Appendix D

Polynomial Fits in WPtools

D.1 Polynomial Regression

In this technical Appendix we sketch the formalism used in the polynomial regression method for fitting data. This is a generalization of the method of linear regression.

We start with a set of data (x_j, y_j), $j = 1, ...m$, and we wish to fit these data to the nth-order polynomial

$$y(x) = \sum_{i=0}^{n} a_i x^i. \tag{D.1}$$

In general each measurement y_j has a corresponding uncertainty σ_j. That is, if the measurements were repeated many times at coordinate x_j the values of y_j would follow a gaussian distribution of standard deviation σ_j. We indicate in sec. D.2 how the program WPtools proceeds in the absence of input data as to the σ_j.

Because of the uncertainties in the measurements y_j we cannot expect to find the ideal values of the coefficients a_i, but only a set of best estimates we will call $\hat{a}_i$. However, we will also obtain estimates of the uncertainties in these best-fit parameters which we will label as $\sigma_{\hat{a}_i}$.

The best-fit polynomial is then

$$\hat{y}(x) = \sum_{i=0}^{n} \hat{a}_i x^i. \tag{D.2}$$

The method to find the $\hat{a}_i$ is called least-squares fitting as well as polynomial regression because we minimize the square of the deviations. We introduce the famous chi square:

$$\chi^2 = \sum_{j=1}^{m} \frac{[y_j - \hat{y}(x_j)]^2}{\sigma_j^2} = \sum_{j=1}^{m} \frac{\left(y_j - \sum_{i=0}^{n} \hat{a}_i x_j^i\right)^2}{\sigma_j^2}. \tag{D.3}$$

Fact: $\exp(-\chi^2/2)$ is the (un-normalized) probability distribution for observing a set of variables $\{y_j(x_j)\}$ supposing the true relation of y to x is given by eq. (D.2).

A great insight is that $\exp(-\chi^2/2)$ can be thought of another way. It is also the (un-normalized) probability distribution that the polynomial coefficients have values a_i when their best-fit values are $\hat{a}_i$ with uncertainties due to the measurements $\{y_j\}$. Expressing this in symbols,

$$\exp(-\chi^2/2) = \text{const} \times \exp\left(-\sum_{k=0}^{n}\sum_{l=0}^{n} \frac{(a_k - \hat{a}_k)(a_l - \hat{a}_l)}{2\sigma_{kl}^2}\right), \tag{D.4}$$

or equivalently

$$\chi^2/2 = \text{const} + \sum_{k=0}^{n}\sum_{l=0}^{n}\frac{(a_k-\hat{a}_k)(a_l-\hat{a}_l)}{2\sigma_{kl}^2}. \tag{D.5}$$

The uncertainty on $\hat{a}_k$ is σ_{kk} in this notation. In eqs. (D.4) and (D.5) we have introduced the important concept that the uncertainties in the coefficients $\hat{a}_k$ are correlated. That is, the quantity σ_{kl}^2 is a measure of the probability that the values of a_k and a_l both have positive fluctuations at the same time. In fact, σ_{kl}^2 can be negative indicating that when a_k has a positive fluctuation then a_l has a correlated negative one.

One way to see the merit of minimizing the χ^2 is as follows. According to eq. (D.5) the derivative of χ^2 with respect to a_k is

$$\frac{\partial\chi^2/2}{\partial a_k} = \sum_{l=0}^{n}\frac{a_l-\hat{a}_l}{\sigma_{kl}^2}, \tag{D.6}$$

so that all first derivatives of χ^2 vanish when all $a_l = \hat{a}_l$. That is, χ^2 is a minimum when the coefficients take on their best-fit values $\hat{a}_i$. A further benefit is obtained from the second derivatives:

$$\frac{\partial^2\chi^2/2}{\partial a_k\partial a_l} = \frac{1}{\sigma_{kl}^2}. \tag{D.7}$$

In practice we evaluate the χ^2 according to eq. (D.3) based on the measured data. Taking derivatives we find

$$\frac{\partial\chi^2/2}{\partial\hat{a}_k} = \sum_{j=1}^{m}\frac{\left(y_j-\sum_{i=0}^{n}\hat{a}_i x_j^i\right)\left(-x_j^k\right)}{\sigma_j^2} = \sum_{i=0}^{n}\sum_{j=1}^{m}\frac{\hat{a}_i x_j^i x_j^k}{\sigma_j^2} - \sum_{j=1}^{m}\frac{y_j x_j^k}{\sigma_j^2}, \tag{D.8}$$

and

$$\frac{\partial^2\chi^2/2}{\partial\hat{a}_k\partial\hat{a}_l} = \sum_{j=1}^{m}\frac{x_j^k x_j^l}{\sigma_j^2} \equiv M_{kl}. \tag{D.9}$$

To find the minimum χ^2 we set all derivatives (D.8) to zero, leading to

$$\sum_{i=0}^{n}\sum_{j=1}^{m}\frac{x_j^i x_j^k}{\sigma_j^2}\hat{a}_i = \sum_{j=1}^{m}\frac{y_j x_j^i}{\sigma_j^2} \equiv V_k. \tag{D.10}$$

Using the matrix M_{kl} introduced in eq. (D.9) this can be written as

$$\sum_{i=0}^{n} M_{ik}\hat{a}_i = V_k. \tag{D.11}$$

We then calculate the inverse matrix M^{-1} and apply it to find the desired coefficients:

$$\hat{a}_k = \sum_{l=0}^{n} M_{kl}^{-1}V_l. \tag{D.12}$$

Comparing eqs. (D.7) and (D.9) we have

$$\frac{1}{\sigma_{kl}^2} = M_{kl}. \tag{D.13}$$

The uncertainty in best-fit coefficient $\hat{a}_i$ is then reported as

$$\sigma_{\hat{a}_i} = \sigma_{ii} = \frac{1}{\sqrt{M_{ii}}}. \tag{D.14}$$

D.2 Procedure When the σ_j Are Not Known

This method can still be used even if the uncertainties σ_j on the measurements y_j are not known. When the functional form (D.1) correctly describes the data we claim that on average the minimum χ^2 has value $m - n - 1$.[1] To take advantage of this remarkable result we suppose that all uncertainties σ_j have a common value, σ. Then

$$\chi^2 = \sum_{j=1}^{m} \frac{[y_j - \hat{y}(x_j)]^2}{\sigma^2} \approx m - n - 1, \tag{D.15}$$

so that

$$\sigma_j = \sigma = \sqrt{\frac{\sum_{j=1}^{m}[y_j - \sum_{i=0}^{n} \hat{a}_i x_j^i]^2}{m - n - 1}}\,. \tag{D.16}$$

In practice it appears that the error estimates from this procedure are more realistic if a fit is made using a polynomial with one order higher than needed for a 'good' fit to the data.

Using eq. (D.16) as the estimate of the uncertainty σ on each of the measurements y_j, the matrix M_{kl} of eq. (D.9) becomes

$$M_{kl} = \frac{m - n - 1}{\sum_{j'=1}^{m}[y_{j'} - \sum_{i'=0}^{n} \hat{a}_{i'} x_{j'}^{i'}]^2} \sum_{j=1}^{m} x_j^k x_j^l. \tag{D.17}$$

The estimate (D.14) of the uncertainty on the fit coefficient $\hat{a}_i$ is now given by

$$\sigma_{\hat{a}_i} = \frac{1}{\sqrt{M_{ii}}} = \sqrt{\frac{\sum_{j'=1}^{m}[y_{j'} - \sum_{i'=0}^{n} \hat{a}_{i'} x_{j'}^{i'}]^2}{(m - n - 1)\sum_{j=1}^{m} x_j^{2i}}} \tag{D.18}$$

When WPtools performs a polynomial regression it generates a plot of the data points and the best-fit curve, along with numerical values of various parameters associated with the fit. Figure D.1 gives an example of a fit to a set of 8 data points of the form $y = x^2$. The fit is to the form $\mathsf{y} = \mathsf{a_0} + \mathsf{a_1 x} + \mathsf{a_2 x^2}$. The fit coefficients are $\mathsf{a_0} = -0.4107$, $\mathsf{a_1} = -0.3274$ and $\mathsf{a_2} = 1.1964$. The uncertainties (standard errors) on the fit coefficients are reported as $\mathsf{SE(a_0)} = 4.0070$, $\mathsf{SE(a_1)} = 2.0429$ and $\mathsf{SE(a_2)} = 0.2216$, as calculated according to eq. (D.18). *Note that the uncertainties on coefficients* $\mathsf{a_1}$ *and* $\mathsf{a_1}$ *are larger than the coefficients themselves, which tells us that these coefficients are indistinguishable from zero.*

Also indicated on the plot are the values $\mathsf{R^2} = 0.9915$ and $\sigma = 2.8721$. The latter is the uncertainty in the data points $\{y_j\}$, calculated according to eq. (D.16) with $m = 8$ and $n = 2$. The quantity $\mathsf{R^2}$ is defined by

$$R^2 = \frac{\sum_{j=1}^{m}[\hat{y}(x_j) - \overline{y}]^2}{\sum_{j=1}^{m}[y(x_j) - \overline{y}]^2}\,, \tag{D.19}$$

where the average $\overline{y} = \sum_{j=1}^{m} y(x_j)/m$. This is a measure of the "goodness of fit". If the fit is perfect then $\hat{y}_j = y_j$ for all j and $R^2 = 1$. It is not obvious, but $R^2 \leq 1$ always. The extreme case of $R^2 = 0$ occurs when the fit has the trivial form $\hat{y}(x) = \overline{y}$ for all x, which in general is a bad fit. The qualitative conclusion is that if R^2 is not close to 1, the fit results are to be regarded with suspicion.

[1] The whole fitting procedure does not make sense unless there are more data points (m) than parameters ($n + 1$) being fitted.

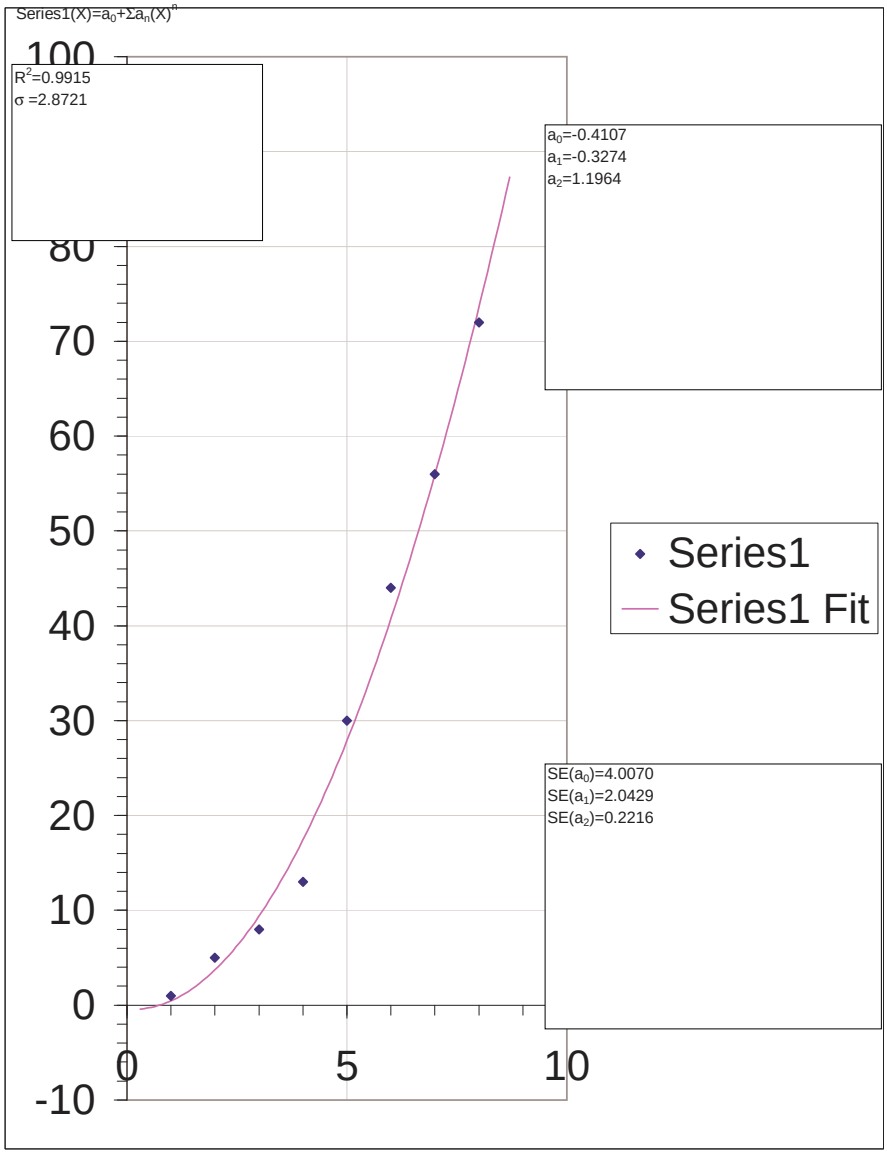

Figure D.1: Sample plot from WPtools Polynomial Fitting.