

# Kapitel 1

## Ausgleich von Messwerten

Bei der Auswertung physikalischer Experimente steht man häufig vor dem Problem, eine Anzahl verschiedener Messungen  $y_i$  (meist mit verschiedener Messgenauigkeit  $\sigma_i$ ) entweder untereinander oder mit einer theoretischen Vorhersage zu vergleichen. Ziel ist es, die Daten untereinander auf Konsistenz zu überprüfen bzw. eine Modellfunktion für  $y$  anzupassen, welche von den anzupassenden Parametern  $a_j$  abhängt.

$$y(x) = y(x; a_1 \dots a_M)$$

Eine oft benutzte Methode, die Parameter  $a_j$  zu bestimmen, ist die Methode der kleinsten Quadrate, bei denen folgende Summe bezüglich der Parameter zu minimieren ist:

$$\sum_{i=1}^N [y_i - y(x_i; a_1 \dots a_M)]^2 \rightarrow \text{Min.}$$

Dieses zunächst willkürlich erscheinende Kochrezept lässt sich begründen:

Die Wahrscheinlichkeit  $P_i$ , dass ein Messpunkt  $y_i$  mit einem Fehler  $\sigma_i$  durch einen theoretischen Wert  $y$  korrekt beschrieben wird, ergibt sich aus der Verteilungsfunktion von  $y_i$ . Der zentrale Grenzwertsatz der Statistik lehrt uns, dass die Summe beliebig verteilter Zufallsgrößen gegen eine Normalverteilung konvergiert. Die Gesamtwahrscheinlichkeit  $P$ , dass die Kurve mit den Parametern  $a_j$  alle Punkte  $y_i$  möglichst gut beschreibt, ergibt sich aus dem Produkt der Einzelwahrscheinlichkeiten  $P_i$  und muss bezüglich der  $a_j$  maximal werden:

$$P = \prod_{i=1}^N P_i \sim \prod_{i=1}^N \exp \left[ -\frac{1}{2} \left( \frac{y_i - y(x_i)}{\sigma_i} \right)^2 \right] \rightarrow \text{Max.}$$

Diese Form der Parameterbestimmung bezeichnet man im Englischen als *Maximum Likelihood Estimation*. Falls wir bei der Streuung der Messwerte  $y_i$  von Gaußverteilungen ausgehen können, so ergibt sich der wahrscheinlichste Satz von Parametern aus der äquivalenten Bedingung:

$$-\ln P \sim \chi^2 = \sum_{i=1}^N \left( \frac{y_i - y(x_i)}{\sigma_i} \right)^2 \rightarrow \text{Min.}$$

Dies ist das Prinzip der kleinsten Quadrate, das bereits von Gauß formuliert wurde. Die Verteilung  $F(\chi^2)^1$  gehorcht folgender Funktion:

$$F(\chi^2) = \frac{1}{\Gamma(\lambda)2^\lambda} \int_0^{\chi^2} u^{\lambda-1} e^{-\frac{1}{2}u} du \quad .$$

Dabei ist

$$\lambda = \frac{N - M}{2}$$

und man kann zeigen, dass der Erwartungswert von  $\chi^2$  gleich der Zahl der Freiheitsgrade  $N - M$  und die Varianz  $\sigma^2(\chi^2)$  gleich  $2(N - M)$  ist. Der Wert von  $\frac{\chi^2}{N - M}$  ist also ein Maß für die Güte der Anpassung: Idealerweise liegt dieser Wert bei 1. Ist er deutlich größer als 1, so sind die Daten nicht kompatibel zur Modellfunktion, meist bedingt durch kleine Varianzen  $\sigma_i^2$  oder unberücksichtigte systematische Fehler. Ist er kleiner als 1, so sind die Messfehler meist zu konservativ angegeben (oder die Daten wurden manipuliert, damit sie besser zum Modell „passen“).

## 1.1 Mittelwert

Als einfachstes Beispiel soll die Ermittlung des Mittelwerts verschiedener Messungen mit verschiedenen Messfehlern das Verfahren illustrieren. Im Jahre 1948 beruhte die Kenntnis der Lichtgeschwindigkeit im Wesentlichen auf folgenden 3 unabhängigen Messungen:

Bergstrand (1948)	$299\,793 \pm 2.0 \text{ km/s}$
Essen et al. (1947)	$299\,792 \pm 4.5 \text{ km/s}$
Jones (1947)	$299\,782 \pm 25 \text{ km/s}$

Mit der Methode der kleinsten Quadrate lässt sich der fehlergewichtete Mittelwert  $\bar{c}$  aus folgender Bedingung finden:

$$\chi^2 = \sum_{i=1}^3 \left( \frac{c_i - \bar{c}}{\Delta c_i} \right)^2 \rightarrow \text{Min.}$$

Bezüglich  $\bar{c}$  muss  $\chi^2$  minimal werden, d.h.

$$\begin{aligned} \frac{d\chi^2}{d\bar{c}} &= \sum_{i=1}^3 2 \frac{c_i - \bar{c}}{\Delta c_i^2} = 0 \\ \longrightarrow \bar{c} &= \frac{\sum_{i=1}^3 \frac{c_i}{\Delta c_i^2}}{\sum_{i=1}^3 \frac{1}{\Delta c_i^2}} \end{aligned}$$

---

<sup>1</sup>Das Symbol  $\chi^2$  wird wie eine gewöhnliche Variable behandelt. Es wurde von K. Pearson eingeführt.

Den Fehler des gewichteten Mittelwerts kann man entweder als internen Fehler, d.h. allein aus der Kombination der Einzelfehler, oder als externen Fehler, der auch noch die Streuung der einzelnen Fehler berücksichtigt, angeben. Der interne Fehler folgt aus der Gaußschen Fehlerfortpflanzung:

$$\sigma_m^2 = \sum_{i=1}^N \frac{\partial \bar{c}}{\partial c_i} \sigma_i^2 = \sum_{i=1}^N \frac{(1/\sigma_i^2)^2}{\sum_{k=1}^N (1/\sigma_k^2)^2} \cdot \sigma_i^2$$

Dies lässt sich vereinfachen zu:

$$\sigma_{m,\text{int}}^2 = \frac{1}{\sum_{i=1}^N 1/\sigma_i^2}$$

Der externe Fehler berücksichtigt auch die Streuung der Messwerte:

$$\sigma_{m,\text{ext}}^2 = \frac{\sum_{i=1}^N (c_i - \bar{c})^2 / \sigma_i^2}{(N-1) \sum_{i=1}^N 1/\sigma_i^2} = \frac{\chi^2}{N-M} \cdot \sigma_{m,\text{int}}^2, \quad ,$$

wobei im vorliegenden Fall mit  $M = 1$  Parametern und  $N = 3$  Messwerten die Zahl der Freiheitsgrade  $N - M = 2$  ist.

## 1.2 Lineare Regression

Eine häufige Anwendung der Methode der kleinsten Quadrate ist die Anpassung einer Geraden an einen Satz von  $(x_i, y_i)$ -Wertepaaren. Die Modellfunktion ist linear:

$$y(x) = y(x; a, b) = a + bx$$

Die Messunsicherheiten  $\sigma_i$  mögen bekannt sein und die Abszissenwerte  $x_i$  seien exakt gegeben. Die Fehlersumme  $\chi^2$  ist dann

$$\chi^2(a, b) = \sum_{i=1}^N \left( \frac{y_i - a - bx_i}{\sigma_i} \right)^2 \quad (1.1)$$

Das Minimum dieser Summe ergibt die optimalen Parameter  $a$  und  $b$  unter der Annahme, dass die Messfehler  $\sigma_i$  normalverteilt sind. Einzelne stark abweichende Punkte (Ausreißer) können das Ergebnis u.U. stark verzerren, so dass dann robustere Verfahren verwendet werden müssen (z.B. Median statt Mittelwert). Die Summe in 1.1 wird bezüglich  $a$  und  $b$  minimiert. Dazu müssen die partiellen Ableitungen jeweils verschwinden:

$$\begin{aligned} 0 &= \frac{\partial \chi^2}{\partial a} = -2 \sum_{i=1}^N \frac{y_i - a - bx_i}{\sigma_i^2} \\ 0 &= \frac{\partial \chi^2}{\partial b} = -2 \sum_{i=1}^N \frac{x_i (y_i - a - bx_i)}{\sigma_i^2} \end{aligned}$$

Der besseren Lesbarkeit wegen sollen folgende Abkürzungen gelten:

$$S \equiv \sum_{i=1}^N \frac{1}{\sigma_i^2}, \quad S_x \equiv \sum_{i=1}^N \frac{x_i}{\sigma_i^2}, \quad S_y \equiv \sum_{i=1}^N \frac{y_i}{\sigma_i^2}$$

$$S_{xx} \equiv \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2}, \quad S_{xy} \equiv \sum_{i=1}^N \frac{x_i y_i}{\sigma_i^2}$$

Damit erhalten wir das Gleichungssystem:

$$\begin{aligned} aS + bS_x &= S_y \\ aS_x + bS_{xx} &= S_{xy} \end{aligned}$$

Als Lösungen ergeben sich dann (mit der Determinante  $D = SS_{xx} - S_x^2$ ):

$$\begin{aligned} a &= \frac{S_{xx}S_y - S_xS_{xy}}{D} \\ b &= \frac{SS_{xy} - S_xS_y}{D} \end{aligned}$$

Nun sollen noch die Fehler der Koeffizienten bestimmt werden. Da wir normalverteilte, unkorrelierte Messwerte  $y_i$  angenommen haben, gilt das Gaußsche Fehlerfortpflanzungsgesetz:

$$\sigma_f^2 = \sum_{i=1}^N \sigma_i^2 \left( \frac{\partial f}{\partial y_i} \right)^2$$

Im vorliegenden Fall sind die partiellen Ableitungen sehr einfach:

$$\begin{aligned} \frac{\partial a}{\partial y_i} &= \frac{S_{xx} - S_x x_i}{\sigma_i^2 D} \\ \frac{\partial b}{\partial y_i} &= \frac{S x_i - S_x}{\sigma_i^2 D} \end{aligned}$$

Die entsprechenden Summen sind dann:

$$\begin{aligned} \sigma_a^2 &= \frac{S_{xx}}{D} \\ \sigma_b^2 &= \frac{S}{D} \end{aligned}$$

Abschließend sollte noch das normierte  $\chi^2/(N - 2)$  bestimmt werden, um die Güte der Anpassung beurteilen zu können.

### 1.3 Lineare Koeffizienten

Lässt sich die Modellfunktion  $y$  in folgender Weise darstellen

$$y(x; a_1, \dots, a_M) = \sum_{j=1}^M a_j f_j(x) \quad ,$$

so lassen sich die linearen Koeffizienten  $a_j$  aus dem linearen Gleichungssystem  $M$ -ter Ordnung bestimmen, das sich aus den  $M$  Bedingungen

$$\frac{\partial}{\partial a_j} \sum_{i=1}^N \frac{\left( y_i - \sum_{j=1}^M a_j f_j(x_i) \right)^2}{\sigma_i^2} = 0$$

ergibt. Wichtige Beispiele sind die Anpassung eines Polynoms  $f_j(x) = x^{j-1}$  oder der Fit spezieller Funktionen, wie z.B. Legendre-Polynome  $P_{j-1}(x) \equiv f_j(x)$  mit  $P_0(x) = 1$ ,  $P_1(x) = x$  und der Rekursionsvorschrift  $(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x)$ .

## 1.4 Linearisierbare Funktionen

In speziellen Fällen lässt sich die Modellfunktion linearisieren, z.B. lässt sich die Anpassung eines exponentiellen Funktionsverlaufs folgendermaßen durchführen:

$$y(x; a, b) = e^{a+bx_i}$$

Die Substitution  $y' = \ln y$  führt dann auf die gesuchte lineare Modellfunktion:

$$y' = \ln y = a + bx_i$$

Zu beachten ist aber, dass auch die Fehler  $\sigma_i$  entsprechend transformiert werden müssen:

$$\sigma_i'^2 = \sigma_i^2 \frac{dy'_i}{dy_i} = \sigma_i^2 \frac{d}{dy_i} \ln y_i = \frac{\sigma_i^2}{y_i}$$

Daher ist folgende Funktion bezüglich  $a'$  (und damit auch  $a$ ) und  $b$  zu minimieren:

$$\chi^2(a, b) = \sum_{i=1}^N \frac{y_i (\ln y_i - a - bx_i)^2}{\sigma_i^2} \rightarrow \text{Min.}$$

## 1.5 Nichtlineare Verfahren

In der Praxis kommen aber auch häufig nichtlineare Probleme vor, bei denen die partiellen Ableitungen  $\frac{\partial \chi^2}{\partial a_j}$  zu einem System nichtlinearer Gleichungen führen, das sich in der Regel nicht mehr analytisch lösen lässt. Hier helfen nur noch sukzessive Suchverfahren, um das Minimum von  $\chi^2$  zu finden. Minimierung oder Maximierung von Funktionen ist ein häufiges Problem, zu dessen Lösung zwei wichtige Verfahren vorgestellt werden sollen, das Gradientenverfahren sowie die Simplexmethode. Der zweite Teil dieser Vorlesung wird sich mit genetischen Algorithmen beschäftigen.

Im Praktikum Computational Physics sollen zunächst die Routinen der „Numerical Recipes“ [3] verwendet werden, deren Methoden sehr gut im entsprechenden Buch beschrieben sind. Da die Lizenzbedingungen sehr restriktiv sind, ist es empfehlenswert, freie Software wie das

Paket Minuit[1] unter Root [2] zu verwenden, das die im Folgenden beschriebenen Verfahren bereits implementiert. Viele weitere bekannte Programme implementieren die unten genannten Verfahren, so z.B. Gnuplot den Levenberg-Marquardt-Algorithmus oder Matlab. Ziel des Praktikums ist also nicht vordergründig die Programmierung der Verfahren selbst (auch wenn der Quellcode verstanden werden sollte!), sondern ihre Anwendung auf konkrete Probleme. Aus diesem Grund sollen die gängigsten Verfahren nur kurz erläutert werden, für weitergehende Informationen wird auf die Literatur verwiesen.

### 1.5.1 Vorbetrachtung

Zu minimieren ist die Funktion  $\chi^2(\mathbf{a})$ , die vom  $M$ -dimensionalen Vektor  $\mathbf{a}$  der anzupassenden Parameter abhängt. Der Einfachheit halber betrachten wir zunächst den eindimensionalen Fall  $\chi^2(a)$ . Man geht sinnvollerweise von einer quadratischen Entwicklung um einen Startwert  $a_o$  aus und benutzt das Minimum der resultierenden Parabel als verbesserten Parameter. Offensichtlich hängt der Erfolg von der Wahl des Startwerts und der Stützweite der Punkte ab; im schlimmsten Fall ist die Parabel konkav und divergiert daher.

Die Situation verbessert sich jedoch deutlich, wenn die erste und zweite Ableitung von  $\chi^2$  nach  $a$  an der Stelle  $a_0$  bekannt sind, d.h.  $\frac{d\chi^2}{da} = b$  und  $\frac{d^2\chi^2}{da^2} = A$ , so dass die quadratische Entwicklung um  $a_0$  den verbesserten Parameter

$$a' = a_o + \frac{b}{A}$$

liefert.

Die Verallgemeinerung der quadratischen Form auf  $M$  Parameter lautet dann analog

$$\begin{aligned}\chi^2(\mathbf{a}) &= c + \mathbf{b} \cdot \mathbf{a} + \frac{1}{2} \mathbf{a}^T \mathbf{A} \mathbf{a} \\ &= c + \sum_k b_k a_k + \frac{1}{2} \sum_{k,l} a_k A_{kl} a_l\end{aligned}$$

Dabei ist

$$\mathbf{b} = \text{grad } \chi^2(\mathbf{a}_0), \quad \text{d.h. } b_i = \left. \frac{\partial \chi^2}{\partial a_i} \right|_{\mathbf{a}=\mathbf{a}_0} \quad \text{und} \quad A_{ik} = \left. \frac{\partial^2 \chi^2}{\partial a_i \partial a_k} \right|_{\mathbf{a}=\mathbf{a}_0}$$

die Hessesche Matrix der zweiten Ableitungen.

### 1.5.2 Simplexverfahren

Ein sehr robustes, aber langsames Verfahren ist das Simplex-Verfahren (Nelder u. Mead). Hierbei müssen nur Funktionswerte  $\chi^2(\mathbf{a})$  bestimmt werden, Gradient und Hessesche Matrix sind nicht notwendig. Ein *Simplex* ist durch  $M+1$  Punkte  $\mathbf{a}_i$  im  $M$ -dimensionalen Raum der anzupassenden Parameter definiert. Ein Simplex in 2 Dimensionen ist also ein Dreieck. Der Simplex wird nun schrittweise verändert. In jedem Teilschritt findet eine von 4 Operationen statt:

**Spiegelung:** Der Punkt mit dem größten  $\chi^2$  wird über die Verbindungslinie der anderen beiden Punkte gespiegelt; dort wird  $\chi^2$  ebenso ermittelt.

**Streckung:** Falls die Spiegelung zu kleineren  $\chi^2$  führt, wird eine Streckung in diese Richtung versucht

**Abflachung:** Der Punkt mit dem größten  $\chi^2$  wird durch einen Punkt auf der gleichen Linie wie oben ersetzt, nur liegt dieser Punkt innerhalb des Simplex. Hat dieser Punkt ein größeres  $\chi^2$ , so wird die Abflachung verworfen und statt dessen eine Kontraktion vorgenommen.

**Kontraktion:** Nur der Punkt mit dem niedrigsten  $\chi^2$  bleibt erhalten, alle anderen werden entlang der Kante des alten Simplex zu diesem Punkt hin verschoben.

### 1.5.3 Gradientenverfahren

Um von einem beliebigen Startwert zum Minimum der Funktion  $\chi^2$  zu gelangen, liegt es nahe, immer genau dem negativen Gradienten  $-\mathbf{b} = -\text{grad } \chi^2$  zu folgen. Verblüffenderweise stehen aber aufeinanderfolgende Richtungen senkrecht aufeinander (warum ist das so?), so dass das Verfahren im Zick-Zack-Kurs verläuft und daher nur langsam konvergiert.

Abhilfe schafft das Verfahren konjugierter Gradientenrichtungen, bei der die neue Richtung die Änderungstärke des Gradienten im vorigen und aktuellen Schritt berücksichtigt.

Die Minimierung nach Levenberg-Marquardt kombiniert das Verfahren mit quadratischer Form (d.h. unter Verwendung der Hesse-Matrix) mit der Robustheit des Gradientenverfahrens.

Abschließend sei hervorgehoben, dass alle diese Verfahren nur dann konvergieren, wenn man sich bereits im Einzugsbereich des absoluten Minimums befindet. Sehr oft besteht die Gefahr, in Nebenminima stecken zu bleiben, wo nur beim Simplexverfahren eine gewisse Chance besteht, wieder heraus zu finden.

Verfahren, die diese Schwierigkeiten auf Kosten langsamerer Konvergenz vermeiden, sind z.B. das simulated annealing (simulierte Abkühlung) bzw. der sog. Sintflut-Algorithmus [5] sowie genetische Verfahren, die im zweiten Teil der Vorlesung behandelt werden.





# Literaturverzeichnis

- [1] F. James, MINUIT Minimization package, CERN Program Library Long Writeup D506, FORTRAN, C++
- [2] ROOT: An Object Oriented Data Analysis Framework,
- [3] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery: Numerical Recipes in C (or FORTRAN), Cambridge University Press 2002,
- [4] S. Brandt: Datenanalyse, BI-Wissenschaftsverlag 1992
- [5] G. Dueck, T. Scheuer und H.-M. Wallmeier: Toleranzschwelle und Sintflut: neue Ideen zur Optimierung, Spektrum der Wissenschaft, Digest 2/1999 „Wissenschaftliches Rechnen“