1 Example

Suppose we want to fit the following linear function, $Y = \sqrt{\theta_1 + \theta_2 X} = f(\theta_1, \theta_2)$, to a given data set. In order to do this we need to estimate the parameters θ_1 and θ_2. Making use of a newton raphson approach with numerical differentiation the following explanation is a possibility.

Remark 1 *There are different ways of obtaining numerical derivatives of any function so please do not take this as written on stone.*

For Newton Raphson we need an objective function that we must optimize, suppose we use SSE as the objective function, that is:

$$SSE = \sum \left(Y - \hat{Y}\right)^2 = \sum \left(Y - (\sqrt{\hat{\theta}_1 + \hat{\theta}_2 X})\right)^2 = \sum \left(Y - f\left(\hat{\theta}_1, \hat{\theta}_2\right)\right)^2.$$

It holds that the best estimates of θ_1 and θ_2 will minimize the SSE. These values will be found by employing the NEWTON RAPHSON iterative approach, so we will use this approach to minimize the SSE. For us to implement the Newton Raphson algorithm we must find the following derivatives:

$$\begin{aligned}
\frac{\partial SSE}{\partial \hat{\theta}_1} &= -2 \sum \left\{ \left(Y - f\left(\hat{\theta}_1, \hat{\theta}_2\right)\right) \frac{\partial f\left(\hat{\theta}_1, \hat{\theta}_2\right)}{\partial \hat{\theta}_1} \right\} \\
\frac{\partial SSE}{\partial \hat{\theta}_2} &= -2 \sum \left\{ \left(Y - f\left(\hat{\theta}_1, \hat{\theta}_2\right)\right) \frac{\partial f\left(\hat{\theta}_1, \hat{\theta}_2\right)}{\partial \hat{\theta}_2} \right\} \\
\frac{\partial^2 SSE}{\partial \hat{\theta}_1^2} &= -2 \sum \left\{ \left(Y - f\left(\hat{\theta}_1, \hat{\theta}_2\right)\right) \frac{\partial^2 f\left(\hat{\theta}_1, \hat{\theta}_2\right)}{\partial \hat{\theta}_1^2} - \left(\frac{\partial f\left(\hat{\theta}_1, \hat{\theta}_2\right)}{\partial \hat{\theta}_1}\right)^2 \right\} \\
\frac{\partial^2 SSE}{\partial \hat{\theta}_2^2} &= -2 \sum \left\{ \left(Y - f\left(\hat{\theta}_1, \hat{\theta}_2\right)\right) \frac{\partial^2 f\left(\hat{\theta}_1, \hat{\theta}_2\right)}{\partial \hat{\theta}_2^2} - \left(\frac{\partial f\left(\hat{\theta}_1, \hat{\theta}_2\right)}{\partial \hat{\theta}_2}\right)^2 \right\} \\
\frac{\partial SSE}{\partial \hat{\theta}_1 \partial \hat{\theta}_2} &= -2 \sum \left\{ \left(Y - f\left(\hat{\theta}_1, \hat{\theta}_2\right)\right) \frac{\partial^2 f\left(\hat{\theta}_1, \hat{\theta}_2\right)}{\partial \hat{\theta}_1 \partial \hat{\theta}_2} - \frac{\partial f\left(\hat{\theta}_1, \hat{\theta}_2\right)}{\partial \hat{\theta}_1} * \frac{\partial f\left(\hat{\theta}_1, \hat{\theta}_2\right)}{\partial \hat{\theta}_2} \right\}
\end{aligned}$$

A close analysis of the derivatives above shows that we need only focus on the derivatives of the linear function $Y = \sqrt{\theta_1 + \theta_2 X} = f(\theta_1, \theta_2)$ in order to implement the Newton Raphson algorithm. This then means that we need only obtain the numerical derivatives of the linear function $Y = \sqrt{\theta_1 + \theta_2 X} = f(\theta_1, \theta_2)$ and then substitute these into the formulas given above. The question remains, "HOW DO WE DO THIS??", well there are various ways of doing this, we will use here the centred finite difference method which is explained on the third page of the paper "*Numerical Partial Derivatives*". *The following holds:*

$$
\begin{aligned}
\frac{\partial f\left(\hat{\theta}_1, \hat{\theta}_2\right)}{\partial \hat{\theta}_1} &\approx \frac{1}{2h}\left\{f\left(\hat{\theta}_1 + h, \hat{\theta}_2\right) - f\left(\hat{\theta}_1 - h, \hat{\theta}_2\right)\right\} \\
\frac{\partial f\left(\hat{\theta}_1, \hat{\theta}_2\right)}{\partial \hat{\theta}_2} &\approx \frac{1}{2k}\left\{f\left(\hat{\theta}_1, \hat{\theta}_2 + k\right) - f\left(\hat{\theta}_1, \hat{\theta}_2 - k\right)\right\} \\
\frac{\partial^2 f\left(\hat{\theta}_1, \hat{\theta}_2\right)}{\partial \hat{\theta}_1^2} &\approx \frac{1}{h^2}\left\{f\left(\hat{\theta}_1 + h, \hat{\theta}_2\right) - 2 * f\left(\hat{\theta}_1, \hat{\theta}_2\right) + f\left(\hat{\theta}_1 - h, \hat{\theta}_2\right)\right\} \\
\frac{\partial^2 f\left(\hat{\theta}_1, \hat{\theta}_2\right)}{\partial \hat{\theta}_2^2} &\approx \frac{1}{k^2}\left\{f\left(\hat{\theta}_1, \hat{\theta}_2 + k\right) - 2 * f\left(\hat{\theta}_1, \hat{\theta}_2\right) + f\left(\hat{\theta}_1, \hat{\theta}_2 - k\right)\right\} \\
\frac{\partial^2 f\left(\hat{\theta}_1, \hat{\theta}_2\right)}{\partial \hat{\theta}_1 \partial \hat{\theta}_2} &\approx \frac{1}{4hk}\left\{f\left(\hat{\theta}_1 + h, \hat{\theta}_2 + k\right) - f\left(\hat{\theta}_1 + h, \hat{\theta}_2 - k\right) - f\left(\hat{\theta}_1 - h, \hat{\theta}_2 + k\right) + f\left(\hat{\theta}_1 - h, \hat{\theta}_2 - k\right)\right\}
\end{aligned}
$$

Now that we know the form of the partial derivatives for the linear function, we can write a subroutine say, NUMERICAL DIFF, that evaluates these derivatives and then substitute the resultant values into the formulas for the partial derivatives of the SSE in the Newton Raphson algorithm. Remember that :

$$
\begin{aligned}
GradientVector &= \left\{ \begin{array}{c} \frac{\partial SSE}{\partial \hat{\theta}_1} \\ \frac{\partial SSE}{\partial \hat{\theta}_2} \end{array} \right\} \\
Hessian\ Matrix &= \left\{ \begin{array}{cc} \frac{\partial^2 SSE}{\partial \hat{\theta}_1^2} & \frac{\partial SSE}{\partial \hat{\theta}_1 \partial \hat{\theta}_2} \\ \frac{\partial SSE}{\partial \hat{\theta}_1 \partial \hat{\theta}_2} & \frac{\partial^2 SSE}{\partial \hat{\theta}_2^2} \end{array} \right\}
\end{aligned}
$$

Now that we have all we need we can then run the Newton Raphson algorithm and be on our merry way!!!

Remark 2 *The values h and k are very small values, how small is a totally different subject but you can play around with these values and see how your estimates are affected. The logic applied here can easily be generalised to the case when there are more than two unknown parameters. All that changes is the number of derivatives that you actually need, the rest is "easy peasy"...GOODLUCK, if you have any questions just drop me an email.*