

Data Reduction Notes

Tobias Marriage

Feb 2011

1 Non-linear Fitting

To interpret a dataset with measurements $\{d_i\}$, we have a model which, given n model parameters \vec{p} , gives corresponding values $\{m_i(\vec{p})\}$. The chi-squared for this model is

$$\chi^2(\vec{p}) = \sum_i \frac{[d_i - m_i(\vec{p})]^2}{\sigma_i^2}, \quad (1)$$

where σ_i^2 is the estimate for the variance on d_i .

The errors on d_i are assumed uncorrelated. For linear models, the chi-squared function was a quadratic in the parameters, so we can write

$$\chi^2(\vec{p}) = \chi^2(\vec{p}_0) + \frac{d\chi^2}{d\vec{p}}(\vec{p}_0)(\vec{p} - \vec{p}_0) + \frac{1}{2} \frac{d^2\chi^2}{d^2\vec{p}}(\vec{p}_0)(\vec{p} - \vec{p}_0)^2, \quad (2)$$

where \vec{p}_0 is an arbitrary starting point in parameter space. The minimizing “jump” in parameter space, $\delta\vec{p} = (\vec{p}_{min} - \vec{p}_0)$, is given by

$$\begin{aligned} 0 &= \frac{d\chi^2}{d\vec{p}}(\vec{p}_0) + \frac{d^2\chi^2}{d^2\vec{p}}(\vec{p}_0)\delta\vec{p} \\ \delta\vec{p} &= - \left(\frac{d^2\chi^2}{d^2\vec{p}}(\vec{p}_0) \right)^{-1} \frac{d\chi^2}{d\vec{p}}(\vec{p}_0) \end{aligned} \quad (3)$$

Note that in general the first derivative of chi-squared is a vector (the gradient), and the second derivative is a matrix (the hessian).

Many times, however, the chi-squared function is not purely quadratic. This is the case for non-linear models. In these cases, the above prescription for a purely quadratic function can fail. In these cases, we must follow our nose, potentially taking many steps around parameter space to find the minimum of the chi-squared. One easy prescription is to follow the gradient of chi-squared. The step in parameter space is then

$$\delta\vec{p} = C \frac{d\chi^2}{d\vec{p}}(\vec{p}_0), \quad (4)$$

where C is a diagonal matrix of dimension $n \times n$. In general, we can use both step formulae (Eq 3 and Eq 4) to explore the parameter space in search of the best fit values.

1.1 Gradient and Hessian

The next step towards an algorithm for non-linear model fitting is constructing the gradient vector and hessian matrix of the chi-squared function (Eq 1). The gradient is straight forward. The j^{th} element of this vector is

$$\frac{\partial \chi^2}{\partial p_j}(\vec{p}_0) = -2 \sum_i \frac{[d_i - m_i(\vec{p}_0)]}{\sigma_i^2} \left(\frac{\partial m_i}{\partial p_j}(\vec{p}_0) \right). \quad (5)$$

And the $(j, k)^{\text{th}}$ element of the hessian is

$$\begin{aligned} \frac{\partial^2 \chi^2}{\partial p_j \partial p_k}(\vec{p}_0) &= -2 \sum_i -\sigma_i^{-2} \left(\frac{\partial m_i}{\partial p_j}(\vec{p}_0) \right) \left(\frac{\partial m_i}{\partial p_k}(\vec{p}_0) \right) + \frac{[d_i - m_i(\vec{p}_0)]}{\sigma_i^2} \left(\frac{\partial^2 m_i}{\partial p_j \partial p_k}(\vec{p}_0) \right) \\ &\approx 2 \sum_i \sigma_i^{-2} \left(\frac{\partial m_i}{\partial p_j}(\vec{p}_0) \right) \left(\frac{\partial m_i}{\partial p_k}(\vec{p}_0) \right). \end{aligned} \quad (6)$$

The second term in the sum on the first line tends to cancel out for a good model near the chi-squared minimum: the quantity $d_i - m_i(\vec{p}_0)$ will be randomly positive and negative across d_i such that its sum over i should be small (Why isn't this the case for Eq 5? My hypothesis is that both of these terms are small near the minimum and it is the first term in the hessian which drives the final steps of the search.). The second term is also supposedly a numerical nuisance, so we drop it for the effective expression of the hessian.

1.2 Levenburg-Marquardt

The final step towards solving a non-linear least squares system is providing a prescription for stepping through parameter space and in particular choosing the constant in Eq 4. The Levenburg-Marquardt approach is the classic way to do this [1]. In this approach the constant for gradient decent is chosen to be proportional to the inverse of the second derivative with respect to the parameter in question. Following this prescription, Eq 3 and Eq 4 can be written

$$- \begin{bmatrix} \frac{\partial^2 \chi^2}{\partial p_1 \partial p_1} & \cdots & \frac{\partial^2 \chi^2}{\partial p_1 \partial p_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \chi^2}{\partial p_n \partial p_1} & \cdots & \frac{\partial^2 \chi^2}{\partial p_n \partial p_n} \end{bmatrix} \begin{bmatrix} \delta p_1 \\ \vdots \\ \delta p_n \end{bmatrix} = \begin{bmatrix} \frac{\partial \chi^2}{\partial p_1} \\ \vdots \\ \frac{\partial \chi^2}{\partial p_n} \end{bmatrix}, \quad (7)$$

$$\lambda \begin{bmatrix} \frac{\partial^2 \chi^2}{\partial p_1 \partial p_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\partial^2 \chi^2}{\partial p_n \partial p_n} \end{bmatrix} \begin{bmatrix} \delta p_1 \\ \vdots \\ \delta p_n \end{bmatrix} = \begin{bmatrix} \frac{\partial \chi^2}{\partial p_1} \\ \vdots \\ \frac{\partial \chi^2}{\partial p_n} \end{bmatrix}, \quad (8)$$

where λ is a constant. Eq 7 and Eq 8 specify, respectively, the quadratic (Eq 3) and gradient (Eq 4) contributions to $\delta \vec{p}$. These equations can be rewritten as a single operation on the gradient

$$\begin{bmatrix} \delta p_1 \\ \vdots \\ \delta p_n \end{bmatrix} = \left(\begin{bmatrix} \frac{\partial^2 \chi^2}{\partial p_1 \partial p_1}(-1 + \lambda) & \cdots & \frac{\partial^2 \chi^2}{\partial p_1 \partial p_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \chi^2}{\partial p_n \partial p_1} & \cdots & \frac{\partial^2 \chi^2}{\partial p_n \partial p_n}(-1 + \lambda) \end{bmatrix} \right)^{-1} \begin{bmatrix} \frac{\partial \chi^2}{\partial p_1} \\ \vdots \\ \frac{\partial \chi^2}{\partial p_n} \end{bmatrix}, \quad (9)$$

Setting the matrix on the right hand side to $\alpha(\vec{p}, \lambda)$, the expression simplifies to

$$\delta\vec{p} = \alpha(\vec{p}, \lambda)^{-1} \nabla_p \chi^2. \quad (10)$$

As a prescription for choosing λ and stepping around parameter space, here is a suggestion from the ever useful Numerical Recipes books [2]

1. Solve Eq 10 for $\delta\vec{p}$.
2. If $\chi^2(\vec{p} + \delta\vec{p}) > \chi^2(\vec{p})$, do *not* set $\vec{p} = \vec{p} + \delta\vec{p}$ (reject the step) and increase λ by a factor of 10. Return to start.
3. If $\chi^2(\vec{p} + \delta\vec{p}) < \chi^2(\vec{p})$, set $\vec{p} = \vec{p} + \delta\vec{p}$ and decrease λ by a factor of 10. Return to start.

You stop iterating this sequence when the quantity $\chi^2(\vec{p} + \delta\vec{p}) - \chi^2(\vec{p})$ is significantly less than one.

References

- [1] Kenneth Levenberg. “A Method for the Solution of Certain Non-Linear Problems in Least Squares”. *The Quarterly of Applied Mathematics* 2: 164-168 (1944).
- [2] William Press et al. *Numerical Recipes in C*. Cambridge: Cambridge University Press, 1992.