Kapitel 2

Lineare Ausgleichsprobleme

2.1 Einführung

Bemerkung 2.1 *Aufgabenstellung, Motivation.* Bei linearen Ausgleichsproblemen handelt es sich ebenfalls um Bestapproximations-Probleme. Allerdings ist hier keine Funktion gegeben, welche optimal durch eine andere Funktion approximiert werden soll. Stattdessen hat man eine endliche Menge von Punkten, die durch eine (einfache) Funktion optimal approximiert werden soll.

Oft will man in Anwendungen etwa den Verlauf einer abhängigen Größe $y = f(t)$, wobei t die Zeit ist, durch Linearkombinationen spezieller Funktionen $f_j(t)$ beschreiben, also

$$f(t) = \sum_{j=1}^{n} f_j(t) x_j,$$

wobei die Funktionen $f_j(t)$ vorgegeben sind. Nun führt man zu gewissen Zeitpunkten t_i Messungen durch und erhält die Messwerte b_i, $i = 1, \ldots, m$. Die m Werte könnten auch durch Wiederholungen des Experiments gewonnen werden, wobei man im Allgemeinen damit rechnen muss, dass man für dieselbe Zeit t_i (leicht) unterschiedliche Werte erhält. Aus dem Ansatz folgt

$$b_i = \sum_{j=1}^{n} f_j(t_i) x_j =: \sum_{j=1}^{n} a_{ij} x_j, \quad i = 1, \ldots, m.$$

In der Praxis ist es wegen der unvermeidlichen Mess- und Modellfehler sinnvoll, dass die Anzahl der Messungen m groß ist, insbesondere größer als die Anzahl der zu bestimmenden Parameter, also $m > n$ oder sogar $m \gg n$. Man erhält also ein lineares Gleichungssystem der Gestalt

$$A\boldsymbol{x} = \boldsymbol{b}, \quad A \in \mathbb{R}^{m \times n},\ \boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{b} \in \mathbb{R}^m, \tag{2.1}$$

wobei man mehr Gleichungen als Unbekannte hat. Dieses System besitzt in der Regel keine Lösung. Trotzdem möchte man aus den Messungen in geeigneter Weise definierte, optimale Werte für die Parameter x_j, $j = 1, \ldots, n$, bestimmen. Dieses Vorgehen nennt man Regression und die damit berechnete Funktion $f(t)$ Ausgleichskurve oder Regressionskurve.

Zur Bestimmung der Parameter minimiert man in der Praxis eine Norm des Residuums $\boldsymbol{b} - A\hat{\boldsymbol{x}}$, zum Beispiel die Euklidische Norm

$$\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\|_2^2 := \min_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2^2 . \tag{2.2}$$

Eine wichtige Frage ist, ob $\hat{\boldsymbol{x}}$ eindeutig bestimmt ist. Allein durch die Bedingung (2.2) ist dies nicht der Fall. Falls zum Beispiel A nicht spaltenregulär ist, $\mathrm{rg}(A) < n$, gibt es unendlich viele Lösungen, nämlich mit $\hat{\boldsymbol{x}}$ auch $\hat{\boldsymbol{x}} + \ker(A)$. □

2.2 Die Methode der kleinsten Quadrate

Definition 2.2 Kleinste-Quadrate-Lösung. Seien $A \in \mathbb{R}^{m\times n}$ und $\boldsymbol{b} \in \mathbb{R}^m$. Ein Vektor $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ wird Kleinste-Quadrate-Lösung von (2.1) genannt, falls er (2.2) erfüllt. Unter diesen Vektoren wird mit $\boldsymbol{x}^+ \in \mathbb{R}^n$ derjenige mit kleinster Euklidischer Norm bezeichnet, das heißt

$$\left\|\boldsymbol{x}^+\right\|_2 := \min\left\{\|\hat{\boldsymbol{x}}\|_2 \ : \ \hat{\boldsymbol{x}} \in \mathbb{R}^n,\ \hat{\boldsymbol{x}} \text{ erfüllt } (2.2)\right\}. \tag{2.3}$$

□

Bemerkung 2.3 *Methode der kleinsten Quadrate.* Bei der Kleinste-Quadrate-Lösung minimiert man also das Residuum (oder den Defekt) in der Euklidischen Norm. In der Praxis sind auch andere Normen von Bedeutung, zum Beispiel $\|\cdot\|_1$ in den Wirtschaftswissenschaften oder $\|\cdot\|_\infty$ in der Tschebyscheffschen Ausgleichsrechnung.

Die Minimierung in der Euklidischen Norm nennt man Methode der kleinsten Quadrate. □

Bemerkung 2.4 *Kern und Bild von A.* Jetzt müssen die Lösungen $\hat{\boldsymbol{x}}$ und $\boldsymbol{x}^+$ genauer charakterisiert werden. Sei $A \in \mathbb{R}^{m\times n}$ und bezeichne

$$\ker(A) = \{\boldsymbol{x} \in \mathbb{R}^n \ : \ A\boldsymbol{x} = \boldsymbol{0}\}$$

den Kern oder Nullraum von A sowie

$$\mathrm{im}(A) = \{\boldsymbol{y} \in \mathbb{R}^m \ : \ \exists \boldsymbol{x} \in \mathbb{R}^n \text{ mit } \boldsymbol{y} = A\boldsymbol{x}\}$$

das Bild von A. Aus der linearen Algebra ist bekannt, dass[1]

$$\mathrm{im}(A)^\perp = \ker\left(A^T\right) \subset \mathbb{R}^m,$$

wobei $^\perp$ das orthogonale Komplement bezeichnet. Des Weiteren sei $P \ : \ \mathbb{R}^m \to \mathrm{im}(A) \subset \mathbb{R}^m$ der sogenannte Orthogonal-Projektor auf $\mathrm{im}(A)$. Dieser zerlegt jeden Vektor $\boldsymbol{u} \in \mathbb{R}^m$ in

$$\boldsymbol{u} = P\boldsymbol{u} + (I - P)\,\boldsymbol{u}, \quad P\boldsymbol{u} \in \mathrm{im}(A),\ (I - P)\,\boldsymbol{u} \in \mathrm{im}(A)^\perp.$$

Aus der linearen Algebra ist bekannt, dass diese Zerlegung eindeutig ist. Der Projektor hat die Eigenschaften

$$P^2 = P \quad \text{und} \quad P^T = P.$$

□

[1] $\boldsymbol{y} \in \mathbb{R}^m, \boldsymbol{y} \in \ker\left(A^T\right) \iff A^T\boldsymbol{y} = \boldsymbol{0} \in \mathbb{R}^n \iff (A^T\boldsymbol{y}, \boldsymbol{x}) = 0\ \forall\, \boldsymbol{x} \in \mathbb{R}^n \iff$
$(\boldsymbol{y}, A\boldsymbol{x}) = 0\ \forall\, \boldsymbol{x} \in \mathbb{R}^n \iff \boldsymbol{y} \in \mathrm{im}(A)^\perp$

Bemerkung 2.5 *Formulierung als Bestapproximations-Aufgabe im Sinne von Kapitel 1.2.* Sei $\boldsymbol{x}^+$ eine Lösung von (2.2) mit kleinster Euklidischer Norm. Dann ist $\boldsymbol{u} = A\boldsymbol{x}^+ \in \mathrm{im}(A) \subset \mathbb{R}^m$ eine Lösung von: Finde $\boldsymbol{u} \in \mathrm{im}(A)$, so dass

$$\|\boldsymbol{b} - \boldsymbol{u}\|_2 \leq \|\boldsymbol{b} - \boldsymbol{v}\|_2 \quad \forall\, \boldsymbol{v} \in \mathrm{im}(A)$$

und umgekehrt. Dies ist eine Bestapproximations-Aufgabe im Sinne von (1.4) mit $V = \mathbb{R}^m$, $U = \mathrm{im}(A)$ und $\|\cdot\| = \|\cdot\|_2$. Da $\mathbb{R}^m$ mit dem Euklidischen Skalarprodukt und der damit induzierten Euklidischen Norm ein Hilbert-Raum ist, erhält man aus Folgerung 1.24 sofort die Existenz und Eindeutigkeit der Lösung $\boldsymbol{u}$ der obigen Bestapproximations-Aufgabe. Allerdings interessiert in der Praxis nicht $\boldsymbol{u} \in \mathbb{R}^m$ sondern $\boldsymbol{x}^+ \in \mathbb{R}^n$. Entsprechende Aussagen für $\boldsymbol{x}^+$ werden im Folgenden bewiesen. □

Satz 2.6 Zur Existenz und Eindeutigkeit der Kleinste-Quadrate-Lösung. *Seien $A \in \mathbb{R}^{m\times n}$ und $\boldsymbol{b} \in \mathbb{R}^m$.*

i) Es gibt Kleinste-Quadrate-Lösungen $\hat{\boldsymbol{x}}$ von (2.1)*. Die Bedingung* (2.2) *ist äquivalent zur Lösung des Systems*

$$A\hat{\boldsymbol{x}} = P\boldsymbol{b} \tag{2.4}$$

sowie zur Lösung des Systems

$$A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b} \quad \Longleftrightarrow \quad A^T\left(A\hat{\boldsymbol{x}} - \boldsymbol{b}\right) = \boldsymbol{0}. \tag{2.5}$$

ii) Die allgemeine Kleinste-Quadrate-Lösung hat die Gestalt $\hat{\boldsymbol{x}}+\ker(A)$. Die Lösung $\boldsymbol{x}^+$ mit der kleinsten Euklidischen Norm ist eindeutig bestimmt. Es gilt $\boldsymbol{x}^+ \in \ker(A)^\perp$.

Beweis: *i).* Wegen $A\boldsymbol{x} \in \mathrm{im}(A)$ gilt $PA\boldsymbol{x} = A\boldsymbol{x}$ für alle $\boldsymbol{x} \in \mathbb{R}^n$. Damit und mit $P^T = P$ folgt

$$\begin{aligned}
\|A\boldsymbol{x} - \boldsymbol{b}\|_2^2 &= \|A\boldsymbol{x} - P\boldsymbol{b} - (I-P)\boldsymbol{b}\|_2^2 \\
&= \left(A\boldsymbol{x} - P\boldsymbol{b} - (I-P)\boldsymbol{b}, A\boldsymbol{x} - P\boldsymbol{b} - (I-P)\boldsymbol{b}\right) \\
&= \|A\boldsymbol{x} - P\boldsymbol{b}\|_2^2 - 2\left(A\boldsymbol{x} - P\boldsymbol{b}, (I-P)\boldsymbol{b}\right) + \|(I-P)\boldsymbol{b}\|_2^2 \\
&= \|A\boldsymbol{x} - P\boldsymbol{b}\|_2^2 - 2\left(P(A\boldsymbol{x} - \boldsymbol{b}), (I-P)\boldsymbol{b}\right) + \|(I-P)\boldsymbol{b}\|_2^2 \\
&= \|A\boldsymbol{x} - P\boldsymbol{b}\|_2^2 - 2(A\boldsymbol{x} - \boldsymbol{b}, \underbrace{P(I-P)\boldsymbol{b}}_{=(P-P^2)\boldsymbol{b}=\boldsymbol{0}}) + \|(I-P)\boldsymbol{b}\|_2^2 \\
&= \|A\boldsymbol{x} - P\boldsymbol{b}\|_2^2 + \|(I-P)\boldsymbol{b}\|_2^2 .
\end{aligned}$$

Der letzte Summand hängt nicht von $\boldsymbol{x}$ ab. Also bekommt man ein Minimum, wenn $\|A\boldsymbol{x} - P\boldsymbol{b}\|_2$ möglichst klein ist. Da $P\boldsymbol{b} \in \mathrm{im}(A)$ ist, gibt es eine Lösung von $A\boldsymbol{x} = P\boldsymbol{b}$. Damit sind die Existenz einer Kleinste-Quadrate-Lösung sowie die Äquivalenz von (2.2) und (2.4) gezeigt.

Äquivalenz von (2.4) *und* (2.5). Sei $\hat{\boldsymbol{x}}$ eine Lösung von (2.4). Einsetzen in (2.5) ergibt

$$A^T A\hat{\boldsymbol{x}} = A^T P\boldsymbol{b} = A^T\boldsymbol{b} + A^T\left(P - I\right)\boldsymbol{b} = A^T\boldsymbol{b},$$

da $\left(P - I\right)\boldsymbol{b} \in \mathrm{im}(A)^\perp = \ker(A^T)$. Sei andererseits $\hat{\boldsymbol{x}}$ eine Lösung von (2.5). Dann gilt

$$\boldsymbol{0} = A^T\left(A\hat{\boldsymbol{x}} - \boldsymbol{b}\right) = A^T\left(A\hat{\boldsymbol{x}} - P\boldsymbol{b}\right) + A^T\left((P-I)\boldsymbol{b}\right) = A^T\left(A\hat{\boldsymbol{x}} - P\boldsymbol{b}\right),$$

da $(P-I)\boldsymbol{b} \in \ker(A^T)$. Es folgt $(A\hat{\boldsymbol{x}} - P\boldsymbol{b}) \in \ker(A^T) = \mathrm{im}(A)^\perp$. Wegen $PA\boldsymbol{x} = A\boldsymbol{x}$ für alle $\boldsymbol{x} \in \mathbb{R}^n$ folgt daraus

$$A\hat{\boldsymbol{x}} - P\boldsymbol{b} = PA\hat{\boldsymbol{x}} - P\boldsymbol{b} = P\left(A\hat{\boldsymbol{x}} - \boldsymbol{b}\right) \in \mathrm{im}(A)^\perp.$$

Nach Definition gilt aber auch $P\,(A\hat{\boldsymbol{x}} - \boldsymbol{b}) \in \mathrm{im}(A)$. Aus diesen beiden Eigenschaften folgt $P\,(A\hat{\boldsymbol{x}} - \boldsymbol{b}) = \boldsymbol{0}$ und damit $PA\hat{\boldsymbol{x}} = A\hat{\boldsymbol{x}} = P\boldsymbol{b}$.

ii). Aus der linearen Algebra ist bekannt, dass die Lösungsmenge von (2.4) die Gestalt $\hat{\boldsymbol{x}} + \ker(A)$ besitzt. In dieser affinen Mannigfaltigkeit existiert genau ein Element kleinster Norm $\boldsymbol{x}^+$. Dieses ist die Projektion des Nullpunktes auf die Mannigfaltigkeit und es gilt $\boldsymbol{x}^+ \in \ker(A)^\perp$. ■

Bemerkung 2.7 *Einschränkung des Definitionsbereiches von A.* Satz 2.6 besagt, dass die Lösung von (2.2) eindeutig ist, wenn man den Definitionsbereich von A auf $\ker(A)^\perp$ einschränkt

$$\tilde{A} \;:\; \ker(A)^\perp \to \mathrm{im}(A).$$

Die Abbildung $\tilde{A}$ ist sogar bijektiv. Damit existiert die Umkehrabbildung A^+. □

Definition 2.8 Verallgemeinerte Inverse, Pseudo-Inverse. Die Abbildung

$$A^+ \;:\; \mathbb{R}^m \to \mathbb{R}^n, \quad \boldsymbol{b} \mapsto A^+\boldsymbol{b} := \tilde{A}^{-1}P\boldsymbol{b} = \boldsymbol{x}^+$$

wird verallgemeinerte Inverse oder Pseudo-Inverse der Matrix A genannt. □

Bemerkung 2.9 *Zur Pseudo-Inversen.* Die Pseudo-Inverse existiert für jede Matrix A. Sie ist eindeutig bestimmt und stimmt bei nicht-singulären quadratischen Matrizen mit A^{-1} überein, siehe Übungsaufgaben zur Berechnung konkreter Pseudo-Inversen. Zur Berechnung der Kleinsten-Quadrate-Lösung wird die Pseudo-Inverse nicht gebraucht, genauso wenig wie A^{-1} für reguläre Gleichungssysteme. □

Beispiel 2.10 *Lineare Regression.* Man soll mit der Methode der kleinsten Quadrate die Regressionsgerade durch die Punkte

$$(0,1),\ (1,3),\ (2,4),\ (3,4)$$

bestimmen.

Die allgemeine Form der Regressionsgeraden ist $u(x) = u_0 + u_1 x$. Einsetzen der Punkte liefert vier Bestimmungsgleichungen

$$A\boldsymbol{u} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} u_0 \\ u_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix} = \boldsymbol{b}. \tag{2.6}$$

Jetzt wird (2.5) angewandt. Es gilt

$$A^TA = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 4 & 6 \\ 6 & 14 \end{pmatrix}.$$

Diese Matrix ist invertierbar, da $\det(A^TA) = 4 \cdot 14 - 6 \cdot 6 = 20 \neq 0$. Man erhält

$$(A^TA)^{-1} = \frac{1}{10} \begin{pmatrix} 7 & -3 \\ -3 & 2 \end{pmatrix}.$$

Aus (2.5) erhält man jetzt die kleinste-Quadrate-Lösung von (2.6)

$$\boldsymbol{u}^+ = \begin{pmatrix} u_0 \\ u_1 \end{pmatrix} \quad = \quad (A^TA)^{-1}A^T\boldsymbol{b}$$

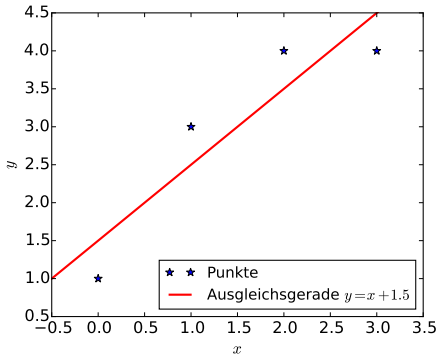

Abbildung 2.1: Ausgleichsgerade im Beispiel 2.10.

$$= \frac{1}{10}\begin{pmatrix} 7 & -3 \\ -3 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix}$$
$$= \frac{1}{10}\begin{pmatrix} 7 & -3 \\ -3 & 2 \end{pmatrix}\begin{pmatrix} 12 \\ 23 \end{pmatrix} = \frac{1}{10}\begin{pmatrix} 15 \\ 10 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 1 \end{pmatrix}.$$

Die Ausgleichgerade lautet somit $u(x) = 1.5 + x$, siehe Abbildung 2.1. □

Bemerkung 2.11 *Normalgleichungen.* Die Gleichung (2.5) werden auch Normalgleichungen genannt. Ist $\mathrm{rg}(A) = n$, so ist $\mathrm{rg}(A^T A) = n$ und (2.5) ist ein reguläres quadratisches lineares Gleichungssystem mit symmetrischer Systemmatrix. Das kann man im Prinzip mit bereits behandelten Verfahren lösen, zum Beispiel mit dem Cholesky[2]-Verfahren oder einem iterativen Verfahren (später in Numerik II). Ist A eine nicht-singuläre quadratische Matrix, dann gilt jedoch

$$\kappa_2\left(A^T A\right) = \left(\kappa_2(A)\right)^2.$$

(*Übungsaufgabe*) Die Konditionszahl von $A^T A$ ist das Quadrat der Konditionszahl von A ! Wenn $\kappa_2(A)$ schon groß ist, dann ist $\kappa_2\left(A^T A\right)$ noch sehr viel größer. Daraus folgt, dass die Rundungsfehler bei direkten Verfahren unnötig groß werden oder dass die Anzahl der Iterationen bei iterativen Verfahren sehr groß sein wird.

Aus diesem Grunde benötigt man für die Praxis ein anderes Verfahren als die Normalgleichungen, welches diese negativen Eigenschaften nicht besitzt, siehe Abschnitt 2.3. □

2.3 Die QR-Zerlegung

Bemerkung 2.12 *Motivation.* Man benötigt zum Finden der Kleinsten-Quadrate-Lösung Verfahren, welche die ursprüngliche Kondition des Problems nicht erhöhen. Betrachtet man insbesondere die Spektralkonditionszahl, dann sollte man Verfahren mit orthogonalen (unitären) Umformungen verwenden, da diese die ursprüngliche Konditionszahl nicht verändern. Das bedeutet, dass man nicht wie bei der LU-Zerlegung die Matrix A in ein Produkt von zwei Dreiecksmatrizen zerlegt, sondern in ein Produkt einer orthogonalen Matrix und einer Dreiecksmatrix. □

[2] André-Louis Cholesky (1875 – 1918)

Definition 2.13 QR-Zerlegung. Seien $A \in \mathbb{R}^{m\times n}$, $Q \in \mathbb{R}^{m\times m}$ eine orthogonale Matrix und $R \in \mathbb{R}^{m\times n}$ eine obere Dreiecksmatrix. Die Zerlegung

$$A = QR, \quad \text{mit} \quad Q^T Q = I$$

und

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ 0 & \cdots & 0 & r_{nn} \\ 0 & \cdots & \cdots & 0 \\ & \cdots & \cdots & \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \text{ oder } R = \begin{pmatrix} r_{11} & \cdots & r_{1m} & \cdots & r_{1n} \\ & \ddots & \vdots & & \vdots \\ 0 & & r_{mm} & \cdots & r_{mn} \end{pmatrix},$$

heißt QR-Zerlegung von A. □

Lemma 2.14 Überblick über Eigenschaften von Orthogonalmatrizen. *Sei*

$$\mathbb{O}_m(\mathbb{R}) := \left\{ Q \in \mathbb{R}^{m\times m} \ : \ Q^T = Q^{-1} \right\}$$

die Menge der Orthogonalmatrizen in $\mathbb{R}^{m\times m}$. Dann gelten:

i) Ist $Q \in \mathbb{O}_m(\mathbb{R})$, so ist $Q^T \in \mathbb{O}_m(\mathbb{R})$.

ii) Ist $Q \in \mathbb{O}_m(\mathbb{R})$, so ist $|\det(Q)| = 1$.

iii) Für $Q_1, Q_2 \in \mathbb{O}_m(\mathbb{R})$ gilt $Q_1 Q_2 \in \mathbb{O}_m(\mathbb{R})$.

iv) Es gilt $\|Q\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$ für alle $\boldsymbol{x} \in \mathbb{R}^m$. Orthogonale Matrizen beschreiben Drehungen oder Spiegelungen von Vektoren.

v) Für jede Matrix $A \in \mathbb{R}^{m\times m}$ gilt $\|A\|_2 = \|QA\|_2 = \|AQ\|_2$.

vi) Für jede reguläre Matrix $A \in \mathbb{R}^{m\times m}$ gilt $\kappa_2(A) = \kappa_2(QA) = \kappa_2(AQ)$.

vii) Für alle $Q \in \mathbb{O}_m(\mathbb{R})$ gilt $\kappa_2(Q) = 1$.

Beweis: Übungsaufgabe. ■

Bemerkung 2.15 *Lösung eines linearen Gleichungssystems mit QR-Zerlegung.* Sei eine QR-Zerlegung $A = QR$ gegeben. Dann folgt

$$A\boldsymbol{x} = \boldsymbol{b} \quad \Longleftrightarrow \quad QR\boldsymbol{x} = \boldsymbol{b} \quad \Longleftrightarrow \quad R\boldsymbol{x} = Q^T\boldsymbol{b}.$$

Das heißt, die Berechnung der Lösung reduziert sich auf die Multiplikation der rechten Seite mit Q^T und eine Rückwärtssubstitution (falls A nicht-singulär ist). □

Bemerkung 2.16 *Prinzipielle Herangehensweise zur Berechnung einer QR-Zerlegung.* Jetzt muss geklärt werden, wie man eine QR-Zerlegung von A berechnet. Die Berechnung erfolgt sukzessive mit geeigneten Orthogonalmatrizen $Q_i \in \mathbb{O}_m(\mathbb{R})$. Mit diesen wird A von links multipliziert und schrittweise auf obere Dreiecksgestalt gebracht. Man setzt

$$Q = \prod_i Q_i^T.$$

Zur Konstruktion der Faktoren Q_i gibt es unterschiedliche Herangehensweisen. Die zwei wichtigsten beruhen auf den beiden geometrischen Interpretationen von Orthogonalmatrizen, siehe Lemma 2.14 iv). Sie sind

- Householder[3]-Spiegelungen,

[3] Alston Scott Householder (1904 – 1993)

- Givens[4]-Drehungen.

Durch diese Verfahren wird die Existenz einer QR-Zerlegung auch konstruktiv bewiesen, siehe Satz 2.37.

Das aus der Algebra bekannte Gram–Schmidtsche Orthogonalisierungsverfahren ist numerisch nicht gutartig. Es existiert eine Modifikation dieses Verfahrens, welche diese Schwierigkeiten überwindet. In der Praxis werden aber die Verfahren von Housholder und Givens vorgezogen. Man wendet diese Verfahren auch bei der numerischen Approximation von Eigenwerten und Eigenvektoren von Matrizen an, siehe Abschnitt 5.6. □

2.3.1 Householder-Spiegelungen

Bemerkung 2.17 *Householder-Transformation, Householder-Spiegelung.* Das Prinzip der Householder-Spiegelungen besteht darin, die Spalten der Matrix A sukzessive auf ein Vielfaches des „ersten“ Einheitsvektors zu transformieren, wobei die genaue Position des Nichtnulleintrags von der aktuellen Spalte abhängt. Sei $\boldsymbol{u} \in \mathbb{R}^m$ mit $\|\boldsymbol{u}\|_2 = 1$, dann besitzt diese Transformation die Gestalt

$$H = I - 2\boldsymbol{u}\boldsymbol{u}^T.$$

Diese Transformation wird Householder-Transformation genannt. Das Produkt

$$\boldsymbol{u}\boldsymbol{u}^T = \begin{pmatrix} u_1u_1 & \cdots & u_1u_m \\ \vdots & \ddots & \vdots \\ u_mu_1 & \cdots & u_mu_m \end{pmatrix}$$

nennt man dyadisches Produkt. □

Lemma 2.18 Eigenschaften der Householder-Transformation. *Seien $\boldsymbol{u} \in \mathbb{R}^m$ mit $\|\boldsymbol{u}\|_2 = 1$ und $H = I - 2\boldsymbol{u}\boldsymbol{u}^T \in \mathbb{R}^{m\times m}$. Dann gelten*

i) Symmetrie: $H = H^T$,
ii) $H^2 = I$,
iii) Orthogonalität: $H^TH = I$,
iv) $H\boldsymbol{y} = \boldsymbol{y}$, $\boldsymbol{y} \in \mathbb{R}^m$, ist äquivalent zu $\boldsymbol{y}^T\boldsymbol{u} = 0$,
v) $H\boldsymbol{u} = -\boldsymbol{u}$.

Beweis: Übungsaufgabe. ■

Bemerkung 2.19 *Geometrische Deutung.* Es gilt

$$\boldsymbol{y} = H\boldsymbol{x} = \boldsymbol{x} - 2\boldsymbol{u}\underbrace{\left(\boldsymbol{u}^T\boldsymbol{x}\right)}_{\in\mathbb{R}} = \boldsymbol{x} - 2\left(\boldsymbol{u}^T\boldsymbol{x}\right)\boldsymbol{u}. \tag{2.7}$$

Das heißt, von $\boldsymbol{x}$ wird ein Vektor in Richtung $\boldsymbol{u}$ subtrahiert. Aus der linearen Algebra ist bekannt, dass sich die Projektion von $\boldsymbol{x}$ in das orthogonale Komplement von $\boldsymbol{u}$ wie folgt berechnet

$$\boldsymbol{x} - \left(\boldsymbol{u}^T\boldsymbol{x}\right)\boldsymbol{u}.$$

Somit ergibt sich für (2.7), dass $\boldsymbol{x}$ am orthogonalen Komplement von $\boldsymbol{u}$ gespiegelt wird, siehe Abbildung 2.2. □

[4] James Wallace Givens, Jr. (1910 – 1993)

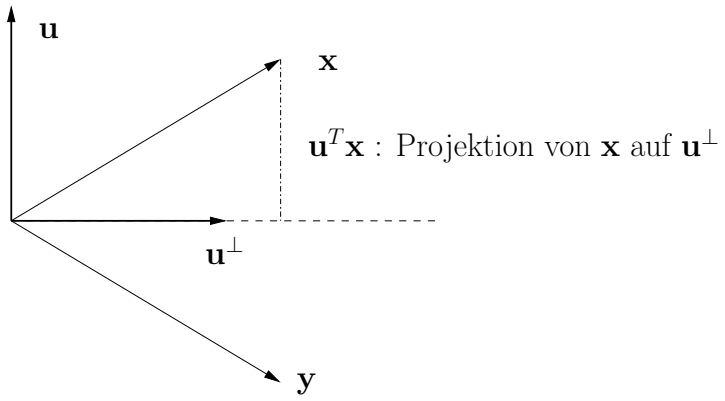

Abbildung 2.2: Geometrische Deutung einer Householder-Spiegelung.

Bemerkung 2.20 *Erster Schritt einer Householder-Transformation.* Für den ersten Schritt ist ein Spiegelvektor $\boldsymbol{u}$ so zu bestimmen, dass die erste Spalte von A auf ein Vielfaches des Einheitsvektors $\boldsymbol{e}_1$ gespiegelt wird, damit die erste Spalte von R die Form $(r_{11}, 0, \ldots, 0)^T$ erhält. Sei $\boldsymbol{x}$ die erste Spalte von A, dann ist die Aufgabenstellung, $\boldsymbol{u} \in \mathbb{R}^m$ so zu finden, dass

$$H\boldsymbol{x} = \boldsymbol{x} - 2\left(\boldsymbol{u}^T\boldsymbol{x}\right)\boldsymbol{u} = \beta\boldsymbol{e}_1.$$

Zunächst gilt, Lemma 2.14 iv),

$$\|H\boldsymbol{x}\|_2 = \|\beta\boldsymbol{e}_1\|_2 \quad \Longrightarrow \quad \|\boldsymbol{x}\|_2 = |\beta|\,\|\boldsymbol{e}_1\|_2 = |\beta| \quad \Longrightarrow \quad \beta = \pm\|\boldsymbol{x}\|_2 .$$

Aus dem Ansatz folgt $2\left(\boldsymbol{u}^T\boldsymbol{x}\right)\boldsymbol{u} = \boldsymbol{x} - \beta\boldsymbol{e}_1$, das heißt, $\boldsymbol{u}$ ist ein Vielfaches von $\boldsymbol{x} - \beta\boldsymbol{e}_1$. Da $\|\boldsymbol{u}\|_2 = 1$ sein soll, folgt

$$\boldsymbol{u} = \frac{\boldsymbol{x} - \beta\boldsymbol{e}_1}{\|\boldsymbol{x} - \beta\boldsymbol{e}_1\|_2} = \frac{1}{\|\boldsymbol{x} - \beta\boldsymbol{e}_1\|_2}\begin{pmatrix} x_1 - \beta \\ x_2 \\ \vdots \\ x_m \end{pmatrix}.$$

Bei der Differenz $\boldsymbol{x} - \beta\boldsymbol{e}_1$ wird nur die erste Komponente x_1 von $\boldsymbol{x}$ verändert. Um dabei Auslöschung zu vermeiden, wählt man das Vorzeichen von β so, dass eine Addition der Beträge stattfindet. Sei $\sigma \in \{-1, 1\}$ das Vorzeichen von x_1. Dann wählt man $\beta = -\sigma\|\boldsymbol{x}\|_2$ und es folgt

$$x_1 - \beta = \sigma|x_1| + \sigma\|\boldsymbol{x}\|_2 = \sigma\left(|x_1| + \|\boldsymbol{x}\|_2\right).$$

Nutzt man die Definition von β, so findet man für den Nenner

$$\begin{aligned} \|\boldsymbol{x} - \beta\boldsymbol{e}_1\|_2^2 &= \left[\left(|x_1| + \|\boldsymbol{x}\|_2\right)^2 + x_2^2 + \ldots + x_m^2\right] \\ &= \left[\|\boldsymbol{x}\|_2^2 + 2|x_1|\,\|\boldsymbol{x}\|_2 + \|\boldsymbol{x}\|_2^2\right] = 2\|\boldsymbol{x}\|_2\left(|x_1| + \|\boldsymbol{x}\|_2\right). \end{aligned}$$

Damit wurde das folgende Lemma 2.21 bewiesen. □

Lemma 2.21 Erster Schritt der Householder-Transformation. *Sei $\boldsymbol{x} \in \mathbb{R}^m$, $\boldsymbol{x} \neq \boldsymbol{0}$, mit $x_1 = \sigma|x_1|$. Dann wird $\boldsymbol{x}$ durch die Matrix*

$$H = I - \frac{2\tilde{\boldsymbol{u}}\tilde{\boldsymbol{u}}^T}{2\|\boldsymbol{x}\|_2\left(|x_1| + \|\boldsymbol{x}\|_2\right)}, \quad \tilde{\boldsymbol{u}} = \boldsymbol{x} + \sigma\|\boldsymbol{x}\|_2\,\boldsymbol{e}_1$$

auf $-\sigma \|\boldsymbol{x}\|_2 \boldsymbol{e}_1$ *abgebildet. Der Vektor* $\tilde{\boldsymbol{u}}$ *ist im Gegensatz zum Vektor* $\boldsymbol{u}$ *nicht normiert.*

Beispiel 2.22 *Erster Schritt der Householder-Transformation.* Sei

$$\boldsymbol{x} = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} \in \mathbb{R}^3 \quad \Longrightarrow \quad \begin{matrix} \|\boldsymbol{x}\|_2 = 3 \\ x_1 = 2 \\ \sigma = 1 \end{matrix} \quad \Longrightarrow \quad \tilde{\boldsymbol{u}} = \begin{pmatrix} 2+3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix}.$$

Damit kann man gemäß Lemma 2.21 sofort das Ergebnis angeben

$$\boldsymbol{x} \mapsto -\sigma \|\boldsymbol{x}\|_2 \boldsymbol{e}_1 = \begin{pmatrix} -3 \\ 0 \\ 0 \end{pmatrix}.$$

Man hat gesehen, dass man die Matrix H nicht explizit braucht. Zur Kontrolle wird in diesem Beispiel die Matrix H berechnet. Es gilt

$$H = I - \frac{2}{6\,(2+3)} \begin{pmatrix} 25 & 10 & 5 \\ 10 & 4 & 2 \\ 5 & 2 & 1 \end{pmatrix} = \frac{1}{15} \begin{pmatrix} -10 & -10 & -5 \\ -10 & 11 & -2 \\ -5 & -2 & 14 \end{pmatrix}.$$

Damit folgt

$$H\boldsymbol{x} = \frac{1}{15} \begin{pmatrix} -10 & -10 & -5 \\ -10 & 11 & -2 \\ -5 & -2 & 14 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} -3 \\ 0 \\ 0 \end{pmatrix} = -\sigma \|\boldsymbol{x}\|_2 \boldsymbol{e}_1.$$

□

Bemerkung 2.23 *Fortsetzung der QR-Zerlegung.* Die Elimination in den folgenden Spalten kann analog ausgeführt werden. Dabei dürfen aber die bereits behandelten Spalten nicht wieder verändert werden. Sei

$$A^{(k)} = \left(\begin{array}{ccc|ccc} r_{11} & \dots & r_{1,k-1} & & & \\ & \ddots & & & * & \\ 0 & \dots & r_{k-1,k-1} & & & \\ \hline 0 & \dots & 0 & a_{kk}^{(k)} & \dots & a_{kn}^{(k)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & a_{mk}^{(k)} & \dots & a_{mn}^{(k)} \end{array}\right) =: \begin{pmatrix} R^{(k)} & * \\ 0 & B^{(k)} \end{pmatrix}$$

mit $R^{(k)} \in \mathbb{R}^{(k-1)\times(k-1)}$, $B^{(k)} \in \mathbb{R}^{(m-k+1)\times(n-k+1)}$. Mit der Wahl

$$\begin{aligned} H^{(k)} &:= \begin{pmatrix} I_{k-1} & 0 \\ 0 & \tilde{H}^{(k)} \end{pmatrix}, \quad \tilde{H}^{(k)} \in \mathbb{R}^{(m-k+1)\times(n-k+1)}, \\ \tilde{\boldsymbol{u}}^{(k)} &:= (\underbrace{0,\dots,0}_{k-1},\underbrace{*,\dots,*}_{m-k+1})^T \quad \Longrightarrow \quad \tilde{\boldsymbol{u}}^{(k)} \left(\tilde{\boldsymbol{u}}^{(k)}\right)^T = \begin{pmatrix} 0_{k-1} & 0 \\ 0 & * \end{pmatrix} \end{aligned}$$

für die k-te Transformation bleiben beim Produkt $H^{(k)}A^{(k)} =: A^{(k+1)}$ die ersten $(k-1)$ Zeilen und Spalten von $A^{(k)}$ unverändert. Nun wird Lemma 2.21 auf den Teilvektor

$$\tilde{H}^{(k)} \begin{pmatrix} a_{kk}^{(k)} \\ \vdots \\ a_{mk}^{(k)} \end{pmatrix} = \begin{pmatrix} \beta \\ \vdots \\ 0 \end{pmatrix}$$

angewandt.

In $l := \min\{m-1, n\}$ Schritten erhält man auf diese Art und Weise die Zerlegung

$$\begin{aligned} A^{(l+1)} &= R^{(l+1)} =: R = H^{(l)} H^{(l-1)} \cdots H^{(1)} A \quad \Longleftrightarrow \\ A &= H^{(1)} H^{(2)} \cdots H^{(l)} R =: QR \end{aligned}$$

mit der orthogonalen Matrix Q und der oberen Dreiecksmatrix R. Bei der Umstellung wurde die Symmetrie der Householder-Matrizen genutzt. Man hat

$$\begin{aligned} R &= \begin{pmatrix} r_{11} & \cdots & r_{1m} & \cdots & r_{1n} \\ & \ddots & \vdots & & \vdots \\ 0 & & r_{mm} & \cdots & r_{mn} \end{pmatrix}, \quad m \le n, l = m-1, \\ R &= \begin{pmatrix} r_{11} & \cdots & r_{1n} \\ & \ddots & \vdots \\ 0 & & r_{nn} \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}, \quad m > n, l = n. \end{aligned}$$

□

Beispiel 2.24 *Householder-Transformation.* Sei

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 2 & 0 \end{pmatrix}.$$

Nach Lemma 2.21 gilt

$$\tilde{\boldsymbol{u}}^{(1)} = \boldsymbol{a}_{*,1} + \sigma \left\| \boldsymbol{a}_{*,1} \right\|_2 \boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} + 1 \cdot 3\boldsymbol{e}_1 = \begin{pmatrix} 4 \\ 2 \\ 2 \end{pmatrix}.$$

Für die Transformationsmatrix gilt

$$H^{(1)} = I - 2\frac{\tilde{\boldsymbol{u}}^{(1)} \left(\tilde{\boldsymbol{u}}^{(1)}\right)^T}{\left\| \tilde{\boldsymbol{u}}^{(1)} \right\|_2^2} = I - \frac{1}{3} \begin{pmatrix} 4 & 2 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} -1 & -2 & -2 \\ -2 & 2 & -1 \\ -2 & -1 & 2 \end{pmatrix},$$

woraus

$$H^{(1)} A = A^{(2)} = \begin{pmatrix} -3 & -1/3 \\ 0 & -2/3 \\ 0 & -2/3 \end{pmatrix}$$

folgt. Für $\tilde{\boldsymbol{u}}^{(2)}$ erhält man

$$\tilde{\boldsymbol{u}}^{(2)} = \begin{pmatrix} 0 \\ -2/3 \\ -2/3 \end{pmatrix} - \frac{\sqrt{8}}{3} \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ -2(1+\sqrt{2})/3 \\ -2/3 \end{pmatrix}.$$

Nach Lemma 2.21 gilt

$$H^{(2)} \begin{pmatrix} 0 \\ -2/3 \\ -2/3 \end{pmatrix} = \begin{pmatrix} 0 \\ 2\sqrt{2}/3 \\ 0 \end{pmatrix}.$$

Insgesamt folgt

$$H^{(2)}H^{(1)}A = \begin{pmatrix} -3 & -1/3 \\ 0 & 2\sqrt{2}/3 \\ 0 & 0 \end{pmatrix} = R.$$

□

Bemerkung 2.25 *Praktische Durchführung.* Die Matrix Q wird in der Praxis nicht explizit berechnet. Stattdessen speichert man die Vektoren $\tilde{\boldsymbol{u}}^{(k)}$ (ohne die oberen Nulleinträge) im unteren Dreieck der Matrix bis einschließlich zur Hauptdiagonalen. Im oberen Dreieck speichert man die Matrix R, ohne die Hauptdiagonaleinträge. Für diese braucht man noch einen zusätzlichen Vektor.

Im Beispiel 2.24 wird das Ergebnis in der Form

$$\begin{pmatrix} 4 & -1/3 \\ 2 & -2(1+\sqrt{2})/3 \\ 2 & -2/3 \end{pmatrix}, \quad \begin{pmatrix} -3 \\ 2\sqrt{2}/3 \end{pmatrix}$$

abgespeichert. □

Bemerkung 2.26 *Kosten.* Der aufwändigste Schritt bei der Berechnung der QR-Zerlegung mit Householder-Spiegelungen ist

$$H^{(k)}A^{(k)} = A^{(k)} - 2\frac{\tilde{\boldsymbol{u}}^{(k)}\left(\tilde{\boldsymbol{u}}^{(k)}\right)^T}{\left\|\tilde{\boldsymbol{u}}^{(k)}\right\|_2^2}A^{(k)} = A^{(k)} - 2\frac{\tilde{\boldsymbol{u}}^{(k)}}{\left\|\tilde{\boldsymbol{u}}^{(k)}\right\|_2^2}\left(\left(\tilde{\boldsymbol{u}}^{(k)}\right)^T A^{(k)}\right) \quad (2.8)$$

zur Berechnung der neuen Einträge in den Spalten $k+1, \ldots, n$[5]. Alle anderen Kosten sind für große m und n vernachlässigbar.

Nutzt man alle Informationen, die bekannt sind, so muss man $(n-k+1)$ Skalarprodukte der Länge $(m-k+1)$ für $(\tilde{\boldsymbol{u}}^{(k)})^T A^{(k)}$ berechnen, da $\tilde{\boldsymbol{u}}^{(k)}$ $(m-k+1)$ Nichtnulleinträge hat und $A^{(k)}$ noch $(n-k+1)$ Zeilen bzw. Spalten besitzt, die transformiert werden müssen. Man benötigt also für die Skalarprodukte

$$(m-k+1)\,(n-k+1) \text{ Multiplikationen.}$$

Dann muss das dyadische Produkt des Ergebnissvektors der Skalarprodukte mit $\tilde{\boldsymbol{u}}^{(k)}$ gebildet werden. Bevor man dies tut, muss man den kürzeren der beiden Vektoren mit $2/\left\|\tilde{\boldsymbol{u}}^{(k)}\right\|_2^2$ skalieren. Diese Kosten werden vernachlässigt. Die Berechnung des dyadischen Produkts bedeutet, dass jede Komponente des einen Vektors mit jeder Komponente des anderen Vektors multipliziert werden muss. Dafür benötigt man prinzipiell ebenfalls

$$(m-k+1)\,(n-k+1) \text{ Multiplikationen.}$$

Im konkreten Fall braucht man sogar die erste Spalte des dyadischen Produkts nicht zu berechnen, da dass Ergebnis der ersten Spalte von $A^{(k+1)} = H^{(k)}A^{(k)}$ klar ist. Diese Einsparung ändert jedoch nichts an den asymptotischen Kosten.

Summiert man im Fall $m > n$ die Kosten aller Schritte von $k = 1, \ldots, n$, dann benötigt man zur Berechnung von (2.8)

$$2\sum_{k=1}^{n}(m-k+1)\,(n-k+1) \text{ Multiplikationen.}$$

[5] Bei der Form auf der rechten Seite müssen ein Matrix-Vektor-Produkt und ein dyadisches Produkt berechnet werden. Würde man von links vorgehen, müsste erst ein dyadisches Produkt und dann ein Matrix-Matrix-Produkt berechnet werden, was ungünstiger ist.

Eine ähnliche Aussage erhält man für $m \leq n$, wobei in diesem Fall die Summation bis $m-1$ geht.

Damit ergibt sich, dass der Rechenaufwand für die QR-Zerlegung mit Householder-Transformationen im Wesentlichen (nur Terme dritten Grades)[6]

$$\min\{m,n\}mn - \frac{1}{3}\min\left\{n^3, m^3\right\} \text{ Multiplikationen/Divisionen}$$

ist. □

Bemerkung 2.27 *Reguläre Systeme.* Man kann die QR-Zerlegung natürlich auch zur Lösung regulärer linearer Gleichungssysteme mit quadratischer Matrix verwenden. Eine Pivotisierung ist dabei nicht erforderlich. Der Vorteil gegenüber der LU-Zerlegung besteht darin, dass sich die Spektralkonditionszahl von A bei der QR-Zerlegung nicht vergrößert. Der Nachteil ist, dass die QR-Zerlegung ungefähr doppelt so teuer ist. □

Bemerkung 2.28 *Berechnung der Kleinste-Quadrate-Lösung $\boldsymbol{x}^+$.* Habe $A \in \mathbb{R}^{m\times n}$ Vollrang, das heißt $\mathrm{rg}(A) = n$ für $m > n$ und seien

$$A = QR, \quad R = \begin{pmatrix} R_0 \\ 0 \end{pmatrix},\ R_0 \in \mathbb{R}^{n\times n}.$$

Dann ist R_0 regulär. Mit Hilfe von Q^T wird die rechte Seite von (2.1) zerlegt

$$Q^T\boldsymbol{b} = \begin{pmatrix} \overline{\boldsymbol{b}} \\ \tilde{\boldsymbol{b}} \end{pmatrix},\ \overline{\boldsymbol{b}} \in \mathbb{R}^n.$$

Das Produkt $Q^T\boldsymbol{b}$ lässt sich aus den gespeicherten Informationen von Q berechnen. Wegen der Invarianz der Euklidischen Vektornorm gegenüber orthogonalen Transformationen erhält man

$$\begin{aligned} \|A\boldsymbol{x}-\boldsymbol{b}\|_2^2 &= \|QR\boldsymbol{x}-\boldsymbol{b}\|_2^2 = \left\|Q\left(R\boldsymbol{x}-Q^T\boldsymbol{b}\right)\right\|_2^2 = \left\|R\boldsymbol{x}-Q^T\boldsymbol{b}\right\|_2^2 \\ &= \left\|\begin{pmatrix} R_0\overline{\boldsymbol{x}}-\overline{\boldsymbol{b}} \\ -\tilde{\boldsymbol{b}} \end{pmatrix}\right\|_2^2 = \|R_0\overline{\boldsymbol{x}}-\overline{\boldsymbol{b}}\|_2^2 + \left\|\tilde{\boldsymbol{b}}\right\|_2^2. \end{aligned}$$

Der zweite Summand ist von $\boldsymbol{x}$ unabhängig. Also erhält man das Minimum von $\|A\boldsymbol{x}-\boldsymbol{b}\|_2^2$ mit der Lösung von

$$R_0\overline{\boldsymbol{x}} = \overline{\boldsymbol{b}} \quad\Longrightarrow\quad \boldsymbol{x}^+ = R_0^{-1}\overline{\boldsymbol{b}}.$$

Auch für den Fall $\mathrm{rg}(A) < n$ kann man $\boldsymbol{x}^+$ mit Hilfe der QR-Zerlegung von A berechnen. □

2.3.2 Givens-Drehungen

Bemerkung 2.29 *Prinzipielle Idee.* Das Prinzip von Givens-Drehungen besteht darin, die Spalten von A durch ebene Drehungen sukzessive in Achsenrichtung zu

[6] Sei $m \geq n$, dann ist

$$\begin{aligned} 2\sum_{k=1}^{n}(m-k+1)(n-k+1) &= 2\sum_{l=1}^{n}(m-n+l)\,l = 2m\sum_{l=1}^{n}l - 2n\sum_{l=1}^{n}l + 2n\sum_{l=1}^{n}l^2 \\ &\approx mn^2 - n^3 + 2n^3/3 = mn^2 - n^3/3. \end{aligned}$$

bringen. Im Gegensatz zur Householder-Transformation werden nicht alle Elemente einer Spalte, die unterhalb der Diagonalen sind, auf einmal auf Null abgebildet, sondern jedes Element wird einzeln abgebildet. Das wird dann von Vorteil sein, wenn es unterhalb der Diagonalen schon viele Nullen gibt, wie bei schwach besetzten Matrizen. □

Bemerkung 2.30 *Ebene Drehungen.* Sei $\boldsymbol{x} = (x_1, x_2)^T = (r\cos(\phi), r\sin(\phi))^T \in \mathbb{R}^2$, $r > 0$, $\phi \in [0, 2\pi)$, gegeben. Gesucht ist eine Matrix, die diesen Vektor in Richtung von $\boldsymbol{e}_1$ dreht. Aus der linearen Algebra ist bekannt, dass der geeignete Ansatz zur Lösung dieser Aufgabe die folgende Gestalt besitzt (Drehungsmatrix)

$$\begin{pmatrix} c & s \\ -s & c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} c & s \\ -s & c \end{pmatrix} \begin{pmatrix} r\cos(\phi) \\ r\sin(\phi) \end{pmatrix} = \begin{pmatrix} r \\ 0 \end{pmatrix}$$

mit $r = \|\boldsymbol{x}\|_2$ und $\phi = \arg(\boldsymbol{x})$. Als Lösung erhält man (man hat zwei Gleichungen für zwei Unbekannte)

$$c = \cos(\phi) = \frac{x_1}{r}, \quad s = \sin(\phi) = \frac{x_2}{r}.$$

Man rechnet leicht nach, dass

$$\begin{pmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{pmatrix} = \frac{1}{\|\boldsymbol{x}\|_2} \begin{pmatrix} x_1 & x_2 \\ -x_2 & x_1 \end{pmatrix}$$

eine orthogonale Matrix ist. Damit besitzt sie alle Eigenschaften aus Lemma 2.14. Insbesondere folgt aus Eigenschaft ii), dass $c^2 + s^2 = 1$. □

Bemerkung 2.31 *Berechnung von c und s unter Vermeidung von Rundungsfehlern.* Um Rundungsfehler infolge der Multiplikation und Division großer Zahlen zu vermeiden, implementiert man die Berechnung von c und s wie folgt:

```
if x1 == 0 and x2 == 0
  c=1; s=0;                % Einheitsmatrix
else
  if abs(x2) >= abs(x1)
    t = x1/x2;             % |t|<=1
    s = 1/sqrt(1+t*t);     % s ist positiv
    c = s*t;
  else
    t = x2/x1;             % |t|<1
    c = 1/sqrt(1+t*t);     % c ist positiv
    s = c*t;
  end
end
```

Dabei nutzt man zum Beispiel die Darstellungen, falls $x_2 > 0$,

$$s = \frac{x_2}{r} = \frac{x_2}{\sqrt{x_1^2 + x_2^2}} = \frac{1}{\sqrt{x_1^2/x_2^2 + 1}}, \quad s\ t = \frac{x_2}{r}\frac{x_1}{x_2} = \frac{x_1}{r} = c.$$

Man beachte, dass im obigen Algorithmus je nach Schleife s oder c positiv ist, obwohl x_2 oder x_1 auch negativ sein können. Damit erhält man gegebenenfalls eine Drehung in die Richtung $-\boldsymbol{e}_1$. Das ist aber für praktische Rechnungen nicht von Belang, da man an den Nullen in den anderen Richtungen interessiert ist. □

Bemerkung 2.32 *Erweiterung auf höhere Dimensionen, Givens-Matrix.* Eine Givens-Matrix ist eine Orthogonalmatrix, die durch

$$G_{ik} = \begin{pmatrix} 1 & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & c & \dots & \dots & \dots & s & & & \leftarrow i \\ & & & & 1 & & & & & & \\ & & & \vdots & & \ddots & & \vdots & & & \\ & & & & & & 1 & & & & \\ & & & -s & \dots & \dots & \dots & c & & & \leftarrow k \\ & & & & & & & & 1 & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{pmatrix} \in \mathbb{R}^{n\times n}$$

gegeben ist. Sie bewirkt eine Rotation in der durch $\boldsymbol{e}_i$ und $\boldsymbol{e}_k$ aufgespannten Ebene. Sei $\boldsymbol{x} \in \mathbb{R}^n$, dann folgt mit

$$r = \sqrt{x_i^2 + x_k^2},\ c = \frac{x_i}{r},\ s = \frac{x_k}{r},$$

dass

$$G_{ik}\boldsymbol{x} = G_{ik}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_{i-1} \\ cx_i + sx_k \\ x_{i+1} \\ \vdots \\ x_{k-1} \\ -sx_i + cx_k \\ x_{k+1} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_{i-1} \\ \frac{1}{r}(x_i^2 + x_k^2) \\ x_{i+1} \\ \vdots \\ x_{k-1} \\ \frac{1}{r}(-x_i x_k + x_i x_k) \\ x_{k+1} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_{i-1} \\ r \\ x_{i+1} \\ \vdots \\ x_{k-1} \\ 0 \\ x_{k+1} \\ \vdots \\ x_n \end{pmatrix}.$$

Der k-te Eintrag von $\boldsymbol{x}$ wird also auf Null transformiert. □

Beispiel 2.33 *Givens-Drehung eines Vektors.* Aufgabe sei es, den Vektor

$$\boldsymbol{x} = \begin{pmatrix} 4 \\ -3 \\ 1 \end{pmatrix}$$

in $\boldsymbol{e}_1$-Richtung mittels Givens-Drehungen zu drehen. Man erhält im ersten Schritt

$$G_{12} = \begin{pmatrix} 4/5 & -3/5 & 0 \\ 3/5 & 4/5 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \Longrightarrow \quad G_{12}\boldsymbol{x} = \begin{pmatrix} 5 \\ 0 \\ 1 \end{pmatrix}.$$

Im zweiten Schritt wird die Givens-Matrix bezüglich des Ergebnissvektors des ersten Schrittes aufgestellt und es ergibt sich

$$G_{13} = \begin{pmatrix} 5/\sqrt{26} & 0 & 1/\sqrt{26} \\ 0 & 1 & 0 \\ -1/\sqrt{26} & 0 & 5/\sqrt{26} \end{pmatrix} \quad \Longrightarrow \quad G_{13}G_{12}\boldsymbol{x} = \begin{pmatrix} \sqrt{26} \\ 0 \\ 0 \end{pmatrix}.$$

□

Bemerkung 2.34 *QR-Zerlegung mit Givens-Drehungen.* Zur Reduktion von A auf obere Dreiecksgestalt geht man folgendermaßen vor. Man wendet nacheinander Givens-Drehungen an, um die Einträge unterhalb der Diagonalen zu Null zu machen. Bei diesem Vorgehen braucht man an den Einträgen unterhalb der Diagonalen nichts zu tun, die bereits Null sind. Man erhält

$$G_{i_N,k_N} \dots G_{i_1,k_1} A = R \quad \Longrightarrow \quad A = G_{i_1,k_1}^T \dots G_{i_N,k_N}^T R =: QR, \ R \in \mathbb{R}^{m\times n}.$$

Je nach Verhältnis von m und n und dem Rang von A kann die Dreiecksmatrix unterschiedlich aussehen, vergleiche Bemerkung 2.23. □

Beispiel 2.35 *Givens-Drehung einer Matrix.* Betrachte (nachrechnen: Übungsaufgabe)

$$A = \begin{pmatrix} 3 & 5 \\ 0 & 2 \\ 0 & 0 \\ 4 & 5 \end{pmatrix} \xrightarrow{G_{14}} \begin{pmatrix} 5 & 7 \\ 0 & 2 \\ 0 & 0 \\ 0 & -1 \end{pmatrix} \xrightarrow{G_{24}} \begin{pmatrix} 5 & 7 \\ 0 & \sqrt{5} \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Man beachte, dass die Anwendung von G_{ik} die gesamte i-te und k-te Zeile von A verändert. □

Bemerkung 2.36 *Zur QR-Zerlegung und Givens-Drehungen.*

- Bei der QR-Zerlegung mittels Givens-Drehungen überschreibt man auch die Einträge der Originalmatrix in geeigneter Art und Weise. Man merkt sich natürlich nicht die Givens-Matrizen, sondern nur die jeweiligen Zahlen c und s.
- Die QR-Zerlegung mit Givens-Drehungen ist sehr stabil. Eine Pivotsuche ist nicht erforderlich.
- Bei voll besetzter Matrix $A \in \mathbb{R}^{n\times n}$ sind die Kosten der QR-Zerlegung mit Givens-Drehungen $\mathcal{O}(4n^3/3)$ Multiplikationen/Divisionen und $\mathcal{O}(\frac{1}{2}n^2)$ Wurzelberechnungen. Das ist doppelt so teuer wie mit Householder-Spiegelungen.
- Bei Givens-Drehungen kann man jedoch bereits vorhandene Nulleinträge in der Matrix gezielt ausnutzen. Mit jedem Nulleintrag sinken die Kosten. Deshalb verwendet man Givens-Drehungen vor allem dort, wo nur wenige Matrixeinträge zu Null gemacht werden müssen, siehe beispielsweise Abschnitt 5.6.

□

Satz 2.37 Existenz einer QR-Zerlegung. *Sei $A \in \mathbb{R}^{m\times n}$. Dann gibt es eine orthogonale Matrix $Q \in \mathbb{R}^{m\times m}$ und eine obere Dreiecksmatrix $R \in \mathbb{R}^{m\times n}$ mit $A = QR$.*

Beweis: Das wurde in diesem Abschnitt durch Konstruktion, mittels Householder-Spiegelungen oder Givens-Drehungen, bewiesen. ■