Mathematics 23: Errors in least-Squares fitting

It is a good idea to reread 'Maths 3' now in order to familiarise yourself with the idea of least-squares fitting. In general there is no easy way to assess the accuracy of a least-squares fit. However, in the special case that we have knowledge about the anticipated errors in our points then it is possible to provide some error analysis. If the *exact* solution is:

$$y(x) = \sum_{j=1}^{N} a_j f_j(x)$$

then the errors that appear may be represented as:

$$y_i = \sum_{j=1}^{N} a_j f_j(x_i) + \epsilon_i \qquad (*)$$

The sort of additional 'assumption' that we make (or happen to know) is that the errors ϵ_i are independently *Gaussianly* distributed with vanishing mean but variance $< \epsilon_i^2 >= \mu_i^2$ which is 'known'. Since under this assumption some points will give larger errors than others, it makes sense to *weight* the contributions so that all contributions are equally important:

$$E = \frac{1}{n}\sum_{i=1}^{n} \frac{\epsilon_i^2}{\mu_i^2} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\mu_i^2}\left[y_i - \sum_{j=1}^{N} a_j f_j(x_i)\right]^2$$

The least-squares optimisation analysis follows a similar course to previously, with the inclusion of these weights, leading to:

$$E = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i^2}{\mu_i^2} - 2\mathbf{a}^T.\mathbf{B} + \mathbf{a}^T A\mathbf{a}$$

where the coefficients of the vector $\mathbf{a}$ are the a_i and for which the vector $\mathbf{B}$ and matrix A have components:

$$B_j = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i f_j(x_i)}{\mu_i^2}$$

$$A_{jk} = \frac{1}{n}\sum_{i=1}^{n} \frac{f_j(x_i) f_k(x_i)}{\mu_i^2}$$

in terms of which the optimum solution satisfies:

$$A\mathbf{a}^* = \mathbf{B}$$

in terms of the optimal variables a_i^*, as before.

We may now try to assess these estimates by using our knowledge of the errors. Let us consider the variance of our least-squares estimate away from the exact values:

$$< (a_j^* - a_j)(a_{j'}^* - a_{j'}) >$$

where the a_j are the exact values, the a_j^* are the least-squares estimates, and the average is over all possible realisations of the errors, ie what you would get if you did the analysis many times over with different sets of errors. Now equation (*) allows us to control the errors:

$$B_j = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i f_j(x_i)}{\mu_i^2} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j'=1}^{N} \frac{a_{j'} f_{j'}(x_i) f_j(x_i)}{\mu_i^2} + \sum_{i=1}^{n} \frac{\epsilon_i f_j(x_i)}{\mu_i^2}$$

$$= \sum_{j'=1}^{N} A_{jj'} a_{j'} + \frac{1}{n}\sum_{i=1}^{n} \frac{\epsilon_i f_j(x_i)}{\mu_i^2}$$

Multiplying through by A^{-1} and recalling that $\mathbf{a}^* = A^{-1}\mathbf{B}$:

$$a_j^* - a_j = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{N} A_{jk}^{-1} \frac{\epsilon_i f_k(x_i)}{\mu_i^2}$$

and so:

$$< (a_j^* - a_j)(a_{j'}^* - a_{j'}) >= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{N} \frac{1}{n}\sum_{i'=1}^{n}\sum_{k'=1}^{N} A_{jk}^{-1} A_{j'k'}^{-1} f_k(x_i) f_{k'}(x_{i'}) \frac{< \epsilon_i \epsilon_{i'} >}{\mu_i^2 \mu_{i'}^2}$$

and using the independence of the errors, ie $< \epsilon_i \epsilon_{i'} >= 0$ if $i \neq i'$, and the variance of the error, $< \epsilon_i^2 >= \mu_i^2$, we are immediately led to:

$$< (a_j^* - a_j)(a_{j'}^* - a_{j'}) >= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{k=1}^{N}\sum_{k'=1}^{N} A_{jk}^{-1} A_{j'k'}^{-1} \frac{f_k(x_i) f_{k'}(x_i)}{\mu_i^2}$$

$$= \frac{1}{n}\sum_{k=1}^{N}\sum_{k'=1}^{N} A_{jk}^{-1} A_{j'k'}^{-1} A_{kk'} = \frac{1}{n} A_{jj'}^{-1}$$

and the errors are controlled directly by the inverse to the matrix A, and scale as $1/\sqrt{n}$.