

# Lineare Regression

Gerhard Wesp

**KISS**

Technologies

<http://www.kisstech.ch/>

April 2015

## Zusammenfassung

Dieses Skript und Tutorial richtet sich an Studenten oder Ingenieure, die die Methode der Linearen Regression kennenlernen bzw. anwenden wollen. Wir behandeln die der Linearen Regression zugrundeliegende Methode der kleinsten Fehlerquadrate sowie deren Implementierung mittels der Singulärwertzerlegung. Wir beschreiben die Herleitung der Parameter der Ausgleichsgerade sowie allgemeiner Ausgleichskurven. Wir zeigen die wichtigsten numerischen Fallstricke bei der Implementierung auf und beschreiben die relevanten Matlab-Funktionen. Schliesslich rundet ein Abschnitt mit sechs Übungsbeispielen das Skript ab.

## Inhaltsverzeichnis

<b>1 Einführung</b>	<b>1</b>
<b>2 Die Ausgleichsgerade</b>	<b>3</b>
<b>3 Die Methode der kleinsten Quadrate</b>	<b>4</b>
<b>4 Allgemeine lineare Regression und die Singulärwertzerlegung</b>	<b>5</b>
<b>5 Numerische Aspekte</b>	<b>11</b>
<b>6 Matlab-Funktionen</b>	<b>12</b>
<b>7 Übungsbeispiele</b>	<b>12</b>

## 1 Einführung

Die Lineare Regression ist eine Methode, um zwischen zwei gegebenen Datensatz einen *funktionalen Zusammenhang* zu finden und sichtbar zu machen. Die Methode funktioniert auch dann, wenn der gesuchte Zusammenhang nur *annähernd* besteht bzw. die Daten mit *Messfehlern* behaftet sind. Ein Zusammenhang kann dann für *Vorhersagen* bzw. Extrapolation der Daten genutzt werden.

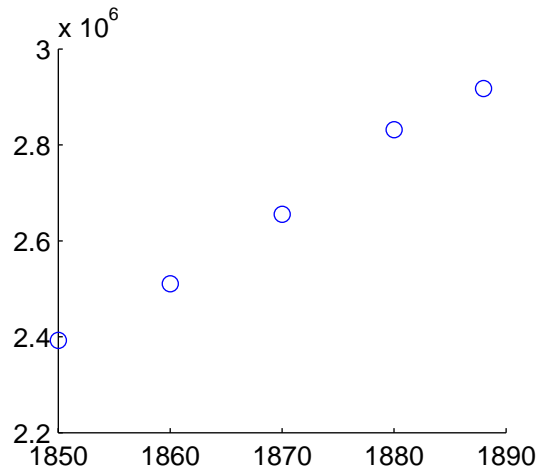


Abbildung 1: Bevölkerung der Schweiz im 19. Jahrhundert

Als Beispiel betrachten wir die Bevölkerungsentwicklung der Schweiz in der zweiten Hälfte des 19. Jahrhunderts. Die entsprechenden Daten stehen auf der Web-Seite des Bundesamtes für Statistik zur Verfügung und betragen:

Jahr ( $x$ )	Bevölkerung ( $y$ )
1850	2392740
1860	2510494
1870	2655001
1880	2831787
1888	2917754

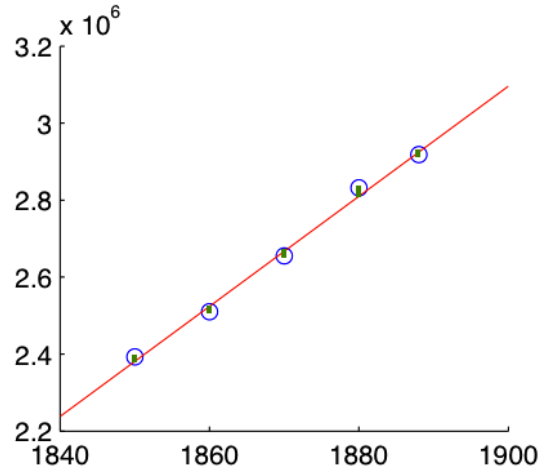
Betrachtung von **Abbildung 1** lässt uns ein annähernd lineares Bevölkerungswachstum vermuten (zumindest im betrachteten Zeitraum!).

Aber wie können wir diesen graphisch offensichtlichen linearen Zusammenhang am besten rechnerisch quantifizieren? In dieser Lektion lernen wir die *Ausgleichsgerade* und die Methode zu ihrer Berechnung kennen.

Die Ausgleichsgerade ist in einem gewissen Sinn die beste *geradlinige* Annäherung an die gegebenen Daten. Für Daten, die nicht auf einer Linie liegen, lernen wir die allgemeine Methode der *linearen Regression* zur Berechnung von *Ausgleichskurven* kennen. Als Werkzeug zur Lösung des Regressionsproblems lernen wir die *Singulärwertzerlegung* von Matrizen kennen.

Schliesslich behandeln wir numerische Aspekte sowie die Implementierung der beschriebenen Methoden in Matlab.

Bitte beachten Sie, dass das Wort "linear" in linearer Regression *nicht* bedeutet, dass die Ausgleichskurve eine Linie sein muss! Es bezieht sich hingegen darauf, dass die Ausgleichskurve eine Linearkombination gewisser Basisfunktionen ist.

Abbildung 2: Ausgleichsgerade mit Fehlern  $e_i$  (grün)

## 2 Die Ausgleichsgerade

Seien  $x = (x_1, \dots, x_m) \in \mathbf{R}^m$  und  $y = (y_1, \dots, y_m) \in \mathbf{R}^m$  die gegebenen Datenvektoren. Angenommen, wir haben festgestellt, dass die Punkte  $(x_i, y_i)$  annähernd auf einer Geraden liegen wie in Abbildung 1.

Wir suchen also eine Gerade

$$y = f(x) = kx + d,$$

die unseren Datenpunkten  $(x_i, y_i)$  möglichst nahe kommt. Dieses “möglichst nahe” müssen wir jetzt mathematisch genau formulieren. Daraus wird sich dann eine Berechnungsvorschrift ergeben, die uns erlaubt, die Geradenparameter  $k$  und  $d$  aus den Daten zu bestimmen.

Dazu definieren wir zunächst die vertikale Abweichungen der Geraden von den Datenpunkten (vgl. Abbildung 2):

$$e_i = f(x_i) - y_i = kx_i + d - y_i$$

Wenn wir wollen, dass die Gerade den Datenpunkten möglichst nahe kommt, müssen *alle* diese Abweichungen—oder Fehler— $e_i$  möglichst klein sein. Natürlich können wir nicht jedes einzelne  $e_i$  minimieren, da dann die Gerade einfach durch den entsprechenden Punkt gehen würde, aber dafür wieder andere Punkte verfehlt.

Wir könnten aber z.B. versuchen, die *Summe* der Abweichung

$$S(k, d) = \sum_{i=1}^m |e_i|$$

zu minimieren. In der Tat, ist diese Summe klein, sind auch alle einzelnen  $e_i$  klein (zumindest nicht grösser als die Summe) und umgekehrt.

Die Minimierung von  $S(k, d)$  ist möglich, jedoch algorithmisch ziemlich aufwändig.<sup>1</sup> Es stellt sich jedoch heraus, dass das entsprechende Optimierungsproblem viel ein-

<sup>1</sup>Es resultiert ein sogenanntes lineares Optimierungsproblem, das z.B. mit der Simplexmethode gelöst werden kann.

facher zu lösen ist, wenn wir anstatt der Summe der Beträge der Abweichungen die Summe der *Quadrate* der Abweichungen minimieren:

$$Q(k, d) = \sum_{i=1}^m e_i^2 \quad (1)$$

Genauso wie  $S(k, d)$  ist auch  $Q(k, d)$  genau dann klein, wenn alle einzelnen  $e_i$  klein sind.

### 3 Die Methode der kleinsten Quadrate

Die Bestimmung der Werte von  $k$  und  $d$ , wo  $Q(k, d)$  sein Minimum annimmt, ist ein zweidimensionales Beispiel für die *Methode der kleinsten (Fehler-)Quadrate*. Zweidimensional deswegen, weil wir *zwei* Parameter, nämlich  $k$  und  $d$  suchen, die in diesem Fall unsere Ausgleichsgerade bestimmen werden. Später werden wir die allgemeine Methode der kleinsten Quadrate kennenlernen, wo eine beliebige Anzahl  $n$  von Parametern gesucht ist.

Um das Minimum—einen Extremwert—von  $Q(k, d)$  bestimmen, bedienen wir uns eines wohlbekanntes Werkzeugs aus der Analysis, nämlich der Ableitung. Wir wissen, wie wir Extremwerte von Funktionen einer Variablen  $g: \mathbf{R} \rightarrow \mathbf{R}$  bestimmen können: Wir setzen die Ableitung gleich Null und lösen die resultierende Gleichung. Genau das können wir aber auch bei Funktionen zweier Variablen tun, denn es gilt folgender Satz:

**Satz 3.1** *Hat eine differenzierbare Funktion  $F: \mathbf{R}^2 \rightarrow \mathbf{R}$  ein (lokales oder globales) Minimum bei  $(x, y)$ , so sind die partiellen Ableitungen in diesem Punkt Null, d.h.,*

$$\frac{\partial F}{\partial x}(x, y) = 0$$

und

$$\frac{\partial F}{\partial y}(x, y) = 0.$$

Hierbei ist die *partielle Ableitung* von  $F$  nach  $x$ , (Notation:  $\partial F/\partial x$ ) nichts weiteres als die Ableitung von  $F$  nach  $x$ , wenn  $y$  als Konstante betrachtet wird und ansonsten die bekannten Ableitungsregeln verwendet werden. Umgekehrt wird bei  $\partial F/\partial y$   $x$  als Konstante betrachtet. Man nennt  $F$  *partiell differenzierbar*, wenn beide partiellen Ableitungen (nach  $x$  und nach  $y$ ) für alle  $(x, y)$  im Definitionsbereich existieren.

Der Beweis dieses Satzes soll hier ausgelassen werden. Er besteht im Wesentlichen in einer zweimaligen Anwendung des entsprechenden Satzes für eindimensionale Funktionen.

Wir versuchen also nun, obigen Satz auf unsere Funktion  $Q(k, d)$  anzuwenden und so das Minimum zu bestimmen. Durch Einsetzen der Definition von  $e_i$  in (1) ergibt sich:

$$\begin{aligned} Q(k, d) &= \sum e_i^2 \\ &= \sum (kx_i + d - y_i)^2 \\ &= k^2 \sum x_i^2 + md^2 + \sum y_i^2 + 2kd \sum x_i - 2k \sum x_i y_i - 2d \sum y_i. \end{aligned}$$

Die Summationsgrenzen ( $i = 1, \dots, m$ ) wurden zwecks besserer Lesbarkeit weggelassen. In Anwendung von Satz 3.1 berechnen wir die partiellen Ableitungen  $\partial Q/\partial k$  und  $\partial Q/\partial d$  und setzen beide gleich Null:

$$\begin{aligned} k \sum x_i^2 + d \sum x_i - \sum x_i y_i &= 0 \\ k \sum x_i + dm - \sum y_i &= 0. \end{aligned}$$

Alle Summen sowie  $m$  in den obigen Gleichungen sind Konstanten, also handelt es sich hier um ein simples lineares Gleichungssystem für  $k$  und  $d$ ! Wir können dieses problemlos mittels Umformungen oder der Cramerschen Regel lösen und es ergibt sich:

$$\begin{aligned} k &= \frac{mS_{xy} - S_x S_y}{mS_{x^2} - S_x^2} \\ d &= \frac{S_y - kS_x}{m}, \end{aligned} \quad (2)$$

wobei wir folgende Abkürzungen verwendet haben:

$$\begin{aligned} S_x &= \sum x_i \\ S_y &= \sum y_i \\ S_{x^2} &= \sum x_i^2 \\ S_{xy} &= \sum x_i y_i \end{aligned}$$

(Achtung:  $S_{x^2}$  ist die Summe der Quadrate der  $x_i$ , hingegen  $S_x^2 = (S_x)^2$  das Quadrat der Summe der  $x_i$ !)

Es kann leicht nachgeprüft werden, dass obiges Gleichungssystem genau dann eindeutig lösbar ist, wenn nicht alle  $x_i$  gleich sind (vgl. Übung 1).

## 4 Allgemeine lineare Regression und die Singulärwertzerlegung

Nicht immer lassen sich Datensätze durch Geraden approximieren. Glücklicherweise können wir jedoch für diese Fälle die Ausgleichsgerade zu einer Ausgleichskurve als Funktion von  $x$  verallgemeinern. Die Ausgleichskurve ist eine Linearkombination

$$F = a_1 f_1 + \dots + a_n f_n$$

von *Basisfunktionen*

$$f_1, \dots, f_n: \mathbf{R} \rightarrow \mathbf{R}$$

und *Regressionskoeffizienten*

$$a_1, \dots, a_n \in \mathbf{R}.$$

Die Basisfunktionen müssen aufgrund der Anwendung bzw. Annahmen oder Vermutungen über den Zusammenhang zwischen den Daten gewählt werden. Oft können Polynome moderaten Grades verwendet werden (etwa bis zum Grad 5). Die Methode der kleinsten Fehlerquadrate liefert die Regressionskoeffizienten für gegebene Daten und Basisfunktionen.

Die Basisfunktionen sollten linear unabhängig sein, d.h., keine von ihnen darf sich als Linearkombination der anderen darstellen lassen. Hier einige Beispiele:

- Die Polynome  $1, x, x^2, \dots, x^{n-1}$  sind linear unabhängig.
- Die Funktionen

$$f_1(x) = 1,$$

$$f_2(x) = x$$

und

$$f_3(x) = kx + d$$

sind linear abhängig, da sich offensichtlich  $f_3$  als Linearkombination von  $f_1$  und  $f_2$  darstellen lässt:

$$f_3(x) = kf_2(x) + df_1(x).$$

Dieses Beispiel zeigt auch, dass das die Ausgleichsgerade ein Spezialfall der allgemeinen Ausgleichskurve ist, wobei die Basisfunktionen 1 und  $x$  verwendet werden.

- Die Funktionen  $\sin(kx)$  und  $\cos(kx)$  für  $k = 1, \dots, N$  sind linear unabhängig. Sie eignen sich besonders für die Approximation *periodischer* Funktionen, zum Beispiel Audiosignalen. Die Approximation durch diese Funktionen nennt man *Fourierreihen*.

Wir setzen für die weiteren Untersuchungen voraus, dass wir mehr Datenpunkte als Basisfunktionen haben, also  $m \geq n$ . Andernfalls wäre die Regression unterbestimmt und die Lösung nicht eindeutig.

Um die Ausgleichskurve zu berechnen, minimieren wir wie für die Ausgleichsgerade die Summe der (Approximations)-Fehlerquadrate, jetzt eine Funktion der Regressionskoeffizienten  $a_1, \dots, a_n$ :

$$\begin{aligned} Q(a_1, \dots, a_n) &= \sum_{i=1}^m (F(x_i) - y_i)^2 \\ &= \sum_{i=1}^m \left( \sum_{j=1}^n a_j f_j(x_i) - y_i \right)^2 \end{aligned}$$

Mit den Matrizen und Vektoren

$$M = \begin{pmatrix} f_1(x_1) & \cdots & f_n(x_1) \\ \vdots & \ddots & \vdots \\ f_1(x_m) & \cdots & f_n(x_m) \end{pmatrix},$$

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix},$$

und

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix},$$

können wir  $Q$  schreiben als

$$Q(a_1, \dots, a_n) = \|Ma - y\|^2. \quad (3)$$

Dieser Wert soll nun minimiert werden.

Falls das lineare Gleichungssystem

$$Ma = y$$

lösbar ist, dann ist  $Ma - y$  und somit der Minimalwert des Optimierungsproblems gleich Null. In diesem Fall sind auch alle einzelnen Fehler gleich Null, d.h., die Ausgleichskurve ist sozusagen "exakt" und geht durch *alle* Datenpunkte. Dies ist jedoch im Allgemeinen weder möglich noch erwünscht (Beispiel: Overfitting bei Approximation durch ein Polynom von Grad 9 in Abbildung 4).

Es bleibt uns also nichts anderes übrig, als das Optimierungsproblem zu lösen. Dafür gibt es verschiedene Methoden. Wir beschreiben diejenige, die die sogenannte *Singulärwertzerlegung* (engl.: *singular value decomposition* oder *SVD*) der Matrix  $M$  benutzt. Die SVD ist nicht nur für die Lösung des Problems der kleinsten Fehlerquadrate nützlich, sondern auch für eine Reihe anderer Probleme aus der Linearen Algebra und deren Anwendungen.

**Satz 4.1** Sei  $M \in \mathbf{R}^{(m \times n)}$  eine beliebige Matrix. Dann existieren orthogonale Matrizen  $U$  und  $V$  sowie eine eindeutige Diagonalmatrix

$$\Sigma = \begin{pmatrix} \sigma_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & \sigma_r & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} \in \mathbf{R}^{m \times n},$$

sodass

$$M = U\Sigma V^T$$

und

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0.$$

Die Diagonalmatrix  $\Sigma$  enthält die sogenannten *Singulärwerte* von  $M$  der Größe nach geordnet. Eine Matrix vom Rang  $r$  hat genau  $r$  Singulärwerte grösser als Null.

Im Gegensatz zu den Eigenwerten sind die Singulärwerte von reellen Matrizen immer reell, und die SVD kann immer gefunden werden. Die Singulärwerte haben eine intuitive geometrische Interpretation als Halbachsen eines Ellipsoids, vgl. Abbildung 3.

Mit Hilfe der Singulärwertzerlegung können wir  $Q(a_1, \dots, a_n)$  wie folgt umformen:

$$\begin{aligned} \|Ma - y\|^2 &= \|U\Sigma V^T a - y\|^2 \\ &= \|U\Sigma b - y\|^2 \quad (\text{mit } b = V^T a) \\ &= \|U^T U \Sigma b - U^T y\|^2 \\ &= \|\Sigma b - U^T y\|^2 \\ &= \|\Sigma b - z\|^2 \quad (\text{mit } z = U^T y) \\ &= (\sigma_1 b_1 - z_1)^2 + \cdots + (\sigma_r b_r - z_r)^2 + z_{r+1}^2 + \cdots + z_m^2 \end{aligned}$$

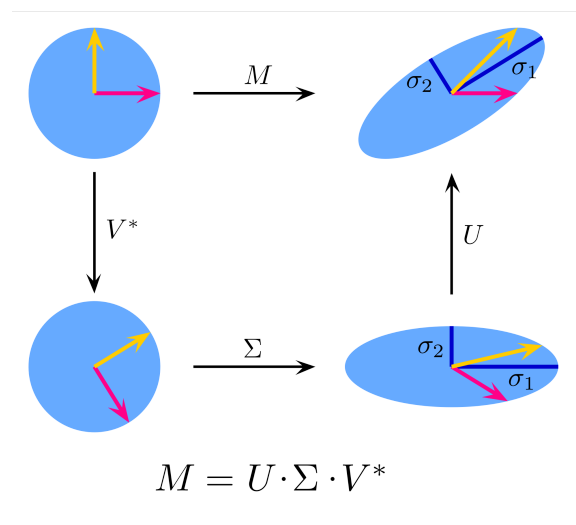


Abbildung 3: Graphische Darstellung der SVD.  
 Die Matrix dieses Beispiels ist die Scherung

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

Die Grafik zeigt, wie die involvierten Matrizen die Einheitsvektoren des  $\mathbf{R}^2$  (gelb und rot) sowie den Einheitskreis (blau) abbilden. Die Matrix  $\Sigma$  ist als Diagonalmatrix eine Kombination von Stauchung und Streckung parallel zu den Koordinatenachsen. Die Matrizen  $U$  und  $V$  sind orthogonal, hier Drehungen.  $\sigma_1$  und  $\sigma_2$  sind die Längen der Halbachsen des Bildes des Einheitskreises unter  $M$  (dieses ist immer ein Ellipsoid).  
 Quelle der Grafik: Wikipedia



Dabei haben wir neben den Substitutionen durch

$$b = (b_1, \dots, b_n)^T$$

und

$$z = (z_1, \dots, z_m)^T$$

die Tatsache verwendet, dass Multiplikation mit der orthogonalen Matrix  $U^T$  die Norm unverändert lässt, sowie dass  $U^T U = I$ .

Wir müssen nun  $b$  so wählen, dass der Ausdruck

$$(\sigma_1 b_1 - z_1)^2 + \dots + (\sigma_r b_r - z_r)^2 + z_{r+1}^2 + \dots + z_m^2 \quad (4)$$

minimal wird. Dank der nun "diagonalisierten" Form ist dies jetzt sehr einfach: Für  $1 \leq i \leq r$  setzen wir

$$b_i = z_i / \sigma_i.$$

Die restlichen Werte  $b_{r+1}, \dots, b_n$  kommen nicht in (4) vor, können also beliebig gewählt werden. Wir setzen sie Null. Somit ist  $b$  bestimmt und wir können durch Rückeinsetzen die gesuchten Koeffizienten  $a$  berechnen:

$$a = Vb$$

**Bemerkung** Für die Berechnung von  $a$  brauchen wir nicht den vollen  $m$ -Vektor  $z$ , sondern nur dessen Komponenten  $z_1$  bis  $z_n$  (Wir erinnern uns, dass i.A.  $n \ll m$ ). Da  $z = U^T y$  und die Matrix  $U$  ausschliesslich in dieser Substitution gebraucht wird, brauchen wir also nur die ersten  $n$  Spalten von  $U$ . Dies kann eine signifikante Einsparung von Speicherplatz und Rechenzeit bedeuten.

**Bemerkung** Wenn die  $n$  Basisfunktionen linear unabhängig sind, wird die Matrix  $M$  in den allermeisten Fällen vollen Rang haben, d.h.  $r = n$ .

**Beispiel 4.2** In *Abbildung 4* sehen Sie eine Sinuskurve und 10 Punkte, die durch zufällige kleine Abweichungen von dieser Sinuskurve auf der  $y$ -Achse erzeugt wurden. Diese künstlich erzeugten Daten wurden dann mit Polynomen verschiedenen Grades approximiert.

Man sieht, dass sich ab Grad 3 eine passable Approximation ergibt. Bei zu hohem Grad kommt es allerdings zu sogenanntem "Overfitting": Das Polynom vom Grad 9 geht zwar durch alle Punkte (der Approximationsfehler ist Null), oszilliert jedoch dazwischen wild hin und her.

Overfitting ist meistens unerwünscht. Um es zu vermeiden, kann man entweder die Anzahl der Datenpunkte erhöhen oder die Anzahl der Basisfunktionen verringern. In *Abbildung 5* sieht man die Approximation von 100 Datenpunkten durch ein Polynom vom Grad 9.

**Beispiel 4.3** In *Abbildung 6* approximieren wir die Bevölkerungsdaten der Schweiz von 1850–2009. Konnte man im 19. Jahrhundert noch annehmen, das Wachstum sei linear (vgl. *Abbildung 1*), so zeigt diese Grafik, dass die quadratische Ausgleichskurve die Daten wesentlich besser approximiert als die Ausgleichsgerade.

Für 2050 sagt die quadratische Kurve eine Bevölkerung von über 10 Millionen vorher.

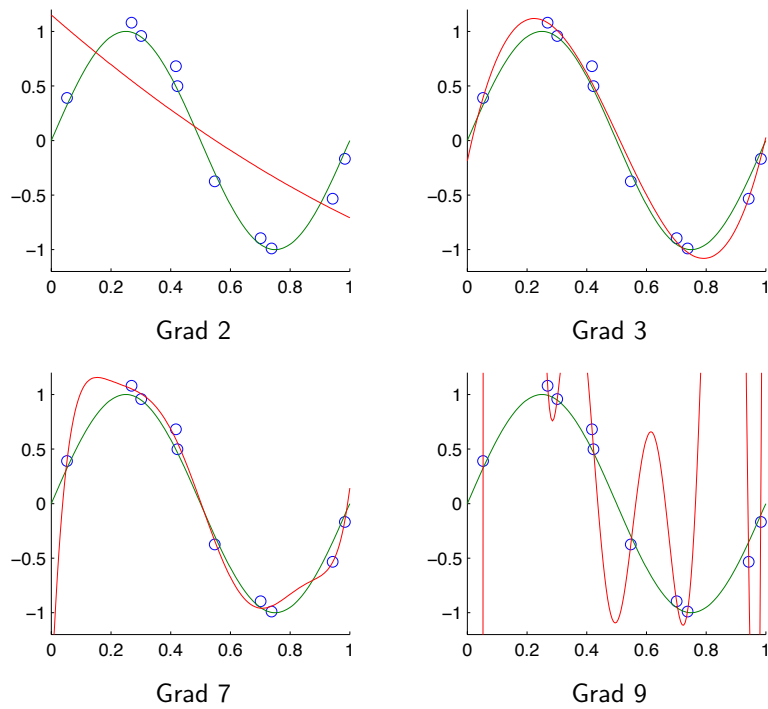


Abbildung 4: Approximation einer Sinuskurve (grün) durch Polynome verschiedenen Grades (rot).

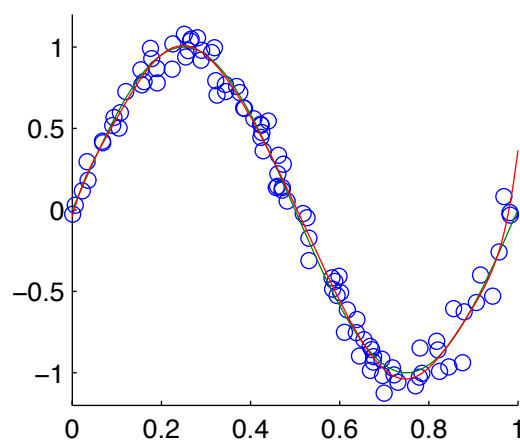


Abbildung 5: Approximation von 100 Datenpunkten durch ein Polynom von Grad 9.

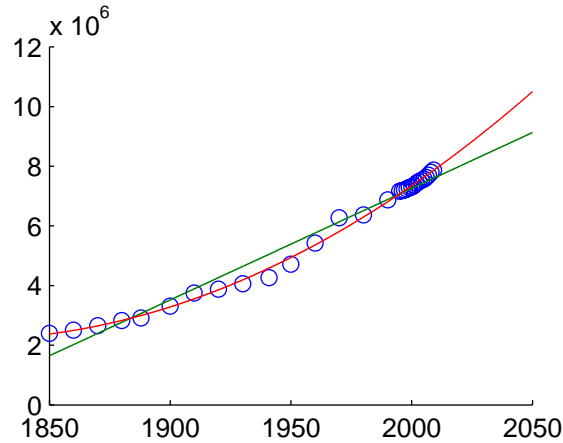


Abbildung 6: Bevölkerungsdaten der Schweiz (blaue Kreise) mit linearer (grün) und quadratischer (rot) Ausgleichskurve.

## 5 Numerische Aspekte

Der bei weitem wichtigste numerische Aspekt bei der Lösung des Problems der kleinsten Fehlerquadrate (vgl. (3)) ist die Konditionszahl der Matrix  $M$ . Im Fall von Polynomen als Basisfunktionen kann diese schon bei relativ geringem Grad sehr hoch werden.

Wir wollen dies am Bevölkerungsbeispiel von Abschnitt 1 erläutern. Wir approximieren die 5 Datenpunkte durch ein quadratisches Polynom. Die Zeilen von  $M$  sind dann die Vektoren  $(1, x, x^2)$ , wobei  $x$  die Jahre durchläuft,

$$M = \begin{pmatrix} 1 & 1850 & 3422500 \\ 1 & 1860 & 3459600 \\ 1 & 1870 & 3496900 \\ 1 & 1880 & 3534400 \\ 1 & 1888 & 3564544 \end{pmatrix}$$

Man sieht hier bereits, dass recht grosse Zahlen mit recht kleinen zusammen in einer Matrix vorkommen. In der Tat ist die Konditionszahl von  $M$  bereits in diesem sehr einfachen Beispiel sehr hoch, nämlich  $\approx 8.2 \cdot 10^{10}$ . Käme das Jahr  $x$  in noch höheren Potenzen vor, würde sich das Problem sogar noch verschärfen.

Um das Problem zu entschärfen, verschieben wir die  $x$ -Komponenten der Daten, sodass deren Mittelwert gleich Null wird. Dazu subtrahieren wir den Mittelwert der ursprünglichen  $x$ -Daten (hier 1869.6) von allen  $x$ -Werten und erhalten neu die Matrix

$$M' = \begin{pmatrix} 1 & -19.6 & 384.16 \\ 1 & -9.6 & 92.16 \\ 1 & 0.4 & 0.16 \\ 1 & 10.4 & 108.16 \\ 1 & 18.4 & 338.56 \end{pmatrix}$$

mit der wesentlich besseren Konditionszahl  $\approx 379$ .

Da eine Verschiebung auf der  $x$ -Achse nichts an den Abweichungen der Ausgleichskurve von den Daten ändert, ist die Lösung—wenn man sie nach der numerischen Berechnung wieder in den Originalbereich zurückverschiebt—exakt die gleiche wie diejenige des ursprünglichen Problems.

Wir empfehlen, diese Verschiebung immer anzuwenden, sofern nicht aufgrund der Natur der Daten der Mittelwert ohnehin Null ist.

## 6 Matlab-Funktionen

Die folgenden Matlab-Funktionen sind im Zusammenhang mit Linearer Regression von besonderem Interesse:

- $P = \text{POLYFIT}(X, Y, N)$  berechnet ein Polynom von Grad  $N$ , das die Daten  $(X, Y)$  im Sinne der kleinsten Fehlerquadrate approximiert.
- $Y = \text{POLYVAL}(P, X)$  wertet ein Polynom  $P$  an den Punkten im Vektor  $X$  aus.
- $[U, S, V] = \text{SVD}(X)$  berechnet die Singulärwertzerlegung der Matrix  $X$ .
- $[U, S, V] = \text{SVD}(X, 0)$  berechnet die “Sparvariante” (“economy size”) der SVD, in der nur die ersten  $n$  Spalten von  $U$  berechnet werden ( $X$  ist eine  $(m \times n)$ -Matrix).
- $x = A \backslash b$  oder  $x = \text{MLDIVIDE}(A, x)$  berechnet für eine Matrix  $A$  und einen Vektor  $b$  die Lösung  $x$  des Gleichungssystems

$$Ax = b.$$

Falls diese Lösung nicht existiert, liefert  $A \backslash b$  die Lösung des Optimierungsproblems

$$\|Ax - b\| \rightarrow \min.$$

Dieser Operator kann also benutzt werden, um direkt das Problem der kleinsten Fehlerquadrate zu lösen.

- $\text{SCATTER}(X, Y)$  plottet Datenvektoren, d.h. die Punkte  $(X_i, Y_i)$ .  $X$  und  $Y$  müssen gleich lange Zeilen- oder Spaltenvektoren sein.
- $\text{COND}(X)$  berechnet die Konditionszahl der Matrix  $X$ .
- Die *Curve Fitting Toolbox* (als Option zu Matlab erhältlich) bietet Funktionen zur linearen und nichtlinearen Regression (die hier nicht behandelt wurde), sowie eine graphische Benutzeroberfläche zu deren Durchführung (`cftool`).

Wir verweisen auf die Matlab-Dokumentation für die vollständige Beschreibung obiger Funktionen.

## 7 Übungsbeispiele

*Musterlösungen sowie der verwendete Datensatz der Schweizer Bevölkerungsentwicklung stehen auf der Webseite des Autors zur Verfügung (<http://gwesptx0.org/>).*

1. Zeigen Sie, dass die Formel für die Ausgleichsgerade (2) genau dann “funktioniert” (d.h. keine Division durch Null auftritt), wenn der Datenvektor mindestens zwei unterschiedliche  $x_i$  enthält.
2. Implementieren Sie eine Funktion `p = linefit(x,y)` analog zu Matlabs Funktion `polyfit`, die mittels (2) die Ausgleichsgerade zu den Daten  $x$  und  $y$  berechnet.
3. Informieren Sie sich über die Berechnungskomplexität der SVD und geben Sie die Berechnungskomplexität des allgemeinen linearen Regressionsproblems mit  $n$  Basisfunktionen und  $m$  Datenpunkten an. Wie würde es sich auswirken, würde man die volle  $(m \times m)$ -Matrix  $U$  anstelle der “Sparvariante” verwenden?
4. Implementieren Sie eine Funktion `mypolyfit` analog zur Matlab-Funktion `polyfit`. Verwenden Sie dabei die Singulärwertzerlegung (nicht den ‘\’-Operator).
5. Implementieren Sie die Verschiebung der  $x$ -Daten zum Mittelwert wie in Abschnitt 5 beschrieben in der Funktion `mypolyfit`. Die Verschiebung soll transparent für den Benutzer erfolgen, das von der Funktion gelieferte Polynom soll also wieder auf den ursprünglichen  $x$ -Bereich bezogen sein.
6. Approximieren Sie nun den *Logarithmus* der Bevölkerungszahlen durch eine Gerade. Approximieren Sie die Bevölkerungszahlen aufgrund dieser Gerade dann durch eine Exponentialfunktion. Schätzen Sie damit die Bevölkerung in den Jahren 2050, 2100 und 2200!