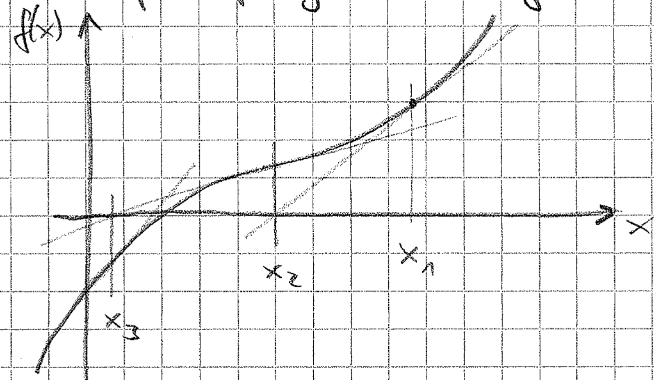
Gauss-Newton method:

Newton method: Linearize function in origin

zeros of tangent gives new origin



Newton's method in optimization: Approximate f(x) by a quadratic function around x_n

Take a step towards minimum

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}$$

Gauss-Newton: Can only be used to minimize a sum of squared-fcts.

No second derivatives required

We can use the GN method to compute the parameters of a completely general model function

$$f(x; a_1, a_2, a_3, \dots, a_q)$$

This fct is plugged into χ^2

$$\chi^2 = \sum_{k=1}^{n} g_k \left(y_k - f(x_k; \vec{a}) \right)^2$$

Iterative procedure:

1.) Choose starting values for parameters

$$\vec{a} = \vec{a}_0$$

2.) Linearize $f(x,\vec{a})$ around these $\vec{a}_0$ using Taylor expansion

$$f(x,\vec{a}) \approx f(x,\vec{a}_0) + \sum_{k=1}^{q} \left.\frac{\partial f(x,\vec{a})}{\partial a_k}\right|_{\vec{a}=\vec{a}_0} (a_k - a_k^0)$$

3.) This linearized fct is plugged into χ^2

$$\chi^2 = \sum_{k=1}^{n} g_k \left(y_k - f(x_k,\vec{a}_0) - \sum_{l=1}^{q} \left.\frac{\partial f(x_k,\vec{a})}{\partial a_l}\right|_{\vec{a}=\vec{a}_0} (a_l - a_l^0)\right)^2$$

4.) Take the derivative of χ^2 with respect to a_j

$$\frac{\partial \chi^2}{\partial a_j} = -2 \sum_{k=1}^{n} g_k \left(y_k - f(x_k,\vec{a}_0) - \sum_{l=1}^{q} \left.\frac{\partial f(x_k,\vec{a})}{\partial a_l}\right|_{\vec{a}=\vec{a}_0} (a_l - a_l^0)\right) \frac{\partial f(x_k,\vec{a})}{\partial a_j}$$

This leads to the set of equations

$$\alpha_{ij} := \sum_{k=1}^{n} g_k \left.\frac{\partial f(x_k,\vec{a})}{\partial a_i}\right|_{\vec{a}=\vec{a}_0} \left.\frac{\partial f(x_k,\vec{a})}{\partial a_j}\right|_{\vec{a}=\vec{a}_0} , \quad A_{ij} = \alpha_{ij}$$

$$\beta_i := \sum_{k=1}^{n} g_k (y_k - f_k) \left.\frac{\partial f(\vec{x}_k,\vec{a})}{\partial a_i}\right|_{\vec{a}=\vec{a}_0}$$

$$A(\vec{a} - \vec{a}_0) = \vec{\beta}$$

5.) The solution of this system can be interpreted as difference between guessed values and improved values

$$a_l - a_l^0 = \Delta a_l$$

$$a_l^1 = a_l^0 + \Delta a_l$$

6.) Repeat steps 2-5 to iteratively improve parameters

When to stop?

All parameters have to satisfy a condition for relative precision

$$|a_l^t - a_l^{t-1}| < |a_l^t| \cdot \varepsilon \quad \forall a_l \quad , \varepsilon > 0 \text{ and very small}$$

Maximum numbers of iterations reached

Once convergence is reached, we want to calculate the normal matrix

$$N_{ij} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_i \partial a_j}$$

$$\left(\begin{array}{lll} \text{Diagonal: standard deviations} & \sigma_{a_i} = \sqrt{C_{ii}} & \\ \text{off diagonal: correlation } \cancel{\text{functions}} \text{ coefficients} & r_{ij} = \frac{C_{ij}}{\sqrt{C_{ii} C_{jj}}} & \text{of } C = N^{-1} \end{array} \right)$$

For our method, this means for a general nonlinear fct $f(x, \vec{a})$

$$\cancel{\frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_j \partial a_i} = + \sum_{k=1}^{n} g_k \frac{\partial^2 f(x_k, \vec{a})}{\partial a_i} \frac{\partial f(x_k, \vec{a})}{\partial a_j}}$$

$$\chi^2 = \sum_{k=1}^{n} g_k \left(y_k - f(x_k, \vec{a}) \right)^2$$

$$\frac{\partial \chi^2}{\partial a_j} = -2 \sum_{k=1}^{n} g_k \left(y_k - f(x_k, \vec{a}) \right) \frac{\partial f(x_k, \vec{a})}{\partial a_j}$$

$$\frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_i \partial a_j} = - \sum_{k=1}^{n} g_k \left(\left(y_k - f(x_k, \vec{a}) \right) \frac{\partial^2 f(x_k, \vec{a})}{\partial a_i \partial a_j} - \frac{\partial f(x_k, \vec{a})}{\partial a_i} \frac{\partial f(x_k, \vec{a})}{\partial a_j} \right)$$

$$= \sum_{k=1}^{n} g_k \frac{\partial f}{\partial a_i} \frac{\partial f}{\partial a_j} - \left(y_k - f(x_k, a) \right) \frac{\partial^2 f}{\partial a_i \partial a_j}$$

If the method works and we converge towards the right solution, the second term $(y_k - f(x_k, \vec{a}))$ gets smaller and smaller.

It is thus justified to approximate N_{ij} by just taking the first term.

This is, by the way, also the solution for N_{ij} if we use our linearized model

$$\frac{1}{2}\frac{\partial^2 \chi^2}{\partial a_i \partial a_j} = \sum_{k=1}^{n} g_k \frac{\partial f(x_k,\vec{a})}{\partial a_i}\frac{\partial f(x_k,\vec{a})}{\partial a_j}$$

In this case the normal matrix N coincides with the matrix of the coefficients A.

The last step ist to invert [redacted] to get the covariance matrix $C = N^{-1}$

Convergence is of vital importance!
Does the GN method always converge? No

Example see script

Levenberg-Marquardt algorithm
This algorithm is an extension to the GN method by using regularization.
Starting point is again our system of equations

$$A \cdot \Delta\vec{a} = \vec{\beta}$$

Here however we use a regularization parameter λ, that is introduced in the following way

$$(A + \lambda D)\,\Delta\vec{a} = \vec{\beta} \quad \text{with } D_{ii} = A_{ii}$$

This we do because of the following reason:

The GN method could yield the result

$$\chi_t^2 < \chi_{t+1}^2$$

i.e. the square error sum of the $t+1$ iteration is bigger than the one of the t^{th} iteration. In this case, convergence can not be guaranteed.

With the λ parameter however it is always possible to get

$$\chi_t^2 > \chi_{t+1}^2$$

Let's look at a 2×2 system:

$$\begin{pmatrix} \alpha_{11}(1+\lambda) & \alpha_{12} \\ \alpha_{12} & \alpha_{22}(1+\lambda) \end{pmatrix} \begin{pmatrix} \Delta a_1 \\ \Delta a_2 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

$$\alpha_{11}(1+\lambda)\,\Delta a_1 + \alpha_{12}\,\Delta a_2 - \beta_1 = 0$$

$$\alpha_{12}\,\Delta a_1 + \alpha_{22}(1+\lambda)\,\Delta a_2 - \beta_2 = 0$$

$$\Delta a_2 = \frac{\beta_1 - \alpha_{11}(1+\lambda)\Delta a_1}{\alpha_{12}}$$

$$\alpha_{12}\,\Delta a_1 + \alpha_{22}(1+\lambda)\frac{\beta_1 - \alpha_{11}(1+\lambda)\Delta a_1}{\alpha_{12}} - \beta_2 = 0$$

$$\alpha_{12}^2\,\Delta a_1 + \alpha_{22}(1+\lambda)\beta_1 - \alpha_{11}\alpha_{22}(1+\lambda)^2\Delta a_1 - \beta_2\alpha_{12} = 0$$

$$\Delta a_1 = \frac{\alpha_{12}\beta_2 - \alpha_{22}(1+\lambda)\beta_1}{\alpha_{12}^2 - \alpha_{11}\alpha_{22}(1+\lambda)^2}$$

$$\Delta a_2 = \frac{\alpha_{12}\beta_1 - \alpha_{11}(1+\lambda)\beta_2}{\alpha_{12}^2 - \alpha_{11}\alpha_{22}(1+\lambda)^2}$$

As we see, for $\lambda \to 0$, the original GN method is recovered

For $\lambda \to \infty$, Δa_1 and Δa_2 go to zero

Δa_1 and Δa_2 are functions of λ, i.e. the magnitude of the correction can be regulated.

An increase of λ leads to a reduction of the correction

The calculation becomes more careful, but the convergence speed decreases

A sufficient increase of λ leads to a monotonic decrease of the error sums

$$\chi_1^2 \geq \chi_2^2 \geq \ldots \geq \chi_t^2 \geq \ldots$$

Marquardt's strategy:

Start iteration with a small, positive λ

Before each new step, λ is reduced by a fixed factor to ensure fast convergence.

If $\chi_{t+1}^2 > \chi_t^2$, λ is increased by a fixed factor and the calculation is repeated until $\chi_{t+1}^2 < \chi_t^2$

If parameters have to meet constraints (there can only be $g-1$ constraints)

$$\sum_{l=1}^{g} r_{t,l}\, a_l \overset{!}{=} \delta_t \qquad (t = 1, 2, \ldots, L < g)$$

Introduction of additional Lagrange parameters

$$\chi^2 = \sum_{k=1}^{n} g_k \left(y_k - f(x_k, a_1, \ldots, a_g)\right)^2 + \sum_{t=1}^{L} \mu_t \left(\sum_{l=1}^{g} r_{t,l}\, a_l - \delta_t\right) \rightarrow \mathrm{Min}$$