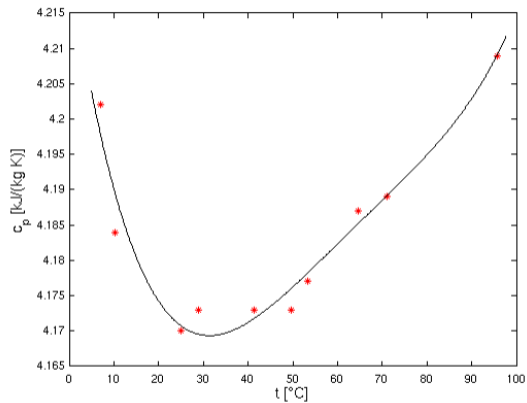


Kapitel 3

Ausgleichsrechnung



Inhalt

3 Ausgleichsrechnung

- Ausgleichsproblem
- Lineare Ausgleichsprobleme
- Orthogonalisierungsverfahren
- Modellauswahl

Ausgleichsproblem

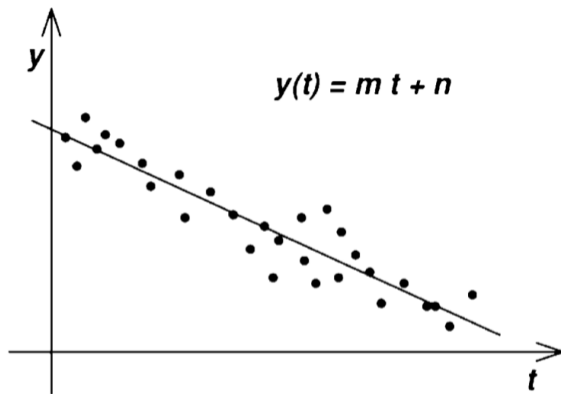
Gegeben:

- n Wertepaare (x_i, y_i) , $i = 1, \dots, n$ mit $x_i \neq x_j$ für $i \neq j$.

Gesucht:

- Eine stetige Funktion f , die in einem gewissen Sinne die Wertepaare bestmöglich annähert,
- d. h. es soll möglichst genau $f(x_i) \approx y_i$ für alle $i = 1, \dots, n$ gelten.

Beispiel: Messwerte durch eine Gerade annähern



Wie sollen wir m und n wählen?

Bemerkungen zum Ausgleichsproblem

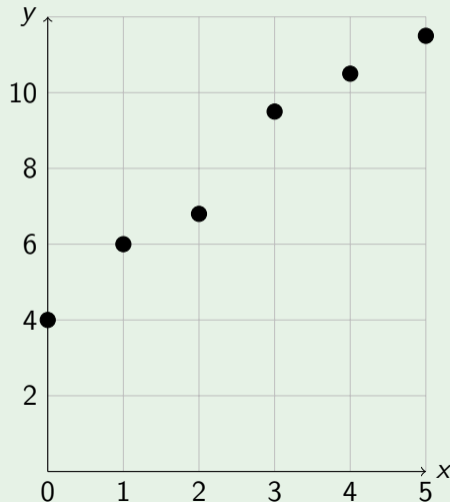
- Wenn wir die Menge der zulässigen stetigen Funktionen nicht weiter einschränken, können wir ein **Interpolationspolynom mit Grad $n - 1$** finden, das $f(x_i) = y_i$ für $i = 1, \dots, n$ erfüllt..
- Solche Interpolationspolynome sind aber **für die Praxis meist untauglich**.
- Die **Wertepaare sind oft empirischer Natur**, sie entstammen z. B. Messungen, die **fehlerbehaftet** sind.
- Daher ist die **Forderung einer exakten Anpassung nicht sinnvoll**.
- Weiteres Problem: Die **Anzahl der Parameter würde linear mit der Anzahl der Messungen wachsen**.
- Der tatsächliche funktionale Zusammenhang ist aber natürlich unabhängig von der Anzahl der Messungen.

Beispiel

Beispiel 3.1

Wir haben als Wertepaare ($n = 6$):

x_i	0	1	2	3	4	5
y_i	4	6	6.8	9.5	10.5	11.5



Ausgleichsfunktion, Fehlerfunktional

Definition 3.2

Gegeben sei:

- eine Menge \mathcal{F} von stetigen Funktionen auf $[a, b]$, also $\mathcal{F} \subseteq C([a, b])$,
- sowie n Wertepaare (x_i, y_i) , $i = 1, \dots, n$.

Eine Funktion $f \in \mathcal{F}$ heißt **Ausgleichsfunktion** von \mathcal{F} zu den gegebenen Wertepaaren, falls f das **Fehlerfunktional**

$$E(f) := \sum_{i=1}^n (y_i - f(x_i))^2$$

minimiert, d. h. $E(f) = \min\{E(g) | g \in \mathcal{F}\}$.

Die Menge \mathcal{F} heißt auch **Menge der Ansatzfunktionen**.

Bemerkungen

Typische Situation:

- Wir haben eine Anzahl fehlerbehafteter Messdaten vorliegen und versuchen nun,
- die Parameter eines Modells so anzupassen, dass das Modell die beobachteten Messdaten möglichst gut wiedergibt.
- Das Modell besteht meist aus einer parameterbehafteten Kombination von Funktionen.
- Üblicherweise haben wir viel mehr Messdaten als Modellparameter.

Fehlerfunktionale

- In Definition 3.2 haben wir das Fehlerfunktional $E(f)$ auf Basis der **euklidische Norm** $\| \cdot \|_2$ definiert. Prinzipiell könnten wir auch andere Normen verwenden.
- Für die **Summennorm** $\| \cdot \|_1$ würde sich

$$E(f) := \sum_{i=1}^n |y_i - f(x_i)|$$

als Fehlerfunktional ergeben.

- Auch möglich: **Gewichtung der Fehler**. Damit entsteht

$$E(f) := \sum_{i=1}^n w_i (y_i - f(x_i))^2$$

mit $w_i > 0$.

- Anwendung: Berücksichtigung der Qualität oder **Zuverlässigkeit einer Messung**.

- Das Fehlerfunktional auf Basis der Maximumsnorm $\| \cdot \|_{\infty}$ ist:

$$E(f) = \max_{i=1}^n |y_i - f(x_i)|$$

- Die Fehlerfunktionale unterscheiden sich insbesondere in ihrer **Empfindlichkeit gegenüber Ausreißern**. Dies können wir auch als **Robustheit** betrachten.
- Wir nutzen in der Regel das Fehlerfunktional basierend auf der Norm $\| \cdot \|_2$. Dies führt zu speziellen **linearen Gleichungssystemen**.
- Ausgleichsprobleme mit Fehlerfunktional basierend auf $\| \cdot \|_1$ oder $\| \cdot \|_{\infty}$ führen dagegen zu **linearen Programmen**.
☞ nächstes Kapitel

Ansatzfunktionen

- Im einfachsten Fall wählt man als \mathcal{F} die Menge aller Geraden, also

$$\mathcal{F} = \{f(x) = c x + d \mid c, d \in \mathbb{R}\}$$

Dies nennt man **lineare Regression**.

- In diesem Fall heißt die Ausgleichsfunktion **Ausgleichsgerade** oder auch **Regressionsgerade**.

Lagrangesche Interpolationsformel

Bereits erwähnt: Ein **Interpolationspolynom** könnte das Ausgleichsproblem exakt (mit Fehler 0) lösen.

Definition 3.3

Für $x_1 < x_2 < \dots < x_n$ heißt das Polynom

$$L_i(x) = \prod_{j=1, j \neq i}^n \frac{x - x_j}{x_i - x_j}$$

i -tes Lagrange-Polynom (zu den Stützstellen x_1, \dots, x_n).

Offensichtlich gilt:

$$L_i(x_j) = \begin{cases} 1 & \text{für } i = j \\ 0 & \text{für } i \neq j \end{cases}$$

Aus den Lagrange-Polynomen bilden wir nun ein geeignetes **Interpolationspolynom** $P(x)$ wie folgt:

$$P(x) = \sum_{j=1}^n y_j L_j(x)$$

Damit gilt dann:

$$\begin{aligned} P(x_i) &= \sum_{j=1}^n y_j L_j(x_i) \\ &= y_i \underbrace{L_i(x_i)}_{=1} \\ &= y_i \end{aligned}$$

Sowohl manuell als auch numerisch ist dieses Vorgehen aber **umständlich**.

Vandermonde-Matrix

- Wir modellieren

$$P(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \dots + a_0$$

und setzen die Stützstellen x_1, \dots, x_n ein.

- Dann entsteht das LGS:

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ 1 & x_3 & x_3^2 & \cdots & x_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{n-1} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$$

Definition 3.4

Für $x_1, x_2, \dots, x_n \in \mathbb{R}$ heißt die Matrix

$$V(x_1, \dots, x_n) = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ 1 & x_3 & x_3^2 & \cdots & x_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{pmatrix}$$

Vandermonde-Matrix.

Lemma 3.5

Es gilt:

$$\det(V(x_1, \dots, x_n)) = \prod_{1 \leq i < j \leq n} (x_j - x_i)$$

Folgerung 3.6

Für $x_1 < x_2 < \dots < x_n$ ist die Vandermonde-Matrix *regulär*.

Damit können wir unsere Methoden zur Lösung eines LGS einsetzen.

Beispiel 3.7

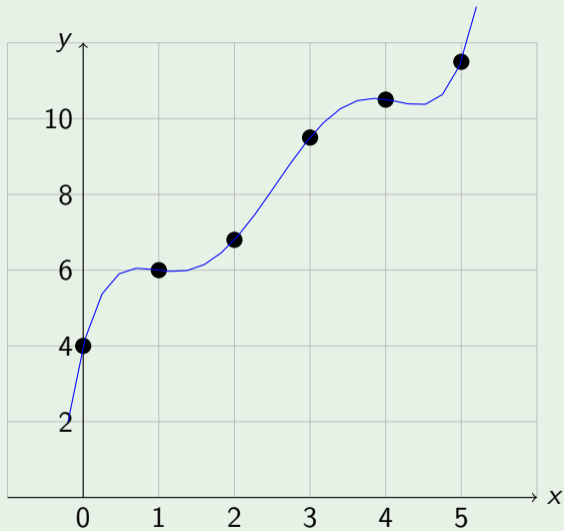
Für Beispiel 3.1 erhalten wir:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 & 16 & 32 \\ 1 & 3 & 9 & 27 & 81 & 243 \\ 1 & 4 & 16 & 64 & 256 & 1024 \\ 1 & 5 & 25 & 125 & 625 & 3125 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \\ 6.8 \\ 9.5 \\ 10.5 \\ 11.5 \end{pmatrix}$$

Es ergibt sich (Koeffizienten gerundet):

$$P(x) = 0.1x^5 - 1.28x^4 + 5.69x^3 - 10.22x^2 + 7.71x + 4$$

Fortsetzung Beispiel.



Lineare Regression

- Die Grafik von Beispiel 3.7 zeigt das **Problem der Polynominterpolation**: Die gefundene Funktion ist **zu wellig**,
- zumal die y_i üblicherweise fehlerbehaftet sind.
- Polynominterpolation führt zu einer **Überanpassung (Overfitting)**, siehe Übungen).
- Besserer Ansatz: Lineare Funktionen:

$$\mathcal{F} = \{af_1 + bf_2 \mid a, b \in \mathbb{R}\}$$

mit den **Ansatzfunktionen** $f_1(x) = x$ und $f_2(x) = 1$.

- Das **Fehlerfunktional** hat dann die Form:

$$E(f) = E(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

Extremwertaufgabe

Die Minimierung der Funktion

$$E(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

ist eine **Extremwertaufgabe** im \mathbb{R}^2 .

Hierfür benötigen wir Differentialrechnung im \mathbb{R}^n .

☞ **Einführung in die Analysis, Kapitel 7**

Lineare Regression revisited

Zur Minimierung von

$$E(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

müssen die partiellen Ableitungen gleich 0 sein.

$$\frac{\partial E(a, b)}{\partial a} = -2 \sum_{i=1}^n (y_i - (ax_i + b))x_i \stackrel{!}{=} 0$$

$$\frac{\partial E(a, b)}{\partial b} = -2 \sum_{i=1}^n (y_i - (ax_i + b)) \stackrel{!}{=} 0$$

Umformung ergibt das LGS:

$$a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i$$
$$a \sum_{i=1}^n x_i + b \sum_{i=1}^n 1 = \sum_{i=1}^n y_i$$

bzw.

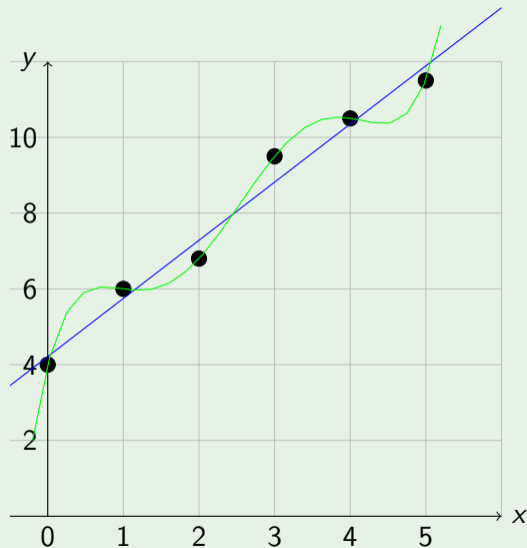
$$\begin{pmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{pmatrix}$$

Beispiel 3.8

Für Beispiel 3.1 erhalten wir:

$$\begin{pmatrix} 55 & 15 \\ 15 & 6 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 147.6 \\ 48.3 \end{pmatrix}$$

mit der Lösung $a = 1.5343$ und $b = 4.2143$.



Lineares Ausgleichsproblem

Definition 3.9

Gegeben seien:

- Ansatzfunktionen f_1, \dots, f_m ,
- $\mathcal{F} = \{\lambda_1 f_1 + \lambda_2 f_2 + \dots + \lambda_m f_m \mid \lambda_i \in \mathbb{R} \text{ für } i = 1, \dots, m\}$,
- n Wertepaare (x_i, y_i) , $i = 1, \dots, n$.

Weiterhin sei $f := \sum_{j=1}^m \lambda_j f_j \in \mathcal{F}$

Dann besteht das **lineare Ausgleichsproblem** darin, die Parameter $\lambda_1, \dots, \lambda_m$ so zu bestimmen, dass das folgende **Fehlerfunktional minimal wird**:

$$E(\lambda_1, \lambda_2, \dots, \lambda_m) = \sum_{i=1}^n (y_i - f(x_i))^2 = \sum_{i=1}^n \left(y_i - \sum_{j=1}^m \lambda_j f_j(x_i) \right)^2.$$

Fehlergleichungssystem

Im Folgenden gelte:

$$\mathbf{A} = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_m(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_m(x_2) \\ \vdots & \vdots & & \vdots \\ f_1(x_n) & f_2(x_n) & \cdots & f_m(x_n) \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix}$$

Definition 3.10

Das Gleichungssystem

$$\mathbf{A}\boldsymbol{\lambda} = \mathbf{y}$$

heißt **Fehlergleichungssystem**.

Folgerung 3.11

$$E(\boldsymbol{\lambda}) = E(\lambda_1, \dots, \lambda_m) = \|\mathbf{y} - \mathbf{A}\boldsymbol{\lambda}\|_2^2$$

Normalgleichungen

Definition 3.12

Die Gleichungen

$$\frac{\partial E(\lambda_1, \dots, \lambda_m)}{\partial \lambda_i} = 0, \quad \text{für } i = 1, \dots, m$$

heißen **Normalgleichungen** des linearen Ausgleichsproblem und das zugehörige Gleichungssystem heißt **Normalgleichungssystem**.

Satz 3.13

Das Normalgleichungssystem lässt sich in der Form:

$$\mathbf{A}^T \mathbf{A} \boldsymbol{\lambda} = \mathbf{A}^T \mathbf{y}$$

schreiben.

Beweis.

Es gilt

$$\frac{\partial E(\lambda_1, \dots, \lambda_m)}{\partial \lambda_k} = \sum_{i=1}^n 2 \left(y_i - \sum_{j=1}^m \lambda_j f_j(x_i) \right) f_k(x_i) \cdot (-1)$$

Damit folgt:

$$\begin{aligned} \frac{\partial E(\lambda_1, \dots, \lambda_m)}{\partial \lambda_k} &= 0 \\ \Leftrightarrow \sum_{i=1}^n \left(y_i - \sum_{j=1}^m \lambda_j f_j(x_i) \right) f_k(x_i) &= 0 \\ \Leftrightarrow \sum_{i=1}^n y_i f_k(x_i) &= \sum_{i=1}^n f_k(x_i) \sum_{j=1}^m \lambda_j f_j(x_i) \end{aligned}$$

Fortsetzung Beweis.

$$\Leftrightarrow \sum_{i=1}^n y_i f_k(x_i) = \sum_{j=1}^m \lambda_j \sum_{i=1}^n f_k(x_i) f_j(x_i)$$

Mit der Notation

$$f_l(\mathbf{x}) := \begin{pmatrix} f_l(x_1) \\ f_l(x_2) \\ \vdots \\ f_l(x_n) \end{pmatrix} \in \mathbb{R}^n$$

können wir diese Gleichung schreiben als:

$$\langle f_k(\mathbf{x}), \mathbf{y} \rangle = \sum_{j=1}^m \lambda_j \langle f_k(\mathbf{x}), f_j(\mathbf{x}) \rangle$$

Fortsetzung Beweis.

Weiterhin gilt:

$$\mathbf{A}^T = \begin{pmatrix} f_1(x_1) & f_1(x_2) & \dots & f_1(x_n) \\ f_2(x_1) & f_2(x_2) & \dots & f_2(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ f_m(x_1) & f_m(x_2) & \dots & f_m(x_n) \end{pmatrix}$$

Sei $\mathbf{B} := \mathbf{A}^T \mathbf{A}$. Dann folgt:

$$\begin{aligned} b_{ij} &= (f_i(x_1) \cdots f_i(x_n)) \begin{pmatrix} f_j(x_1) \\ \vdots \\ f_j(x_n) \end{pmatrix} \\ &= \langle f_i(\mathbf{x}), f_j(\mathbf{x}) \rangle \end{aligned}$$

Fortsetzung Beweis.

Wir schauen uns die k -te Zeile von $\mathbf{A}^T \mathbf{A} \boldsymbol{\lambda}$ an. Es gilt:

$$\text{Zeile } k \text{ von } \mathbf{A}^T \mathbf{A} \boldsymbol{\lambda} = \sum_{j=1}^m b_{kj} \lambda_j = \sum_{j=1}^m \lambda_j \langle f_k(\mathbf{x}), f_j(\mathbf{x}) \rangle$$

Dies ist die **rechte Seite** der Gleichung auf Folie 154 unten.

Jetzt schauen wir uns die k -te Zeile von $\mathbf{A}^T \mathbf{y}$ an.

$$\text{Zeile } k \text{ von } \mathbf{A}^T \mathbf{y} = \sum_{i=1}^n f_k(x_i) y_i = \langle f_k(\mathbf{x}), \mathbf{y} \rangle$$

Dies ist die **linke Seite** der Gleichung auf Folie 154 unten. Also:

$$\frac{\partial E(\lambda_1, \dots, \lambda_m)}{\partial \lambda_k} = 0 \text{ f\"ur } k = 1, \dots, m \iff \mathbf{A}^T \mathbf{A} \boldsymbol{\lambda} = \mathbf{A}^T \mathbf{y}$$

Bemerkungen

- Das Normalgleichungssystem hat die $m \times m$ Koeffizientenmatrix $\mathbf{A}^T \mathbf{A}$. Die **Größe** ist also durch die **Anzahl der Parameter** bestimmt.
- Üblicherweise gilt, dass die **Anzahl der Messpunkte deutlich größer** ist als die **Anzahl der Parameter**.
- Die Matrix $\mathbf{A}^T \mathbf{A}$ ist **positiv definit**.
- Daher können wir eine Lösung mit einer **Cholesky-Zerlegung** berechnen oder eine LR-Zerlegung verwenden.
- Später besprechen wir mit der **QR-Zerlegung** noch eine alternative Zerlegung.
- Der Vektor $\mathbf{r} := \mathbf{A}\boldsymbol{\lambda} - \mathbf{y}$ heißt **Residuumsvektor**.

Beispiel 3.14

Wir betrachten das Normalgleichungssystem für die Regressionsgerade von Beispiel 3.1. Unsere Ansatzfunktionen sind bekanntlich $f_1(x) = x$ und $f_2(x) = 1$. Damit erhalten wir

$$\mathbf{A} = \begin{pmatrix} f_1(x_1) & f_2(x_1) \\ f_1(x_2) & f_2(x_2) \\ f_1(x_3) & f_2(x_3) \\ f_1(x_4) & f_2(x_4) \\ f_1(x_5) & f_2(x_5) \\ f_1(x_6) & f_2(x_6) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{pmatrix}$$

und somit

$$\mathbf{A}^T \mathbf{A} = \begin{pmatrix} 55 & 15 \\ 15 & 6 \end{pmatrix}$$

sowie

$$\mathbf{A}^T \mathbf{y} = \begin{pmatrix} 147.6 \\ 48.3 \end{pmatrix}.$$

Fortsetzung Beispiel.

Damit stellt das LGS

$$\mathbf{A}^T \mathbf{A} \boldsymbol{\lambda} = \mathbf{A}^T \mathbf{y}$$

genau das LGS dar, welches wir schon aus Beispiel 3.7 kennen.

Dementsprechend ist die Lösung $\lambda_1 = 1.5343$ und $\lambda_2 = 4.2143$.

Bemerkung:

- Bei einem linearen Ausgleichsproblem suchen wir nach einer **optimalen Linearkombination von Ansatzfunktionen**,
- d. h. die Kombination der Funktionen kann nur linear sein.
- **Die Ansatzfunktionen selbst können aber auch nicht linear sein.**

Nichtlineare Ansatzfunktionen

Beispiel 3.15

Gegeben seien die Daten

x_i	0	1	2	3	4
y_i	6	12	30	80	140

Wir wollen eine Funktion der Form

$$f(x) = \lambda_1 e^x + \lambda_2$$

finden, die die Daten bestmöglich approximiert.

Unsere Ansatzfunktionen sind also $f_1(x) = e^x$ und $f_2(x) = 1$.

Fortsetzung Beispiel.

Das Fehlergleichungssystem lautet damit

$$\begin{pmatrix} e^0 & 1 \\ e^1 & 1 \\ e^2 & 1 \\ e^3 & 1 \\ e^4 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 6 \\ 12 \\ 30 \\ 80 \\ 140 \end{pmatrix}.$$

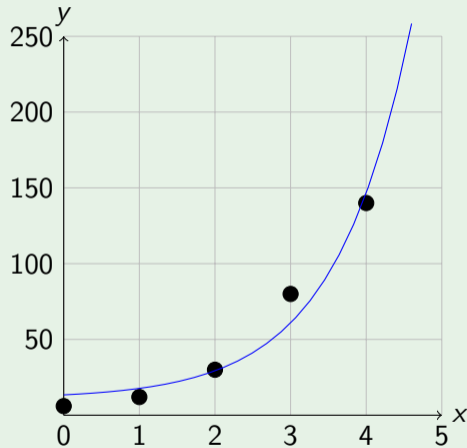
Für das Normalgleichungssystem entsteht

$$\begin{pmatrix} 3447.373987 & 85.79102488 \\ 85.79102488 & 5.0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 9510.875023 \\ 268.0 \end{pmatrix}$$

mit Lösung

$$\lambda = \begin{pmatrix} 2.4869 \\ 10.9295 \end{pmatrix}.$$

Fortsetzung Beispiel.



Logarithmisch-lineare Regression

Wählt man den Ansatz

$$f(x) = \lambda_1 e^{\lambda_2 x}$$

dann spricht man von **logarithmisch linearer Regression**. Hierbei handelt es sich zunächst um ein nichtlineares Regressionsproblem.

Wir können diesen Ansatz aber linearisieren, indem wir **zum Logarithmus übergehen**:

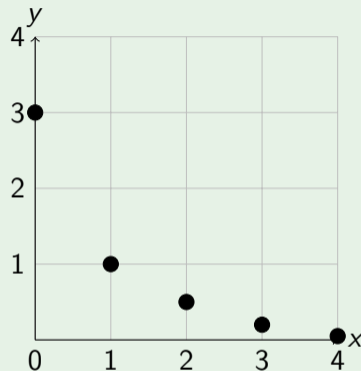
$$\log(f(x)) = \log(\lambda_1) + \lambda_2 x$$

Voraussetzung: $\mathbf{y} > \mathbf{0}$

Beispiel 3.16

Gegeben seien die Daten

x_i	0	1	2	3	4
y_i	3	1	0.5	0.2	0.05



Wir suchen eine Funktion der Form

$$f(x) = \lambda_1 e^{\lambda_2 x},$$

die die Daten bestmöglich approximiert.

Fortsetzung Beispiel.

Wir gehen zum Logarithmus über:

x_i	0	1	2	3	4
y_i	3	1	0.5	0.2	0.05
$\log(y_i)$	1.099	0	-0.693	-1.609	-2.996

und berechnen die Regressionsgerade für die x_i und die $\log(y_i)$. Wir erhalten:

$$\lambda_1 = 1.11968, \quad \lambda_2 = -0.9798$$

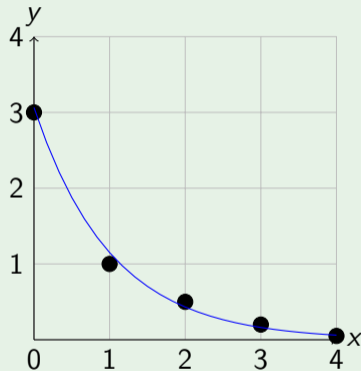
also

$$\log(f(x)) = -0.9798x + 1.11968.$$

Daraus ergibt sich

$$f(x) = e^{1.11968} e^{-0.9798x} = 3.0639 e^{-0.9798x}.$$

Fortsetzung Beispiel.



Multiple Regression

- Anschaulich betrachtet sprechen wir von **multipler Regression**, wenn die x_j -Werte der Wertepaare selber wieder Vektoren $\in \mathbb{R}^d$ sind.
- Mit anderen Worten: Die approximierende Funktion f ist von der Form $f : D \rightarrow \mathbb{R}$ mit $D \subseteq \mathbb{R}^d$.
- Wir haben also **nicht nur eine Einflussgröße sondern d** .
- Zur Übersicht:
 - n Anzahl an Wertepaaren/Stützstellen
 - m Anzahl Parameter für die Ansatzfunktionen
 - d Anzahl der Werte/Dimensionen für eine einzelne Stützstelle

Beispiel 3.17

Gegeben seien die Daten

$x_{i,1}$	1	1	1	2	2	2	3	3	3
$x_{i,2}$	1	2	3	1	2	3	1	2	3
y_i	1.2	1.4	2.0	2.5	3.2	3.4	3.7	3.9	4.3

Wir haben also als Stützstellen ein **äquidistantes Gitter der Größe 3×3** .


Wir vermuten einen funktionalen Zusammenhang der Form

$$f(x_1, x_2) = \lambda_0 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_1^2 + \lambda_4 x_2^2 + \lambda_5 x_1 x_2.$$

Fortsetzung Beispiel.

Allgemeine Form des Fehlergleichungssystems:

$$\begin{pmatrix} 1 & x_{1,1} & x_{1,2} & x_{1,1}^2 & x_{1,2}^2 & x_{1,1}x_{1,2} \\ 1 & x_{2,1} & x_{2,2} & x_{2,1}^2 & x_{2,2}^2 & x_{2,1}x_{2,2} \\ \vdots & & & & & \vdots \\ 1 & x_{9,1} & x_{9,2} & x_{9,1}^2 & x_{9,2}^2 & x_{9,1}x_{9,2} \end{pmatrix} \begin{pmatrix} \lambda_0 \\ \lambda_1 \\ \vdots \\ \lambda_5 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_9 \end{pmatrix}$$

Konkretes Fehlergleichungssystems: Tafel 

Fortsetzung Beispiel.

Lösung:

$$\lambda_0 = -1.44$$

$$\lambda_1 = 2.45$$

$$\lambda_2 = 0.42$$

$$\lambda_3 = -0.28$$

$$\lambda_4 = 0.016$$

$$\lambda_5 = -0.05$$

Orthogonale Matrix

- Ausgleichsprobleme sind häufig **schlecht konditioniert**.
- Matrix des Normalgleichungssystems kann **“nahezu singulär”** sein.
- Spezielle Matrixzerlegung für höhere numerische Stabilität: **QR-Zerlegung**

Definition 3.18

Es sei $\mathbf{Q} \in \mathbb{R}^{n \times n}$ eine quadratische Matrix. Gilt

$$\mathbf{Q}^{-1} = \mathbf{Q}^T,$$

dann heißt \mathbf{Q} **orthogonal**.

Beispiele orthogonaler Matrizen

Beispiel 3.19

(i) Die Matrix

$$\mathbf{Q} = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

ist orthogonal. Sie beschreibt als lineare Abbildung eine **Drehung um 45°** .

(ii) Die Verallgemeinerung der Matrix aus (i) ist

$$\mathbf{R} = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix}.$$

Sie beschreibt eine **Drehung um den Winkel α** .

(iii) **Permutationsmatrizen** sind orthogonal.

Eigenschaften orthogonaler Matrizen

Satz 3.20

Es sei $\mathbf{Q} \in \mathbb{R}^{n \times n}$ eine orthogonale Matrix. Dann gilt:

- (i) Die Spaltenvektoren $\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \dots, \mathbf{q}^{(n)}$ bilden eine *orthonormale Basis des \mathbb{R}^n* , d. h.
 - ▶ die Vektoren $\mathbf{q}^{(1)}, \dots, \mathbf{q}^{(n)}$ sind linear unabhängig,
 - ▶ $\text{Span}(\mathbf{q}^{(1)}, \dots, \mathbf{q}^{(n)}) = \mathbb{R}^n$,
 - ▶ $\|\mathbf{q}^{(i)}\| = 1$ für $i = 1, \dots, n$ und
 - ▶ $\langle \mathbf{q}^{(i)}, \mathbf{q}^{(j)} \rangle = 0$ für $i \neq j$.
- (ii) \mathbf{Q}^T ist auch orthogonal.
- (iii) \mathbf{Q} entspricht einer *längentreuen linearen Abbildung*, d. h. $\|\mathbf{Q}\mathbf{x}\| = \|\mathbf{x}\|$ für alle $\mathbf{x} \in \mathbb{R}^n$.
- (iv) $\det(\mathbf{Q}) = \pm 1$
- (v) Das *Produkt* $\mathbf{Q} \cdot \mathbf{R}$ zweier orthogonaler Matrizen \mathbf{Q} und \mathbf{R} ist wieder *orthogonal*.

QR-Zerlegung

Definition 3.21

Es sei $\mathbf{A} \in \mathbb{R}^{n \times m}$. Gilt

$$\mathbf{A} = \mathbf{QR}$$

- (i) für eine Matrix $\mathbf{Q} \in \mathbb{R}^{n \times m}$ mit orthonormierten Spalten und eine obere Dreiecksmatrix $\mathbf{R} \in \mathbb{R}^{m \times m}$, dann heißt diese Zerlegung **reduzierte QR-Zerlegung**.
- (ii) für eine Orthogonalmatrix $\mathbf{Q} \in \mathbb{R}^{n \times n}$ und eine Matrix $\mathbf{R} \in \mathbb{R}^{n \times m}$ mit oberer Dreiecksgestalt, dann heißt diese Zerlegung **volle QR-Zerlegung**.

Bemerkungen zu Definition 3.21:

(i) \mathbf{Q} hat orthonormierte Spalten heißt

$$\langle \mathbf{q}^{(i)}, \mathbf{q}^{(j)} \rangle = \begin{cases} 1 & \text{für } i = j \\ 0 & \text{für } i \neq j \end{cases}$$

und $1 \leq i, j \leq m$ und $\mathbf{q}^{(i)}, \mathbf{q}^{(j)} \in \mathbb{R}^n$.

(ii) Obere Dreiecksgestalt bedeutet $r_{i,j} = 0$ für $i > j$.

$$\mathbf{R} = \begin{pmatrix} * & * & \cdots & * \\ 0 & * & \cdots & * \\ 0 & 0 & \ddots & \vdots \\ \vdots & & \ddots & * \\ & & & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & & 0 \end{pmatrix} \in \mathbb{R}^{n \times m}$$

Eindeutigkeit der reduzierten QR-Zerlegung

Satz 3.22

Sei $\mathbf{A} \in \mathbb{R}^{n \times m}$ mit $r(\mathbf{A}) = m \leq n$ und seien

$$\mathbf{Q}_1 \mathbf{R}_1 = \mathbf{A} = \mathbf{Q}_2 \mathbf{R}_2$$

zwei reduzierte QR-Zerlegungen von \mathbf{A} .

Dann existiert eine Diagonalmatrix $\mathbf{D} \in \mathbb{R}^{m \times m}$ mit $d_{ii} \pm 1$ für alle $i = 1, \dots, m$, so dass gilt:

$$\mathbf{Q}_1 = \mathbf{Q}_2 \mathbf{D} \quad \text{und} \quad \mathbf{R}_2 = \mathbf{D} \mathbf{R}_1$$

Folgerung 3.23

Es existiert genau eine QR-Zerlegung mit $r_{ii} > 0$ für alle $i = 1, \dots, m$.

QR-Zerlegung und Normalgleichungssystem

Die Lösung eines linearen Ausgleichsproblems ergibt sich durch Lösung des Normalgleichungssystems

$$\mathbf{A}^T \mathbf{A} \boldsymbol{\lambda} = \mathbf{A}^T \mathbf{y}.$$

Sei nun $\mathbf{A} = \mathbf{QR}$ eine reduzierte QR-Zerlegung. Wegen $\mathbf{Q}^T \mathbf{Q} = \mathbf{E}$ folgt

$$\mathbf{R}^T \mathbf{R} \boldsymbol{\lambda} = \mathbf{R}^T \mathbf{Q}^T \mathbf{Q} \mathbf{R} \boldsymbol{\lambda} = \mathbf{A}^T \mathbf{A} \boldsymbol{\lambda} = \mathbf{A}^T \mathbf{y} = \mathbf{R}^T \mathbf{Q}^T \mathbf{y}.$$

Da \mathbf{R} und damit auch \mathbf{R}^T regulär ist, ergibt sich

$$\mathbf{R} \boldsymbol{\lambda} = \mathbf{Q}^T \mathbf{y}.$$

Dieses Gleichungssystem können wir durch Rückwärtssubstitution lösen.

Lösung von $\mathbf{A}\boldsymbol{\lambda} \approx \mathbf{y}$ mittels QR-Zerlegung

Algorithmus 3.24

- 1 Bestimme eine reduzierte QR-Zerlegung $\mathbf{A} = \mathbf{QR}$.
- 2 Setze $\mathbf{c} := \mathbf{Q}^T \mathbf{y}$.
- 3 Löse $\mathbf{R}\boldsymbol{\lambda} = \mathbf{c}$ durch Rückwärtssubstitution.

Bemerkung:

- Für $n = m$ besteht kein Unterschied zwischen einer reduzierten und einer vollen QR-Zerlegung.
- In diesem Fall können wir eine QR-Zerlegung auch zur Lösung eines LGS verwenden.
- Dies wird aus Gründen der numerischen Stabilität auch verwendet.

Gram-Schmidt-Orthogonalisierung

Wir schauen uns $\mathbf{A} = \mathbf{QR}$ näher an.

$$\left(\begin{array}{c|c|c} \mathbf{a}^{(1)} & \mathbf{a}^{(2)} & \dots & \mathbf{a}^{(m)} \\ \hline & & & \end{array} \right) = \left(\begin{array}{c|c|c} \mathbf{q}^{(1)} & \mathbf{q}^{(2)} & \dots & \mathbf{q}^{(m)} \\ \hline & & & \end{array} \right) \begin{pmatrix} r_{1,1} & r_{1,2} & \dots & r_{1,m} \\ & r_{2,2} & \dots & r_{2,m} \\ & & \ddots & \vdots \\ & & & r_{m,m} \end{pmatrix}$$

Dies ist äquivalent zu:

$$\begin{aligned} \mathbf{a}^{(1)} &= r_{1,1} \mathbf{q}^{(1)} \\ \mathbf{a}^{(2)} &= r_{1,2} \mathbf{q}^{(1)} + r_{2,2} \mathbf{q}^{(2)} \\ &\vdots \\ \mathbf{a}^{(m)} &= r_{1,m} \mathbf{q}^{(1)} + r_{2,m} \mathbf{q}^{(2)} + \dots + r_{m,m} \mathbf{q}^{(m)} \end{aligned}$$

In Kurzform:

$$\mathbf{a}^{(j)} = \sum_{i=1}^j r_{i,j} \mathbf{q}^{(i)}$$

Hieraus können wir nun die $\mathbf{q}^{(j)}$ und die $r_{i,j}$ induktiv herleiten.

Für $j = 1$ ergibt sich:

$$r_{1,1} = \pm \|\mathbf{a}^{(1)}\| \quad \text{und} \quad \mathbf{q}^{(1)} = \frac{1}{r_{1,1}} \mathbf{a}^{(1)}$$

Sei nun $j \geq 2$, sowie $\mathbf{q}^{(1)}, \dots, \mathbf{q}^{(j-1)}$ ein orthonormiertes System von Vektoren, so dass die ersten $j - 1$ Gleichungen mit bekannten Zahlen $r_{i,k}$, $1 \leq i \leq k \leq j - 1$ erfüllt sind.

Wir machen den Ansatz:

$$\mathbf{q}^{(j)} = \frac{1}{r_{j,j}} \left(\mathbf{a}^{(j)} - \sum_{i=1}^{j-1} r_{i,j} \mathbf{q}^{(i)} \right)$$

Da $\mathbf{q}^{(j)}$ normiert sein soll, folgt

$$r_{j,j} = \left\| \mathbf{a}^{(j)} - \sum_{i=1}^{j-1} r_{i,j} \mathbf{q}^{(i)} \right\|$$

Die noch unbekanntenen $r_{i,j}$ für $i < j$ ergeben sich aus der geforderdeten Orthogonalität:

$$0 = \langle \mathbf{q}^{(i)}, \mathbf{q}^{(j)} \rangle = \frac{1}{r_{j,j}} \left(\langle \mathbf{q}^{(i)}, \mathbf{a}^{(j)} \rangle - r_{i,j} \right)$$

für alle $i = 1, \dots, j - 1$. Damit folgt

$$r_{i,j} = \langle \mathbf{q}^{(i)}, \mathbf{a}^{(j)} \rangle$$

Gram-Schmidt-Verfahren

Algorithmus 3.25

$$r_{1,1} := \|\mathbf{a}^{(1)}\|$$

$$\mathbf{q}^{(1)} := \frac{1}{r_{1,1}}\mathbf{a}^{(1)}$$

for $j := 2$ **to** m **do**

$$\mathbf{q}^{(j)} := \mathbf{a}^{(j)}$$

for $i := 1$ **to** $j - 1$ **do**

$$r_{i,j} := \langle \mathbf{q}^{(i)}, \mathbf{a}^{(j)} \rangle$$

$$\mathbf{q}^{(j)} := \mathbf{q}^{(j)} - r_{i,j}\mathbf{q}^{(i)}$$

end

$$r_{j,j} := \|\mathbf{q}^{(j)}\|$$

$$\mathbf{q}^{(j)} := \frac{1}{r_{j,j}}\mathbf{q}^{(j)}$$

end

Existenz der QR-Zerlegung

- In Folgerung 3.23 können wir aus Satz 3.22 nur die **Eindeutigkeit der QR-Zerlegung** herleiten, aber nicht deren Existenz.
- Das Gram-Schmidt-Verfahren liefert aber eine **Konstruktion für eine QR-Zerlegung**.
- Wenn wir zeigen können, dass dieser Algorithmus immer durchführbar ist, haben wir auch den **Existenzbeweis**.
- Das Gram-Schmidt-Verfahren kann nur durch $r_{j,j} = 0$ scheitern.
- Wäre aber $r_{j,j} = 0$, dann würde

$$\mathbf{a}^{(j)} \in \text{Span}\{\mathbf{q}^{(1)}, \dots, \mathbf{q}^{(j-1)}\} = \text{Span}\{\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(j-1)}\}$$

gelten.

- Dies ist ein Widerspruch zur linearen Unabhängigkeit der Spaltenvektoren von $\mathbf{A} \in \mathbb{R}^{n \times m}$, denn n. V. gilt $r(\mathbf{A}) = m$.

Modifiziertes Gram-Schmidt-Verfahren

- Wenn die Spalten von \mathbf{A} nahezu linear abhängig sind, ist das Gram-Schmidt-Verfahren immer noch **numerisch instabil**. Die berechneten Vektoren $\mathbf{q}^{(j)}$ sind also nicht orthogonal.
- Besser wird die Stabilität, wenn wir die Berechnung der $r_{i,j}$ ersetzen durch

$$r_{i,j} = \langle \mathbf{q}^{(i)}, \mathbf{q}^{(j)} \rangle.$$

- Diese zum klassischen Gram-Schmidt-Verfahren **äquivalente Methode** heißt **modifiziertes Gram-Schmidt-Verfahren**.

Algorithmus 3.26

$$r_{1,1} := \|\mathbf{a}^{(1)}\|$$

$$\mathbf{q}^{(1)} := \frac{1}{r_{1,1}}\mathbf{a}^{(1)}$$

for $j := 2$ **to** m **do**

$$\mathbf{q}^{(j)} := \mathbf{a}^{(j)}$$

for $i := 1$ **to** $j - 1$ **do**

$$r_{i,j} := \langle \mathbf{q}^{(i)}, \mathbf{q}^{(j)} \rangle$$

$$\mathbf{q}^{(j)} := \mathbf{q}^{(j)} - r_{i,j}\mathbf{q}^{(i)}$$

end

$$r_{j,j} := \|\mathbf{q}^{(j)}\|$$

$$\mathbf{q}^{(j)} := \frac{1}{r_{j,j}}\mathbf{q}^{(j)}$$

end

Beispiel 3.27

Wir lösen Beispiel 3.1 mittels Algorithmus 3.24.

Schritt 1: Wir berechnen eine QR-Zerlegung mit dem Gram-Schmidt-Verfahren. Es gilt

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{pmatrix}$$

und damit $\|\mathbf{a}^{(1)}\| = \sqrt{55} = r_{1,1}$, sowie

$$\mathbf{q}^{(1)} = \frac{1}{\sqrt{55}} \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}.$$

Fortsetzung Beispiel.

Daraus ergibt sich

$$r_{1,2} = \langle \mathbf{q}^{(1)}, \mathbf{a}^{(2)} \rangle = \frac{15}{\sqrt{55}}.$$

und für den Vektor $\mathbf{q}^{(2)}$ vor der Normierung

$$\mathbf{q}^{(2)} = \mathbf{a}^{(2)} - r_{1,2}\mathbf{q}^{(1)} = \frac{1}{11} \begin{pmatrix} 11 \\ 8 \\ 5 \\ 2 \\ -1 \\ -4 \end{pmatrix}.$$

Fortsetzung Beispiel.

Die Normierung von $\mathbf{q}^{(2)}$ liefert $r_{2,2} = \frac{\sqrt{231}}{11}$ und

$$\mathbf{q}^{(2)} = \frac{1}{\sqrt{231}} \begin{pmatrix} 11 \\ 8 \\ 5 \\ 2 \\ -1 \\ -4 \end{pmatrix}.$$

Schritt 2: Wir berechnen $\mathbf{c} = \mathbf{Q}^T \mathbf{y}$.

$$\mathbf{c} = \mathbf{Q}^T \mathbf{y} = \begin{pmatrix} \langle \mathbf{q}^{(1)}, \mathbf{y} \rangle \\ \langle \mathbf{q}^{(2)}, \mathbf{y} \rangle \end{pmatrix} = \begin{pmatrix} \frac{147.6}{\sqrt{55}} \\ \frac{88.5}{\sqrt{231}} \end{pmatrix}$$

Fortsetzung Beispiel.

Schritt 3: Wir lösen $\mathbf{R}\boldsymbol{\lambda} = \mathbf{c}$.

$$\begin{pmatrix} \sqrt{55} & \frac{15}{\sqrt{55}} \\ 0 & \frac{\sqrt{231}}{11} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \frac{147.6}{\sqrt{55}} \\ \frac{88.5}{\sqrt{231}} \end{pmatrix}$$

Daraus ergibt sich

$$\begin{aligned} \lambda_2 &= \frac{11 \cdot 88.5}{231} = 4.2143 \\ \lambda_1 &= \frac{147.6 - 15\lambda_2}{55} = 1.5343 \end{aligned}$$

Householder-Matrix


Definition 3.28

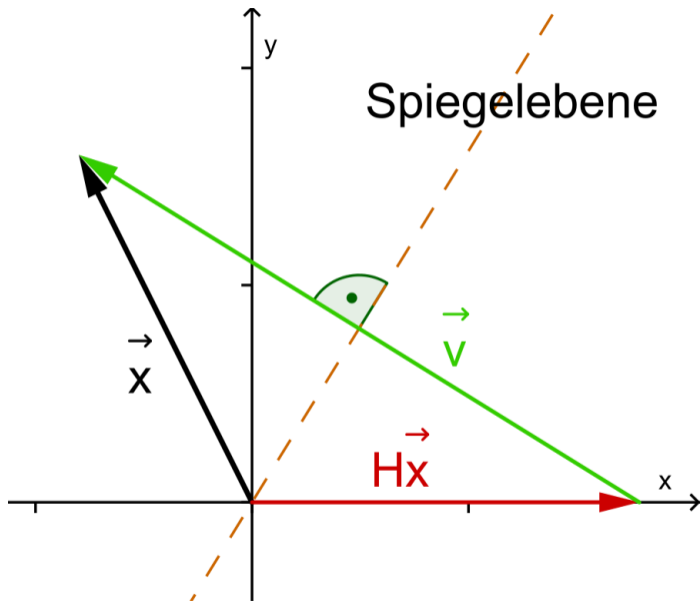
Es sei $\mathbf{v} \in \mathbb{R}^n$ mit $\mathbf{v} \neq \mathbf{0}$.

Dann heißt die Matrix

$$\mathbf{H} = \mathbf{E} - \frac{2}{\mathbf{v}^T \mathbf{v}} \mathbf{v} \mathbf{v}^T$$

Householder-Matrix zum Vektor \mathbf{v} .

- Ist \mathbf{v} auf die Länge 1 normiert, ergibt sich $\mathbf{H} = \mathbf{E} - 2\mathbf{v}\mathbf{v}^T$.
- Householder-Matrizen beschreiben **Spiegelungen** an einer durch den Ursprung gehenden Spiegelebene mit Normalenvektor \mathbf{v} . 



Eigenschaften von Householder-Matrizen

- 1 Jede Householder-Matrix \mathbf{H} ist **orthogonal**.
- 2 Householder-Matrizen sind **symmetrisch**, es gilt also

$$\mathbf{H} = \mathbf{H}^T.$$

- 3 Aus (1) und (2) folgt

$$\mathbf{H} = \mathbf{H}^{-1}.$$

- 4 Die Multiplikation von beliebig vielen Householder-Matrizen liefert wieder eine orthogonale Matrix.

Beispiel 3.29

Wir berechnen die Householder-Matrix \mathbf{H} zum Vektor

$$\mathbf{v} = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}$$

und überprüfen die genannten Eigenschaften.



Schritt zur Erzeugung einer oberen Dreiecksmatrix

Lemma 3.30

Es sei

- $\mathbf{A} \in \mathbb{R}^{n \times n}$ eine quadratische Matrix,
- $k \in \{1, \dots, n\}$,
- $\mathbf{e}^k \in \mathbb{R}^n$ der k -te Einheitsvektor,
- $\mathbf{a}^k = (0, \dots, 0, a_{k,k}, \dots, a_{n,k})^T \in \mathbb{R}^n$ der k -te Spaltenvektor von \mathbf{A} , wobei die Elemente oberhalb der Diagonalen durch Null ersetzt wurden,
- $\mathbf{v}_k = \mathbf{a}^k + \|\mathbf{a}^k\|_2 \mathbf{e}^k \in \mathbb{R}^n$ und
- $\mathbf{H}_k = \mathbf{E} - \frac{2}{\mathbf{v}_k^T \mathbf{v}_k} \mathbf{v}_k \mathbf{v}_k^T \in \mathbb{R}^{n \times n}$.

Dann besitzt die Matrix

$$\mathbf{B} = \mathbf{H}_k \mathbf{A}$$

unterhalb der Diagonalen in Spalte k nur Nulleinträge.

Herleitung einer QR-Zerlegung

- Wir multiplizieren eine Matrix \mathbf{A} mit Householder-Matrizen \mathbf{H}_k , so dass wir jeweils in Spalte k unterhalb der Diagonalen Nullen erhalten.
- Wiederholtes Anwenden mit $k = 1, \dots, n - 1$ ergibt eine obere Dreiecksmatrix \mathbf{R} mit

$$\mathbf{R} = \mathbf{H}_{n-1} \cdot \dots \cdot \mathbf{H}_1 \cdot \mathbf{A}.$$

- Wegen $\mathbf{H}_k^{-1} = \mathbf{H}_k$ folgt

$$\mathbf{A} = \mathbf{H}_1 \cdot \dots \cdot \mathbf{H}_{n-1} \cdot \mathbf{R}.$$

- Die Matrix

$$\mathbf{Q} = \mathbf{H}_1 \cdot \dots \cdot \mathbf{H}_{n-1}$$

ist eine orthogonale Matrix.

Algorithmus zur QR-Zerlegung

Algorithmus 3.31

Gegeben sei eine Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$.

Sei $\mathbf{B}_1 = \mathbf{A}$ mit den Matrixeinträgen $b_{i,j}^1$.

Für $k = 1, \dots, n - 1$ führe die folgenden Schritte aus:

- 1 Setze $\mathbf{a}_k = (0, \dots, 0, b_{k,k}^k, \dots, b_{n,k}^k)^T$ und damit

$$\mathbf{v}_k = \mathbf{a}_k + \|\mathbf{a}_k\|_2 \cdot \mathbf{e}_k.$$

- 2 Berechne die Householder-Matrix

$$\mathbf{H}_k = \mathbf{E} - \frac{2}{\mathbf{v}_k^T \mathbf{v}_k} \mathbf{v}_k \mathbf{v}_k^T.$$

Fortsetzung Algorithmus.

- 3 Berechne $\mathbf{B}_{k+1} = \mathbf{H}_k \cdot \mathbf{B}_k$ mit den Matrixeinträgen $b_{i,j}^{k+1}$.

Die QR-Zerlegung von \mathbf{A} ist dann gegeben durch

$$\mathbf{Q} = \mathbf{H}_1 \cdot \dots \cdot \mathbf{H}_{n-1}$$

und $\mathbf{R} = \mathbf{B}_n$.

Beispiel zur QR-Zerlegung

Beispiel 3.32

Es sei

$$\mathbf{A} = \begin{pmatrix} 2 & 4 & -9 \\ -2 & -10 & 3 \\ 1 & -1 & 6 \end{pmatrix}.$$

Es gilt $\mathbf{v}_1 = (5, -2, 1)^T$ (warum?) und damit erhalten wir

$$\mathbf{H}_1 \cdot \mathbf{B}_1 = \begin{pmatrix} -\frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{2}{3} & \frac{11}{15} & \frac{2}{15} \\ -\frac{1}{3} & \frac{2}{15} & \frac{14}{15} \end{pmatrix} \cdot \begin{pmatrix} 2 & 4 & -9 \\ -2 & -10 & 3 \\ 1 & -1 & 6 \end{pmatrix} = \begin{pmatrix} -3 & -9 & 6 \\ 0 & -\frac{24}{5} & -3 \\ 0 & -\frac{18}{5} & 9 \end{pmatrix} = \mathbf{B}_2.$$

Fortsetzung Beispiel.


Wir nutzen $\mathbf{v}_2 = (0, \frac{6}{5}, -\frac{18}{5})^T$ und erhalten

$$\mathbf{H}_2 \cdot \mathbf{B}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{4}{5} & \frac{3}{5} \\ 0 & \frac{3}{5} & -\frac{4}{5} \end{pmatrix} \begin{pmatrix} -3 & -9 & 6 \\ 0 & -\frac{24}{5} & -3 \\ 0 & -\frac{18}{5} & 9 \end{pmatrix} = \begin{pmatrix} -3 & -9 & 6 \\ 0 & -6 & 3 \\ 0 & 0 & -9 \end{pmatrix} = \mathbf{B}_3 = \mathbf{R}.$$

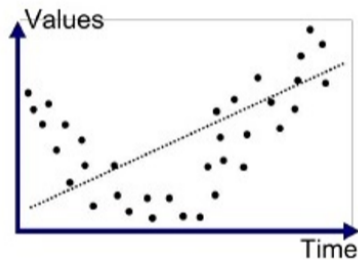
Die orthogonale Matrix \mathbf{Q} ergibt sich durch

$$\mathbf{Q} = \mathbf{H}_1 \cdot \mathbf{H}_2 = \begin{pmatrix} -\frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{2}{3} & \frac{11}{15} & \frac{2}{15} \\ -\frac{1}{3} & \frac{2}{15} & \frac{14}{15} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{4}{5} & \frac{3}{5} \\ 0 & \frac{3}{5} & -\frac{4}{5} \end{pmatrix} = \begin{pmatrix} -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \end{pmatrix}.$$

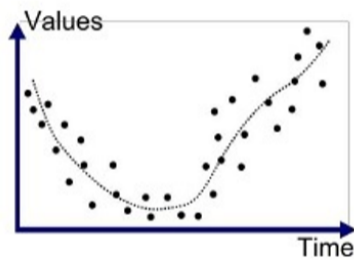
Diskussion QR-Zerlegung

- Komplexität $O(n^3)$
- Diese Komplexität ist **nur mit einer effizienten Matrixmultiplikation erreichbar**, die aber existiert (Übungsaufgabe .
- Eine QR-Zerlung kann **für jede beliebige Matrix durchgeführt werden**, es ist keine Pivotisierung notwendig.
- Die Verwendung einer QR-Zerlegung zur Lösung eines LGS ist i. A. numerisch stabiler als eine LR-Zerlegung.

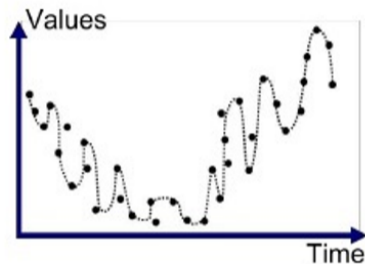
Underfitting und Overfitting



Underfitted



Good Fit/Robust



Overfitted

Prognosemodelle bewerten

- Geht es um das Finden von Ansatzfunktionen, so sollte das **Prognosemodell so genau aber auch so einfach wie möglich** sein.
- Hat man ein sparsames Modell, also mit wenigen Einflussvariablen, ermittelt, so benötigt man auch für die Prognose nur diese wenigen Informationen.
- Als Maßzahl für die Güte der Anpassung einer Funktion an die Daten haben wir das Fehlerfunktional kennengelernt. Dieses Maß ist jedoch für die Suche nach einem sparsamen Modell ungeeignet, da es nicht kleiner wird, wenn ein Merkmal hinzugefügt wird.
- Andererseits führen komplexe Modelle dazu, dass $E(f)$ immer kleiner wird, aber **wo beginnt die Überanpassung?**
- Besser geeignet ist das **Akaike Informationskriterium (AIC)**.

Akaike Informationsindex

Definition 3.33

Gegeben sei:

- p die Anzahl der Koeffizienten und damit der Variablen im Modell und
- n Wertepaare $(x_i, y_i), i = 1, \dots, n$.

$$AIC = \log \left(\frac{1}{n-p} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \right) + 2 * \frac{p}{n}$$

Dabei ist \hat{y}_i der für x_i prognostizierte Wert.

Bemerkung: In der Literatur findet man verschiedene Formulierungen des AIC's, alle folgen jedoch der allgemeinen Formel:

$$AIC = 2p - 2 \log(E(f)).$$

Auf die Theorie, die der Formel zugrunde liegt, wird hier nicht eingegangen. Man kann aber erkennen, dass

- eine **gute Anpassung der Funktion an die Daten** zu einer kleinen Summe $E(f)$ führt, und damit das Gesamtergebnis reduziert;
- eine **große Anzahl von Beobachtungen** n das Ergebnis ebenfalls reduziert;
- jedoch eine große Anzahl von Parametern wie z.B. Koeffizienten von Variablen oder Polynomtermen das Ergebnis vergrößert, die Anzahl p wird also **strafend** berücksichtigt.

Das heißt:

- wenn der AIC-Wert kleiner wird, wenn eine Variable, ein Polynomterm o. ä. hinzugefügt wird, kann man von einer Verbesserung des Modells ausgehen.
- tritt keine Reduktion ein, so wird die Vergrößerung von p nicht durch die Reduktion der Summe der quadrierten Abweichungen kompensiert, der Term verbessert das Modell nicht.

Parameterauswahl

Zur Auswahl des nach diesem Kriterium besten Modells geht man wie folgt vor:

- 1 Man erstellt ein Modell, das die Funktionen enthält, die man für sinnvoll erachtet, am besten nach Visualisierung der Zusammenhänge anhand einer Grafik.
- 2 Man entfernt jeweils einen Term (ein Merkmal oder einen Polynomterm o. ä.) und vergleicht die AIC-Werte.
- 3 Hat sich der AIC-Wert verringert, so wird der Term aus dem Modell endgültig entfernt.

Beispiel: in der Übung

Zusammenfassung

- **Lineares Ausgleichsproblem:** Finde eine Linearkombination $f = \sum_{j=1}^m \lambda_j f_j$ von Funktionen, so dass das **Fehlerfunktional**

$$E(f) = \sum_{i=1}^n (y_i - f(x_i))^2$$

minimiert wird.

- Interpolation liefert i. d. R. keine befriedigende Lösung.
- Lösung des Ausgleichsproblems auf Basis des **Normalgleichungssystems** mittels Cholesky-Zerlegung.
- Numerisch stabiler: **QR-Zerlegung**
- Berechnung einer QR-Zerlegung mit dem **(modifizierten) Gram-Schmidt-Verfahren**
- Lösung für andere Fehlerfunktionale mittels linearer Programmierung