

Kapitel 7

Methode der kleinsten Quadrate

Im Folgenden wird die Methode der kleinsten Quadrate (LS = ‘least square’), die auf dem χ^2 -Test beruht, für die Anpassung von parametrisierten Funktionen an normalverteilte (oder annähernd normalverteilte) Messwerte eingeführt. Im vorigen Kapitel hatten wir bereits darauf hingewiesen, dass diese Methode der Maximum-Likelihood-Methode im Falle normalverteilter Wahrscheinlichkeiten entspricht.

7.1 Prinzip der Methode der kleinsten Quadrate

Gegeben sei eine Stichprobe mit folgenden Messwerten und der parametrisierten Beschreibung der Messwerte:

y_i : Messwerte an den (ohne Fehler) bekannten Punkten x_i (unabhängige Variable, kann auch ein Vektor sein, $i = 1, \dots, n$);

σ_i : Fehler von y_i , Standardabweichung;

η_i : $\eta_i = f(x_i|\theta)$ ist der Erwartungswert von y_i , wenn die Abhängigkeit von x_i durch $f(x|\theta)$ beschrieben wird;

θ_j : Parameter der Funktion f , die so optimiert werden sollen, dass $f(x_i|\theta) = \eta_i$ die Messwerte y_i möglichst gut beschreibt ($j = 1, \dots, m$).

Das LS-Prinzip lautet: Bestimme die Schätzwerte $\hat{\theta}$ der Parameter $\theta = (\theta_1, \dots, \theta_m)$ durch Minimierung der Summe der Quadrate der auf die Fehler normierten Abweichungen:

$$S = \sum_{i=1}^n \frac{(y_i - \eta_i)^2}{\sigma_i^2} = \sum_{i=1}^n \frac{(y_i - f(x_i|\theta))^2}{\sigma_i^2} \quad (7.1)$$

Wenn die Messwerte korreliert sind, $\text{cov}(y_i, y_j) \neq 0$, muss man die gesamte Kovarianzmatrix $V_{ij}(y)$ der y -Werte benutzen:

$$S = \sum_{i=1}^n \sum_{j=1}^n (y_i - \eta_i) V_{ij}^{-1}(y) (y_j - \eta_j) \quad (7.2)$$

Wenn die Messwerte y_i einer Normalverteilung mit einer Breite σ_i um den wahren Wert $\eta_i = f(x_i|\theta)$ folgen, dann folgt die LS-Funktion S einer χ^2 -Verteilung

mit $n_F = n - m$ Freiheitsgraden (Anzahl der Messungen minus Anzahl der aus den Messungen bestimmten Parametern). Da der Erwartungswert von $E(\chi^2) = n_F$ ist, ist die Erwartung für die Verminderung von χ^2 bei Hinzunahme eines Parameters $E(\Delta\chi^2) \geq 1$. Das heisst χ^2 vermindert sich im Mittel um 1, selbst wenn der zusätzliche Parameter nicht notwendig ist. Die Signifikanz für die Notwendigkeit eines Parameters ergibt sich aus $\Delta\chi^2$.

Für den betrachteten Fall normalverteilter Messwerte ergibt sich die Likelihood-Funktion:

$$L = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(y_i - \eta_i)^2}{2\sigma_i^2}} \quad (7.3)$$

Daraus berechnet sich die Log-Likelihood-Funktion:

$$\mathcal{L} = - \sum_{i=1}^n \frac{(y_i - \eta_i)^2}{2\sigma_i^2} - \sum_{i=1}^n \ln \sqrt{(2\pi\sigma_i^2)} = -\frac{1}{2} S + \text{const.} \quad (7.4)$$

In diesem Fall entspricht also die Parameteroptimierung durch Maximierung von \mathcal{L} genau der Optimierung durch Minimierung der LS-Funktion S , das heisst die ML- und LS-Methoden sind für normalverteilte Messwerte äquivalent. Das LS-Prinzip wird allerdings häufig auch für andere Verteilungen der Messwerte benutzt, weil die formelmässige Behandlung des Problems in der Regel einfacher ist.

7.2 Lineare Anpassung

In der Praxis kommt häufig der Fall vor, dass die Anpassungsfunktion $f(x|\theta)$ eine lineare Funktion der Parameter $\theta = (\theta_1, \dots, \theta_m)$ ist:

$$f(x|\theta) = \theta_1 f_1(x) + \dots + \theta_m f_m(x) \quad (7.5)$$

Die f_j können beliebige (also auch nicht-lineare) Funktionen von x sein.

7.2.1 Anpassung der Messwerte an eine Gerade

Für die Hypothese, dass die Messwerte auf einer Geraden liegen sollen, ergibt sich die Anpassungsfunktion ($f_1(x) = 1$, $f_2(x) = x$):

$$f(x|\theta) = \theta_1 + x\theta_2 \quad (7.6)$$

Die Messungen ergeben die n Tripel (x_i, y_i, σ_i) (Abb. 7.1). Wenn die y_i unabhängig sind erhält man die LS-Funktion:

$$S = \sum_{i=1}^n \frac{(y_i - \eta_i)^2}{\sigma_i^2} = \sum_{i=1}^n \frac{(y_i - \theta_1 - x_i\theta_2)^2}{\sigma_i^2} \quad (7.7)$$

Die Minimierung von S als Funktion der Parameter fordert:

$$\begin{aligned} \frac{\partial S}{\partial \theta_1} &= \sum \frac{-2}{\sigma_i^2} (y_i - \theta_1 - x_i\theta_2) = 0 \\ \frac{\partial S}{\partial \theta_2} &= \sum \frac{-2x_i}{\sigma_i^2} (y_i - \theta_1 - x_i\theta_2) = 0 \end{aligned} \quad (7.8)$$

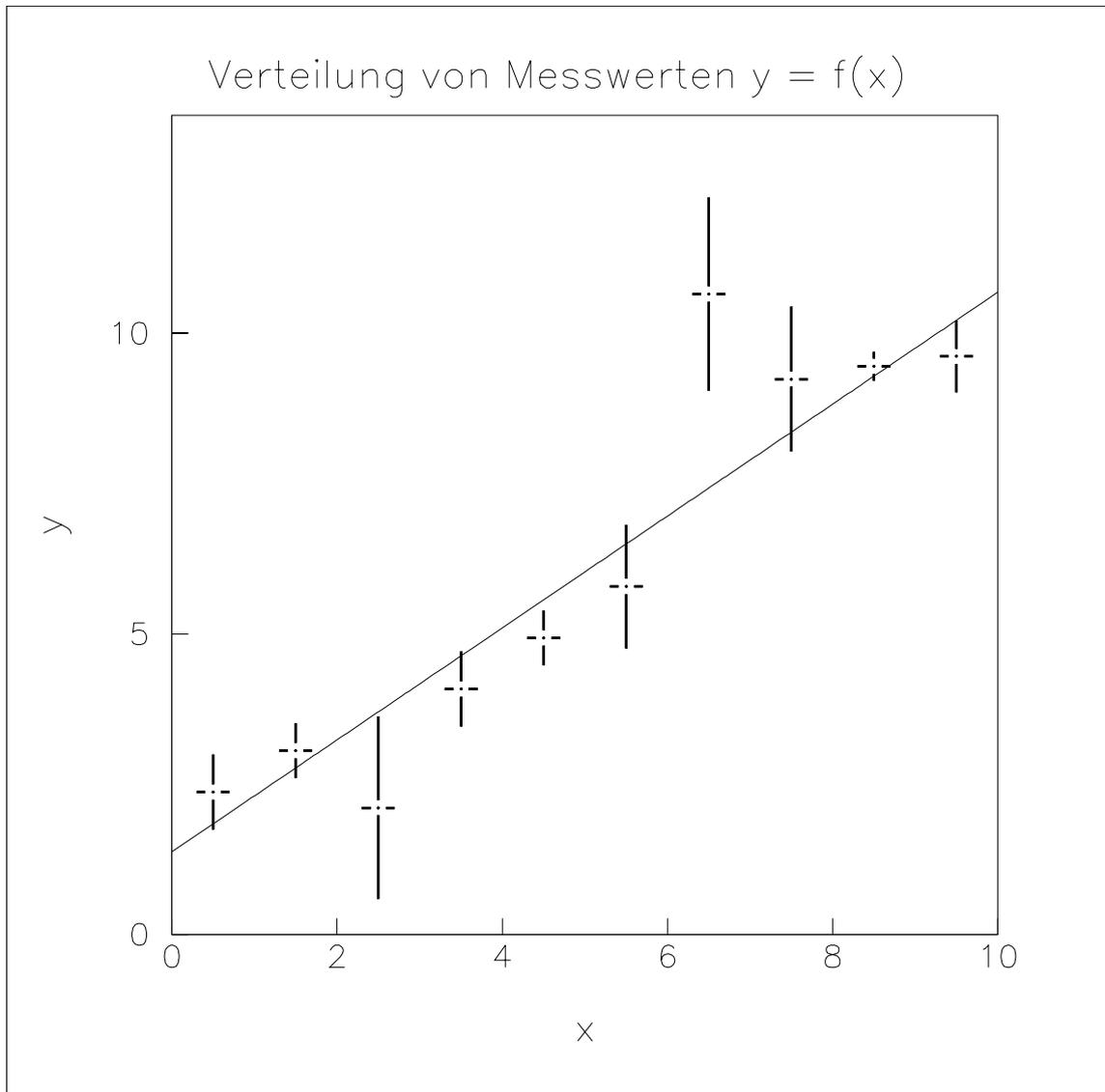


Abbildung 7.1: Messwerte y_i als Funktion von x mit normalverteilten Fehlern. Die Anpassung einer Geraden an die 10 Datenpunkte liefert für Achsenabschnitt und Steigung: $\theta_1 = 1.37 \pm 0.36$, $\theta_2 = 0.93 \pm 0.05$ und $\chi^2 = 11.4$ bei 8 Freiheitsgraden, entsprechend einem Vertrauensniveau von etwa 20%. Die Anpassung wurde mit dem CERN-Programm MINUIT durchgeführt.

Aus der Minimierungsbedingung ergibt sich ein lineares inhomogenes Gleichungssystem für die θ_i . Zur weiteren Behandlung bilden wir folgende Summen, die zum Beispiel auch in entsprechenden Computer-Programmen gebildet werden:

$$\begin{aligned} S_1 &= \sum \frac{1}{\sigma_i^2} \\ S_x &= \sum \frac{x_i}{\sigma_i^2} \\ S_y &= \sum \frac{y_i}{\sigma_i^2} \\ S_{xx} &= \sum \frac{x_i^2}{\sigma_i^2} \\ S_{xy} &= \sum \frac{x_i y_i}{\sigma_i^2} \end{aligned} \quad (7.9)$$

Damit folgt aus (7.8) für die LS-Schätzung $\hat{\theta}$:

$$\begin{aligned} S_1 \cdot \hat{\theta}_1 + S_x \cdot \hat{\theta}_2 &= S_y \\ S_x \cdot \hat{\theta}_1 + S_{xx} \cdot \hat{\theta}_2 &= S_{xy} \end{aligned} \quad (7.10)$$

Mit der Determinante der Koeffizientenmatrix

$$D = S_1 S_{xx} - S_x^2 \quad (7.11)$$

ergeben sich durch Auflösung von (7.10) die **LS-Schätzwerte der Parameter**:

$$\begin{aligned} \hat{\theta}_1 &= \frac{1}{D} (S_{xx} S_y - S_x S_{xy}) \\ \hat{\theta}_2 &= \frac{1}{D} (S_1 S_{xy} - S_x S_y) \end{aligned} \quad (7.12)$$

Kovarianzmatrix der Parameter: Die Fehler der Parameter ergeben sich aus der Relation (6.42):

$$V_{ij}^{-1} = - \left. \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j} \right|_{\theta=\hat{\theta}} = + \frac{1}{2} \left. \frac{\partial^2 S}{\partial \theta_i \partial \theta_j} \right|_{\theta=\hat{\theta}} \quad (7.13)$$

Die einzelnen Matrixelemente sind:

$$\begin{aligned} \frac{1}{2} \frac{\partial^2 S}{\partial \theta_1^2} &= \sum \frac{1}{\sigma_i^2} = S_1 \\ \frac{1}{2} \frac{\partial^2 S}{\partial \theta_1 \partial \theta_2} &= \sum \frac{x_i}{\sigma_i^2} = S_x \\ \frac{1}{2} \frac{\partial^2 S}{\partial \theta_2^2} &= \sum \frac{x_i^2}{\sigma_i^2} = S_{xx} \end{aligned} \quad (7.14)$$

Die inverse Kovarianzmatrix ist also:

$$V^{-1}(\theta) = \begin{pmatrix} S_1 & S_x \\ S_x & S_{xx} \end{pmatrix} \quad (7.15)$$

Die Kovarianzmatrix erhält man aus der Inversion:

$$V(\theta) = \frac{1}{D} \begin{pmatrix} S_{xx} & -S_x \\ -S_x & S_1 \end{pmatrix} \quad (7.16)$$

Extrapolationsfehler: Damit lässt sich der y -Wert zu jedem beliebigen x -Wert berechnen:

$$y = \hat{\theta}_1 + x \hat{\theta}_2 \quad (7.17)$$

Der Fehler von y ergibt sich durch Fehlerfortpflanzung:

$$\sigma^2(y) = V_{11} + x^2 V_{22} + 2x V_{12} = \frac{1}{D}(S_{xx} + x^2 S_1 - 2x S_x) \quad (7.18)$$

Güte der Anpassung: Die Größe

$$\chi^2 = S_{min} = S(\hat{\theta}) \quad (7.19)$$

folgt einer χ^2 -Verteilung mit $n_F = n - m = n - 2$ Freiheitsgraden (Anzahl der Messungen minus Anzahl der Parameter) mit dem Erwartungswert

$$E(\chi^2) = n_F. \quad (7.20)$$

Für das Ergebnis der Anpassung (oder des ‘Fits’) kann man dann das Vertrauensniveau α wie in Abschnitt 4.3 (Gl. 4.36 und Abb. 4.2) bestimmen, wenn die Messwerte normalverteilt sind. Zum Beispiel ist bei 12 Messungen $n_F = 10$ und man liest folgende Vertrauensniveaus α für $\chi^2 = S_{min}$ ab:

S_{min}	α [%]
8	62.9
10	44.0
12	28.5
16	10.0

Mit dem Folgenden Python-Skript kann diese Tabelle reproduziert werden:

```
from scipy import *
for x in [8.,10.,12.,16] :
    print x, stats.chi2.sf(x,10.)
```

Geringe Vertrauensniveaus können die gleichen Gründe haben, wie in Abschnitt 4.3 angeführt (falsches Modell, falsche Fehler, Untergrund). Wenn das Gauss-Modell nicht zutrifft, kann die Variation von χ^2 um das Minimum immer noch ein gutes Mass für die Bestimmung der Parameter sein. Wie in Abschnitt 6.4.3 ausgeführt, erhält man eine **Schätzung der Standardabweichung eines Parameters**, wenn man diesen Parameter so variiert (die anderen Parameter bleiben fest), dass sich χ^2 um

$$\Delta\chi^2 = 1 \quad (7.21)$$

ändert.

7.2.2 Anpassung einer allgemeinen linearen Funktion der Parameter

Wir wollen jetzt die LS-Anpassung einer allgemeinen linearen Funktion von m Parametern betrachten:

$$f(x|\theta) = \theta_1 f_1(x) + \dots + \theta_m f_m(x) \quad (7.22)$$

Die LS-Anpassung an n Messwerte y_i an den Punkten x_i hat $n_F = n - m$ Freiheitsgrade. Es wird zugelassen, dass die Messwerte nicht unabhängig sind, dass also die Kovarianzmatrix $V(y)$ nicht-verschwindende $\text{cov}(y_i, y_j)$ -Terme hat.

Die Erwartungswerte für die n Messwerte y_i sind dann:

$$\eta_i = \theta_1 f_1(x_i) + \dots + \theta_m f_m(x_i) = \sum_{j=1}^m \theta_j f_j(x_i) \quad (7.23)$$

Um eine kompakte Schreibweise zu erhalten, definieren wir die $n \times m$ -Matrix H :

$$H_{ij} = f_j(x_i) \quad (7.24)$$

Damit wird (7.23):

$$\eta_i = \sum_{j=1}^m H_{ij} \theta_j \Rightarrow \vec{\eta} = H \theta \quad (7.25)$$

Mit der Kovarianzmatrix $V(y)$ der Messwerte ergibt sich dann die LS-Funktion (zur Abkürzung soll im Folgenden $V(y) = V$ gesetzt werden; die Kovarianzmatrix der Parameter wird dann $V(\theta)$ genannt):

$$S = (\vec{y} - H \theta)^T V^{-1} (\vec{y} - H \theta) \quad (7.26)$$

Die **Minimierungsbedingung** fordert, dass der Gradient von S bezüglich der Parameter verschwindet:

$$\vec{\nabla}_{\theta} S = -2 H^T V^{-1} (\vec{y} - H \theta) = 0 \quad (7.27)$$

Daraus ergibt sich ein lineares Gleichungssystem für θ :

$$H^T V^{-1} H \theta = H^T V^{-1} \vec{y} \quad (7.28)$$

Wenn $H^T V^{-1} H$ nicht singulär und damit invertierbar ist, ist die Lösung:

$$\hat{\theta} = (H^T V^{-1} H)^{-1} H^T V^{-1} \vec{y} \quad (7.29)$$

Durch Matrixinversionen lassen sich die Lösungen im Prinzip exakt bestimmen. Allerdings wird man bei $m > 3$ auf numerische Methoden für die Matrixinversionen zurückgreifen müssen.

Kovarianzmatrix der Parameter: Nach (7.29) ergeben sich die Parameter θ aus einer linearen Transformation der Messwerte:

$$\hat{\theta} = (H^T V^{-1} H)^{-1} H^T V^{-1} \vec{y} = A \vec{y} \quad (7.30)$$

Dann ergibt sich nach (3.63) die Kovarianzmatrix der Parameter θ durch Fehlerfortpflanzung als lineare Transformation der Kovarianzmatrix der Messwerte \vec{y} :

$$V(\theta) = A \cdot V(y) \cdot A^T \quad (7.31)$$

Nach Einsetzen von A erhält man:

$$\begin{aligned} V(\theta) &= A \cdot V(y) \cdot A^T = (H^T V^{-1} H)^{-1} H^T V^{-1} V [(H^T V^{-1} H)^{-1} H^T V^{-1}]^T = (H^T V^{-1} H)^{-1} \\ &V(\theta) = (H^T V^{-1} H)^{-1} \end{aligned} \quad (7.32)$$

Der Ausdruck $(H^T V^{-1} H)^{-1}$ ist bereits zur Lösung der Gleichung (7.29) für die Parameter berechnet worden.

Zusammenfassung der Formeln für die lineare Anpassung: Die beste Anpassung ergibt sich fuer die Parameter nach (7.29):

$$\hat{\theta} = (H^T V^{-1} H)^{-1} H^T V^{-1} \vec{y} \quad (7.33)$$

Die Parameter haben die Kovarianzmatrix (7.32)

$$V(\theta) = (H^T V^{-1} H)^{-1} \quad (7.34)$$

Der χ^2 -Wert der Anpassung ist:

$$\chi_{min}^2 = S(\hat{\theta}) = (\vec{y} - H \hat{\theta})^T V^{-1} (\vec{y} - H \hat{\theta}) \quad (7.35)$$

In MATLAB (oder mit Python) lassen sich diese Formeln mit den Matrixoperationen sehr einfach programmieren. Ein Beispiel ist in Abb. 7.2 gezeigt.

Beispiel: Wir betrachten den Fall, den wir im vorigen Abschnitt 7.2.1 bereits speziell behandelt haben: Geradengleichung ($m = 2$), unabhängige Messungen y_i :

$$H = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{pmatrix}, \quad V^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & 0 & 0 & 0 \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & 0 & 0 & 0 & \frac{1}{\sigma_n^2} \end{pmatrix} \quad (7.36)$$

Die benötigten Produkte dieser Matrizen sind:

$$V^{-1} H = \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{x_1}{\sigma_1^2} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \frac{1}{\sigma_n^2} & \frac{x_n}{\sigma_n^2} \end{pmatrix} = (H^T V^{-1})^T \Rightarrow H^T V^{-1} \vec{y} = \begin{pmatrix} \sum \frac{y_i}{\sigma_i^2} \\ \sum \frac{x_i y_i}{\sigma_i^2} \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix} \quad (7.37)$$

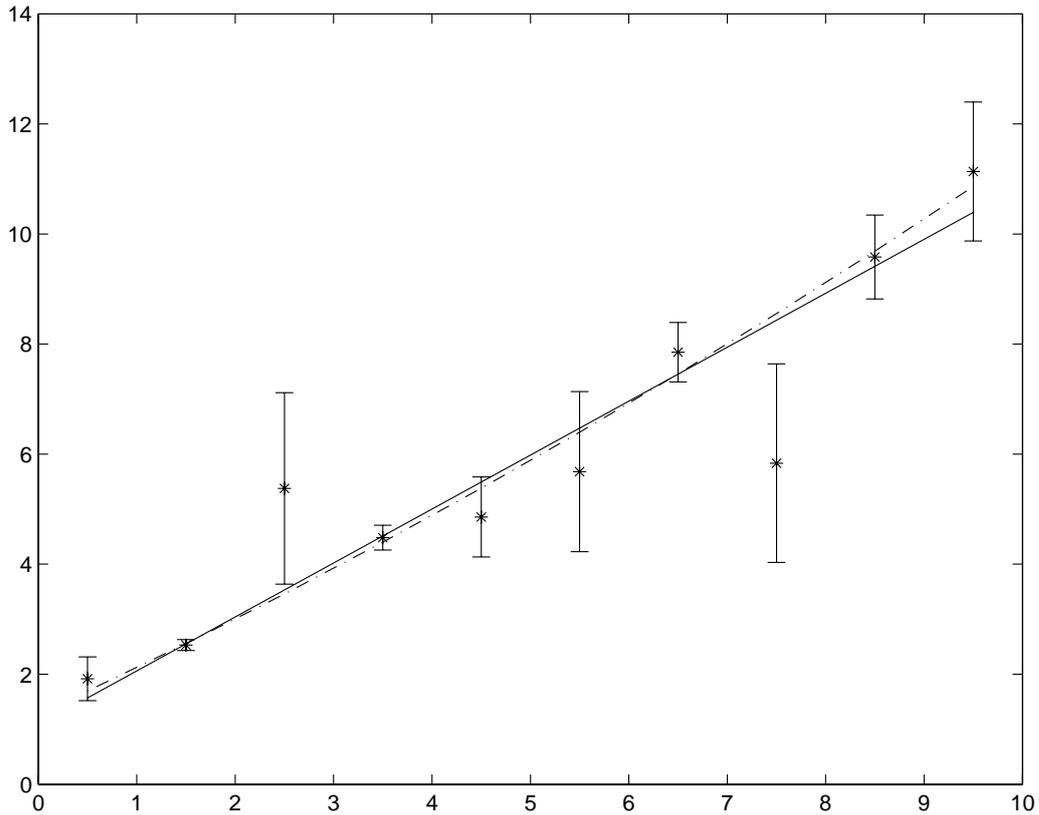
$$H^T V^{-1} H = \begin{pmatrix} \sum \frac{1}{\sigma_i^2} & \sum \frac{x_i}{\sigma_i^2} \\ \sum \frac{x_i}{\sigma_i^2} & \sum \frac{x_i^2}{\sigma_i^2} \end{pmatrix} = \begin{pmatrix} S_1 & S_x \\ S_x & S_{xx} \end{pmatrix} \quad (7.38)$$

Damit wird also die Gleichung (7.10) reproduziert:

$$\begin{pmatrix} S_1 & S_x \\ S_x & S_{xx} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix} \quad (7.39)$$

Anpassung an ein orthogonales Funktionensystem: Wenn die Messwerte y_i unabhängig sind, also ihre Kovarianzmatrix diagonal,

$$V_{ij}(y) = \sigma_i^2(y) \cdot \delta_{ij}, \quad (7.40)$$



Fit mit 2 Parametern $y = a_0 + a_1 \cdot x$

Parameter mit Fehlern:

1.0787 +/- 0.1482
0.9805 +/- 0.0589

Kovarianzmatrix:

0.0220 -0.0071
-0.0071 0.0035

chi2 = 6.0240 chi2/nf = 0.7530

Fit mit 3 Parametern $y = a_0 + a_1 \cdot x + a_2 \cdot x^2$

Parameter mit Fehlern:

1.2780 +/- 0.2883
0.8239 +/- 0.2032
0.0195 +/- 0.0242

Kovarianzmatrix:

0.0831 -0.0552 0.0060
-0.0552 0.0413 -0.0047
0.0060 -0.0047 0.0006

chi2 = 5.3751 chi2/nf = 0.7679

Abbildung 7.2: Beispiel für eine lineare Anpassung (mit einem MATLAB-Programm): Ein Polynom 1. Grades (durchgezogene Linie) oder 2. Grades (gestrichelt) wird an 10 Messwerte mit normalverteilten Fehlern angepasst. Die Messwerte sind ursprünglich entlang einer Geraden erzeugt worden. Man sieht an dem Fehler des Koeffizienten a_2 des quadratischen Terms, dass dieser Beitrag nicht signifikant ist.

ergibt sich aus (7.32) für die inverse Kovarianzmatrix der Parameter:

$$V_{ij}^{-1}(\theta) = \sum_{k=1}^n \sum_{l=1}^n H_{ki} V_{kl}^{-1} H_{lj} = \sum_{k=1}^n \sum_{l=1}^n f_i(x_k) \frac{1}{\sigma_k^2(y)} \delta_{kl} f_j(x_l) = \sum_{k=1}^n \frac{f_i(x_k) f_j(x_k)}{\sigma_k^2(y)} \quad (7.41)$$

Die Parameter sind unkorreliert, wenn die Diagonalelemente von $V^{-1}(\theta)$ Null sind:

$$V_{ij}^{-1}(\theta) = \sum_{k=1}^n \frac{f_i(x_k) f_j(x_k)}{\sigma_k^2(y)} = \frac{1}{\sigma_i^2(\theta)} \delta_{ij}. \quad (7.42)$$

Wenn die Fehler der Messwerte alle gleich sind, $\sigma_k^2(y) = \sigma^2(y)$, folgt aus (7.42) die Orthogonalität der Funktionen f_i in Bezug auf die Messwerte:

$$\sum_{k=1}^n f_i(x_k) f_j(x_k) = \frac{\sigma^2(y)}{\sigma_i^2(\theta)} \delta_{ij}. \quad (7.43)$$

Im Grenzfall einer unendlich großen Stichprobe geht die Summe in (7.43) in ein Integral über den Definitionsbereich Ω der f_i über:

$$\int_{\Omega} f_i(x) f_j(x) dx \sim \delta_{ij}. \quad (7.44)$$

Dieses Integral definiert ein Skalarprodukt in dem Raum der Funktionen f_i und (7.44) bedeutet, dass die f_i orthogonale Basisvektoren sind. Eine Anpassung mit orthogonalen Funktionen, erlaubt die sukzessive Hinzunahme weiterer Terme, ohne die bisher bestimmten Parameter wesentlich zu verändern. Das ist zum Beispiel wichtig für die Beurteilung der Signifikanz des Beitrags eines f_i -Terms. Orthogonale Funktionen sind zum Beispiel die sin- und cos-Funktionen einer Fourier-Zerlegung, die Legendre-Polynome, die Kugelflächenfunktionen usw.

Beispiel: Für eine Geradengleichung mit $f_1 = 1$; $f_2 = x$ ergibt sich:

$$\sum_k f_1(x_k) f_2(x_k) = \sum_k x_k = n \bar{x} \quad (7.45)$$

Mit der Transformation

$$f_2 \rightarrow f_2' = x - \bar{x} \quad (7.46)$$

ergibt sich:

$$\sum_k f_1(x_k) f_2'(x_k) = \sum_k (x_k - \bar{x}) = n \bar{x} - n \bar{x} = 0 \quad (7.47)$$

Daraus folgt, dass man den Ursprung der x -Koordinate am günstigsten in den Schwerpunkt \bar{x} zwischen den Messwerten legt (siehe Übungsaufgabe).

Extrapolationsfehler: Mit den Anpassungsfunktionen kann man nun y für beliebige x -Werte berechnen:

$$y = \sum_{j=1}^m \hat{\theta}_j f_j(x) \quad (7.48)$$

Der Fehler in y ergibt sich durch Fehlerfortpflanzung:

$$\sigma^2(y) = \sum_{i=1}^n \sum_{j=1}^m \frac{\partial y}{\partial \theta_i} \frac{\partial y}{\partial \theta_j} V_{ij}(\theta) = \sum_{i=1}^n \sum_{j=1}^m f_i(x) f_j(x) V_{ij}(\theta) \quad (7.49)$$

Güte der Anpassung: Die Güte der Anpassung wird wieder über das minimale $\chi^2 = S_{min}$, wie in im vorigen Abschnitt 7.2.1 besprochen, abgeschätzt.

7.3 Anpassung nicht-linearer Funktionen der Parameter

Wir betrachten jetzt die Anpassung einer beliebigen Funktion $f(x|\theta)$ an die n Messwerte y_i . Die LS-Funktion lautet wie in (7.2):

$$S = \sum_{i=1}^n \sum_{j=1}^n (y_i - \eta_i) V_{ij}^{-1}(y) (y_j - \eta_j) \quad (7.50)$$

Diese Funktion soll wieder als Funktion der Parameter minimalisiert werden. Im allgemeinen muss die Lösung $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m)$, die S minimiert, mit numerischen Methoden iterativ gesucht werden.

Iterationsverfahren: Es sei im ν -ten Iterationsschritt eine Näherung von $\hat{\theta}$ gefunden:

$$\theta^\nu = (\theta_1^\nu, \dots, \theta_m^\nu). \quad (7.51)$$

Gesucht ist ein Inkrement $\Delta\theta^\nu$, das zu der nächsten Näherung für die Parameter führt,

$$\theta^{\nu+1} = \theta^\nu + \Delta\theta^\nu, \quad (7.52)$$

und das die Näherung verbessert:

$$S(\theta^{\nu+1}) < S(\theta^\nu) \quad (7.53)$$

Das Verfahren wird abgebrochen, wenn Konvergenz erreicht ist. Als **Konvergenzkriterium** verlangt man in der Regel, dass S sich von einem Schritt zum nächsten um weniger als einen kleinen Betrag ϵ ändert:

$$|S(\theta^{\nu+1}) - S(\theta^\nu)| < \epsilon \quad (7.54)$$

Es gibt verschiedenen Verfahren, die Inkremente $\Delta\theta^\nu$ zu bestimmen, um das Minimum von S zu finden. Bei vielen Parametern und etwas komplexer strukturierten LS-Funktionen können solche multi-dimensionalen Optimierungsprobleme zu einer mathematischen Herausforderung werden. In der Teilchenphysik wird sehr viel das beim CERN entwickelte Programm MINUIT benutzt, das verschiedene Verfahren zur Auswahl anbietet (Abb. 7.4). Bei komplexen Problemen ist es notwendig, dass der Benutzer die verschiedenen Möglichkeiten kennt und steuernd eingreift. Wichtig sind gute Startwerte θ^0 , die man häufig nur durch ein gutes Gespür erhält, um eventuelle Nebenminima im Parameterraum zu vermeiden (Abb. 7.3). Man muss deshalb immer überprüfen, ob die Lösung von den Startwerten abhängt.

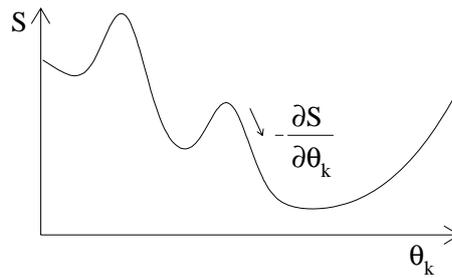


Abbildung 7.3: Beispiel für den Verlauf einer LS-Funktion im Parameterraum.

Gradientenverfahren: Eine naheliegende Möglichkeit, Extremwerte einer Funktion zu finden, ist das Gradientenverfahren: Man geht mit einer vorgewählten Schrittweite $\Delta\theta$ in Richtung des Gradienten der Funktion, im Fall der Minimierung in Richtung des negativen Gradienten (Abb. 7.3). Häufig wird die Schrittweite proportional dem Gradienten gewählt:

$$\Delta\theta^{\nu+1} = -\eta \left(\vec{\nabla}_{\theta} S \right)_{|\theta^{\nu}} \quad (7.55)$$

Die Wahl der Schrittweite proportional zum Gradienten von S scheint vernünftig zu sein, weil im Minimum von S Konvergenz erreicht wird und die Schrittweite dann tatsächlich gegen Null geht. Häufig wird der Schrittparameter η aber auch dynamisch angepasst, um zum Beispiel nicht zu lange in Gebieten mit flachem Funktionsverlauf zu verweilen (große Schritte) oder in Bereichen steiler Gradienten auch das Minimum finden zu können (kleine Schritte). Wenn sich in einem Iterationsschritt das Vorzeichen des Gradienten ändert, das Extremum also überschritten wurde, sollte man die Schrittweite verkleinern.

Linearisierung der Anpassungsfunktion: Durch Entwicklung der Anpassungsfunktion nach den Parametern bis zu den linearen Termen, kann man das Problem auf lineare Anpassungen mit Iterationen zurückführen:

$$\eta_i(\theta) = \eta_i(\theta^{\nu}) + \left(\vec{\nabla}_{\theta} \eta_i \right)_{|\theta=\theta^{\nu}} \Delta\theta^{\nu} + \dots \quad (7.56)$$

In der ν -ten Iteration sind die Abweichungen der Messwerte von dem Anpassungswert ('Residuen'):

$$\Delta y_i^{\nu} = y_i - \eta_i(\theta^{\nu}) \quad (7.57)$$

Mit der Definition der Matrix H

$$H_{ij} = \frac{\partial \eta_i}{\partial \theta_j} \quad (7.58)$$

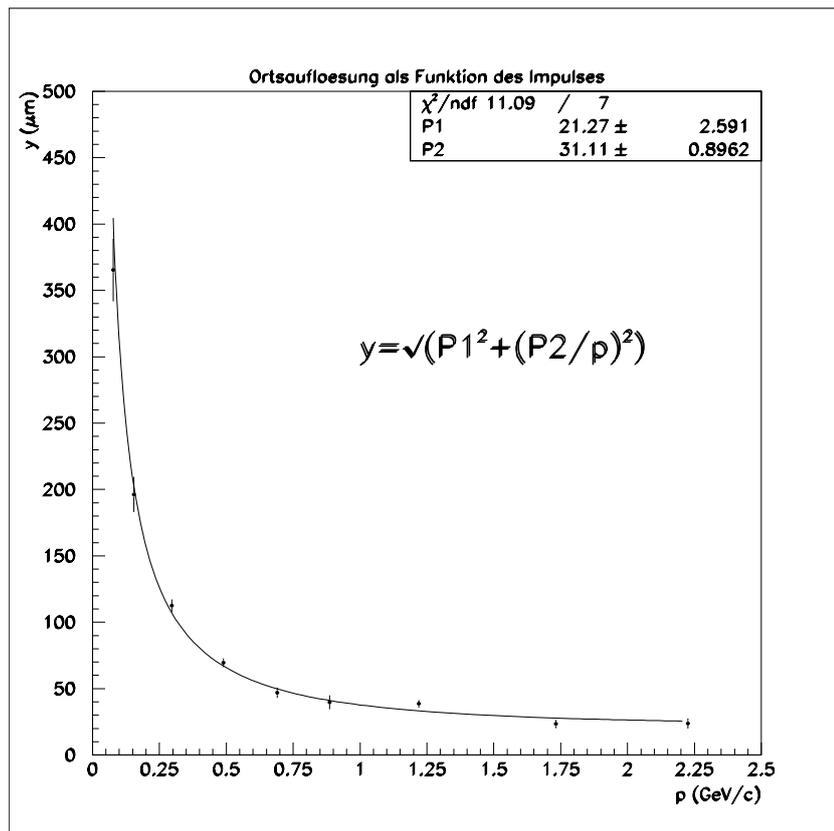
ergibt sich dann die LS-Funktion in der ν -ten Iteration:

$$S^{\nu} = (\Delta \vec{y}^{\nu} - H \Delta \theta^{\nu})^T V^{-1} (\Delta \vec{y}^{\nu} - H \Delta \theta^{\nu}) \quad (7.59)$$

Diese LS-Funktion entspricht völlig derjenigen für die lineare Anpassung (7.26), wenn man die Ersetzung

$$\vec{y} \rightarrow \Delta \vec{y}^{\nu}; \quad \theta \rightarrow \Delta \theta^{\nu} \quad (7.60)$$

macht.



```

*****
*                                                                    *
* Function minimization by SUBROUTINE HFITV *
* Variable-metric method *
* ID =          0  CHOPT = S *
*                                                                    *
*****
Convergence when estimated distance to minimum (EDM) .LT.  0.10E+01

FCN=  11.08658      FROM MIGRAD      STATUS=CONVERGED      37 CALLS      38 TOTAL
EDM=  0.32E-05      STRATEGY= 1      ERROR MATRIX ACCURATE

EXT PARAMETER
NO.  NAME      VALUE      ERROR      STEP      FIRST
    1   P1      21.272     2.5912     0.28605   0.10326E-02
    2   P2      31.111     0.89618   0.99125E-01 0.37551E-03

CHISQUARE = 0.1584E+01  NPFIT =      9
    
```

Abbildung 7.4: Beispiel für die Anwendung des Programmes MINUIT. Unter der graphischen Darstellung ist der Ausdruck des Programmes MINUIT gezeigt. Eine nicht-lineare Funktion der Parameter, angegeben in der Graphik, wird an Messwerte angepasst.