear Regression - 1st Try

Introduction

Power series

Differentiation in Higher Dimensions

Limits and Continuity in $\mathbb{R}^n$

Differentiability in $\mathbb{R}^n$

Optimization in $\mathbb{R}^n$

Introduction into Linear Regression

Integration in Higher Dimensions

Further Topics in Calculus

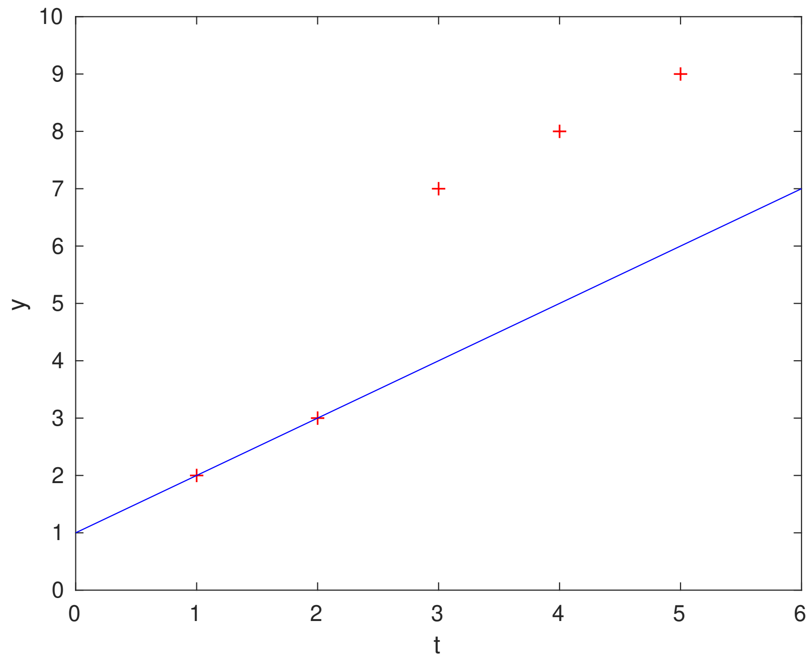
Summary - Outlook and Review

Theory: $y(t) = 1 + t$

Data: (t_i, y_i)

Error (residuum): $r_i = y_i - y(t_i) = y_i - 1 - t_i$

i	t_i	y_i	$y(t_i)$	r_i	r_i^2
1	1	2	2	0	0
2	2	3	3	0	0
3	3	7	4	3	9
4	4	8	5	3	9
5	5	9	6	3	9
				$\Sigma_{i=1}^5 r_i^2$	$= 27$



Linear Regression - 2nd Try

Introduction

Power series

Differentiation in Higher Dimensions

Limits and Continuity in $\mathbb{R}^n$

Differentiability in $\mathbb{R}^n$

Optimization in $\mathbb{R}^n$

Introduction into Linear Regression

Integration in Higher Dimensions

Further Topics in Calculus

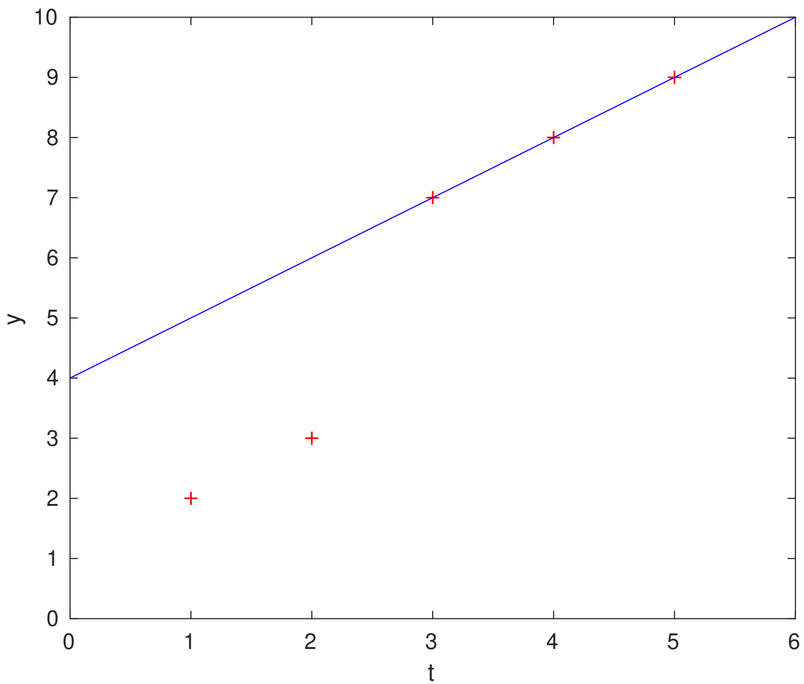
Summary - Outlook and Review

Theory: $y(t) = 4 + t$

Data: (t_i, y_i)

Error (residuum): $r_i = y_i - y(t_i) = y_i - 4 - t_i$

i	t_i	y_i	$y(t_i)$	r_i	r_i^2
1	1	2	5	−3	9
2	2	3	6	−3	9
3	3	7	7	0	0
4	4	8	8	0	0
5	5	9	9	0	0
				$\Sigma_{i=1}^5 r_i^2$	$= 18$



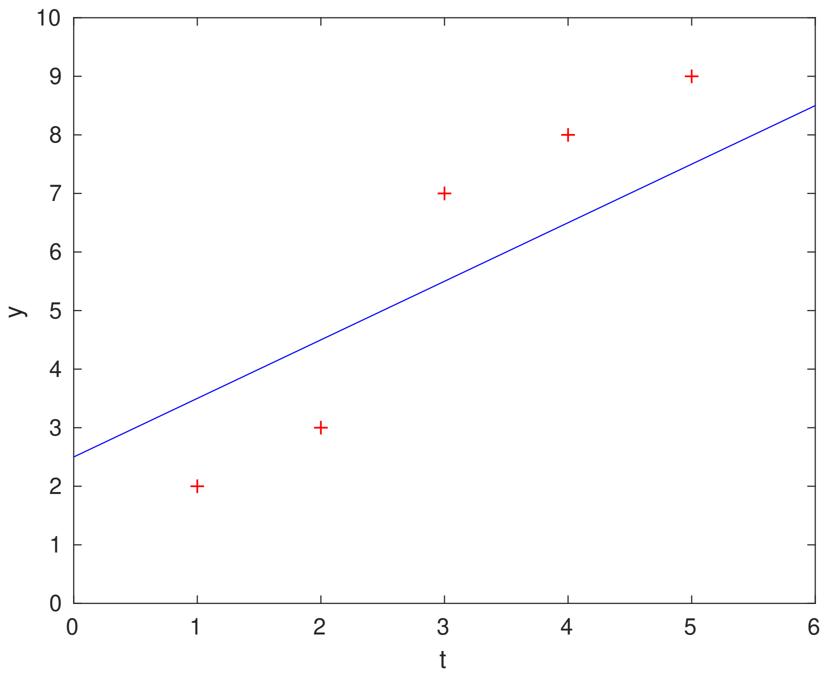
Linear Regression - 3rd Try

$$\text{Theory: } y(t) = 2.5 + t$$

$$\text{Data: } (t_i, y_i)$$

$$\text{Error (residuum): } r_i = y_i - y(t_i) = y_i - 2.5 - t_i$$

i	t_i	y_i	$y(t_i)$	r_i	r_i^2
1	1	2	3.5	-1.5	2.25
2	2	3	4.5	-1.5	2.25
3	3	7	5.5	1.5	2.25
4	4	8	6.5	1.5	2.25
5	5	9	7.5	1.5	2.25
				$\Sigma_{i=1}^5 r_i^2 = 11.25$	



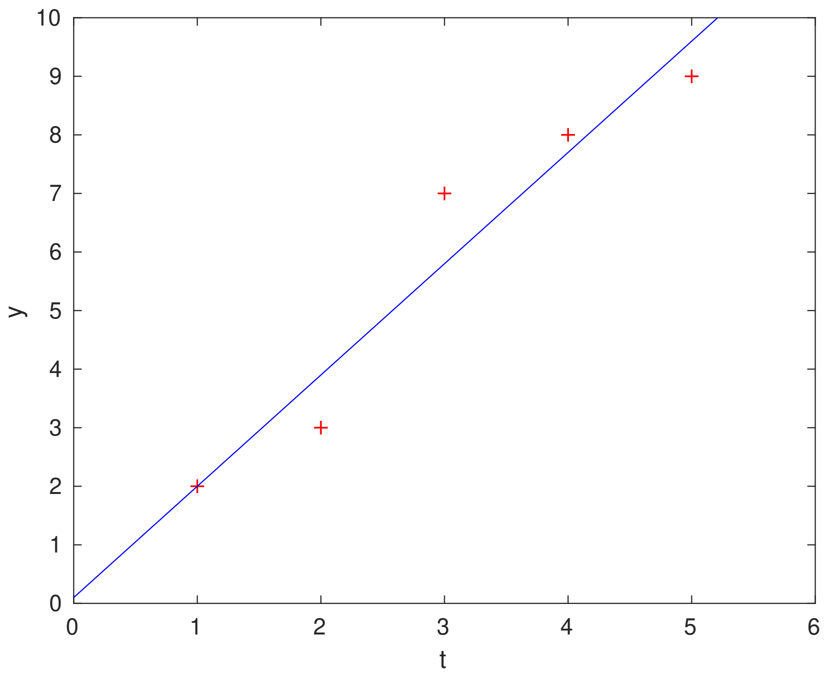
Linear Regression - Solved

$$\text{Theory: } y(t) = 0.1 + 1.9\,t$$

$$\text{Data: } (t_i, y_i)$$

$$\text{Error (residuum): } r_i = y_i - y(t_i) = y_i - 0.1 - 1.9\,t_i$$

i	t_i	y_i	$y(t_i)$	r_i	r_i^2
1	1	2	2	0	0
2	2	3	3.9	−0.9	0.81
3	3	7	5.8	1.2	1.44
4	4	8	7.7	0.3	0.09
5	5	9	9.6	−0.6	0.36
				$\Sigma_{i=1}^5 r_i^2 = 2.7$	



Linear Regression: Normal Equations

Introduction

Power series

Differentiation in Higher Dimensions

Limits and Continuity in $\mathbb{R}^n$

Differentiability in $\mathbb{R}^n$

Optimization in $\mathbb{R}^n$

Introduction into Linear Regression

Integration in Higher Dimensions

Further Topics in Calculus

Summary - Outlook and Review

In general (m arbitrary):

$$A^\top A = \begin{pmatrix} m & \sum_{i=1}^m t_i \\ \sum_{i=1}^m t_i & \sum_{i=1}^m t_i^2 \end{pmatrix}, \quad A^\top \mathbf{y} = \begin{pmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m t_i y_i \end{pmatrix}$$

$$x_1 = \frac{\sum_{i=1}^m t_i^2 \sum_{j=1}^m y_j - \sum_{i=1}^m t_i \sum_{j=1}^m t_j y_j}{m \sum_{i=1}^m t_i^2 - \left(\sum_{i=1}^m t_i\right)^2}$$

$$x_2 = \frac{m \sum_{i=1}^m t_i y_i - \sum_{i=1}^m t_i \sum_{j=1}^m y_j}{m \sum_{i=1}^m t_i^2 - \left(\sum_{i=1}^m t_i\right)^2}$$

Linear Regression: Motivation & Aims

Introduction

Power series

Differentiation in Higher Dimensions

Limits and Continuity in $\mathbb{R}^n$

Differentiability in $\mathbb{R}^n$

Optimization in $\mathbb{R}^n$

Introduction into Linear Regression

Integration in Higher Dimensions

Further Topics in Calculus

Summary - Outlook and Review

We conclude:

- In general more data points y_i (here 5) as parameters x_i (here 2)
- Measurement of process/data collection afflicted with uncertainties (w/o systematical errors)
- Overdetermined LES $\rightsquigarrow$ in general no solution exists

Aims:

- Determine $\mathbf{x} = (x_1, \ldots, x_n)^\top$ "optimally" from data
- General method to do that

We call this **regression** (in particular in stochastics) or **curve fitting**.

Special case of a mathematical optimization method

Linear Regression: Stochastic Uncertainties

Problem (Linear Regression)

Assume: linear relation (e.g. from physics)

$$y(t) = x_1 + x_2 t$$

Given: data points (t_i, y_i), $i = 1, \ldots, m$, afflicted with uncertainties (errors) ε_i. The uncertainties are random variables and 0 in average.

Searched for: $x_1, x_2 \in \mathbb{R}$, such that

$$\begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_i \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_m \end{pmatrix}$$

We write

$$A\mathbf{x} = \mathbf{y} + \boldsymbol{\varepsilon}$$

with $A \in \mathbb{R}^{m \times 2}$, $\mathbf{x} \in \mathbb{R}^2$, $\mathbf{y}, \boldsymbol{\varepsilon} \in \mathbb{R}^m$.

Minimization Problem

Introduction

Power series

Differentiation in Higher Dimensions

Limits and Continuity in $\mathbb{R}^n$

Differentiability in $\mathbb{R}^n$

Optimization in $\mathbb{R}^n$

Introduction into Linear Regression

Integration in Higher Dimensions

Further Topics in Calculus

Summary - Outlook and Review

General situation: $A \in \mathbb{R}^{m\times n}$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$

Idea: determine $\mathbf{x} = (x_1, \ldots, x_n)^\top$, such that the error in the LES

$$\|A\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum_{i=1}^{m}\left(\sum_{j=1}^{n} A_{ij}x_j - y_i\right)^2} \quad \text{bzw.} \quad \frac{1}{2}\,\|A\mathbf{x} - \mathbf{y}\|_2^2$$

is minimized.

The minimum is denoted by $\hat{\mathbf{x}}$.

This minimization problem is called a **linear regression problem** or **least-squares problem**.

Method of least squares, better **method of least residual sum of squares**
By C. F. Gauß and A.-M. Legendre (1805)

General problem

Introduction
Power series
Differentiation in Higher Dimensions
Limits and Continuity in $\mathbb{R}^n$
Differentiability in $\mathbb{R}^n$
Optimization in $\mathbb{R}^n$
Introduction into Linear Regression
Integration in Higher Dimensions
Further Topics in Calculus
Summary - Outlook and Review

Problem ($\star$) (Lin. regression as minimization problem)

Let be given a matrix $A \in \mathbb{R}^{m\times n}$ and a vector $\mathbf{y} \in \mathbb{R}^m$ with $m, n \in \mathbb{N}$.

We search for the solution $\hat{\mathbf{x}} \in \mathbb{R}^n$ of the minimization problem

$$\frac{1}{2}\|A\hat{\mathbf{x}} - \mathbf{y}\|_2^2 = \min_{x\in\mathbb{R}^n} \frac{1}{2}\|A\mathbf{x} - \mathbf{y}\|_2^2.$$

- If $m = n$ and A invertible, then $A\mathbf{x} = \mathbf{y}$ has a unique solution.
- The case $m > n$ is of uttermost importance in applications.
- In the case $m < n$ the LES $A\mathbf{x} = \mathbf{y}$ is underdetermined. It may be unsolvable, if $Rank(A) \neq Rank(A \mid \mathbf{y})$.

Moreover, other norms as $\|\cdot\|_2$ could be considered. Then the determination of solutions is harder in general, since differentiability might not be given.

Gaussian normal equations

Introduction
Power series
Differentiation in Higher Dimensions
Limits and Continuity in $\mathbb{R}^n$
Differentiability in $\mathbb{R}^n$
Optimization in $\mathbb{R}^n$
Introduction into Linear Regression
Integration in Higher Dimensions
Further Topics in Calculus
Summary - Outlook and Review

Theorem (Gaussian normal equations)

$\hat{x}$ solves Problem $(\star)$ if and only iff the **normal equations**

$$A^\top A \hat{\mathbf{x}} = A^\top \mathbf{y}.$$

hold true.

$A^\top A \in \mathbb{R}^{n \times n}$ is symmetric and positive semi-definite.

Theorem (Uniqueness)

Let $m \geq n$. Suppose $A \in \mathbb{R}^{m \times n}$ has maximal rank, d.h. $rank(A) = n$.
Exactly then the minimization problem $(\star)$ or the normal equations, resp., are uniquely solvable.

Then $A^\top A \in \mathbb{R}^{n \times n}$ is invertible and positive definite.

(Non-)Linear Regression: TTF under Temperature Stress

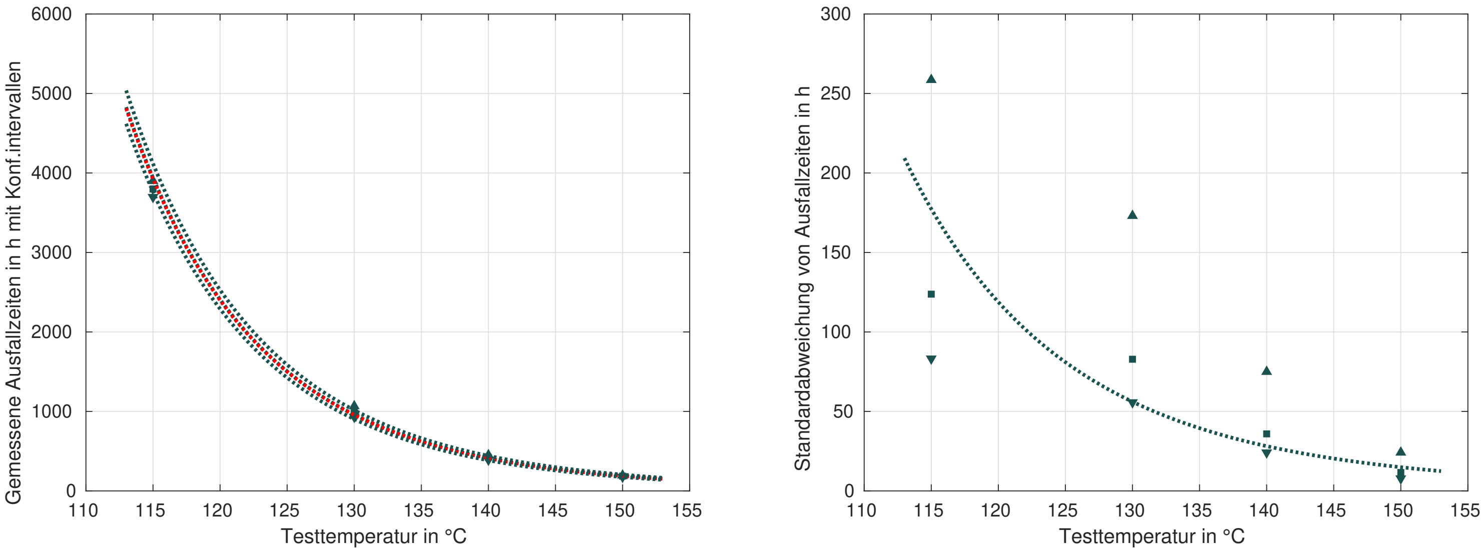


Figure: Curve fitting over different temperatures T compared with measured data points (squares) with confidence intervals for $q = 90\%$ (triangles). Left-hand side for $\mu_{krit}(T) \pm \sigma_{krit}(T)$, in red the fitted curve, right-hand side $\sigma_{krit}(T)$. [K., Dvorsky, Ließ, Avenhaus 2019]

Model based approach

$$f_{krit}(T) = t_\ominus + t_0 \exp\left(\left(\frac{T_a}{T - T_\infty}\right)^d\right)$$

Lin. Regression: Transformation

Introduction

Power series

Differentiation in Higher Dimensions

Limits and Continuity in $\mathbb{R}^n$

Differentiability in $\mathbb{R}^n$

Optimization in $\mathbb{R}^n$

Introduction into Linear Regression

Integration in Higher Dimensions

Further Topics in Calculus

Summary - Outlook and Review

Example (Measurement of TTF (Time-To-Failure) of Electrical Automotive Components)

T	[°C]	115	130	140	155
μ_{krit}	$[h]$	3791.62	987.74	439.66	189.94

Conjecture (model based): Arrhenius law

$$\mu_{krit}(T) = t_0 \exp\left(\frac{T_a}{T - T_\infty}\right)$$

(μ_{krit} TTF in h, t_0 reaction-kinetic period in h, T temperature in K, T_a activation temperature in K, $T_\infty = 173.15$ consolidation temperature in K)

We may transform this into an affine-linear relation

$$ln(\mu_{krit}(T)) = ln(t_0) + \frac{T_a}{T - T_\infty} \qquad \Longleftrightarrow \qquad y(t) = x_1 + x_2 t$$

By insertion of measured data we obtain a linear equation system (LES)

$$y_i = x_1 + x_2 t_i, \quad i = 1, \dots, 4.$$

Linear Regression: Example - Result

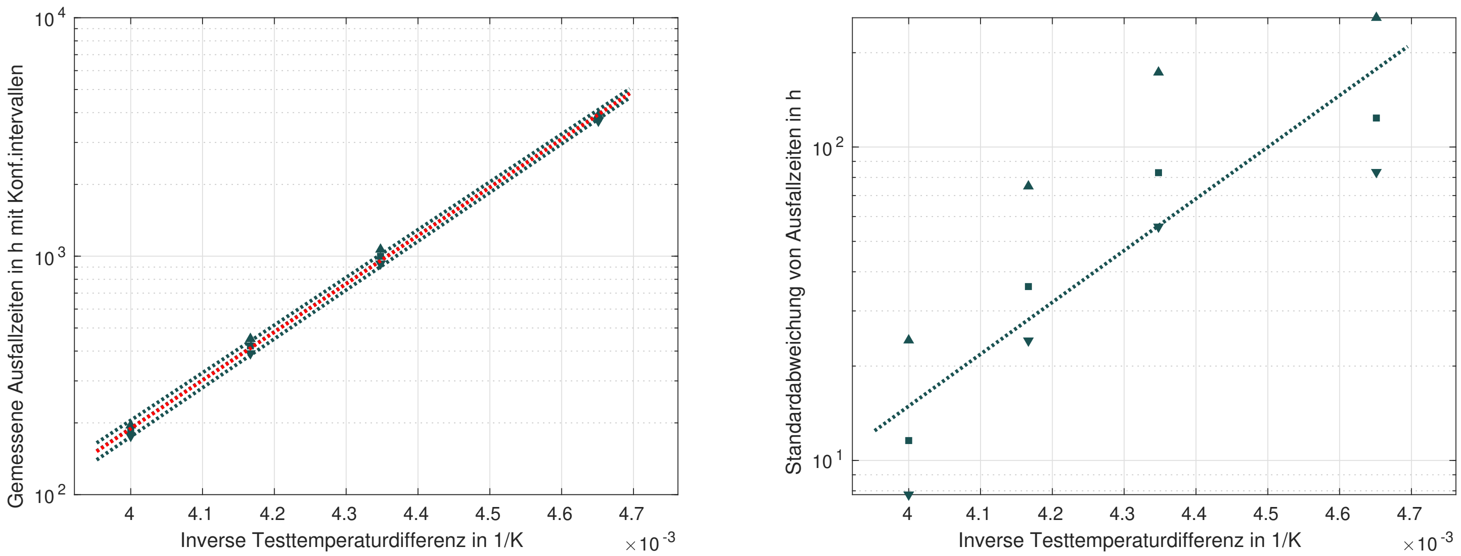


Figure: Curve fitting over different inverse temperature differences t, in semi-logarithmic representation and compared with measured data points (squares) together with confidence intervals for $q = 90\%$ (triangles). Left-hand side for $t_{krit}(1/t) \pm \sigma_{krit}(1/t)$, in red the fitted curve. Right-hand side $\sigma_{krit}(T)$. [K., Dvorsky, Ließ, Avenhaus 2019]

Introduction

Power series

Differentiation in Higher Dimensions

Limits and Continuity in $\mathbb{R}^n$

Differentiability in $\mathbb{R}^n$

Optimization in $\mathbb{R}^n$

Introduction into Linear Regression

Integration in Higher Dimensions

Further Topics in Calculus

Summary - Outlook and Review

Linear Regression: Example - Re-transformed Result

Introduction

Power series

Differentiation in Higher Dimensions

Limits and Continuity in $\mathbb{R}^n$

Differentiability in $\mathbb{R}^n$

Optimization in $\mathbb{R}^n$

Introduction into Linear Regression

Integration in Higher Dimensions

Further Topics in Calculus

Summary - Outlook and Review

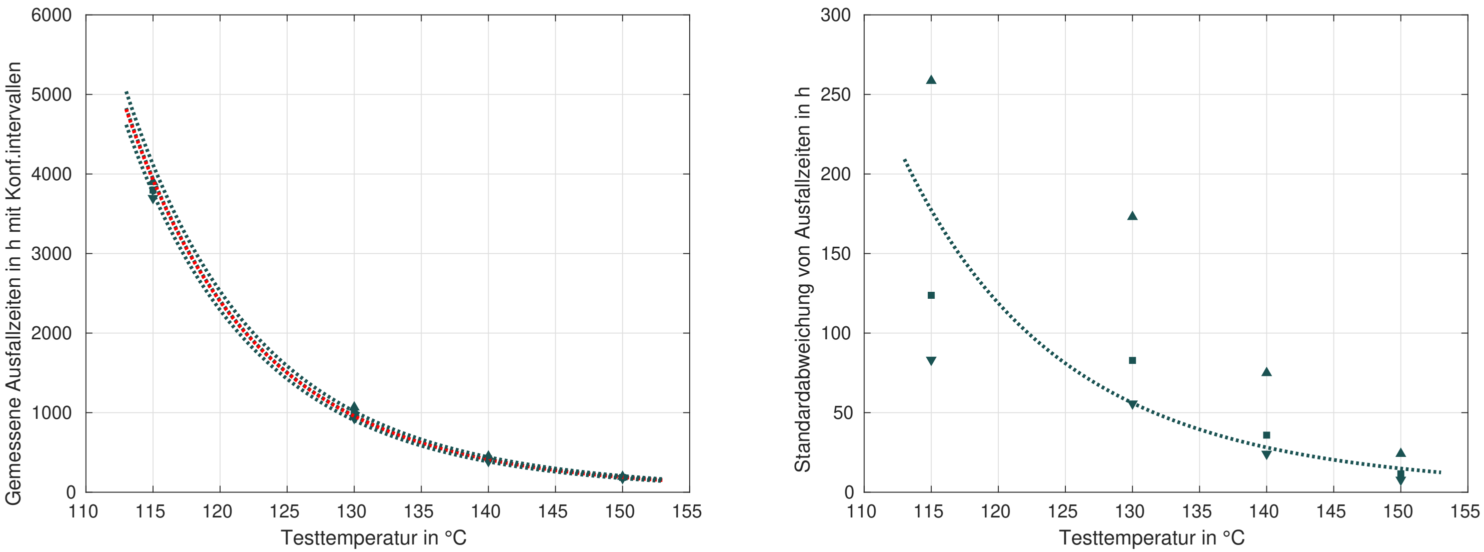


Figure: Curve fitting over different temperatures T, re-transformed in exponential representation and compared with measured data points (squares) together with confidence intervals for $q = 90\%$ (triangles). Left-hand side for $t_{krit}(T) \pm \sigma_{krit}(T)$, in red the fitted curve. Right-hand side $\sigma_{krit}(T)$. [K., Dvorsky, Ließ, Avenhaus 2019]