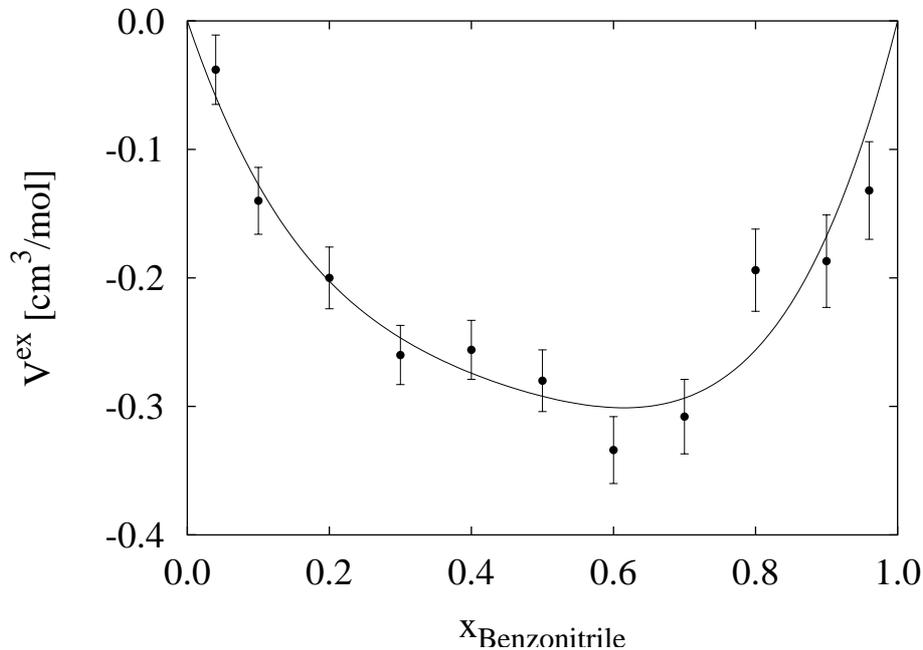


Lineare Ausgleichsprobleme

Datenreduktion und Parameteranpassung



Sowohl bei der Auswertung von Laborexperimenten als auch bei der Analyse von numerischen Berechnungen ergibt sich oft die Situation, daß man für eine Reihe von Werten x_1, x_2, \dots, x_N einer unabhängigen Variablen x zugehörige Werte y_1, y_2, \dots, y_N einer abhängigen Variablen y gemessen (oder berechnet) hat und an diese Daten eine analytische Funktion $y(x)$ anpassen will.

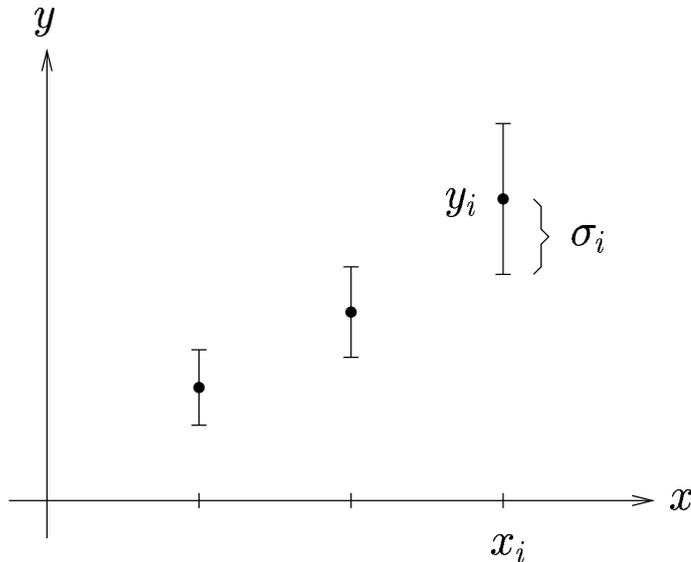
Der Zweck eines solchen “Fits” kann es z.B. sein, die Streuung von mit Meßfehlern behafteten Daten auszugleichen, die Meßwerte für beliebige x zwischen den x_i zu interpolieren, oder einen umfangreichen Datensatz auf einige wenige Parameter zu reduzieren, die zur Beschreibung der Funktion $y(x)$ notwendig sind. In all diesen Fällen muß die Fit-Funktion zunächst keine physikalische Bedeutung haben, und man wird für $y(x)$ die einfachste mathematische Form wählen, die es erlaubt, die Daten im Rahmen der Meßfehler zu reproduzieren. Beliebte Funktionen dieser Art sind Polynome, Exponentialfunktionen, rationale Funktionen, usw.

Andererseits kann es sein, daß man, auf der Grundlage eines konkreten physikalischen Modells, sehr wohl eine Vorstellung von der funktionalen Form von $y(x)$ hat. Meist hängt dieses Modell von einer Reihe von unbekanntem Parametern ab, und die Aufgabe des Fits ist es, die Werte dieser Parameter aus den Messungen zu bestimmen. Zusätzlich möchte man auch noch beurteilen können, ob das theoretische Modell überhaupt mit den Daten verträglich ist. Es muß also ein Gütekriterium für den Fit formuliert werden.

Beide Fragestellungen sind insofern äquivalent, als sie im allgemeinen mit demselben Verfahren gelöst werden, nämlich der sogenannten “Methode der kleinsten Quadrate”. Während sich

dieses Verfahren im Fall der Parameteranpassung unmittelbar aus den Vorstellungen der Maximum Likelihood-Schätzung ableiten läßt, könnte man bei bloßen Datenreduktionsproblemen aber natürlich auch nach alternativen Strategien vorgehen.

Maximum Likelihood-Schätzung



Es seien N Paare von Meßwerten mit zugehörigen Meßfehlern

$$\{x_i, y_i, \sigma_i\}_{i=1, \dots, N}$$

gegeben. Dabei wird angenommen, daß die x_i exakt bekannt, die y_i aber mit einem statistischen Meßfehler σ_i behaftet sind. (Die Größenordnung von σ_i kann man z.B. abschätzen, indem man für jedes x_i mehrere Messungen von y durchführt und den Standardfehler des Mittelwerts bestimmt.) Weiters wird angenommen, daß die Meßwerte y_i *statistisch unabhängig* und *normalverteilt* sind, d.h. die Wahrscheinlichkeit (eigentlich Wahrscheinlichkeitsdichte), als i -ten Meßpunkt den Wert y_i zu erhalten, ist

$$p(y_i) = \left(\frac{1}{2\pi\sigma_i^2} \right)^{1/2} \exp \left\{ -\frac{[y_i - y(x_i)]^2}{2\sigma_i^2} \right\}$$

Es wird hier also die Gültigkeit eines “theoretischen Modells” $y = y(x)$ angenommen, d.h. $y(x_i)$ wäre der “wahre” Funktionswert an der Stelle x_i , bei einer tatsächlichen Messung wird aber ein davon abweichender Wert y_i beobachtet. Zusätzlich wird noch vorausgesetzt, daß die σ_i unabhängig von den im konkreten Fall gemessenen Werten y_i sein sollen.

Das theoretische Modell hängt im allgemeinen von M Parametern $\lambda_1, \lambda_2, \dots, \lambda_M$ ab

$$y = y(\lambda_1, \dots, \lambda_M, x)$$

Nach dem Prinzip der Maximum Likelihood-Schätzung werden die λ_α so bestimmt, daß die Wahrscheinlichkeit, bei einer Messung genau jene Werte zu erhalten, die man tatsächlich gemessen hat, maximal wird.

Da vorausgesetzt wurde, daß die y_i statistisch unabhängig sein sollen, ist die Wahrscheinlichkeitsdichte für das Auftreten der Meßwerte y_1, y_2, \dots, y_N das Produkt der Einzelwahrscheinlichkeitsdichten

$$\begin{aligned} p(y_1, y_2, \dots, y_N) &= \left(\frac{1}{2\pi\sigma_1^2} \right)^{1/2} \exp \left\{ -\frac{[y_1 - y(x_1)]^2}{2\sigma_1^2} \right\} \left(\frac{1}{2\pi\sigma_2^2} \right)^{1/2} \exp \left\{ -\frac{[y_2 - y(x_2)]^2}{2\sigma_2^2} \right\} \dots \\ &\quad \times \left(\frac{1}{2\pi\sigma_N^2} \right)^{1/2} \exp \left\{ -\frac{[y_N - y(x_N)]^2}{2\sigma_N^2} \right\} \\ &= \prod_{i=1}^N \left(\frac{1}{2\pi\sigma_i^2} \right)^{1/2} \exp \left\{ -\sum_{i=1}^N \frac{[y_i - y(x_i)]^2}{2\sigma_i^2} \right\} \end{aligned}$$

Diese Wahrscheinlichkeitsdichte wird maximal, wenn die Summe im Exponenten (die als Summe von Quadraten ja nicht negativ sein kann) minimal wird. Man wird also die Parameter $\lambda_1, \lambda_2, \dots, \lambda_M$ so bestimmen, daß die Größe

$$s = \sum_{i=1}^N \frac{1}{\sigma_i^2} [y_i - y(\lambda_1, \lambda_2, \dots, \lambda_M, x_i)]^2$$

ein Minimum wird. Dies ist die *Methode der kleinsten Quadrate*.

Gütekriterium

Die aus dem Maximum Likelihood-Prinzip abgeleitete Methode der kleinsten Quadrate hat den Vorteil, daß man unter den gemachten Voraussetzungen sofort ein Gütekriterium für den Fit formulieren kann. Wären nämlich die exakten Parameter $\lambda_1, \lambda_2, \dots, \lambda_M$ des theoretischen Modells bekannt, dann wären die Variablen

$$\eta_i = [y_i - y(\lambda_1, \lambda_2, \dots, \lambda_M, x_i)] / \sigma_i$$

$N(0,1)$ -verteilt (normalverteilt mit Mittelwert 0 und Varianz 1) und s , als Summe der Quadrate von N unabhängigen $N(0,1)$ -verteilten Zufallsvariablen, wäre χ^2 -verteilt mit N Freiheitsgraden. Die Wahrscheinlichkeit für das Auftreten des beobachteten oder eines noch größeren Wertes der Abweichungsquadratsumme s ist dann

$$P(t > s) = 1 - P(t \leq s) = 1 - \frac{1}{\Gamma(N/2)} \int_0^{s/2} dt t^{N/2-1} e^{-t}$$

wobei $\Gamma(z)$ die Gamma-Funktion ist, d.h. $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$ und $\Gamma(z+1) = z\Gamma(z)$. Damit das theoretische Modell mit den Messungen verträglich ist, sollte diese Wahrscheinlichkeit nicht zu klein sein, z.B. nicht kleiner als 50% (sonst hätten die Meßergebnisse nie zustande kommen können).

In Wirklichkeit kennt man jedoch die λ_α nicht a priori, sondern bestimmt sie erst durch Anpassung an die Daten. In diesem Fall sind zwar die y_i (bzw. η_i) nicht mehr alle voneinander unabhängig, man kann aber für den Fall normalverteilter Variablen zeigen, daß s dann noch immer χ^2 -verteilt ist, allerdings mit

$$\nu = N - M$$

(Anzahl der Beobachtungen minus Anzahl der Parameter) Freiheitsgraden. Da für große ν der Median der χ^2 -Verteilung von der Größenordnung ν ist, geht man in der Praxis oft nach der Faustregel vor, daß man ein "reduziertes χ^2 ", $s/(N - M)$, berechnet. Ist $s/(N - M)$ wesentlich größer als 1, so ist das theoretische Modell nicht mit den Daten verträglich (oder man hat die Meßfehler unterschätzt); ist $s/(N - M)$ wesentlich kleiner als 1, dann hat man zu viele Parameter angepaßt (oder die Meßfehler überschätzt).

Normalgleichungen

Zur tatsächlichen Bestimmung der Parameter müssen die simultanen Extremumsbedingungen

$$\frac{\partial s}{\partial \lambda_\alpha} = 0, \quad \alpha = 1, \dots, M$$

nach $\lambda_1, \lambda_2, \dots, \lambda_M$ aufgelöst werden. Das ist im allgemeinen Fall ein System von M gekoppelten transzendenten Gleichungen, dessen Lösung sich beliebig schwierig gestalten kann.

Wesentlich einfacher ist die Situation, wenn $y(x)$ ein *lineares Modell* ist,

$$y = \sum_{\alpha=1}^M \lambda_\alpha \varphi_\alpha(x)$$

wobei die $\varphi_\alpha(x)$ beliebige Funktionen (häufig Polynome) sein können. D.h. y ist eine Linearkombination der φ_α , in der die Parameter λ_α als Koeffizienten auftreten. Differenzieren von

$$s = \sum_{i=1}^N \frac{1}{\sigma_i^2} \left[y_i - \sum_{\beta=1}^M \lambda_\beta \varphi_\beta(x_i) \right]^2$$

nach λ_α liefert in diesem Fall

$$\frac{\partial s}{\partial \lambda_\alpha} = -2 \sum_{i=1}^N \frac{1}{\sigma_i^2} \left[y_i - \sum_{\beta=1}^M \lambda_\beta \varphi_\beta(x_i) \right] \varphi_\alpha(x_i) = 0$$

oder, umgeformt,

$$\sum_{\beta=1}^M \left[\sum_{i=1}^N \frac{1}{\sigma_i^2} \varphi_\alpha(x_i) \varphi_\beta(x_i) \right] \lambda_\beta = \sum_{i=1}^N \frac{1}{\sigma_i^2} \varphi_\alpha(x_i) y_i$$

Mit den Abkürzungen

$$A_{\alpha\beta} = \sum_{i=1}^N \frac{1}{\sigma_i^2} \varphi_\alpha(x_i) \varphi_\beta(x_i)$$

$$b_\alpha = \sum_{i=1}^N \frac{1}{\sigma_i^2} \varphi_\alpha(x_i) y_i$$

ist das ein lineares Gleichungssystem der Form

$$A\lambda = b$$

Dieses System der *Normalgleichungen* kann mit dem Gaußschen Eliminationsverfahren oder einem anderen gängigen Verfahren zur Lösung linearer Gleichungssysteme nach den λ_α aufgelöst werden:

$$\lambda = A^{-1}b$$

Falls die $\varphi_\alpha(x)$ Polynome sind, kann schon für $M \sim 5-10$ das Lösungsverfahren numerisch problematisch werden. In diesem Fall empfiehlt es sich, zunächst orthogonale Polynome $\psi_\alpha(x)$ mit

$$\sum_{i=1}^N \frac{1}{\sigma_i^2} \psi_\alpha(x_i) \psi_\beta(x_i) = \delta_{\alpha\beta}$$

zu konstruieren, das Ausgleichsproblem in diesen zu lösen (A ist dann eine Diagonalmatrix) und die Lösung auf die $\varphi_\alpha(x)$ umzurechnen.

Fehlerabschätzung

Für ein lineares Modell und unter der Voraussetzung, daß das Modell mit den Messungen verträglich ist, kann man auch noch Aussagen über die Unsicherheit der mit Hilfe eines Fits bestimmten Parameter machen.

Dazu benützt man, daß die Matrix A in den Normalgleichungen zwar von den Stützstellen x_i und den (als von den Daten unabhängig angenommenen) Fehlern σ_i abhängt, aber nicht von den eigentlichen Meßwerten y_i . Für feste x_i und σ_i sind daher die λ_α lineare Funktionen der y_i ,

$$\lambda_\alpha = \sum_{\beta=1}^M (A^{-1})_{\alpha\beta} \sum_{i=1}^N \frac{1}{\sigma_i^2} \varphi_\beta(x_i) y_i$$

und als solche ebenfalls Zufallsvariable.

Stellt man sich nun vor, daß man den Prozeß des Messens und Anpassens viele Male wiederholt, so bekommt man für jeden Satz von Messungen $\{y_1, y_2, \dots, y_N\}$ einen zugehörigen Satz von Parametern $\{\lambda_1, \lambda_2, \dots, \lambda_M\}$. Der Mittelwert des Parameters λ_α über alle diese Meßserien ist dann

$$\langle \lambda_\alpha \rangle = \sum_{\beta=1}^M (A^{-1})_{\alpha\beta} \sum_{i=1}^N \frac{1}{\sigma_i^2} \varphi_\beta(x_i) \langle y_i \rangle$$

Ein Maß für die Abweichung von λ_α in einem einzelnen Fit vom Mittelwert ist die Varianz

$$\begin{aligned} \sigma_{\lambda_\alpha}^2 &= \langle (\lambda_\alpha - \langle \lambda_\alpha \rangle)^2 \rangle \\ &= \sum_{\beta=1}^M (A^{-1})_{\alpha\beta} \sum_{\gamma=1}^M (A^{-1})_{\alpha\gamma} \sum_{i=1}^N \frac{1}{\sigma_i^2} \varphi_\beta(x_i) \sum_{j=1}^N \frac{1}{\sigma_j^2} \varphi_\gamma(x_j) \langle (y_i - \langle y_i \rangle) (y_j - \langle y_j \rangle) \rangle \end{aligned}$$

Wegen der statistischen Unabhängigkeit von y_i und y_j für $i \neq j$ ist

$$\langle (y_i - \langle y_i \rangle) (y_j - \langle y_j \rangle) \rangle = \delta_{ij} \sigma_i^2$$

und daher

$$\begin{aligned}\sigma_{\lambda_\alpha}^2 &= \sum_{\gamma=1}^M (A^{-1})_{\alpha\gamma} \sum_{\beta=1}^M (A^{-1})_{\alpha\beta} \sum_{i=1}^N \frac{1}{\sigma_i^2} \varphi_\beta(x_i) \varphi_\gamma(x_i) \\ &= \sum_{\gamma=1}^M (A^{-1})_{\alpha\gamma} \sum_{\beta=1}^M (A^{-1})_{\alpha\beta} A_{\beta\gamma} \\ &= \sum_{\gamma=1}^M (A^{-1})_{\alpha\gamma} \delta_{\alpha\gamma}\end{aligned}$$

Die Unsicherheit der Parameter ist also durch die Diagonalelemente der Matrix A^{-1} gegeben:

$$\sigma_{\lambda_\alpha}^2 = (A^{-1})_{\alpha\alpha}$$

Eine analoge Rechnung ergibt

$$\text{cov}(\lambda_\alpha, \lambda_\beta) = (A^{-1})_{\alpha\beta}$$

Dabei ist die Kovarianz

$$\text{cov}(\lambda_\alpha, \lambda_\beta) = \langle (\lambda_\alpha - \langle \lambda_\alpha \rangle) (\lambda_\beta - \langle \lambda_\beta \rangle) \rangle$$

ein Maß für die Korrelation zwischen den Parametern.