

## 5 Überbestimmte Systeme

Ein lineares Gleichungssystem  $A\mathbf{x} = \mathbf{b}$  mit mehr Gleichungen als Unbekannten heißt *überbestimmt*.

In so einem Fall ist  $A$  eine rechteckige  $n \times m$ -Matrix mit  $n > m$ , hat also mehr Zeilen als Spalten. In der Regel hat ein solches System keine exakte Lösung, aber eine eindeutig bestimmte „am wenigsten falsche“ *Ausgleichslösung nach der Methode der kleinsten Quadrate*.

Regelfall:  $m$  linear unabhängige Spalten in  $A$ ,  $m + 1$  linear unabhängige Spalten in der erweiterten Koeffizientenmatrix  $[A, \mathbf{b}]$ .

Sonderfälle:

- $\text{rank } A = \text{rank}[A, \mathbf{b}] = m \rightarrow$  eindeutige exakte Lösung.

Systeme ohne vollen Spaltenrang; die nachfolgend beschriebenen Normalgleichungen sind nicht eindeutig lösbar.

- $\text{rank } A = \text{rank}[A, \mathbf{b}] < m \rightarrow$  unendlich viele exakte Lösungen;
- $\text{rank } A < \text{rank}[A, \mathbf{b}] < m \rightarrow$  keine exakte Lösung, unendlich viele gleichberechtigte Ausgleichslösungen.

Überbestimmte Systeme entstehen beispielsweise bei der Auswertung von Messergebnissen, wenn mehr Messungen vorliegen als Parameter gesucht sind.

### 5.1 Normalgleichungen

#### Überbestimmte Systeme

Ausgleichslösung nach der *Methode der kleinsten Quadrate*: suche jenes  $\mathbf{x}$ , für das die euklidische Norm des Residuenvektors

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

minimal wird. Führt auf die *Normalgleichungen*

$$A^T A\mathbf{x} = A^T \mathbf{b}$$

Herleitung 1: Differenzieren von  $(\|\mathbf{r}\|_2)^2 = \mathbf{r}^T \cdot \mathbf{r}$  nach den einzelnen Komponenten von  $\mathbf{x}$ ; Nullsetzen der Ableitung.

Herleitung 2: Nehmen wir an,  $\hat{\mathbf{x}}$  sei ein Vektor, der  $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$  (also die Normalgleichungen) erfüllt. Wir zeigen nun: für jeden anderen Vektor  $\mathbf{x} \neq \hat{\mathbf{x}}$  gilt

$$\|\mathbf{b} - A\mathbf{x}\|_2 \geq \|\mathbf{b} - A\hat{\mathbf{x}}\|_2.$$

Das heißt, für keinen anderen Vektor gibt es einen kleineres Residuum<sup>10</sup>. In diesem Sinn ist  $\hat{\mathbf{x}}$  eine optimale Ausgleichslösung des Systems  $A\mathbf{x} = \mathbf{b}$ .

Beweis: Man setze  $\hat{\mathbf{r}} = \mathbf{b} - A\hat{\mathbf{x}}$  und  $\mathbf{r} = \mathbf{b} - A\mathbf{x}$ . Dann können wir schreiben

$$\mathbf{r} = \mathbf{b} - A\mathbf{x} = (\mathbf{b} - A\hat{\mathbf{x}}) + A(\hat{\mathbf{x}} - \mathbf{x}) = \hat{\mathbf{r}} + A(\hat{\mathbf{x}} - \mathbf{x})$$

<sup>10</sup>Im Regelfall sind alle Spaltenvektoren in  $A$  linear unabhängig. Dann gilt sogar strikt  $\|\mathbf{b} - A\mathbf{x}\|_2 > \|\mathbf{b} - A\hat{\mathbf{x}}\|_2$  und  $\hat{\mathbf{x}}$  ist die eindeutig bestimmte beste Ausgleichslösung.

Damit berechnen wir das innere Produkt  $\mathbf{r}^T \cdot \mathbf{r}$  und den Zusammenhang mit  $\hat{\mathbf{r}}^T \cdot \hat{\mathbf{r}}$ .

$$\begin{aligned} \mathbf{r}^T \cdot \mathbf{r} &= (\hat{\mathbf{r}} + A(\hat{\mathbf{x}} - \mathbf{x}))^T \cdot (\hat{\mathbf{r}} + A(\hat{\mathbf{x}} - \mathbf{x})) \\ &= \hat{\mathbf{r}}^T \cdot \hat{\mathbf{r}} + \hat{\mathbf{r}}^T \cdot (A(\hat{\mathbf{x}} - \mathbf{x})) + (A(\hat{\mathbf{x}} - \mathbf{x}))^T \cdot \hat{\mathbf{r}} + (A(\hat{\mathbf{x}} - \mathbf{x}))^T \cdot (A(\hat{\mathbf{x}} - \mathbf{x})) \\ &= \hat{\mathbf{r}}^T \cdot \hat{\mathbf{r}} + (\hat{\mathbf{r}}^T A) \cdot (\hat{\mathbf{x}} - \mathbf{x}) + (\hat{\mathbf{x}} - \mathbf{x})^T \cdot (A^T \hat{\mathbf{r}}) + (A(\hat{\mathbf{x}} - \mathbf{x}))^T \cdot (A(\hat{\mathbf{x}} - \mathbf{x})) \end{aligned}$$

Weil  $\hat{\mathbf{x}}$  laut Voraussetzung die Normalengleichungen erfüllt, ist  $A^T \hat{\mathbf{r}} = 0$ , ebenso gilt  $\hat{\mathbf{r}}^T A = 0$ . Es bleibt also

$$\mathbf{r}^T \cdot \mathbf{r} = \hat{\mathbf{r}}^T \cdot \hat{\mathbf{r}} + (A(\hat{\mathbf{x}} - \mathbf{x}))^T \cdot (A(\hat{\mathbf{x}} - \mathbf{x})).$$

Der letzte Term ist (als inneres Produkt eines Vektors mit sich selbst) immer  $\geq 0$ <sup>11</sup>. Damit ist die Aussage bewiesen.

Herleitung 3: Geometrische Veranschaulichung im Fall  $\mathbf{x} \in \mathbb{R}^2$  und  $A$  eine  $3 \times 2$ -Matrix. Die Vektoren aus der Menge  $\{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^2\}$  spannen (wenn die Spalten von  $A$  linear unabhängig sind) eine Ebene im Raum auf. Es soll also  $A\hat{\mathbf{x}}$  jener Punkt (Ortsvektor) auf der Ebene sein, der von  $\mathbf{b}$  minimalen Abstand hat. Der kleinstmögliche Abstand ist der Normalabstand. Der Residuumsvektor  $\mathbf{b} - A\hat{\mathbf{x}}$  steht somit normal auf die Ebene, also auf alle Vektoren der Form  $A\mathbf{x}$ . Aus

$$\forall \mathbf{x} : (A\mathbf{x})^T \cdot (\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{x}^T \cdot (A^T(\mathbf{b} - A\hat{\mathbf{x}})) = 0$$

folgt  $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$ , weil nur der Nullvektor auf alle Vektoren  $\mathbf{x} \in \mathbb{R}^2$  orthogonal sein kann.

## 5.2 Weitere Verfahren

Die Lösung überbestimmter Systeme über die Normalengleichungen ist das klassische Standardverfahren, aber deswegen nicht unbedingt der günstigste Algorithmus. Es ist bloß das einzige, das sich bei kleinen Beispielen mit vertrauten Methoden (Matrixmultiplikation, Elimination) händisch durchrechnen lässt. Programmpakete und Rechenumgebungen (wie MATLAB) verwenden zumeist die *QR*-Zerlegung. Bei Systemen *ohne vollen Spaltenrang* (seien sie überbestimmt oder nicht) ist die Singulärwertzerlegung (*singular value decomposition, SVD*) vorteilhaft. In diesem Fall gibt es nämlich unendlich viele Lösungen (exakt oder im Sinn der kleinsten Quadrate), und SVD liefert automatisch die betragskleinste. Aus der *QR*-Zerlegung lassen sich hingegen die Lösungen mit den meisten Null-Komponenten ablesen.

## 5.3 Beispiele zu überbestimmten Systemen

Gegeben sind drei lineare Gleichungen in zwei Unbekannten,

$$\begin{aligned} 2x + y &= 19 \\ -4x + 4y &= 13 \\ 4x - y &= 17 \end{aligned}$$

Das Gleichungssystem in Matrixschreibweise lautet

$$A\mathbf{x} = \mathbf{b} \text{ mit } A = \begin{bmatrix} 2 & 1 \\ -4 & 4 \\ 4 & -1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 19 \\ 13 \\ 17 \end{bmatrix}$$

<sup>11</sup>Sind alle Spaltenvektoren in  $A$  linear unabhängig, dann ist mit  $\mathbf{x} \neq \hat{\mathbf{x}}$  auch  $A(\hat{\mathbf{x}} - \mathbf{x}) \neq 0$ ; der letzte Term ist dann echt  $> 0$ .

## Methode der Normalgleichungen

Bilde  $A^T \cdot A$  und  $A^T \mathbf{b}$ :

$$A^T \cdot A = \begin{bmatrix} 2 & -4 & 4 \\ 1 & 4 & -1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 \\ -4 & 4 \\ 4 & -1 \end{bmatrix} = \begin{bmatrix} 36 & -18 \\ -18 & 18 \end{bmatrix} = 18 \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$
$$A^T \mathbf{b} = \begin{bmatrix} 2 & -4 & 4 \\ 1 & 4 & -1 \end{bmatrix} \cdot \begin{bmatrix} 19 \\ 13 \\ 17 \end{bmatrix} = \begin{bmatrix} 54 \\ 54 \end{bmatrix} = 18 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

Die Normalgleichungen lauten daher (schon durch 18 gekürzt):

$$\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \text{ oder in ausgeschriebener Form } \begin{array}{rcl} 2x & - & y = 3 \\ -x & + & y = 3 \end{array}$$

Addition der beiden Gleichungen liefert sofort  $x = 6$ , und daraus durch Einsetzen  $y = 9$ .

## Minimaler Fehler in verschiedenen Normen

Die Normalgleichungen liefern  $\mathbf{x} = [6; 9]$ . Einsetzen in die ursprünglichen Gleichungen zeigt aber: diese „Lösung“ erfüllt keine der drei Gleichungen exakt. Der Fehlervektor  $\mathbf{r} = \mathbf{b} - A\mathbf{x}$  lautet

$$\begin{bmatrix} 19 \\ 13 \\ 17 \end{bmatrix} - \begin{bmatrix} 2 & 1 \\ -4 & 4 \\ 4 & -1 \end{bmatrix} \begin{bmatrix} 6 \\ 9 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}$$

Der Fehlervektor hat Länge 3, und das ist die kleinstmögliche Länge. Gemeint ist hier die euklidische Länge, also die Zweinorm.

Ändert man beispielsweise den Vektor  $\mathbf{x}$  geringfügig auf

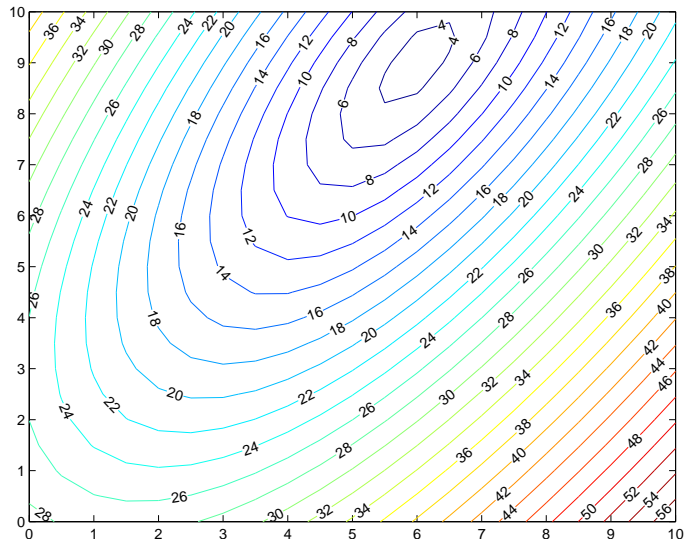
$$\mathbf{x} = \begin{bmatrix} 6 \\ 8.8 \end{bmatrix} = \begin{bmatrix} 6 \\ 44/5 \end{bmatrix}$$

so lautet der Fehlervektor

$$\mathbf{r} = \mathbf{b} - A\mathbf{x} = \begin{bmatrix} -9/5 \\ 9/5 \\ 9/5 \end{bmatrix} = \begin{bmatrix} -1.8 \\ 1.8 \\ 1.8 \end{bmatrix}$$

mit euklidischer Länge  $\|\mathbf{r}\|_2 = \frac{9}{5}\sqrt{3} = 3.1177$ , also deutlich über dem Optimum.

Höhenschichtlinien der Länge des Fehlervektors in der Zweinorm. Das Fehlerminimum bei [6;9] ist deutlich zu erkennen.



In der Unendlichnorm (Maximum der Komponentenbeträge) ist der Fehler nun allerdings geringer: 1.8 statt 2.

Versuchen wir noch einen anderen Vektor  $\mathbf{x}$ :

$$\mathbf{x} = \begin{bmatrix} 6.15 \\ 9.4 \end{bmatrix}, \quad \text{zugehöriger Fehlervektor } \mathbf{r} = \mathbf{b} - A\mathbf{x} = \begin{bmatrix} -2.7 \\ 0 \\ 1.8 \end{bmatrix}$$

$$\|\mathbf{r}\|_2 = 3.2450, \quad \|\mathbf{r}\|_\infty = 2.7$$

Dieser Fehlervektor liegt also sowohl in der 2- als auch der  $\infty$ -Norm über den beiden vorherigen. Misst man aber die Summe der Absolutbeträge (die Einsnorm), so ergibt sich hier der Wert 4.5. Die beiden vorigen Fehlervektoren hatten Einsnormen von 5 bzw. 5.4. In der Einsnorm ist also die hier gewählte Lösung optimal.

### Lösung durch QR-Zerlegung von $A$

$$A = Q \cdot R, \quad \begin{bmatrix} 2 & 1 \\ -4 & 4 \\ 4 & -1 \end{bmatrix} = \begin{bmatrix} -1/3 & 2/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \\ -2/3 & 1/3 & 2/3 \end{bmatrix} \cdot \begin{bmatrix} -6 & 3 \\ 0 & 3 \\ 0 & 0 \end{bmatrix}$$

Das transformierte Gleichungssystem  $R\mathbf{x} = Q^T\mathbf{b}$  ist ebenfalls überbestimmt und lautet ausgeschrieben

$$\begin{aligned} -6x + 3y &= -9 \\ 3y &= 27 \\ 0 &= 3 \end{aligned}$$

Wenn man die ersten beiden Gleichungen exakt löst (wegen Dreiecksform einfach durch Rücksubstitution), liefern sie keinen Beitrag zum Residuum. Die letzte Gleichung hängt nicht von  $\mathbf{x}$  ab. Keine Wahl von  $\mathbf{x}$  kann den Beitrag dieser Gleichungen zum Residuum ändern.

Daher ist die aus den ersten beiden Gleichungen bestimmte Lösung optimal für das transformierte System. Weil die Transformation die Norm des Fehlervektors nicht beeinflusst (wegen Orthogonalität von  $Q$ ), ist diese Lösung auch optimale Lösung von  $A \cdot \mathbf{x} = \mathbf{b}$ .

### Weiteres Beispiel: Lösung mit Singulärwertzerlegung

Ein anderes überbestimmtes System lautet

$$\mathbf{Ax} = \mathbf{b} \text{ mit } A = \begin{bmatrix} 14 & -2 \\ -4 & 22 \\ 16 & -13 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -80 \\ 40 \\ -145 \end{bmatrix}$$

Die Singulärwertzerlegung von  $A$  ist

$$A = U \cdot S \cdot V^T, \quad \begin{bmatrix} 14 & -2 \\ -4 & 22 \\ 16 & -13 \end{bmatrix} = \begin{bmatrix} -1/3 & -2/3 & -2/3 \\ 2/3 & -2/3 & 1/3 \\ -2/3 & -1/3 & 2/3 \end{bmatrix} \cdot \begin{bmatrix} 30 & 0 \\ 0 & 15 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} -3/5 & -4/5 \\ 4/5 & -3/5 \end{bmatrix}^T$$

Das transformierte System lautet in diesem Fall

$$S \cdot \mathbf{y} = U^T \cdot \mathbf{b}, \quad \begin{bmatrix} 30 & 0 \\ 0 & 15 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 150 \\ 75 \\ -30 \end{bmatrix}$$

Wegen der Diagonalgestalt von  $S$  ist die Optimallösung direkt ablesbar:  $\mathbf{y} = [5; 5]$ . Die Lösung des Originalsystems erhält man über die Beziehung

$$\mathbf{x} = V \cdot \mathbf{y} = \begin{bmatrix} -3/5 & -4/5 \\ 4/5 & -3/5 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 5 \end{bmatrix} = \begin{bmatrix} -7 \\ 1 \end{bmatrix}$$

Wenn  $QR$ -Zerlegung oder SVD bereits gegeben sind, ist die Lösung von (überbestimmten) Gleichungssystemen relativ einfach. Der eigentliche Arbeitsaufwand steckt im Berechnen der Zerlegungen. Auf die dabei verwendeten Verfahren kann im Rahmen der Vorlesung nicht eingegangen werden.

### MATLAB-Befehle

$QR$ -Zerlegung einer Matrix $A$ :	<code>[Q R]=qr(A)</code>
SVD, Singulärwertzerlegung $A = U \cdot S \cdot V^T$	<code>[U S V]=svd(A)</code>
Lösung eines überbestimmten Gleichungssystems	
... durch $QR$ -Zerlegung	<code>A\b</code>
... durch SVD	<code>pinv(A)*b</code>

## 5.4 Anpassen eines linearen Modells (einer Ausgleichsebene)

Dieser Abschnitt erläutert anhand eines weiteren Beispiels die Lösung überbestimmter Systeme. Vom Thema her überschneidet er sich mit Kapitel 6; dort werden weitere Methoden zur Approximation von Daten behandelt. Was das Beispiel hier illustrieren soll: die unterschiedlichen Größenordnungen bei den Zwischenergebnissen und der daraus resultierende Einfluss der Rundungsfehler im Vergleich zwischen der Methode der Normalengleichung und der  $QR$ -Zerlegung.

Für die Blies (einen Nebenfluss der Saar) sollen die Hochwasserstände am Pegel Neunkirchen aus den Wasserständen des Pegels Ottweiler und des Pegels Hangard vorhergesagt werden. Es liegen die Daten der Scheitelwasserstände von 12 Winterhochwässern aus den Jahren 1963–1971 vor:

Wasserstand in cm												
Neunkirchen $y$	172	309	302	283	443	298	319	419	361	267	337	230
Ottweiler $x_1$	93	193	187	174	291	184	205	260	212	169	216	144
Hangard $x_2$	120	258	255	238	317	246	265	304	292	242	272	191

Daten-Quelle: U. Maniak, Hydrologie und Wasserwirtschaft, Springer, 1988

Wir unterstellen den Daten das lineare Modell  $a_0 + a_1x_1 + a_2x_2 = y$ . In diesem Ansatz sind  $a_0, a_1$  und  $a_2$  unbekannte Koeffizienten, die aus den zwölf gegebenen Werte-Tripeln möglichst gut bestimmt werden sollen. Geometrisch lassen sich die Tripel  $(x_1|x_2|y)$  als Punkte im Raum interpretieren. Das lineare Modell entspricht dann einer Ebene, die möglichst gut an die Datenpunkte angepasst werden soll.

Einsetzen der Daten in den Ansatz liefert das Gleichungssystem

$$\begin{array}{rclcl}
 a_0 & + & 93a_1 & + & 120a_2 & = & 172 \\
 a_0 & + & 193a_1 & + & 258a_2 & = & 309 \\
 a_0 & + & 144a_1 & + & 191a_2 & = & 230 \\
 & & \vdots & & & & \vdots \\
 a_0 & + & 187a_1 & + & 255a_2 & = & 302
 \end{array}$$

Es liegt somit ein überbestimmtes Gleichungssystem  $A\mathbf{x} = \mathbf{b}$  vor, mit  $12 \times 3$ -Matrix  $A$  und rechter Seite  $\mathbf{b}$ ,

$$A = \begin{bmatrix} 1 & 93 & 120 \\ 1 & 193 & 258 \\ 1 & 187 & 255 \\ \vdots & \vdots & \vdots \\ 1 & 144 & 191 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 172 \\ 309 \\ 230 \\ \vdots \\ 302 \end{bmatrix}$$

Der klassische Weg zur Ausgleichslösung nach der Methode der kleinsten Quadrate führt über die Normalgleichungen  $A^T \cdot A\mathbf{x} = A^T\mathbf{b}$ , in diesem Beispiel

$$\begin{bmatrix} 12 & 2328 & 3000 \\ 2328 & 480202 & 609985 \\ 3000 & 609985 & 780572 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 3740 \\ 766996 \\ 975996 \end{bmatrix}$$

Es treten recht unterschiedlich große Zahlen in diesem System auf (Größenordnung  $10^1$  bis  $10^6$ ). Das ist typisch für Normalgleichungen und ein Indiz dafür, dass dieses System empfindlich gegenüber Rundungsfehlern ist. Numerisch günstiger, aber praktisch nur rechnergestützt durchführbar ist die  $QR$ -Zerlegung. Das transformierte System  $R\mathbf{x} = Q^T\mathbf{b}$  lautet hier

$$\begin{bmatrix} -3.4641 & -672.0357 & -866.0254 \\ 0 & -169.0266 & -165.5656 \\ 0 & 0 & -56.2141 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} -1079.6 \\ -245.14 \\ -7.2659 \\ -2.4029 \\ \vdots \\ -11.003 \end{bmatrix}$$

Die ersten drei Gleichungen liegen als System in oberer Dreiecksform vor und sind durch Rücksubstitution exakt auflösbar. Die restlichen neun Gleichungen von der Form  $0 = -2.4029, \dots$  sind klarerweise unlösbar. Sie liefern den Beitrag zum Restvektor.

locker gesagt: *QR*-Zerlegung transformiert ein überbestimmtes System in ein exakt lösbares System in Dreiecksform und einen unlösbaren Rest. Ignorieren der unlösbaren Gleichungen liefert die bestmögliche Lösung im Sinn der kleinsten Quadrate

Anmerkung: Es ist zwar unmittelbar einsichtig, dass obiges Rezept die Lösung mit betragsmäßig minimalem Restvektor für das *transformierte* System liefert. Allerdings hat das *Originalsystem* einen anderen Restvektor. Die eigentliche Pointe des Verfahrens ist, dass die Transformation von dem einen zum anderen Restvektor durch Multiplikation mit einer *orthogonalen* Matrix passiert. Multiplikation mit einer orthogonalen Matrix lässt den Betrag eines Vektors unverändert. Daher hat auch der Restvektor des Originalsystems minimalen Betrag.

In MATLAB verwendet der Standard-Gleichungslöser-Befehl `A\b` im Falle eines überbestimmten Systems automatisch das *QR*-Verfahren und liefert hier

```
>> A\b
ans =
    22.5505
     1.3237
     0.1293
```

(Die aufwändigere Lösung mit Singulärwertzerlegung (MATLAB: `pinv(A)*b`) liefert hier das gleiche Resultat. )

Das Modell zur Vorhersage der Hochwasser-Scheitelwerte lautet somit

$$y = 22.5505 + 1.3237x_1 + 0.1293x_2$$

Man erkennt, dass die  $x_2$ -Daten gut zehnfach weniger Einfluss auf das Vorhersagemodell haben (weil der entsprechende Koeffizient im linearen Modell nur 0.1293 im Vergleich zu 1.3237 beiträgt). Eine mögliche Interpretation wäre, dass der Wasserstand in Neunkirchen hauptsächlich vom Pegel Ottweiler und nicht wirklich vom Pegel Hangard abhängt.

Wie vertrauenswürdig ist dieses Modell? Einsetzen der Datenpunkte liefert den Restvektor.

Messwert $y$	172	309	302	283	443	298	319	419	361	267	337	230
Modellvorhersage	161	311	303	284	449	298	328	406	341	278	344	238
Residuum	11	-2	-1	-1	-6	0	-9	13	20	-11	-7	-8

Der mittlere absolute Fehler liegt bei 7,3 cm, der maximale Fehler allerdings bei 20 cm.

Eine genauere Beurteilung des Modells im Hinblick auf

- die Richtigkeit des angenommenen linearen Zusammenhangs,
- Einfluss der einzelnen Modellgrößen
- Wahrscheinlichkeit, dass der Prognosewert mit gewisser Genauigkeit zutrifft
- mögliche Ausreißer

erfordert Methoden der multivariaten Statistik und der Regressionsanalyse (und wohl auch eine größere Datenmenge).

## 6 Approximation von Daten, Ausgleichsrechnung

### 6.1 Lineare Datenmodelle

Ein Beispiel dazu hat schon Kapitel 5.4 gebracht. Die Vorlesungsfolien und die Übungen bringen weitere Beispiele dazu. Das best-angepasste Modell ergibt sich dabei immer aus der kleinste-Quadrate-Näherung an ein überbestimmtes Gleichungssystem.

Aber nicht immer liefert die Methode der kleinsten Fehlerquadrate eine plausible Anpassung. Einige wenige grob falsche Werte in den Daten („Ausreißer“) können das Ergebnis gewaltig verzerren. Im Kapitel 6.6 wird eine robuste Methode vorgestellt. MATLAB bietet in seinen Toolboxen verschiedene Methoden zur robusten Anpassung (*robust fit*) an.

### 6.2 Polynomiale Regression

Hier handelt es sich um einen wichtigen Spezialfall der linearen Datenmodelle, für den sich die allgemeinen Formeln etwas vereinfachen.

Eine typische Aufgabe zu diesem Abschnitt könnte lauten: Gegeben sind Messergebnisse für einen Satz von Temperaturwerten  $T$  und die entsprechenden Widerstandswerte  $R$  eines Temperaturfühlers. Der Zusammenhang zwischen  $R$  und  $T$  lässt sich näherungsweise in der Form  $R = a + bT + cT^2$  beschreiben. Wenn für genau drei (verschiedene)  $T$ -Werte Daten vorliegen, lassen sich die drei Parameter  $a, b$  und  $c$  eindeutig bestimmen. Es ist aber sinnvoll, mehr Messungen durchzuführen, damit Messfehler der Einzelmessungen nicht so stark ins Gewicht fallen. Die Parameter  $a, b$  und  $c$  werden dann durch *Ausgleichsrechnung* (man sagt auch *Regression*) bestimmt.

#### Polynomiale Regression (Ausgleich durch ein Polynom)

Gegeben:  $m + 1$  Wertepaare  $(x_i, y_i), i = 0, \dots, m$

Gesucht:  $p(x)$ , ein Polynom  $n$ -ten Grades,  $n < m$ , so dass die Summe der Fehlerquadrate

$$\sum_{i=0}^m (p(x_i) - y_i)^2$$

minimal wird. Locker formuliert:  $y = p(x)$  approximiert möglichst gut die Datenpunkte.

Nicht immer ist ein Polynomansatz ein geeignetes Modell, aber oft lässt sich ein scheinbar komplizierteres Modell auf ein polynomiales Modell zurückführen (Beispiele in den Übungen).

Direkter Lösungsweg: Ansatz des Polynoms mit unbestimmten Koeffizienten,

$$p(x) = a_0 + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1} + a_nx^n.$$

Einsetzen der gegebenen Wertepaare führt auf ein System von  $m + 1$  linearen Gleichungen in den  $n + 1$  unbekanntem Koeffizienten  $a_0, a_1, \dots, a_n$ .

Sofern  $n < m$ , liegt ein *überbestimmtes System* vor.

- klassische Lösung: Näherung nach der Methode der Normalgleichungen: lässt sich bei kleinen Beispielen mit Papier und Stift rechnen; bei großen Datenmengen Gefahr von Rundungsfehlern.

Die Normalgleichungen sind aber nur dann eindeutig lösbar, wenn in den insgesamt  $m + 1$  Wertepaaren mindestens  $n + 1$  der  $x$ -Werte verschieden sind.



- moderner Lösungsweg: *QR*-Zerlegung. Praktisch nur am Rechner durchführbar. Bessere Konditionszahl, weniger anfällig für Rundungsfehler.

Dieser Lösungsweg (Ansatz mit unbestimmten Koeffizienten, Aufstellen des überbestimmten Systems, Bilden der Normalgleichungen) lässt sich für die polynomiale Regression etwas abkürzen. Setzt man

$$s_0 = m + 1, \quad t_0 = \sum_{i=0}^m y_i$$

und

$$s_k = \sum_{i=0}^m x_i^k, \quad t_k = \sum_{i=0}^m x_i^k y_i \quad \text{für } k > 0,$$

so lassen sich, wie man leicht herleiten kann, die Normalgleichungen in folgender Gestalt schreiben:

$$\begin{aligned} s_0 a_0 + s_1 a_1 + \dots + s_n a_n &= t_0 \\ s_1 a_0 + s_2 a_1 + \dots + s_{n+1} a_n &= t_1 \\ &\dots \\ s_n a_0 + s_{n+1} a_1 + \dots + s_{2n} a_n &= t_n \end{aligned}$$

Die Normalgleichungen können für größere  $n$  ziemlich schlecht konditioniert sein. Bessere Resultate lassen sich in solchen Fällen durch Entwicklung nach orthogonalen Polynomen erzielen.

Verwendet man beispielsweise als Datensatz die Punkte

$$(x, \exp(x)) \quad \text{für } x = 0; 0,01; 0,02; \dots; 3,99; 4$$

und approximiert die Daten durch Regressionspolynome verschieden hohen Grades (Rechnung mit vierzehnstelliger Genauigkeit), so wird die Güte der Approximation vorerst mit steigendem Grad besser. Ab dem dreizehnten Grad aber wachsen die Fehler wieder an. Das aus den Normalgleichungen berechnete Polynom 25-ten Grades hat kaum mehr Ähnlichkeit mit der approximierten Funktion. Verwendet man orthogonale Polynome (Tschebyscheff-Polynome), treten diese numerischen Probleme nicht auf.

### 6.3 Ausgleichsgerade

Ein wichtiger Spezialfall der obigen Problemstellung: An  $m + 1$  (mehr als zwei) gegebene Datenpunkte soll eine Gerade mit Gleichung  $y = a + bx$  so angepasst werden, dass die Summe der Fehlerquadrate minimal wird.

$$a = \frac{s_2 t_0 - s_1 t_1}{s_0 s_2 - s_1^2}$$

$$b = \frac{s_0 t_1 - s_1 t_0}{s_0 s_2 - s_1^2}$$

Das Finden einer Ausgleichsgeraden wird oft auch als „lineare Regression“ bezeichnet. Das ist aber ein mißverständlicher Terminus, weil er leicht zur Verwechslung mit „linearen Datenmodellen“ führt.

## 6.4 Nichtlineare Datenmodelle

Dazu gibt es Material auf den Vorlesungsfolien und ausführlicher in den Übungsunterlagen. Kurzfassung: Jacobi-Matrix bilden und iterieren. Im Unterschied zum Newton-Verfahren, das Sie schon kennen, ist das lineare Gleichungssystem mit der Jacobimatrix nun überbestimmt und wird näherungsweise im Sinn der kleinsten Fehlerquadrate gelöst.

Dieses Verfahren wird als *Gauß-Newton-Verfahren* bezeichnet.

Bei nichtlinearen Ausgleichsproblemen gibt es neben dem Gauß-Newton-Verfahren noch weitere Verfahren (z. B. Levenberg-Marquardt-Algorithmus). Damit beschäftigt sich die Optimierung als Teilgebiet der Angewandten Mathematik. Dieses Skript kann nicht näher darauf eingehen.

## 6.5 Warum „kleinste Quadrate“

Die Methode der kleinsten Quadrate minimiert die euklidische Länge des Residuenvektors. Man kann aber die Größe des Residuenvektors durchaus auch anders messen und entsprechend andere Minimalbedingungen fordern. Wichtige Beispiele: Die Summe der *Absolutbeträge* der Fehler soll minimal werden, oder der *maximale Fehler* soll minimal werden („Minimax-Approximation“). Zwei Gründe sprechen für die kleinsten Quadrate:

- Einfache Herleitung und Durchführung: die Minimalwert-Aufgabe lässt sich mit elementarer Differentialrechnung lösen und führt auf ein einfaches algebraisches Problem
- Statistische Überlegungen: Wenn die Daten mit unabhängigen, zufälligen, normalverteilten Fehlern mit gleicher Standardabweichung behaftet sind, sind kleinste Quadrate in gewissem Sinn optimal (genauer: Die Methode liefert eine *maximum likelihood*-Schätzung der Parameter). Umgekehrt gilt aber: wenn die Fehler in den Daten *nicht* normalverteilt etc. sind, dann sind kleinste Quadrate unter Umständen ziemlich schlecht; siehe unten.
- Weitere statistische Eigenschaften: Wenn man eine Abschätzung für die Genauigkeit der Daten hat, kann man auf die Genauigkeit der berechneten Modellparameter schließen.

Angenommen, die Daten sind so skaliert, dass die Varianz der Messfehler gleich 1 ist. Sei  $C = (A^T A)^{-1}$  die inverse Matrix des Systems der Normalgleichungen. Die Diagonalelemente von  $C$  sind die Varianzen der entsprechenden Modellparameter; die Elemente außerhalb der Hauptdiagonale sind die entsprechenden Kovarianzen.

Und was spricht dagegen?

- Die Methode reagiert empfindlich auf „Ausreißer“ in den Daten. Das Quadrieren der Fehler bestraft grosse Abweichungen streng. Deswegen ist die Methode der kleinsten Quadrate bereit, eine Kurve wild zu verzerren, um ein paar weit außen liegende Datenpunkte auch noch annähernd zu erreichen. Ein paar Ausreißer in ansonsten vernünftigen Daten können so eine völlig unsinnige Approximation bewirken.

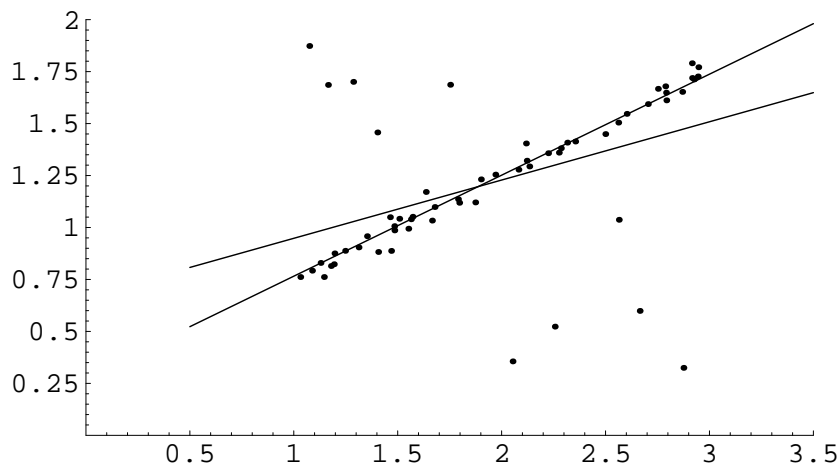


Abbildung 8: Anpassen einer Geraden an Datenpunkte. Die Ausgleichsgerade nach der Methode der kleinsten Quadrate lässt sich von den wenigen Ausreißern stark ablenken. Minimieren des absoluten Fehlers legt eine wesentlich plausiblere Gerade durch die Daten.

## 6.6 Ausgleichsgerade mit Minimieren der absoluten Fehler ( $L_1$ -Norm)

Gegeben:  $m$  Wertepaare  $(x_i, y_i), i = 1, \dots, m$

Gesucht: Eine Gerade in der Form  $y = a + bx$ , so dass die Summe der absoluten Fehler

$$\sum_{i=1}^m |p(x_i) - y_i|$$

minimal wird.

Die Lösung lautet: Für gegebenes  $b$  ergibt sich  $a$  als Median eines Datenfeldes,

$$a = \text{median}\{y_i - bx_i\}$$

Den Parameter  $b$  findet man als Lösung der Gleichung

$$0 = \sum_{i=1}^m x_i \text{sgn}(y_i - a - bx_i)$$

(wobei  $\text{sgn}(0)$  als Null interpretiert werden soll). Wenn man für  $a$  in dieser Gleichung die durch die vorigen Gleichung bestimmte Funktion  $a(b)$  einsetzt, bleibt eine Gleichung in einer Unbekannten übrig. Intervallhalbierung (siehe Kapitel 1.7) ist die geeignete Lösungsmethode dafür.

## 6.7 Ausgleichsgerade mit Minimieren der Normalabstände

Dazu gibt es unter dem Titel *Total Least Squares* Material auf den Vorlesungsfolien und den Übungsunterlagen.