

7.3 Theoretische Grundlagen

Im Folgenden wird davon ausgegangen, dass die Datenpunkte Normalverteilt sind.

Def.: Eine Messreihe sei durch die Menge $\{x_i\}$, $i=0,1,\dots,n-1$ gegeben, dann ist

$$\bar{x} = \frac{1}{n} \sum_{j=0}^{n-1} x_j \quad \text{der Mittelwert}$$

$$s^2 = \frac{1}{n-1} \sum_{j=0}^{n-1} (x_j - \bar{x})^2 \quad \text{die Varianz}$$

Ist ein zweiter Datensatz $\{y_i\}$ gegeben, so kann die Korrelation (Abhängigkeit) zwischen diesen beiden Datensätzen bestimmt werden:

$$c_{xy} = \frac{1}{n-1} \sum_{i=1}^{n-1} (x_i - \bar{x})(y_i - \bar{y})$$

Daraus kann die Korrelationsmatrix/Kovarianzmatrix C_{ov} bestimmt werden:

$$C_{ov} = \begin{pmatrix} c_{xx} & c_{xy} \\ c_{yx} & c_{yy} \end{pmatrix} \quad \text{mit } c_{xx} = s_x^2 \quad \text{und } c_{yy} = s_y^2$$

Diese Matrix ist positiv definit und symmetrisch. Die Matrixelemente werden häufig noch normiert, so dass die Diagonale 1 beträgt.

$$\Gamma_{xx} = \frac{c_{xx}}{s_x^2}, \quad \Gamma_{yy} = \frac{c_{yy}}{s_y^2}, \quad \Gamma_{xy} = \Gamma_{yx} = \frac{c_{xy}}{s_x s_y}$$

Man definiert $R = \begin{pmatrix} r_{xx} & r_{xy} \\ r_{yx} & r_{yy} \end{pmatrix}$. Auf den Nebendiagonalen kann R die Werte von -1 bis 1 annehmen, wobei

- 0 keine Korrelation
- 1 vollständige Korrelation (erhöht man x , erhöht sich auch y)
- (-1) vollständige Anti-Korrelation (erhöht man x , verringert sich y)

bedeutet. Programme wie z.B. Gnuplot geben am Ende die Matrix R für die Fitparameter aus.

7.3.1 Methode der kleinsten Quadrate

Angenommen, wir wollen eine Funktion

$$y(x) = y(x, a_0, a_1, \dots, a_{M-1})$$

mit M anpassbaren Parametern a_j , $j = 0, 1, \dots, M-1$ an N Datenpunkte (x_i, y_i) , $i = 0, \dots, N-1$ fiten, dann bietet sich hierfür die Methode der kleinsten Quadrate an.

$$\sum_{i=0}^{N-1} [y_i - y(x_i, a_0, a_1, \dots, a_{M-1})]^2$$

Wieso?

Die Datenpunkte weichen im Allgemeinen von den wahren Werten um einen bestimmten Betrag Δy ab.

Außerdem haben die Datenpunkte einen Fehler σ , der hier erstmal für alle Datenpunkte als gleich angenommen wird. Dann ist

$$P(\text{Daten} | \text{Modell}) = \prod_{i=0}^{N-1} \left[\exp \left\{ -\frac{1}{2} \left(\frac{y_i - y(x_i)}{\sigma} \right)^2 \right\} \Delta y \right] \quad (*)$$

die Wahrscheinlichkeit, dass das „wahre“ Modell $y(x)$ [die Fitfunktion] die Daten beschreibt.

Angenommen, mehrere Modelle sind vorgegeben, dann ist das wahrscheinlichste Modell jenes, welches (*) maximiert oder den negativen Logarithmus minimiert

$$\left(\sum_{i=0}^{N-1} \left(\frac{y_i - y(x_i)}{\sigma} \right)^2 \right) - N \log(\Delta y)$$

wobei N, σ und Δy Konstanten sind.

Hieraus folgt dann direkt, dass

$$\sum_{i=0}^{N-1} [y_i - y(x_i)]^2$$

minimiert werden muss. Erhält nun jeder Datenpunkt seinen individuellen Fehler, so erweitert sich die Gleichung zu

$$\sum_{i=0}^{N-1} \left[\frac{y_i - y(x_i)}{\sigma} \right]^2 =: \chi^2$$

Diese Größe nennt man Chi-Quadrat, da sie χ^2 -verteilt ist. Definiert man den Vektor:

$$\vec{y} = \begin{pmatrix} y_0 - y(x_0) \\ y_1 - y(x_1) \\ \vdots \\ y_{N-1} - y(x_{N-1}) \end{pmatrix}$$

So kann χ^2 allgemein geschrieben werden

$$\chi^2 = \vec{y}^T C_{ov}^{-1} \vec{y}$$

wobei C_{ov} die Kovarianzmatrix darstellt. Da χ^2 χ^2 -verteilt ist, ist der wahrscheinlichste Wert die Anzahl der Freiheitsgrade

degrees of freedom = # Messpunkte - # Fitparameter

also $\chi^2 / \text{d.o.f.} = 1$. (für große d.o.f.)

Interpretation für Abweichungen:

- χ^2 viel größer als 1:

- i) Fitfunktion schlecht
- ii) Fehler unterschätzt

- χ^2 viel kleiner als 1:

- i) Fehler überschätzt
- ii) Korrelation zwischen Datenpunkten vernachlässigt [$C_{ov} = \text{diag}(\dots)$]

Nur zulässig,
wenn alles
normalverteilt
ist!

7.3.2 Fit an eine Gerade: $a + bx = y(x)$

$$\chi^2(a, b) = \sum_{i=0}^{N-1} \left[\frac{y_i - a - bx_i}{\sigma_i} \right]^2$$

$$\frac{\partial \chi^2}{\partial a} = 0 = -2 \sum_i \frac{y_i - a - bx_i}{\sigma_i^2}$$

$$\frac{\partial \chi^2}{\partial b} = 0 = -2 \sum_i \frac{x_i (y_i - a - bx_i)}{\sigma_i^2}$$

Kurzschreibweise:

$$S := \sum_i \frac{1}{\sigma_i^2} ; S_x := \sum_i \frac{x_i}{\sigma_i^2} ; S_y := \sum_i \frac{y_i}{\sigma_i^2}$$

$$S_{xx} := \sum_i \frac{x_i^2}{\sigma_i^2} ; S_{xy} := \sum_i \frac{x_i y_i}{\sigma_i^2}$$

Für Minimierung folgt:

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

$$\Rightarrow \Delta = SS_{xx} - (S_x)^2$$

$$a = \frac{S_{xx}S_y - S_xS_{xy}}{\Delta} ; b = \frac{SS_{xy} - S_xS_y}{\Delta}$$

Fehlerfortpflanzung:

$$\sigma_f^2 = \sum_{i=0}^{N-1} \sigma_i^2 \left(\frac{\partial f}{\partial y_i} \right)^2$$

$$\text{mit } \frac{\partial a}{\partial y_i} = \frac{S_{xx} - S_x x_i}{\sigma_i^2 \Delta} , \quad \frac{\partial b}{\partial y_i} = \frac{S_x - S x_i}{\sigma_i \Delta}$$

$$\Rightarrow \sigma_a^2 = \frac{S_{xx}}{\Delta} , \quad \sigma_b^2 = \frac{S}{\Delta}$$

$$Cov(a, b) = \frac{S_x}{\Delta} \Rightarrow \Gamma_{ab} = \frac{-S_x}{\sqrt{SS_{xx}}}$$

7.2.2 Minimierungsverfahren

Das Gauß-Newton-Verfahren muss nicht gegen die gesuchte Lösung konvergieren (z.B. schlechte Startwerte oder „kinetische“ Ausgleichsprobleme)

Um die Konvergenz sicherzustellen, fordert man

$$F(\vec{x}^{(k)}) < F(\vec{x}^{(k-1)}) \quad (k=1, 2, \dots) \quad (*)$$

Dazu muss eine Abstiegsrichtung $\vec{v}^{(k)}$ gefunden werden, mit:

$$\vec{x}^{(k)} = \vec{x}^{(k-1)} + t \vec{v}^{(k)}, \quad t > 0$$

Eine Abstiegsrichtung stellt der negative Gradient von $F(\vec{x})$ im Punkt $\vec{x}^{(k-1)}$ dar, welche folgende Form hat:

$$\vec{v}^{(k)} = -[C^{(k-1)}]^+ \vec{r}^{(k-1)}$$

Bei der Methode des kleinsten Abstiegs wird t aus

$$F(\vec{x}^{(k)}) = \min_t [F(\vec{x}^{(k-1)} + t \vec{v}^{(k)})]$$

bestimmt. Aus Gründen des Rechenaufwands wird t häufig nur näherungsweise bestimmt, z.B. wie folgt:

1) Löse für die Startwerte $\vec{x}^{(0)}$ die linearisierte Fehlergleichung

$$C^{(k)} \vec{s}^{(k+1)} - \vec{d}^{(k)} = \vec{s}^{(k)}$$

Der Korrekturvektor $\vec{f}^{(k+1)}$ dient jeweils als Abstiegsrichtung.

2) Prüfe (*), ob für $t = 1, \frac{1}{2}, \frac{1}{4} \dots$ mit $\vec{y} := \vec{x}^{(k)} + t \vec{f}^{(k+1)}$, $F(\vec{y}) < F(\vec{x}^{(k)})$ gilt. Ist dies der Fall, so ist $\vec{x}^{(k+1)} = \vec{y}$ und man kann auf Konvergenz testen. Die Methode kann sehr langsam konvergieren.

Marquardt entwickelte eine effizientere Methode, um die Abstiegsrichtung zu bestimmen. Sei $C = C^{(k)}$ und $\vec{d} = \vec{d}^{(k)}$, dann erhält man \vec{v} durch Lösen des Extremalproblems:

$$|C\vec{v} - \vec{d}|^2 + \lambda^2 |\vec{v}|^2 = \text{Min!}, \quad \lambda > 0$$

Bei gegebenem λ führt dies wieder auf eine Fehlergleichung nach der Methode der kleinsten Quadrate:

$$\tilde{C} \vec{v} - \tilde{\vec{d}} = \tilde{\vec{f}} \quad \text{mit} \quad \tilde{C} = \begin{pmatrix} C \\ \lambda I \end{pmatrix} \in \mathbb{R}^{(N+n), n} \quad (*)$$

$$\tilde{\vec{d}} = \begin{pmatrix} \vec{d} \\ 0 \end{pmatrix} \in \mathbb{R}^{N+n}; \quad \tilde{\vec{f}} \in \mathbb{R}^{N+n}$$

In der Marquardt-Methode wird nun die euklidische Norm von $\vec{v}^{(k+1)}$ durch λ so gesteuert, dass mit

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} + \vec{v}^{(k+1)}$$

$$F(\vec{x}^{(k+1)}) < F(\vec{x}^{(k)}), \quad k = 0, 1, \dots$$

gilt.

Ist dies erfüllt, so kann λ im nächsten Schritt verkleinert (z.B. halbiert) werden. Ist dies nicht erfüllt, so muß λ vergrößert (z.B. verdoppelt) werden, bis das Kriterium wieder erfüllt ist.

Ein problemabhängiger Startwert für λ ist z.B.

$$\lambda^{(0)} = \|C^{(0)}\|_F / \sqrt{nN} = \sqrt{\frac{1}{nN} \sum_{i,j} [C_{ij}^{(0)}]^2}$$

Die Berechnung von $\vec{v}^{(k+1)}$ aus (*) erfolgt in zwei Schritten:

1) Die ersten N Gleichungen, die von λ unabhängig sind, werden mit der orthogonalen Matrix Q_1 transformiert.

$$Q_1^T \tilde{C} = \begin{pmatrix} R_1 \\ O_1 \\ \lambda I \end{pmatrix}, \quad Q_1^T \tilde{d} = \begin{pmatrix} \hat{d}_1 \\ \hat{d}_2 \\ \vec{0} \end{pmatrix} \quad R_1 \in \mathbb{R}^{n,n}, \quad O_1 \in \mathbb{R}^{(N-n),n}$$

2) Eine zweite Transformation mit der orthogonalen Matrix Q_2 wird durchgeführt.

$$Q_2^T Q_1^T \tilde{C} = \begin{pmatrix} R_2 \\ O_1 \\ O_2 \end{pmatrix}, \quad Q_2^T Q_1^T \tilde{d} = \begin{pmatrix} \hat{d}_1 \\ \hat{d}_2 \\ \hat{d}_3 \end{pmatrix}, \quad R_2, O_2 \in \mathbb{R}^{n,n}$$

Der gesuchte Vektor $\vec{v}^{(k+1)}$ ergibt sich dann aus der Rücksubstitution. Muss λ vergrößert werden, muss nur Schritt 2 ausgeführt werden.