

5 Lineare Ausgleichsrechnung

5.1 Die Normalgleichungen

5.2 Singulärwertzerlegung

5.3 Pseudoinverse

5.4 Orthogonale Matrizen und QR-Zerlegung

5.5 Kondition des Ausgleichsproblems

5.6 Ausgleichspolynome

5.1 Die Normalgleichungen

Das **lineare Ausgleichsproblem** (Kleinste-Quadrate-Problem, engl. **least squares problem**):

Gegeben sind $A \in \mathbb{R}^{m \times n}$ und $\mathbf{b} \in \mathbb{R}^m$.

Gesucht ist ein Vektor $\mathbf{x}^* \in \mathbb{R}^n$ mit

$$\|\mathbf{b} - A\mathbf{x}^*\|_2 = \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - A\mathbf{x}\|_2. \quad (\text{LS})$$

Andere Formulierung: Bei gegebenen A und \mathbf{b} ist jedem $\mathbf{x} \in \mathbb{R}^n$ durch

$$\mathbf{r}_x := \mathbf{b} - A\mathbf{x}$$

ein **Residualvektor** (Residuum, Defekt) \mathbf{r}_x zugeordnet, der misst, wie gut \mathbf{x} das Gleichungssystem $A\mathbf{x} = \mathbf{b}$ erfüllt. Wir wollen einen Vektor \mathbf{x}^* bestimmen, dessen Residuum (gemessen in der Euklid-Norm) so klein wie möglich ist.

Beachte: Ist $A\mathbf{x} = \mathbf{b}$ lösbar, dann ist \mathbf{x}^* eine Lösung.

Satz 5.1. Gegeben seien $A \in \mathbb{R}^{m \times n}$ mit $m \geq n$ sowie $\mathbf{b} \in \mathbb{R}^m$.

Dann gilt:

(a) Ein Vektor $\mathbf{x} \in \mathbb{R}^n$ ist genau dann eine Lösung von (LS), wenn

$$A^\top \mathbf{r}_x = \mathbf{0} \quad (\text{äquivalent: } \mathbf{r}_x \perp_2 \mathcal{R}(A))$$

gilt. Folglich sind die Lösungen von (LS) genau die Lösungen der Normalgleichungen

$$A^\top A \mathbf{x} = A^\top \mathbf{b}. \quad (\text{NG})$$

(b) (NG) ist ein LGS mit einer quadratischen Koeffizientenmatrix, die symmetrisch und positiv semidefinit ist. Außerdem sind die Normalgleichungen immer lösbar, d.h. (LS) besitzt mindestens eine Lösung.

(c) Ist $\text{rank}(A) = n$, so ist $A^\top A$ positiv definit, also invertierbar und (LS) besitzt genau eine Lösung $\mathbf{x}_* = (A^\top A)^{-1} A^\top \mathbf{b}$.

Bemerkungen.

(a) Ist x Lösung von (LS) bzw. (NG), so ist

$$b = Ax + r_x$$

die **eindeutige** Zerlegung von b in Komponenten aus $\mathcal{R}(A)$ und $\mathcal{R}(A)^\perp = \mathcal{N}(A^\top)$.

Insbesondere gelten für zwei Lösungen x, y von (LS) bzw. (NG)

$$Ax = Ay \quad \text{und} \quad r_x = r_y.$$

(b) x löst (LS) genau dann, wenn

$$\begin{bmatrix} I_m & A \\ A^\top & O_n \end{bmatrix} \begin{bmatrix} r_x \\ x \end{bmatrix} = \begin{bmatrix} b \\ \mathbf{0} \end{bmatrix}.$$

Zur Lösung eines linearen Ausgleichsproblems bietet sich somit folgendes Vorgehen an:

- 1: Stelle die Normalgleichungen auf.
- 2: Löse die Normalgleichungen durch Gauß-Elimination.

Beachte: $A^\top A$ ist immer symmetrisch. Besitzt A vollen Spaltenrang, so ist $A^\top A$ sogar positiv definit, besitzt also eine Cholesky-Zerlegung.

Allerdings: Weil die Konditionszahl der Koeffizientenmatrix $A^\top A$ von (NG) sehr groß werden kann, auch wenn A selbst nicht extrem schlecht konditioniert ist, ist dieses Verfahren nicht immer zu empfehlen.

Genauer: es gilt $\text{cond}_2(A^\top A) = \text{cond}_2(A)^2$. Hierbei sei die Konditionszahl einer nichtquadratischen Matrizen A durch

$$\text{cond}(A) = \frac{\max_{\|x\|=1} \|Ax\|}{\min_{\|x\|=1, Ax \neq 0} \|Ax\|}$$

(vorläufig) definiert. Man mache sich klar, dass dies eine Verallgemeinerung der bisherigen Definition darstellt.

Beispiel. Bestimme **Ausgleichsgerade** durch $(0, 0)$, $(1, 2)$, $(2, 1)$
bzw. löse das Ausgleichsproblem

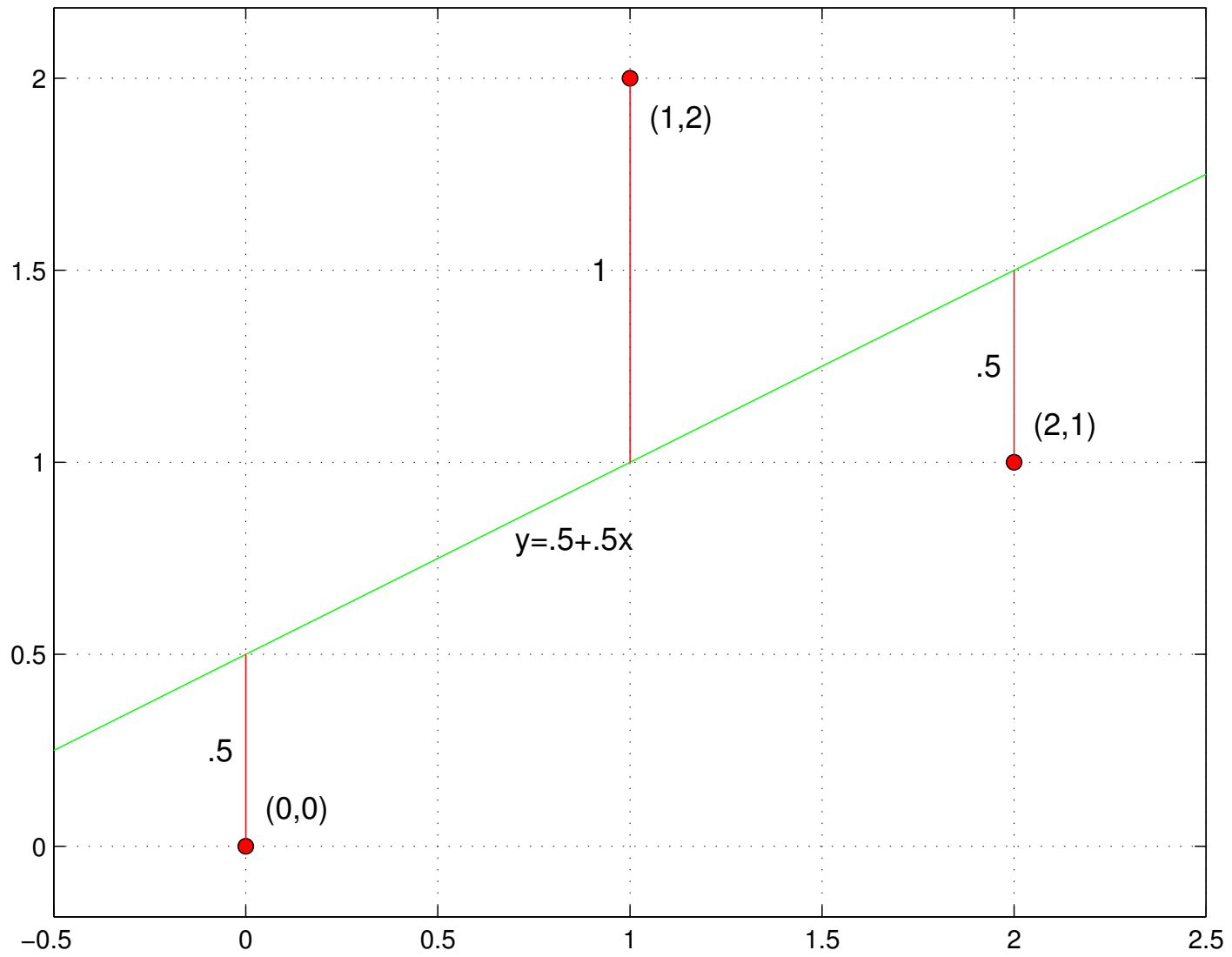
$$\left\| \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} \right\|_2 \rightarrow \min .$$

1. Normalgleichungen: $\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$.

2. Cholesky-Zerlegung: $\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} = LL^T$ mit $L = \begin{bmatrix} \sqrt{3} & 0 \\ \sqrt{3} & \sqrt{2} \end{bmatrix}$.

Ergebnis: $\mathbf{x} = [\alpha_0, \alpha_1]^T = [.5, .5]^T$, die Ausgleichsgerade ist also $y = .5 + .5x$ mit dem **Kleinsten-Quadrate-Fehler**

$$\left[\sum_{i=1}^3 |y_i - (\alpha_0 + \alpha_1 x_i)|^2 \right]^{1/2} = [(0 - .5)^2 + (2 - 1)^2 + (1 - 1.5)^2]^{1/2} = .5 \cdot \sqrt{6}.$$



5.2 Die Singulärwertzerlegung

Satz 5.2. Sei $A \in \mathbb{R}^{m \times n}$ eine Matrix vom Rang r .

Dann gibt es orthogonale Matrizen $U \in \mathbb{R}^{m \times m}$ und $V \in \mathbb{R}^{n \times n}$ sowie eine „Diagonalmatrix“

$$\Sigma = \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix} \in \mathbb{R}^{m \times n} \text{ mit } \Sigma_r = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r) \in \mathbb{R}^{r \times r}$$

und $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$, so dass A die Zerlegung

$$A = U \Sigma V^T \quad (\text{SVD})$$

besitzt.

Die Darstellung (SVD) heißt **Singulärwertzerlegung** von A . Die positiven Zahlen σ_i nennt man die **Singulärwerte** von A . Schreibt man $U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m]$ und $V = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$, so heißen $\mathbf{u}_i \in \mathbb{R}^m$ bzw. $\mathbf{v}_i \in \mathbb{R}^n$ zugehörige **linke** bzw. **rechte Singulärvektoren**.

Bemerkungen.

1. $A = U\Sigma V^\top = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^\top = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r] \Sigma_r [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r]^\top$
 (Darstellung von A als Summe von r Rang-1-Matrizen).

2. Es gelten:

$$A\mathbf{v}_i = \begin{cases} \sigma_i \mathbf{u}_i & (i = 1, 2, \dots, r), \\ \mathbf{0} & (i = r + 1, r + 2, \dots, n) \end{cases} \quad \text{und}$$

$$A^\top \mathbf{u}_i = \begin{cases} \sigma_i \mathbf{v}_i & (i = 1, 2, \dots, r), \\ \mathbf{0} & (i = r + 1, r + 2, \dots, m) \end{cases} .$$

3.

$\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ ist eine ON-Basis von $\mathcal{R}(A)$.

$\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$ ist eine ON-Basis von $\mathcal{N}(A^\top) = \mathcal{R}(A)^\perp$.

$\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ ist eine ON-Basis von $\mathcal{R}(A^\top) = \mathcal{N}(A)^\perp$.

$\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ ist eine ON-Basis von $\mathcal{N}(A)$.

4. $A^\top A = V\Sigma^\top \Sigma V^\top = V \begin{bmatrix} \Sigma_r^2 & O \\ O & O \end{bmatrix} V^\top$, $AA^\top = U\Sigma\Sigma^\top U^\top = U \begin{bmatrix} \Sigma_r^2 & O \\ O & O \end{bmatrix} U^\top$.
 $\sigma_1^2, \dots, \sigma_r^2$ sind die von Null verschiedenen Eigenwerte von $A^\top A$ bzw. AA^\top . Insbesondere sind die Singulärwerte $\sigma_1, \dots, \sigma_r$ durch A eindeutig festgelegt.

Die rechten Singulärvektoren $\mathbf{v}_1, \dots, \mathbf{v}_n$ bilden eine ON-Basis des \mathbb{R}^n aus Eigenvektoren von $A^\top A$:

$$A^\top A \mathbf{v}_i = \begin{cases} \sigma_i^2 \mathbf{v}_i & (i = 1, 2, \dots, r), \\ \mathbf{0} & (i = r + 1, r + 2, \dots, n) \end{cases} .$$

Die linken Singulärvektoren $\mathbf{u}_1, \dots, \mathbf{u}_m$ bilden eine ON-Basis des \mathbb{R}^m aus Eigenvektoren von AA^\top :

$$AA^\top \mathbf{u}_i = \begin{cases} \sigma_i^2 \mathbf{u}_i & (i = 1, 2, \dots, r), \\ \mathbf{0} & (i = r + 1, r + 2, \dots, m) \end{cases} .$$

5. Ist $A \in \mathbb{R}^{n \times n}$ symmetrisch mit von Null verschiedenen Eigenwerten $\lambda_1, \dots, \lambda_r$, $|\lambda_1| \geq \dots \geq |\lambda_r| > 0$, dann sind $\sigma_i = |\lambda_i|$ die Singulärwerte von A .
6. Das Bild der (n -dimensionalen) Einheitskugel unter A ist ein Ellipsoid (im \mathbb{R}^m) mit Mittelpunkt $\mathbf{0}$ und Halbachsen $\sigma_i \mathbf{u}_i$ ($\sigma_i := 0$ für $i > r$).
7. Für $A \in \mathbb{R}^{m \times n}$ gilt $\|A\|_2 = \sigma_1$. Ist $A \in \mathbb{R}^{n \times n}$ invertierbar, gilt außerdem $\|A^{-1}\|_2 = \sigma_n^{-1}$ und $\text{cond}_2(A) = \sigma_1/\sigma_n$.
8. Besitzt $A \in \mathbb{R}^{m \times n}$ die SVD $A = U\Sigma V^\top$, dann besitzt

$$H = \begin{bmatrix} O & A^\top \\ A & O \end{bmatrix} \in \mathbb{R}^{(m+n) \times (m+n)}$$

die von Null verschiedenen Eigenwerte $\pm\sigma_i$ mit zugehörigen (normierten) Eigenvektoren $\frac{1}{\sqrt{2}}[\mathbf{v}_i^\top, \pm\mathbf{u}_i^\top]^\top$.

9. Analoge Aussagen gelten für komplexe Matrizen $A = U\Sigma V^H$ (U, V unitär). (Ersetze in 5. 'symmetrisch' durch 'normal'!)

Geometrische Interpretation der SVD. Besitzt $A \in \mathbb{R}^{m \times n}$ die SVD

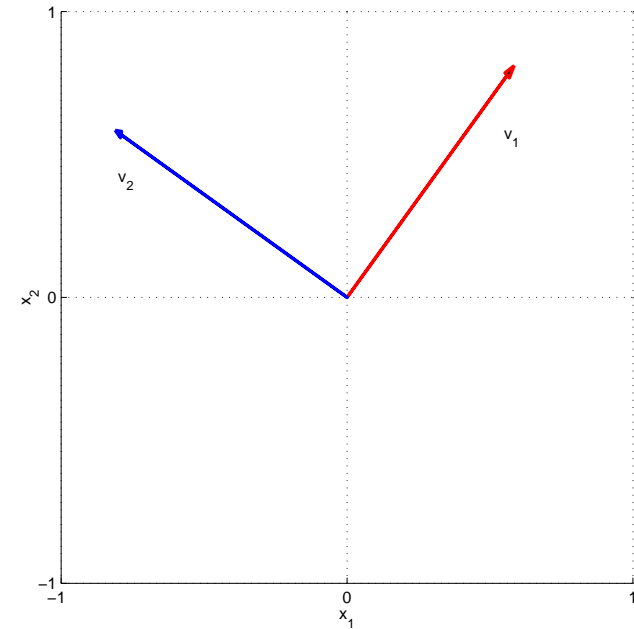
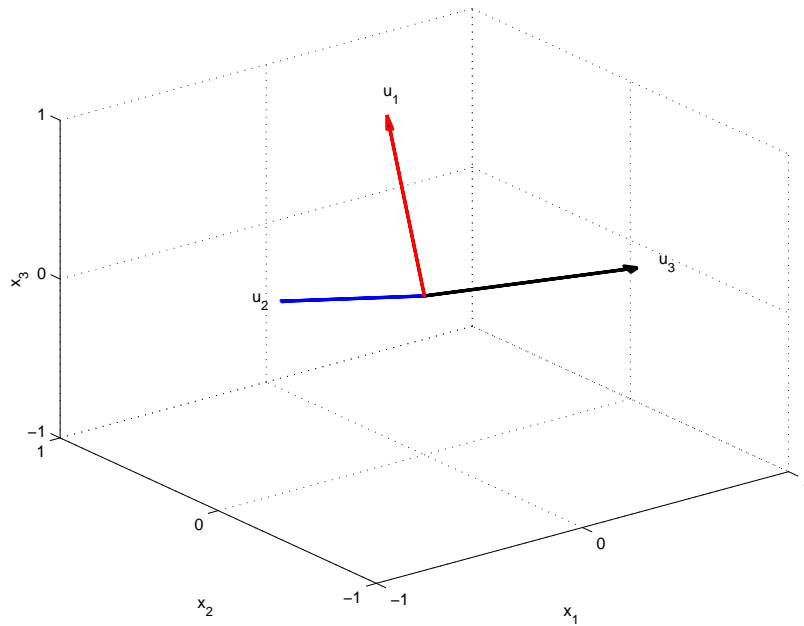
$$A = U \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix} V^\top$$

mit $\Sigma_r = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$, $U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m]$ und $V = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$, so kann man die Abbildungseigenschaften von A (und A^\top) leicht beschreiben (vgl. Bemerkung 2).

Beispiel:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 0.22 & -0.89 & 0.41 \\ 0.52 & -0.25 & -0.82 \\ 0.82 & 0.39 & 0.41 \end{bmatrix} \begin{bmatrix} 2.68 & 0 \\ 0 & 0.92 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0.58 & -0.81 \\ 0.81 & 0.58 \end{bmatrix}^\top$$

(Werte auf 2 Dezimalstellen gerundet).



$$Av_1 = 2.68u_1,$$

$$Av_2 = 0.92u_2,$$

$$A^T u_1 = 2.68v_1,$$

$$A^T u_2 = 0.92v_2,$$

$$A^T u_3 = \mathbf{0}.$$

Satz 5.3. *Es sei $A \in \mathbb{R}^{m \times n}$ eine Matrix vom Rang r mit SVD $A = U \Sigma V^\top = U \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix} V^\top$. Dann löst*

$$\mathbf{x}_* = V \begin{bmatrix} \Sigma_r^{-1} & O \\ O & O \end{bmatrix} U^\top \mathbf{b}$$

die lineare Ausgleichsaufgabe (LS). Darüberhinaus ist \mathbf{x}_ die eindeutig bestimmte Lösung von (LS) mit minimaler Euklid-Norm.*

Satz 5.4 (Schmidt, 1907; Eckart & Young, 1936; Mirsky, 1960). *Für eine Matrix $A \in \mathbb{R}^{m \times n}$ vom Rang r mit SVD $A = U \Sigma V^\top$ besitzt die Approximationsaufgabe*

$$\min\{\|A - B\|_2 : B \in \mathbb{R}^{m \times n} \text{ und } \text{rank}(B) \leq k\}$$

für $k < r$ die Lösung

$$A_k := \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^\top \quad \text{mit} \quad \|A - A_k\|_2 = \sigma_{k+1}.$$

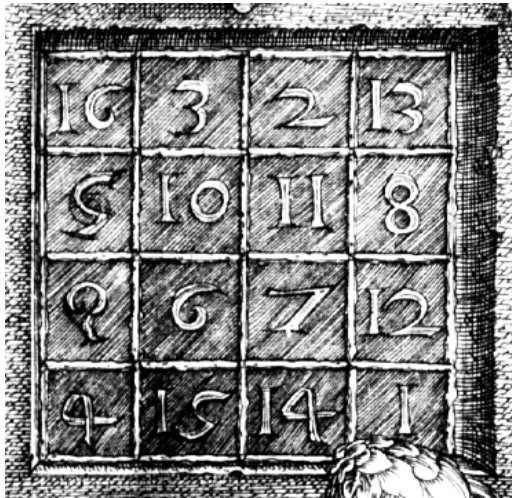
Anwendung der SVD in der **Datenkompression**:

Die folgende Grafik zeigt (links oben) das magische Quadrat aus Albrecht Dürers Melancholie I (1514). Die Bildinformation ist in einer Pixelmatrix X der Dimension 359×371 gespeichert, deren Einträge – ganze Zahlen zwischen 1 und 64 – verschiedene Graustufen repräsentieren. Wir approximieren X durch Matrizen niedrigen Rangs k (vgl. Satz 5.4):

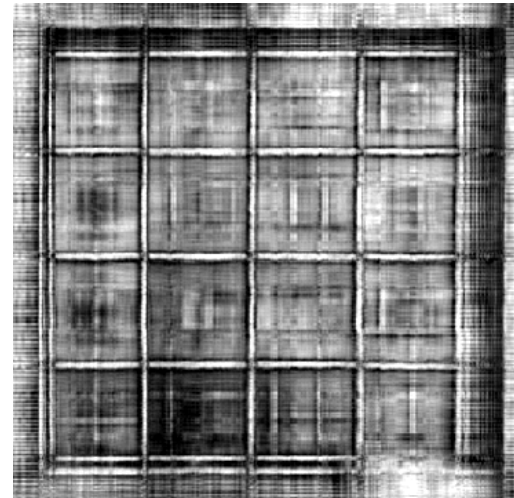
```
load detail.mat;  
[U,S,V]=svd(X);  
X_k=U(:,1:k)*S(1:k,1:k)*V(:,1:k)';  
image(X_k), colormap('gray'), axis('image'), axis('off')
```

Zur Speicherung von X_k sind $k(m + n) = 730k$ Zahlen (statt $mn = 133189$ für X) erforderlich.

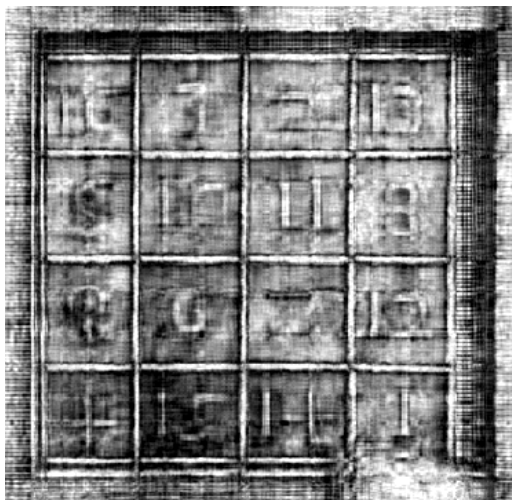
Original



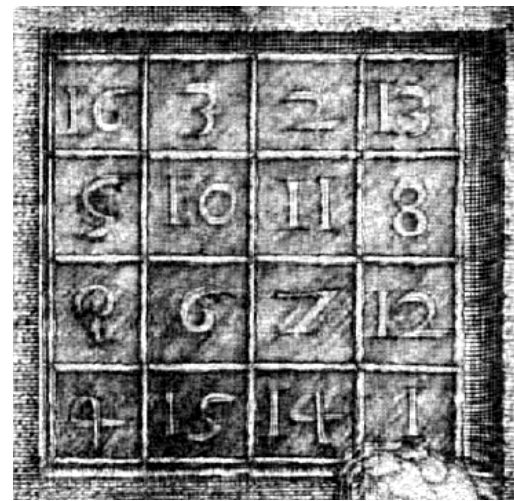
k=10



k=20



k=40



k	Relativer Fehler σ_{k+1}/σ_1	Kompressionsrate
10	0.0666	0.055
20	0.0528	0.110
40	0.0382	0.219

5.3 Die Pseudoinverse

Es sei $A \in \mathbb{R}^{m \times n}$ eine Matrix vom Rang r mit SVD $A = U \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix} V^T$.

Dann heißt

$$A^\dagger := V \begin{bmatrix} \Sigma_r^{-1} & O \\ O & O \end{bmatrix} U^T \in \mathbb{R}^{n \times m}$$

die (Moore-Penrose) Pseudoinverse von A .

Satz 5.5. Für $A \in \mathbb{R}^{m \times n}$ gelten die folgenden Aussagen:

(P1): $A^\dagger A A^\dagger = A^\dagger$,

(P2): $A A^\dagger A = A$,

(P3): $(A A^\dagger)^T = A A^\dagger$,

(P4): $(A^\dagger A)^T = A^\dagger A$.

Darüberhinaus ist $A^\dagger \in \mathbb{R}^{n \times m}$ durch die Eigenschaften (P1) – (P4) eindeutig bestimmt.

Bemerkungen:

- (1) $A^\dagger b$ ist die Kleinste-Quadrate-Lösung von $Ax = b$ mit minimaler Euklid-Norm.
- (2) Für $m \geq n = r$ gilt $A^\dagger = (A^\top A)^{-1} A^\top$.
Ist zusätzlich $m = n$, so folgt $A^\dagger = A^{-1}$.
- (3) Für $n \geq m = r$ gilt $A^\dagger = A^\top (AA^\top)^{-1}$. In diesem Fall löst $A^\dagger b$ das Problem $\|x\|_2 \rightarrow \min$ unter allen x mit $Ax = b$.

Satz 5.6. *Es sei \mathcal{V} ein Unterraum des \mathbb{R}^n . Dann gibt es genau eine Matrix $P = P_{\mathcal{V}} \in \mathbb{R}^{n \times n}$ mit*

$$(a) \mathcal{R}(P) = \mathcal{V}, \quad (b) P^2 = P, \quad \text{und} \quad (c) P^\top = P.$$

P heißt *orthogonale Projektion auf \mathcal{V}* und ist durch

$$Px = \begin{cases} x, & x \in \mathcal{V}, \\ \mathbf{0}, & x \in \mathcal{V}^\perp, \end{cases}$$

eindeutig bestimmt.

Satz 5.7. *Es sei $A \in \mathbb{R}^{m \times n}$ eine Matrix vom Rang r mit SVD*

$$A = U\Sigma V^{\top} = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix} \begin{bmatrix} V_1^{\top} \\ V_2^{\top} \end{bmatrix}$$

($U_1 \in \mathbb{R}^{m \times r}$, $U_2 \in \mathbb{R}^{m \times (m-r)}$, $V_1 \in \mathbb{R}^{n \times r}$, $V_2 \in \mathbb{R}^{n \times (n-r)}$) und der Pseudoinversen A^{\dagger} .

Dann gelten:

$$\begin{aligned} P_{\mathcal{R}(A)} &= AA^{\dagger} = U_1U_1^{\top}, & P_{\mathcal{N}(A)} &= I_n - A^{\dagger}A = V_2V_2^{\top}, \\ P_{\mathcal{R}(A^{\top})} &= A^{\dagger}A = V_1V_1^{\top}, & P_{\mathcal{N}(A^{\top})} &= I_m - AA^{\dagger} = U_2U_2^{\top}. \end{aligned}$$

Beispiel.

$$\begin{aligned}
 \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}^\dagger &= \begin{bmatrix} 0.58 & -0.81 \\ 0.81 & 0.58 \end{bmatrix} \begin{bmatrix} 1/2.68 & 0 & 0 \\ 0 & 1/0.92 & 0 \end{bmatrix} \begin{bmatrix} 0.22 & 0.52 & 0.82 \\ -0.89 & -0.25 & 0.39 \\ 0.41 & -0.82 & 0.41 \end{bmatrix} \\
 &= \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ -3 & 0 & 3 \end{bmatrix}
 \end{aligned}$$

(SVD auf 2 Dezimalstellen gerundet).

$$AA^\dagger = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}, \quad A^\dagger A = I_2.$$

5.4 Orthogonale Matrizen und QR-Zerlegung

Eine Matrix $Q \in \mathbb{R}^{m \times m}$ heißt **orthogonal**, wenn eine der folgenden äquivalenten Bedingungen erfüllt ist:

1. $Q^\top Q = I_m$, d.h. Q ist invertierbar und $Q^{-1} = Q^\top$,
2. $\|Qx\|_2 = \|x\|_2 \forall x \in \mathbb{R}^m$, d.h. Q ist normerhaltend,
3. die Spalten (bzw. Zeilen) von Q bilden eine Orthonormalbasis des \mathbb{R}^m .

Für uns wichtig: Sind $A \in \mathbb{R}^{m \times n}$ beliebig und $Q_1 \in \mathbb{R}^{m \times m}$, $Q_2 \in \mathbb{R}^{n \times n}$ orthogonal, so gilt

$$\|Q_1 A\|_2 = \|A Q_2\|_2 = \|A\|_2$$

und folglich: $\text{cond}_2(Q_1 A) = \text{cond}_2(A Q_2) = \text{cond}_2(A)$.

(Durch orthogonale Transformationen wird die Kondition eines Gleichungssystems nicht verschlechtert!)

Satz 5.8. *Zu jeder Matrix $A \in \mathbb{R}^{m \times n}$ mit $m \geq n$ gibt es eine orthogonale Matrix $Q \in \mathbb{R}^{m \times m}$ und eine obere Dreiecksmatrix $R \in \mathbb{R}^{n \times n}$ mit nichtnegativen Diagonaleinträgen, so dass*

$$A = Q \begin{bmatrix} R \\ O \end{bmatrix} \quad (\text{QR-Zerlegung von } A)$$

(O steht hier für eine Nullmatrix der Dimension $(m - n) \times n$).

Korollar 5.9. *Besitzt die Matrix A aus Satz 5.8 vollen Rang n , dann hat R nur positive Diagonalelemente und ist daher invertierbar. $R^\top R$ ist die Cholesky-Zerlegung von $A^\top A$. Insbesondere ist R dann durch A eindeutig bestimmt. Zerlegt man $Q = [Q_1 \ Q_2]$, $Q_1 \in \mathbb{R}^{m \times n}$, so ist auch $Q_1 = AR^{-1}$ durch A eindeutig festgelegt und es gilt $Q_1 Q_1^\top = P_{\mathcal{R}(A)}$. Für Q_2 kann man jede Matrix wählen, deren Spalten eine ON-Basis von $\mathcal{N}(A^\top)$ bilden.*

Das Beispiel

$$A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} c & -s \\ s & c \end{bmatrix} \begin{bmatrix} 0 & s \\ 0 & c \end{bmatrix}, \quad s^2 + c^2 = 1,$$

zeigt, dass die QR-Zerlegung im Fall $\text{rank}(A) < n$ i.a. nicht eindeutig ist.

Satz 5.10. *Die Matrix $A \in \mathbb{R}^{m \times n}$, $m \geq n$, besitze vollen Rang n und die QR-Zerlegung $A = Q \begin{bmatrix} R \\ O \end{bmatrix}$. Es sei $\mathbf{b} \in \mathbb{R}^m$ und $Q^\top \mathbf{b} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}$ mit $\mathbf{c}_1 \in \mathbb{R}^n$, $\mathbf{c}_2 \in \mathbb{R}^{m-n}$.*

Dann ist $\mathbf{x}^ = R^{-1} \mathbf{c}_1$ die (eindeutig bestimmte) Lösung des Kleinsten-Quadrate-Problems $\|\mathbf{b} - A\mathbf{x}\| \rightarrow \min_{\mathbf{x} \in \mathbb{R}^n}$ und für den Kleinsten-Quadrate-Fehler gilt $\|\mathbf{b} - A\mathbf{x}^*\|_2 = \|\mathbf{c}_2\|_2$.*

Satz 5.11. Die Matrix $A \in \mathbb{R}^{m \times n}$ besitze Rang $r \leq \min\{m, n\}$. Dann gibt es eine Permutationsmatrix $P \in \mathbb{R}^{n \times n}$ und eine orthogonale Matrix $Q \in \mathbb{R}^{m \times m}$ mit

$$AP = Q \begin{bmatrix} R_{1,1} & R_{1,2} \\ O & O \end{bmatrix}.$$

Dabei ist $R_{1,1} \in \mathbb{R}^{r \times r}$ eine obere Dreiecksmatrix mit positiven Diagonalelementen und $R_{1,2} \in \mathbb{R}^{r \times (n-r)}$.

Für $\mathbf{b} \in \mathbb{R}^m$ sei $\begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix} = Q^\top \mathbf{b}$ mit $\mathbf{c}_1 \in \mathbb{R}^r$. Dann ist die Lösungsmenge des Kleinsten-Quadrate-Problems $\|\mathbf{b} - A\mathbf{x}\|_2 \rightarrow \min_{\mathbf{x} \in \mathbb{R}^n}$ durch

$$\left\{ P \begin{bmatrix} R_{1,1}^{-1}(\mathbf{c}_1 - R_{1,2}\mathbf{y}_2) \\ \mathbf{y}_2 \end{bmatrix} : \mathbf{y}_2 \in \mathbb{R}^{n-r} \text{ beliebig} \right\}$$

gegeben. Der zugehörige Kleinste-Quadrate-Fehler beträgt $\|\mathbf{c}_2\|_2$.

In unserem **Beispiel**:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \frac{1}{6} \underbrace{\begin{bmatrix} -2\sqrt{3} & 3\sqrt{2} & \sqrt{6} \\ -2\sqrt{3} & 0 & -2\sqrt{6} \\ -2\sqrt{3} & -3\sqrt{2} & \sqrt{6} \end{bmatrix}}_Q \underbrace{\begin{bmatrix} -\sqrt{3} & -\sqrt{3} \\ 0 & -\sqrt{2} \\ \hline 0 & 0 \end{bmatrix}}_{\begin{bmatrix} R \\ O \end{bmatrix}}$$

$$\mathbf{c} = Q^T \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = -.5 \begin{bmatrix} 2\sqrt{3} \\ \sqrt{2} \\ \sqrt{6} \end{bmatrix}.$$

D.h.: $\mathbf{c}_1 = -.5 [2\sqrt{3}, \sqrt{2}]^T$ und $\mathbf{c}_2 = -.5\sqrt{6}$.

Also ist $\mathbf{x}^* = [\alpha_0, \alpha_1]^T = R^{-1} \mathbf{c}_1 = [.5, .5]^T$ und $\|\mathbf{b} - A\mathbf{x}^*\|_2 = .5\sqrt{6}$.

Bleibt die Frage, wie man eine QR-Zerlegung von A berechnet.

Prinzipielle Vorgehensweise:

Bestimme eine endliche Folge orthogonaler Matrizen Q_1, Q_2, \dots, Q_s :

$$A \xrightarrow{Q_1} A_1 = Q_1 A, \quad A_1 \xrightarrow{Q_2} A_2 = Q_2 A_1, \quad \dots, \quad A_{s-1} \xrightarrow{Q_s} A_s = Q_s A_{s-1},$$

so dass

$$A_s = \begin{bmatrix} R \\ O \end{bmatrix} \quad (\text{mit einer oberen Dreiecksmatrix } R) \text{ gilt.}$$

Dann ist

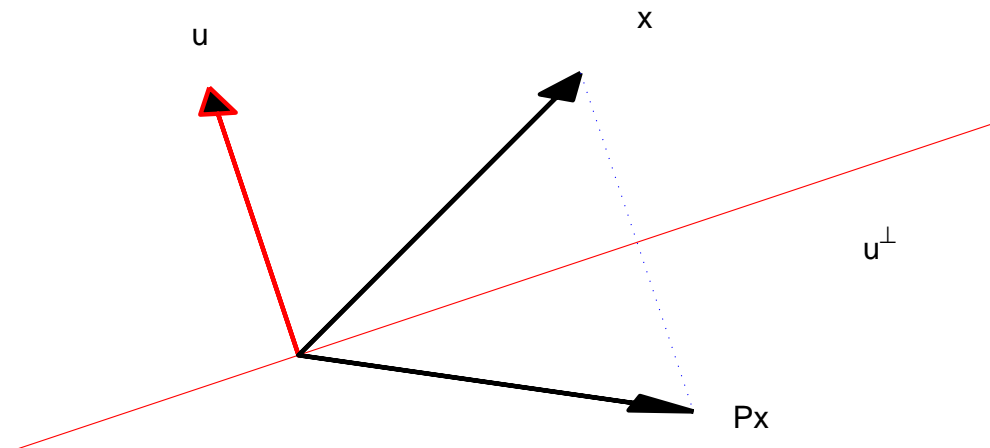
$$A = (Q_s \cdots Q_2 Q_1)^\top A_s = Q_1^\top Q_2^\top \cdots Q_s^\top A_s =: Q \begin{bmatrix} R \\ O \end{bmatrix}$$

die gesuchte QR-Zerlegung von A . (Ein Produkt orthogonaler Matrizen ist orthogonal.)

5.4.1 Householder-Transformationen

Jede Matrix der Form $P = I_m - 2uu^T \in \mathbb{R}^{m \times m}$ mit $u \in \mathbb{R}^m$, $\|u\|_2 = 1$, heißt **Householder-Transformation**.

Wegen $P^T = P$ sowie $P^2 = I_m$ ist P orthogonal. Genauer: P ist die Spiegelung an der Hyperebene $u^\perp := \{v \in \mathbb{R}^m : u^T v = 0\}$ (dem sog. **orthogonalen Komplement** von u).



Gegeben: Ein Vektor $\mathbf{x} = [x_1, \dots, x_m]^\top \in \mathbb{R}^m$, $\mathbf{x} \neq \mathbf{0}$.

Gesucht: Eine Householder-Transformation $P \in \mathbb{R}^{m \times m}$ mit $P\mathbf{x} = \gamma \mathbf{e}_1$
 ($\gamma \in \mathbb{R}$, $\mathbf{e}_1 =$ erster Einheitsvektor).

Lösung: $P = I_m - 2\mathbf{u}\mathbf{u}^\top$, wobei

$$\mathbf{u} = \frac{\tilde{\mathbf{u}}}{\|\tilde{\mathbf{u}}\|_2} \text{ mit } \tilde{\mathbf{u}} = \begin{bmatrix} x_1 + \text{sign}(x_1)\|\mathbf{x}\|_2 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \Rightarrow P\mathbf{x} = \begin{bmatrix} -\text{sign}(x_1)\|\mathbf{x}\|_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Für $A \in \mathbb{R}^{m \times n}$ berechnet man $B = PA = A - \mathbf{u}(2A^\top \mathbf{u})^\top$ durch

$$\mathbf{w} = A^\top \mathbf{u}, \quad \mathbf{w} = 2\mathbf{w}, \quad B = A - \mathbf{u}\mathbf{w}^\top$$

(ohne explizite Kenntnis von P) in $4mn$ flops. P wird niemals explizit gespeichert, sondern nur der Vektor \mathbf{u} .

Wie erzeugt man eine QR-Zerlegung mit Householder-Transformationen?

$$A = \begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \\ x & x & x \end{bmatrix} \xrightarrow{1} \underbrace{\begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & x & x \\ 0 & x & x \end{bmatrix}}_{A_1} \xrightarrow{2} \underbrace{\begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & 0 & x \\ 0 & 0 & x \end{bmatrix}}_{A_2} \xrightarrow{3} \underbrace{\begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & 0 & x \\ 0 & 0 & 0 \end{bmatrix}}_{A_3}$$

Schritt 1. Wähle $P_1 \in \mathbb{R}^{4 \times 4}$ so, dass $P_1 A(:, 1) = \gamma_1 e_1 \in \mathbb{R}^4$.

Bestimme $A_1 = P_1 A$.

Schritt 2. Wähle $P_2 \in \mathbb{R}^{3 \times 3}$ so, dass $P_2 A_1(2 : 4, 2) = \gamma_2 e_1 \in \mathbb{R}^3$.

Bestimme $A_2 = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & P_2 \end{bmatrix} A_1$.

Schritt 3. Wähle $P_3 \in \mathbb{R}^{2 \times 2}$ so, dass $P_3 A_2(3 : 4, 3) = \gamma_3 e_1 \in \mathbb{R}^2$.

Bestimme $A_3 = \begin{bmatrix} I_2 & O \\ O & P_3 \end{bmatrix} A_2$.

Schematisch:

```
for i = 1:n
    bestimme u_i so, dass fuer P_i' = I - 2*u_i*u_i^T
        die Identitaet P_i'*A(i:m,i) = gamma*e_1 gilt
    A(i:m,i:n) = P_i'*A(i:m,i:n)
end
```

A wird hier durch seine QR-Zerlegung überschrieben. Q wird dabei in faktorisierte Form gespeichert: $Q = P_1 P_2 \cdots P_n$. Die einzelnen Matrizen P_i bzw. P_i' werden natürlich nicht explizit gespeichert. Es genügt, den Vektor u_i (genauer \tilde{u}_i) zu speichern. Man speichert \tilde{u}_i in Spalte i der Matrix A unterhalb der Hauptdiagonalen. Ein Zusatzvektor ist erforderlich für die ersten Komponenten der Vektoren \tilde{u}_i .

In unserem **Beispiel**:

$$\tilde{\mathbf{u}}_1 = \begin{bmatrix} A(1, 1) + \mathbf{sign}(A(1, 1)) \|A(:, 1)\|_2 \\ A(2, 1) \\ A(3, 1) \end{bmatrix} = \begin{bmatrix} 1 + \sqrt{3} \\ 1 \\ 1 \end{bmatrix}$$

$\Rightarrow P_1 = I_3 - \frac{2}{\tilde{\mathbf{u}}_1^\top \tilde{\mathbf{u}}_1} \tilde{\mathbf{u}}_1 \tilde{\mathbf{u}}_1^\top$ (wird nicht berechnet). Es folgt

$$A_1(:, 1) = \begin{bmatrix} -\mathbf{sign}(A(1, 1)) \|A(:, 1)\|_2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -\sqrt{3} \\ 0 \\ 0 \end{bmatrix},$$

$$A_2(:, 2) = A(:, 2) - \frac{2\tilde{\mathbf{u}}_1^\top A(:, 2)}{\tilde{\mathbf{u}}_1^\top \tilde{\mathbf{u}}_1} \tilde{\mathbf{u}}_1 = A(:, 2) - \frac{3 - \sqrt{3}}{2} \tilde{\mathbf{u}}_1 = \begin{bmatrix} -\sqrt{3} \\ .5(\sqrt{3} - 1) \\ .5(\sqrt{3} + 1) \end{bmatrix}.$$

$$\tilde{\mathbf{u}}_2 = \begin{bmatrix} A_1(2, 2) + \mathbf{sign}(A_1(2, 2)) \|A_1(2 : 3, 3)\|_2 \\ A_1(3, 2) \end{bmatrix} = \begin{bmatrix} .5(\sqrt{3} - 1) + \sqrt{2} \\ .5(\sqrt{3} + 1) \end{bmatrix}$$

$\Rightarrow P'_2 = I_2 - \frac{2}{\tilde{\mathbf{u}}_2^\top \tilde{\mathbf{u}}_2} \tilde{\mathbf{u}}_2 \tilde{\mathbf{u}}_2^\top$ (wird nicht berechnet). Es folgt

$$A_2(2 : 3, 2) = \begin{bmatrix} -\mathbf{sign}(A_1(2 : 3, 2)) \|A_1(2 : 3, 2)\|_2 \\ 0 \end{bmatrix} = \begin{bmatrix} -\sqrt{2} \\ 0 \end{bmatrix}.$$

Insgesamt

$$\begin{bmatrix} R \\ O \end{bmatrix} = \begin{bmatrix} -\sqrt{3} & -\sqrt{3} \\ 0 & -\sqrt{2} \\ 0 & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & P'_2 \end{bmatrix}^\top P_1^\top \text{ (nicht explizit bekannt).}$$

Benötigt wird auch $Q^T \mathbf{b} = P_n P_{n-1} \cdots P_1 \mathbf{b}$.

```

for i = 1:n
    tau = -2*u_i^T*b(i:m)
    b(i:m) = b(i:m) + tau*u_i
end

```

In unserem **Beispiel**:

$$\mathbf{b} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}, \quad \tilde{\mathbf{u}}_1 = \begin{bmatrix} 1 + \sqrt{3} \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_1 = \frac{1}{\|\tilde{\mathbf{u}}_1\|_2} \tilde{\mathbf{u}}_1 = \frac{1}{\sqrt{6 + 2\sqrt{3}}} \begin{bmatrix} 1 + \sqrt{3} \\ 1 \\ 1 \end{bmatrix}.$$

Es folgt

$$\tau_1 = \frac{-6}{\sqrt{6 + 2\sqrt{3}}}, \quad \mathbf{b}_1 = \mathbf{b} - \tau_1 \mathbf{u}_1 = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} - \frac{3 - \sqrt{3}}{2} \begin{bmatrix} 1 + \sqrt{3} \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\sqrt{3} \\ .5(\sqrt{3} + 1) \\ .5(\sqrt{3} - 1) \end{bmatrix}.$$

$$\mathbf{b}_1(2 : 3) = \begin{bmatrix} .5(\sqrt{3} + 1) \\ .5(\sqrt{3} - 1) \end{bmatrix}, \quad \tilde{\mathbf{u}}_2 = \begin{bmatrix} .5(\sqrt{3} - 1) + \sqrt{2} \\ .5(\sqrt{3} + 1) \end{bmatrix},$$

$$\mathbf{u}_2 = \frac{1}{\|\tilde{\mathbf{u}}_2\|_2} \tilde{\mathbf{u}}_2 = \frac{1}{\sqrt{4 + \sqrt{2}(\sqrt{3} - 1)}} \begin{bmatrix} .5(\sqrt{3} - 1) + \sqrt{2} \\ .5(\sqrt{3} + 1) \end{bmatrix}.$$

Es folgt

$$\tau_2 = \frac{-2 - \sqrt{2} - \sqrt{6}}{\sqrt{4 + \sqrt{2}(\sqrt{3} - 1)}}, \quad \mathbf{b}_2 = \mathbf{b}_1(2 : 3) - \tau_2 \mathbf{u}_2 = \begin{bmatrix} -.5\sqrt{2} \\ -.5\sqrt{6} \end{bmatrix}$$

und insgesamt

$$\mathbf{c} = \mathbf{Q}^T \mathbf{b} = \mathbf{P}_2 \mathbf{P}_1 \mathbf{b} = [-\sqrt{3}, -.5\sqrt{2}, -.5\sqrt{6}]^T.$$

5.4.2 Givens-Rotationen

Eine Givens-Rotation

$$G(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (\theta \in [0, 2\pi))$$

dreht einen Vektor $x \in \mathbb{R}^2$ um den Winkel θ (im Gegenuhrzeigersinn).

Eine Givens-Rotation in der (i, j) -Ebene des \mathbb{R}^n hat (für $i < j$) die Form

$$G(i, j, \theta) = \begin{bmatrix} I_{i-1} & \mathbf{0} & O & \mathbf{0} & O \\ \mathbf{0} & \cos \theta & \mathbf{0} & -\sin \theta & \mathbf{0} \\ O & \mathbf{0} & I_{j-i-1} & \mathbf{0} & O \\ \mathbf{0} & \sin \theta & \mathbf{0} & \cos \theta & \mathbf{0} \\ O & \mathbf{0} & O & \mathbf{0} & I_{n-j} \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

Beachte, dass beim Übergang $x \rightarrow y = G(i, j, \theta)x$ nur die Komponenten i und j verändert werden. Es gilt: $y_k = x_k \forall k \notin \{i, j\}$,

$$\begin{bmatrix} y_i \\ y_j \end{bmatrix} = G(\theta) \begin{bmatrix} x_i \\ x_j \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x_i \\ x_j \end{bmatrix}.$$

Die Matrix $G(i, j, \theta)$ ist orthogonal ($G(i, j, \theta)^{-1} = G(i, j, -\theta) = G(i, j, \theta)^\top$).

Gegeben: i, j, x_i, x_j ($[x_i, x_j] \neq [0, 0]$).

Gesucht: θ mit $\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x_i \\ x_j \end{bmatrix} = \begin{bmatrix} * \\ 0 \end{bmatrix}$.

Lösung:

$$\cos \theta = \frac{x_i}{\sqrt{x_i^2 + x_j^2}}, \quad \sin \theta = \frac{-x_j}{\sqrt{x_i^2 + x_j^2}} \quad \text{mit}$$

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x_i \\ x_j \end{bmatrix} = \begin{bmatrix} \sqrt{x_i^2 + x_j^2} \\ 0 \end{bmatrix}.$$

Die Konstruktion einer QR-Zerlegung von A durch Givens-Rotationen ist jetzt offensichtlich.

Die Reihenfolge, in der Elemente von A 'wegrotiert' werden:

$$\begin{bmatrix} * & * & * \\ 3 & * & * \\ 2 & 5 & * \\ 1 & 4 & 6 \end{bmatrix} .$$

Aufwand: I.a. doppelt so hoch wie bei Householder-Transformationen (eine QR-Zerlegung durch Householder-Transformationen erfordert $2n^2m - n^3/3$ Gleitpunktoperationen). Besitzt A aber bereits viele Nulleinträge unterhalb der Hauptdiagonalen, dann ist eine QR-Zerlegung mit Hilfe von Givens-Rotationen kostengünstiger.

5.5 Die Kondition des linearen Ausgleichsproblems

Satz 5.12. Die Matrix $A \in \mathbb{R}^{m \times n}$, $m \geq n$, besitze vollen Rang n .

\mathbf{x}^* und $\tilde{\mathbf{x}}^*$ seien die Lösungen der beiden Ausgleichsprobleme $\|\mathbf{b} - A\mathbf{x}\|_2 \rightarrow \min_{\mathbf{x} \in \mathbb{R}^n}$ bzw. $\|\tilde{\mathbf{b}} - \tilde{A}\mathbf{x}\|_2 \rightarrow \min_{\mathbf{x} \in \mathbb{R}^n}$. Es sei

$$\varepsilon := \max \left\{ \frac{\|A - \tilde{A}\|_2}{\|A\|_2}, \frac{\|\mathbf{b} - \tilde{\mathbf{b}}\|_2}{\|\mathbf{b}\|_2} \right\} < \frac{1}{\text{cond}_2(A)}$$

(diese Voraussetzung garantiert die Eindeutigkeit von $\tilde{\mathbf{x}}^*$). Dann gilt

$$\frac{\|\mathbf{x}^* - \tilde{\mathbf{x}}^*\|_2}{\|\mathbf{x}^*\|_2} \leq \varepsilon \kappa_{LS} + O(\varepsilon^2) \quad \text{mit } \kappa_{LS} := \frac{2\text{cond}_2(A)}{\cos(\theta)} + \tan(\theta)\text{cond}_2(A)^2.$$

Der Winkel $\theta \in [0, \pi/2]$ ist definiert durch $\sin(\theta) = \|\mathbf{b} - A\mathbf{x}^*\|_2 / \|\mathbf{b}\|_2$.

Interpretation: Ist $\theta = 0$ (oder sehr klein), d.h. der Kleinste-Quadrate-Fehler $\|b - Ax^*\|_2$ ist sehr klein (im Verhältnis zu $\|b\|_2$), dann ist — wegen $\cos(\theta) \approx 1$ und $\tan(\theta) \approx 0$ — $\kappa_{LS} \approx 2\text{cond}_2(A)$.

Ist θ weder sehr klein noch nahe bei $\pi/2$ (z.B. $\theta = \pi/4$), dann ist die Konditionszahl κ_{LS} sehr viel größer: Sie verhält sich ungefähr wie $\text{cond}_2(A)^2$ ($\kappa_{LS} = 2\text{cond}_2(A) + \text{cond}_2(A)^2$ für $\theta = \pi/4$).

Ist schließlich $\theta \approx \pi/2$, d.h. $x^* \approx 0$, so ist κ_{LS} praktisch unbeschränkt ($\tan(\theta) \rightarrow \infty$ für $\theta \rightarrow \pi/2$). Im Grenzfall $\theta = \pi/2$ ist $x^* = 0$ und (fast) jede beliebig kleine Störung von A bzw. b führt auf $\tilde{x}^* \neq 0$ und damit zu einem ‘unendlichen’ relativen Fehler.

5.6 Anwendungen der Ausgleichsrechnung

Hauptanwendung ist das **Anpassen** (Eichen, Kalibrieren) eines **Modells**

$$y = \Phi(x, \alpha)$$

an einen Datensatz $\{(x_j, y_j)\}_{j=1}^m$ (engl. **data fitting**):

$$y_j \approx \Phi(x_j, \alpha), \quad j = 1, \dots, m.$$

Hierbei sind

x : unabhängige Variable (z.B. Ort, Zeit etc.),

y : abhängige Variable,

$\alpha = (\alpha_1, \dots, \alpha_n)$: ein oder mehrere zu bestimmende Parameter.

Typischerweise ist y beobachtbar, die α_i jedoch nicht, bzw. nur indirekt über y .

Hängt das Modell Φ **linear** von den Parametern α ab, d.h. ist

$$\Phi(x, \alpha) = \alpha_1 \phi_1(x) + \cdots + \alpha_n \phi_n(x)$$

mit gegebenen **Regressionsfunktionen** $\{\phi_j\}_{j=1}^n$, so spricht man von einer **linearen Ausgleichsaufgabe**.

Die Forderung

$$\sum_{j=1}^m |y_j - \Phi(x_j, \alpha)|^2 \rightarrow \min_{\alpha \in \mathbb{R}^n}$$

führt auf das lineare Ausgleichsproblem (LS) mit

$$A = \begin{bmatrix} \phi_1(x_1) & \cdots & \phi_n(x_1) \\ \vdots & & \vdots \\ \phi_1(x_m) & \cdots & \phi_n(x_m) \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}.$$

5.6.1 Beispiel: Ausgleichspolynome

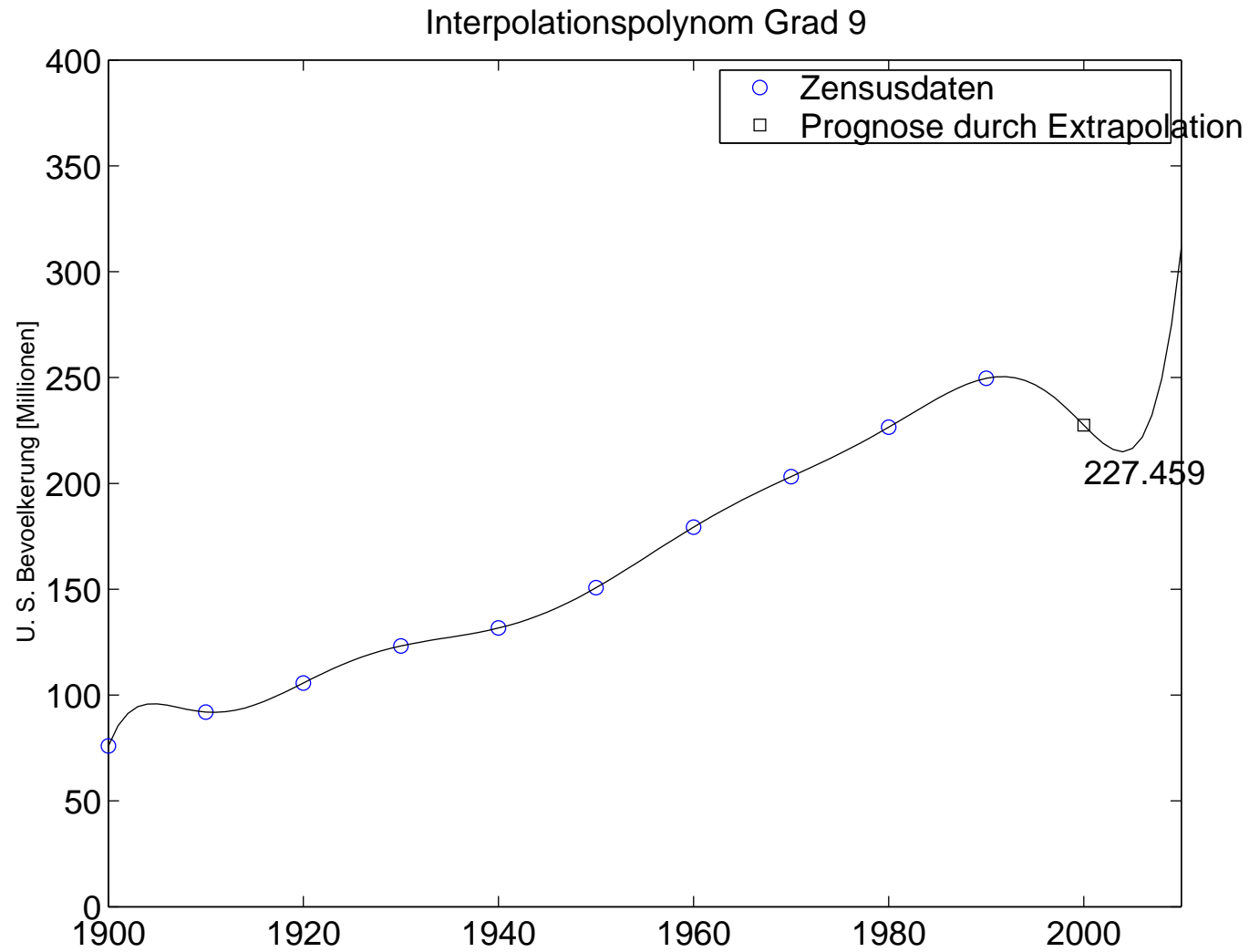
Die folgende Tabelle zeigt die Bevölkerungsentwicklung in den U.S.A.

1900	1910	1920	1930	1940
75.995	91.972	105.711	123.203	131.669
1950	1960	1970	1980	1990
150.697	179.323	203.212	226.505	249.633

(Angaben in Millionen Einwohner).

Wir wollen hieraus eine Vorhersage für die Anzahl der Einwohner im Jahr 2000 bestimmen und werten hierzu das Interpolationspolynom p_9 vom Grad 9 durch die 10 Datenpaare an der Stelle 2000 aus. Wir erhalten

$$p_9(2000) = 227.459 (?!).$$



Extrapolationen, die auf Polynomen hohen Grades beruhen, sind riskant!

Bei Ausgleichspolynomen ist das Modell $\Phi(x, \alpha)$ ein Polynom in x mit Koeffizienten α_i . Nimmt man die Monome $\{x^j\}_{j=0}^n$ als Regressionsfunktionen, so ist

$$\Phi(x, \alpha) = p_n(x) := \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n.$$

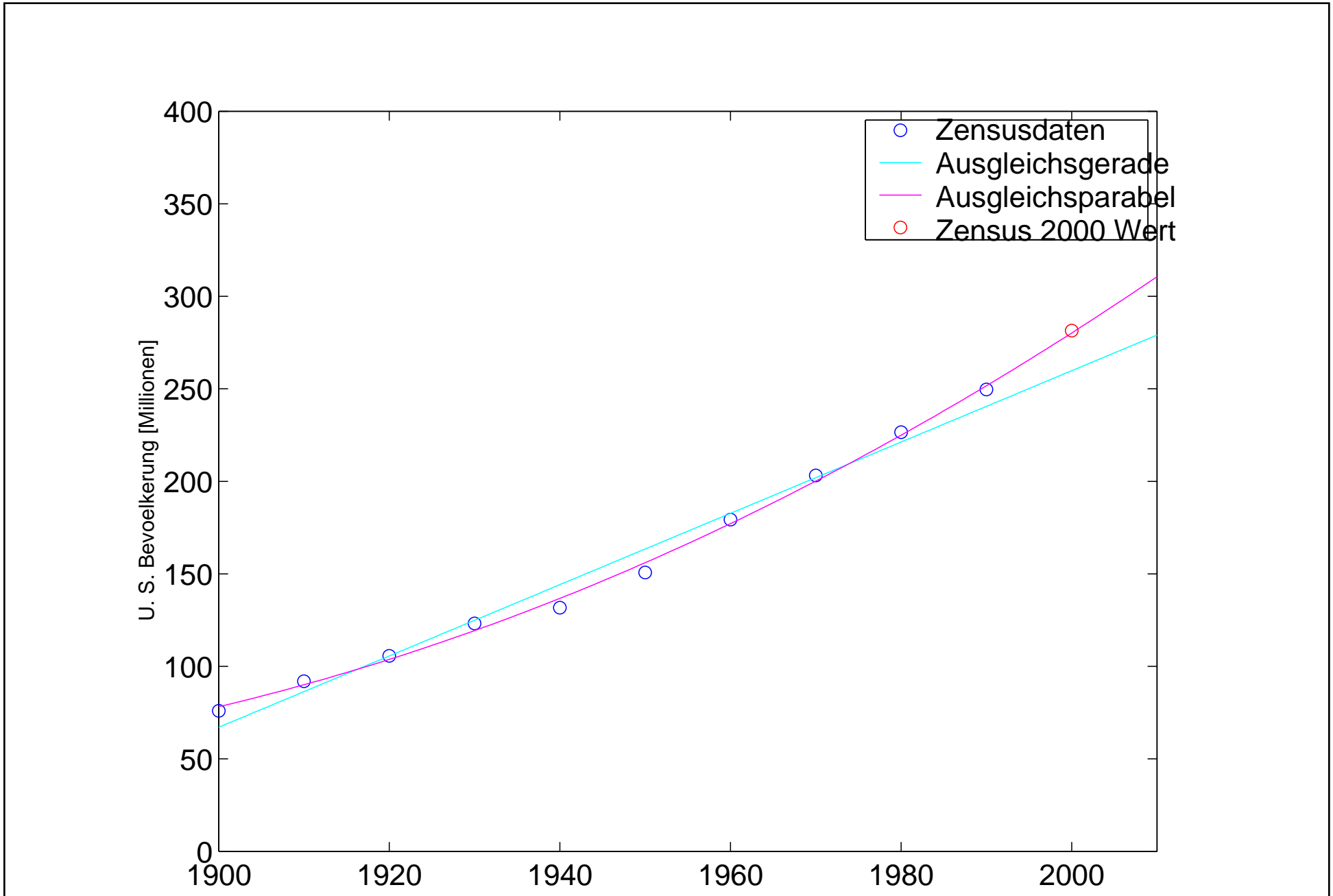
Das Ausgleichspolynom p_n vom Grad n ist somit durch die Forderung

$$\sum_{j=1}^m |y_j - p_n(x_j)|^2 = \min \left\{ \sum_{j=1}^m |y_j - q(x_j)|^2 : q \in \mathcal{P}_n \right\}$$

bestimmt. Dabei bezeichnen $\{(x_j, y_j)\}_{j=1}^m$ die gegebenen Datenpaare.

Für unser Beispiel ergibt sich für p_1 und p_2 sowie die zugehörigen Prognosen

$$p_1(2000) = 259.771 \quad \text{bzw.} \quad p_2(2000) = 280.167.$$



Problem. „Lege“ durch $(m + 1)$ Punkte $\{(x_i, y_i)\}_{i=1}^m$ ein Polynom

$$p_n(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n \in \mathcal{P}_n,$$

wobei $n \leq m$ ($m + 1$ Bedingungen für $n + 1$ Koeffizienten).

Ansatz:

$$\begin{array}{ccccccccccc} \alpha_0 & + & \alpha_1 x_0 & + & \alpha_2 x_0^2 & + & \cdots & \alpha_n x_0^n & = & y_0 \\ \alpha_0 & + & \alpha_1 x_1 & + & \alpha_2 x_1^2 & + & \cdots & \alpha_n x_1^n & = & y_1 \\ \vdots & & & & & & & \vdots & & \vdots \\ \alpha_0 & + & \alpha_1 x_m & + & \alpha_2 x_m^2 & + & \cdots & \alpha_n x_m^n & = & y_m \end{array}$$

oder kürzer

$$A\alpha = y$$

mit der (rechteckigen) Vandermonde-Matrix

$$A = \begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & & & & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix} \in \mathbb{R}^{(m+1) \times (n+1)}.$$

A besitzt Rang $n + 1$, wenn x_0, x_1, \dots, x_n verschieden sind. Dann ist das Ausgleichspolynom **eindeutig** bestimmt. Seine Koeffizienten sind die Lösung des linearen Ausgleichsproblems $\|A\alpha - \mathbf{y}\|_2 \rightarrow \min_{\alpha}$.

Beachte: Ist $m = n$, so ist A quadratisch und invertierbar. Das Polynom p_n **interpoliert** in diesem Fall sämtliche Punkte (x_i, y_i) .

Einfache Spezialfälle:

$n = 0$: (konstantes Polynom)

$$p_0(x) \equiv \alpha_0 \quad \text{mit} \quad \alpha_0 = \frac{1}{m+1} \sum_{i=0}^m y_i.$$

$n = 1$: (Ausgleichsgerade) $p_1(x) = \alpha_0 + \alpha_1 x$

mit

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \frac{1}{D} \begin{bmatrix} \sum_{i=0}^m x_i^2 & -\sum_{i=0}^m x_i \\ -\sum_{i=0}^m x_i & m+1 \end{bmatrix} \begin{bmatrix} \sum_{i=0}^m y_i \\ \sum_{i=0}^m x_i y_i \end{bmatrix}$$

und

$$D = (m+1) \sum_{i=0}^m x_i^2 - \left[\sum_{i=0}^m x_i \right]^2.$$

5.6.2 Statistischer Zugang

Eine gebräuchliche statistische Formulierung der linearen Ausgleichsrechnung schreibt anstelle von

$$b \approx Ax$$

zunächst

$$b = Ax + \xi \quad (5.1)$$

mit einem **Zufallsvektor** $\xi = \xi(\omega) \in \mathbb{R}^m$ („**lineares Modell**“), womit auch der Beobachtungsvektor zu einer Zufallsvariablen $b = b(\omega)$ wird. Ein gemessener Zahlenvektor b wird somit als **Realisierung** (eine von vielen möglichen) dieses Zufallsvektors angesehen.

Die gegebenen Regressionsfunktionen (und damit A) sowie die zu schätzenden Parameter x bleiben deterministische (nicht zufallsbehaftete) Größen.

Durch geeignete Wahl der Wahrscheinlichkeitsverteilung von ξ lassen sich so statistische Annahmen über z.B. Messfehler bei der Beobachtung von b in das Modell einbringen.

Üblich: $\mathbf{E}[\xi] = \mathbf{0}$ (Erwartungswert Null), und somit

$$\mathbf{E}[b] = \mathbf{E}[Ax] + \mathbf{E}[\xi] = Ax,$$

d.h. die Messfehler „mitteln sich von Versuchsreihe zu Versuchsreihe aus“.

Die **Kovarianzmatrix** $C = C_\xi$ von $\xi = [\xi_1, \dots, \xi_m]^\top$ ist definiert durch

$$C = \mathbf{E}[(\xi - \mathbf{E}[\xi])(\xi - \mathbf{E}[\xi])^\top] = \mathbf{E}[\xi\xi^\top] = \left[\mathbf{E}[\xi_i\xi_j] \right]_{i,j=1}^m \in \mathbb{R}^{m \times m}.$$

Wir gehen zunächst von dem einfachen Fall eines **unkorrelierten** ($\mathbf{E}[\xi_i\xi_j] = 0$ für $i \neq j$) und **homoskedastischen** Zufallsvektors, d.h. dass alle Komponenten dieselbe Varianz $\mathbf{E}[\xi_i^2] = \sigma^2$ für alle i besitzen. Insbesondere ist in diesem Fall die Kovarianzmatrix gegeben durch $C = \sigma^2 I$.

Wir bezeichnen die **Varianz** $\mathbf{E}[(\xi - \mathbf{E}[\xi])^2]$ einer Zufallsvariablen ξ mit **Var** $[\xi]$ und die **Kovarianz** eines Zufallsvektors ξ mit **Cov** $[\xi]$.

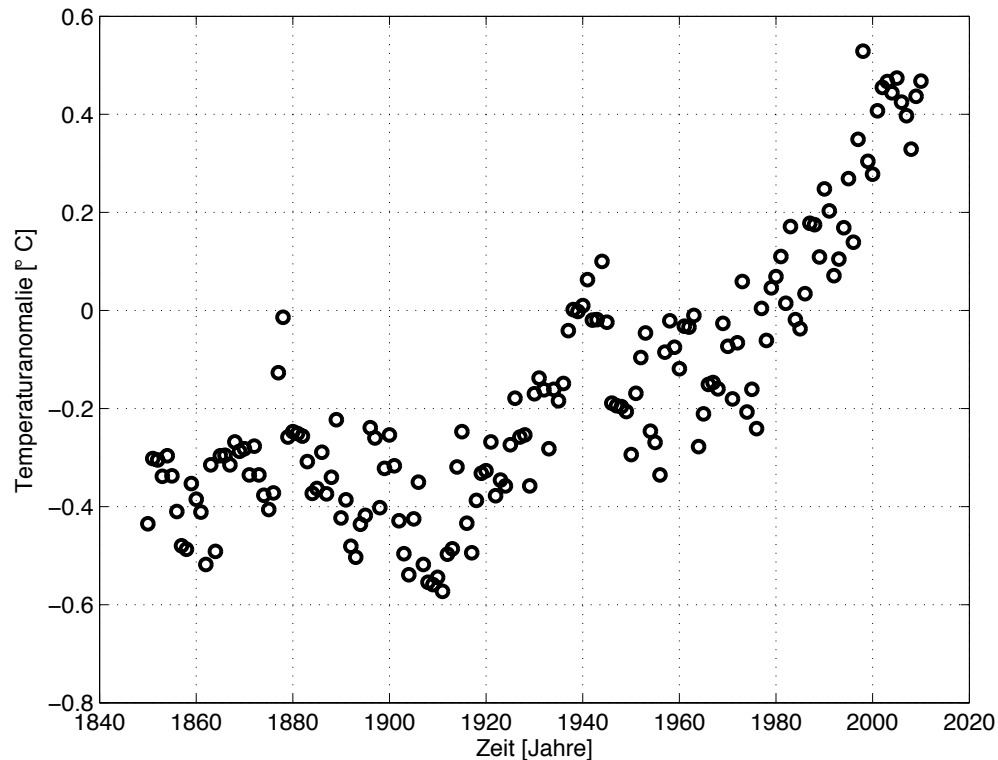
Folgendes Resultat liefert eine statistische Untermauerung der Methode der kleinsten Quadrate.

Definition 5.13. Eine Funktion $\hat{x} = s(\mathbf{b})$ des Zufallsvektors \mathbf{b} heißt *erwartungstreuer Schätzer* für x falls $\mathbf{E}[\hat{x}] = x$. Eine lineare Funktion $s(\mathbf{b}) = S\mathbf{b}$, $S \in \mathbb{R}^{m \times n}$, heißt *linearer Schätzer*.

Definition 5.14. Ein Schätzer \hat{x} heißt *optimal*, wenn seine Varianz minimal ist.

Satz 5.15 (Gauß-Markov-Theorem, [Gauß, 1821]). *Unter allen erwartungstreuen linearen Schätzern für den Vektor x beim linearen Modell (5.1) ist derjenige optimal, der die Euklid-Norm des Datenfehlers $\|\mathbf{b} - A\mathbf{x}\|_2^2$ minimiert, also die kleinste-Quadrate-Lösung.*

Beispiel: globale Temperaturanomalie^a

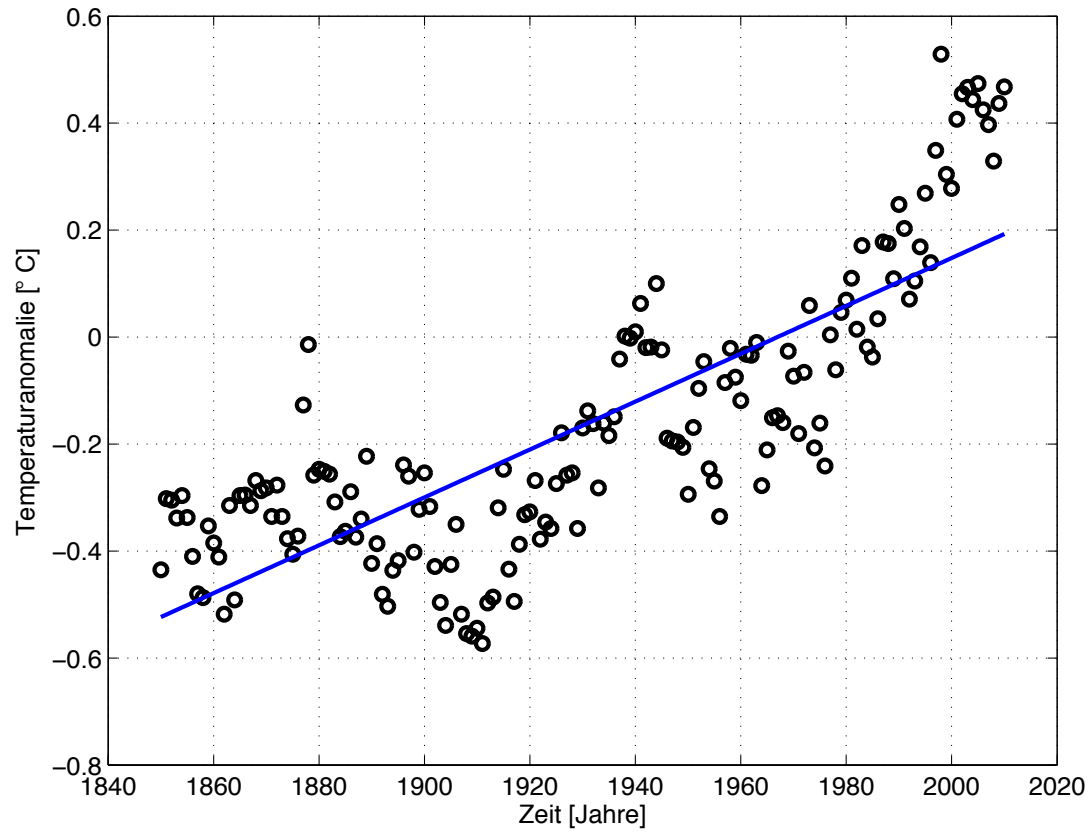


- Abweichung mittlere globale Temperatur vom Mittelwert zwischen 1961 und 1990.
- Lineares Modell:

$$\Phi(t, \alpha) = \alpha_0 + \alpha_1(t - t_0),$$

$$\alpha = (\alpha_0, \alpha_1), t_0 = 1850.$$

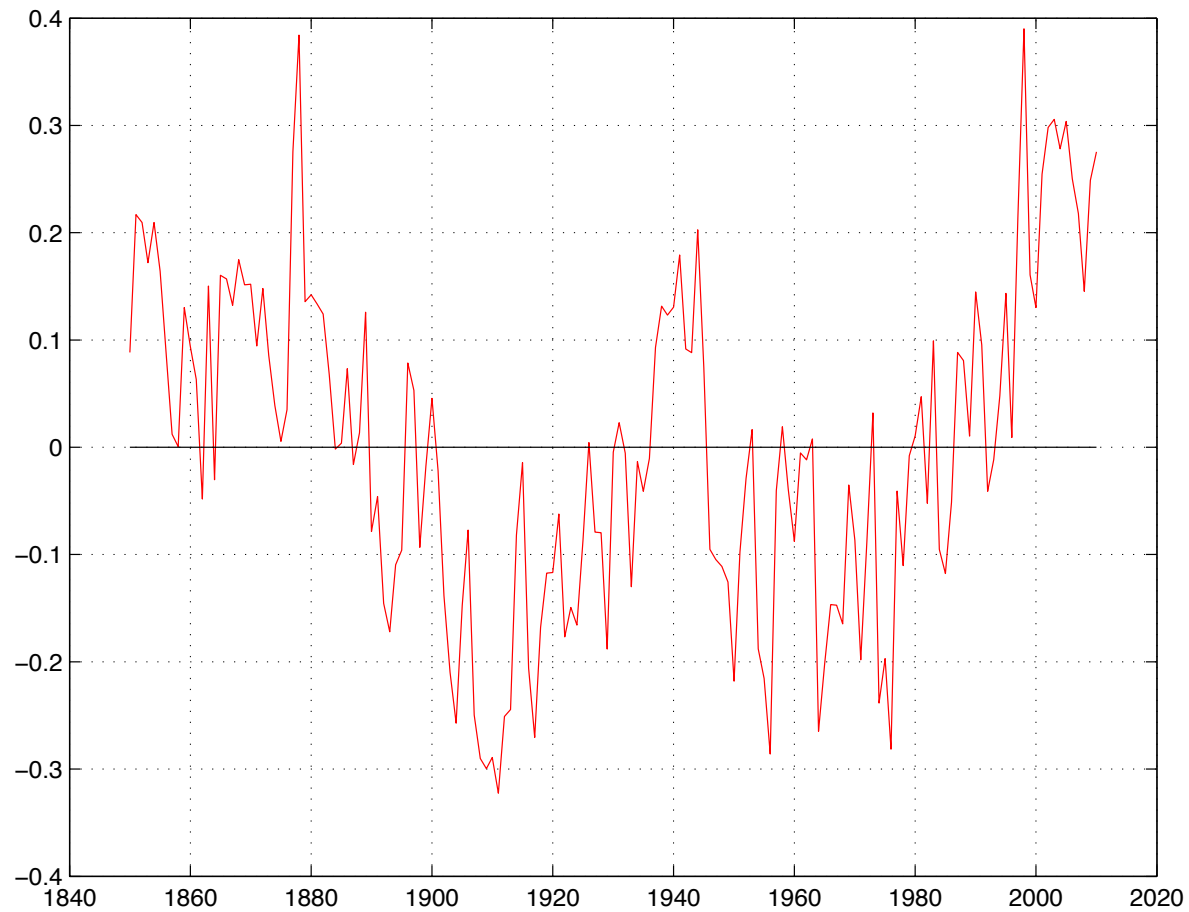
^aQuelle: Bert W. Rust. Fitting Nature's Basic Functions Part I: Polynomials and Linear Least Squares. Computing in Science and Engineering. Sept./Oct. 2001 pp. 84–89.



• Ausgleichsgerade:

$$\alpha_0 = -0.5234,$$

$$\alpha_1 = 0.0045.$$



Residuum: Abweichungen um Null erscheinen nicht zufällig.
Unzulängliches Modell?

- Residuum scheint 4 relative Extrema aufzuweisen.
- Zweiter Versuch: Ausgleichspolynom vom Grad 5:

$$\Phi(t, \alpha) = \sum_{j=0}^5 \alpha_j (t - t_0)^j$$

