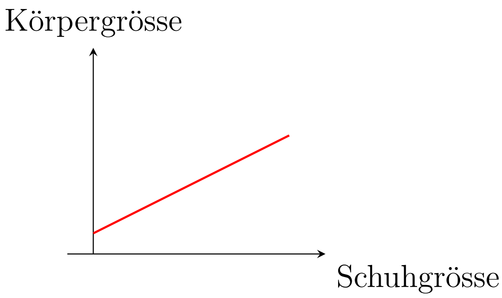
9 Methode der Kleinsten Quadrate und QR-Zerlegung

9.1 Methode der kleinsten Quadrate

Die Methode der kleinsten Quadrate wird immer dann angewendet, wenn wir eine möglichst *gute* Lösung für ein überbestimmtes Gleichungssystem $Ax = c$ finden wollen, für welches es keine eindeutige Lösung gibt. Ein Gleichungssystem ist überbestimmt, wenn es mehr Gleichungen als Unbekannte hat. Für die Matrix $A \in \mathbb{R}^{m \times n}$ gilt dann $m > n$. In realen Anwendungen sind Gleichungssysteme sehr oft Überbestimmt, weswegen es wichtig ist für solche Systeme möglichst *gute* Lösungen zu finden. Was dabei eine *gute* Lösung ausmacht, betrachten wir im folgenden Beispiel.

Bei einer Umfrage werden Menschen bezüglich Körpergrösse und Schuhgrösse befragt. Wir wollen einen Zusammenhang zwischen Körpergrösse und Schuhgrösse finden damit wir nachher basierend auf der Schuhgrösse eine Schätzung für die Körpergrösse machen können. Wir suchen also eine Gleichung der Form

$$\text{Körpergrösse} = x_1 \cdot \text{Schuhgrösse} + x_2, \qquad (\ddagger)$$



wobei die Koeffizienten x_1, x_2 unbekannt sind und experimentell durch eine Umfrage ermittelt werden. In einer Umfrage werden nun 8 Menschen befragt, wodurch wir für zwei Unbekannte x_1 und x_2, 8 Gleichungen erhalten. Wir werden also keine eindeutige Lösung für x_1 und x_2 finden können. Wie wählen wir nun die besten Werte für x_1 und x_2 damit die Gleichung (‡) möglichst gut die Realität repräsentiert? Betrachten wir zunächst die Werte der Umfrage. Aus einer Tabelle mit allen Umfragewerten können wir das überstimmte LGS $Ax = c$ aufstellen.

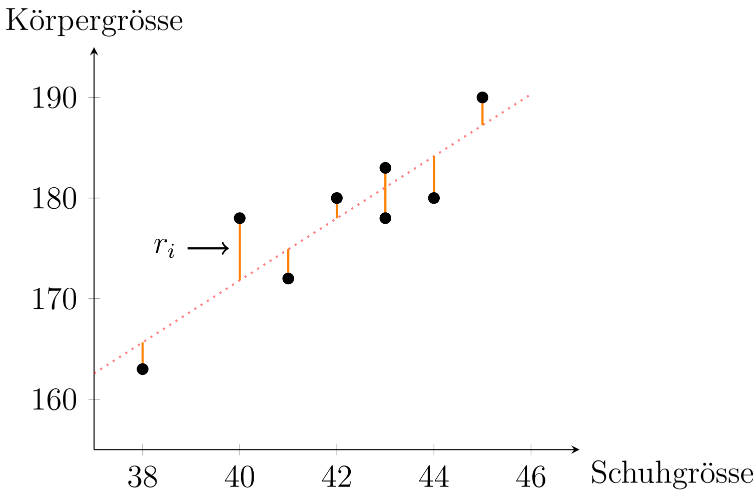
Schuhgrösse	Körpergrösse
41	172
45	190
42	180
43	183
43	178
38	163
44	180
40	178

$$\begin{aligned} 172 &= x_1 \cdot 41 + x_2 \\ 190 &= x_1 \cdot 45 + x_2 \\ 180 &= x_1 \cdot 42 + x_2 \\ 183 &= x_1 \cdot 43 + x_2 \\ 178 &= x_1 \cdot 43 + x_2 \\ 163 &= x_1 \cdot 38 + x_2 \\ 180 &= x_1 \cdot 44 + x_2 \\ 178 &= x_1 \cdot 40 + x_2 \end{aligned}$$

Die Matrix $A \in \mathbb{R}^{8 \times 2}$ und der Vektor $c \in \mathbb{R}^{8 \times 1}$ sind dann wie folgt definiert:

$$A = \begin{pmatrix} 41 & 1 \\ 45 & 1 \\ \vdots & \vdots \\ 40 & 1 \end{pmatrix}, \quad c = \begin{pmatrix} 172 \\ 190 \\ \vdots \\ 178 \end{pmatrix}.$$

Wir können nun alle Messwerte in einem Koordinatensystem auftragen um zu sehen, wie der Zusammenhang zwischen Schuhgrösse und Körpergrösse aussieht. Als Nächstes wollen wir eine Gerade der Form (‡) finden, die möglichst nah an allen Messpunkten ist. Ein Weg dies zu erreichen ist es alle Abstände r_i von der Geraden zu den Messwerten zu minimieren.



$$\begin{aligned}
r_1 &= (x_1 \cdot 41 + x_2) - 172 \\
r_2 &= (x_1 \cdot 45 + x_2) - 190 \\
r_3 &= (x_1 \cdot 42 + x_2) - 180 \\
r_4 &= (x_1 \cdot 43 + x_2) - 183 \\
r_5 &= (x_1 \cdot 43 + x_2) - 178 \\
r_6 &= (x_1 \cdot 38 + x_2) - 163 \\
r_7 &= (x_1 \cdot 44 + x_2) - 180 \\
r_8 &= (x_1 \cdot 40 + x_2) - 178
\end{aligned}$$

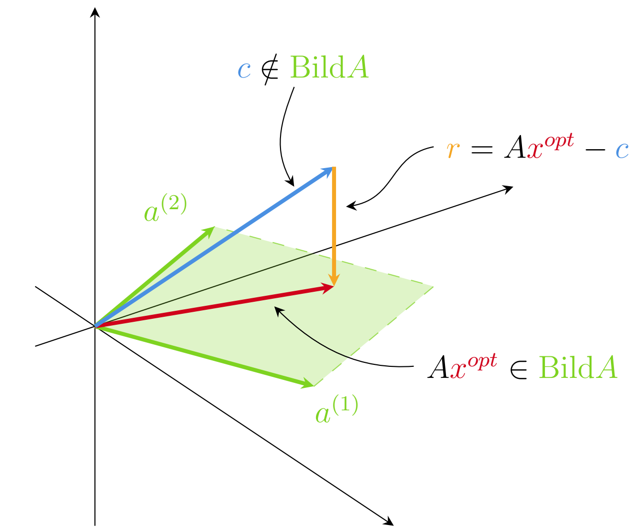
Da alle Abstände gleichzeitig minimiert werden sollen, fassen wir alle r_i im sogenannten Residuenvektor $r = (r_1, \dots, r_m)^\top$ zusammen und minimieren dann die gesamte Länge von r, das entspricht der quadratischen Minimierung des Fehlers r. Wir suchen also das x sodass

$$\|r\|_2 = \|Ax - c\|_2,$$

minimal wird. In unserem Fall suchen wir $x^{opt} = (x_1^{opt} \quad x_2^{opt})^\top$ welches die Länge von r minimiert. Aber wie können wir dieses Problem lösen? Betrachten wir dafür ein reduziertes Gleichungssystem mit nur drei Gleichungen und zwei Unbekannten, dafür nehmen wir einfach die ersten drei Messwerte von oben.

$$\begin{aligned} x_1 \cdot 41 + x_2 &= 172 \\ x_1 \cdot 45 + x_2 &= 190 \\ x_1 \cdot 42 + x_2 &= 180 \end{aligned} \quad \rightarrow \quad \underbrace{\begin{pmatrix} 41 & 1 \\ 45 & 1 \\ 42 & 1 \end{pmatrix}}_{A} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 172 \\ 190 \\ 180 \end{pmatrix}}_{c}$$

Genau wie oben ist dieses LGS überbestimmt und in dieser Form nicht exakt lösbar. Grafisch können wir uns das Problem wie folgt vorstellen.



Der Vektor x^{opt} repräsentiert hier die optimale Lösung, welche den kürzesten Residuenvektor r hat und im Bild von A liegt. Grafisch können wir hier auch sehen, dass der Vektor r minimal ist, wenn er orthogonal zum Bild$A$ ist. Für die optimale Lösung muss also gelten:

$$\underbrace{Ay}_{\text{Bild}A} \perp r\ , \quad \forall y \in \mathbb{R}^n.$$

Mit dem Skalarprodukt können wir das umformulieren zu

$$\langle Ay, r \rangle = 0,$$

$$(Ay)^\top r = 0.$$

Nun können wir unsere Definition für r einsetzen

$$y^\top A^\top \left(Ax^{opt} - c\right) = 0,$$

und alles ausmultiplizieren

$$y^\top A^\top A x^{opt} - y^\top A^\top c = 0.$$

Schliesslich können wir noch umstellen und nach x^{opt} auflösen

$$A^\top c = A^\top A x^{opt},$$

$$(A^\top A)^{-1} A^\top c = x^{opt}.$$

Dieser Ausdruck gibt uns nun die optimale Lösung, indem der quadratische Fehler (Länge des Residuenvektors) minimiert wird. Oft wollen wir jedoch nicht die inverse von $A^\top A$ berechnen und können alternativ das folgende LGS lösen

$$A^\top A x^{opt} = A^\top c.$$

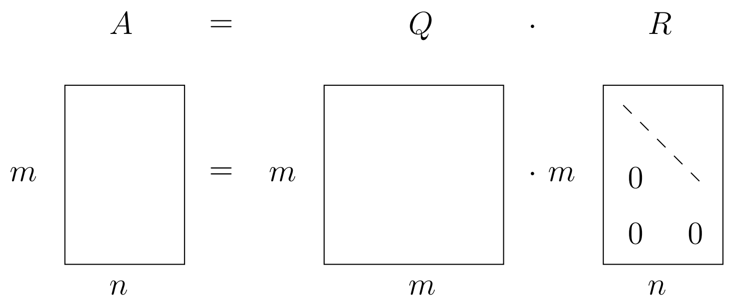
Kleinste Quadrate

Wir wollen die optimale Lösung x^{opt} finden welche den quadratischen Fehler $\|r\|_2 = \|Ax - c\|_2$ minimiert. Oft ist das LGS schon in der Form $Ax - c = r$ gegeben.

1. Bestimme A und c
2. Berechne $A^\top A$ und $A^\top c$
3. Löse das LGS $A^\top A x^{opt} = A^\top c$

9.2 QR-Zerlegung

Bei der QR-Zerlegung werden wir eine Matrix $A \in \mathbb{R}^{m \times n}$ in das Produkt einer orthogonalen Matrix $Q \in \mathbb{R}^{m \times m}$ und einer oberen Dreiecksmatrix $R \in \mathbb{R}^{m \times n}$ zerlegen. Es gilt dann



Später können wir die QR-Zerlegung als ein alternatives Lösungsverfahren der kleinsten Quadrate verwenden. Um die QR-Zerlegung zu berechnen, müssen einige Schritte befolgt werden. Betrachten wir das Ganze an einem Beispiel.

Wir suchen nun Q und R sodass für die Matrix

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

gilt $A = QR$. Dafür eliminieren wir schrittweise alle Elemente a_{ij} unterhalb der Hauptdiagonalen von A. In unserem fall beginnen wir mit dem Eintrag a_{31}. Zuerst notieren wir die Werte für i, j und die Einträge $a_{jj},\ a_{ij}$

$$i = 3,\ j = 1.$$

$$a_{jj} = 1,\ a_{ij} = 1.$$

Anschliessend berechnen wir den Wert einer Zwischenvariable ω gegeben durch

$$\omega = \sqrt{a_{jj}^2 + a_{ij}^2} = \sqrt{1^2 + 1^2} = \sqrt{2}.$$

Als Nächstes suchen wir eine Rotationsmatrix $Q'^\top$. Dafür fangen wir mit einer Identitätsmatrix $I \in \mathbb{R}^{m \times m}$ an und füllen die Einträge wie folgt ein

$$i_{ii} = \cos(\alpha),\ i_{ij} = -\sin(\alpha),\ i_{ji} = \sin(\alpha),\ i_{jj} = \cos(\alpha).$$

Wobei die spezifischen Werte für cos und sin durch das ω gegeben sind

$$\sin(\alpha) = \frac{a_{ij}}{\omega} = \frac{1}{\sqrt{2}}, \quad \cos(\alpha) = \frac{a_{jj}}{\omega} = \frac{1}{\sqrt{2}}.$$

In unserem Fall ist $Q'^\top$ gegeben durch

$$Q'^\top = \begin{pmatrix} \cos(\alpha) & 0 & \sin(\alpha) \\ 0 & 1 & 0 \\ -\sin(\alpha) & 0 & \cos(\alpha) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Nun Berechnen wir die Matrix $A' = Q'^\top A$ und überprüfen ob A' eine obere Dreiecksmatrix ist.

$$A' = Q'^\top A = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & 1 \\ 0 & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Falls, wie in unserem Fall, A' keine obere Dreiecksmatrix ist, wiederholen wir den Prozess und nehmen den nächsten Eintrag unterhalb der Hauptdiagonalen von A'. Hier wäre das der Eintrag a'_{32} dadurch werden $i = 3, j = 2$. Wir lesen wieder die Einträge a'_{ij}, a'_{jj} und ω ab

$$a'_{jj} = 1,\ a'_{ij} = \frac{1}{\sqrt{2}},\ \omega = \sqrt{\frac{3}{2}}.$$

Weiter konstruieren wir wie oben eine Rotationsmatrix $Q''^\top$ welche die folgende Form annimmt

$$Q''^\top = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & \sin(\alpha) \\ 0 & -\sin(\alpha) & \cos(\alpha) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}} \end{pmatrix},$$

wodurch wir die Matrix $A'' = Q''^\top A'$ erhalten

$$A'' = \begin{pmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{3}}{\sqrt{2}} \\ 0 & 0 \end{pmatrix}.$$

Nun ist A'' eine obere Dreiecksmatrix und es gilt $A'' = R$. Um schliesslich noch die Matrix Q zu erhalten, multiplizieren wir alle Rotationsmatrizen zusammen

$$Q^\top = Q''^\top Q'^\top.$$

Somit ist die QR-Zerlegung vollendet. Zusammengefasst müssen also folgende Schritte befolgt werden.

QR-Zerlegung

Bei der QR-Zerlegung Zerlegen wir eine Matrix $A \in \mathbb{R}^{m\times n}$ in das Produkt einer orthogonalen Matrix $Q \in \mathbb{R}^{m\times m}$ und einer oberen Dreiecksmatrix $R \in \mathbb{R}^{m\times n}$. Nach und nach eliminieren wir die Einträge unterhalb der Hauptdiagonalen von A.

1. Wähle das zu eliminierende Element a_{ij} unter der Hauptdiagonalen
2. Notiere die Werte i, j, a_{ij}, a_{jj}
3. Berechne $\omega = \sqrt{a_{ij}^2 + a_{jj}^2}$
4. Finde die Rotationsmatrix $Q'^\top$ basierend auf einer Identitätsmatrix $I \in \mathbb{R}^{m\times m}$ und setze $i_{ii} = \cos(\alpha), i_{ij} = -\sin(\alpha), i_{ji} = \sin(\alpha), i_{jj} = \cos(\alpha)$
5. Für die Werte von $\cos(\alpha)$ und $\sin(\alpha)$ setze $\sin(\alpha) = \frac{a_{ij}}{\omega}, \cos(\alpha) = \frac{a_{jj}}{\omega}$
6. Berechne $A' = Q'^\top A$
7. Überprüfe ob A' eine obere Dreiecksmatrix ist, falls nicht, wiederhole den Prozess mit dem nächsten Eintrag a'_{ij} von A' solange, bis alle Einträge unterhalb der Hauptdiagonalen eliminiert sind
8. Wenn nach k Wiederholungen $A^{(k)} = R$ ist, berechne $Q^\top = Q^{(k)\top} \cdot \ldots \cdot Q''^\top \cdot Q'^\top$
9. Nun ist $A = QR$

9.3 Kleinste Quadrate mit QR-Zerlegung

In gewissen Fällen kann es sein das die oben gezeigte Methode zur Lösung von überbestimmten LGS ungenaue Ergebnisse liefert. Vor allem, wenn Sie numerisch mit dem Computer berechnet werden. Oft kann man für eine numerisch stabilere Lösung die QR-Zerlegung verwenden.

Kleinste Quadrate mit QR-Zerlegung

Ein alternativer Lösungsweg um überbestimmte LGS $Ax = c$ durch kleinste Quadrate zu lösen.

1. Bestimme A und c
2. Berechne die QR-Zerlegung $A = QR$
3. Berechne $d = Q^\top c$
4. Löse das LGS $R_0 x = d_0$, wobei R_0 die extrahierte Dreiecksmatrix von R ist und d_0 die dazugehörigen oberen Einträge von d
5. Die Lösung x ist die optimale Lösung für $Ax = c$