# PART 6: ERROR PROPAGATION AND MODEL FITTING

## 6.1 ERROR PROPAGATION

Rarely are the data we record—the signals in each pixel of a CCD, for example—exactly what we are interested in for a final result. Hence we must use the data as variables in equations to calculate what we really want to learn. When we use data in an equation however, we must learn how the uncertainty (or error) in each quantity in the equation affects the final result. The methods used to do this are termed *error propagation*, and in this section we shall develop these methods by beginning with simple cases and then progressing to the more complicated ones. We shall learn how to combine several measurements of the same quantity into a single value with the minimum possible error.

### Combining Measurements

Suppose that we have a random variable $z$ that is related to two other random variables, $x$ and $y$, and the constants $a$, $b$, and $c$ with the equation:

$$z = ax + by + c \tag{6.1}$$

The random variables $x$ and $y$ are continuous (non-discrete) random variables, described by the probability density functions $p_x(x)$ and $p_y(y)$. They are *independent* random variables, implying that knowing the value of one of them tells us nothing about the other: $p_y(y|x) = p_y(y)$ and $p_x(x|y) = p_x(x)$. We can write the probability density function for $z$ as:

$$p_z(z) = \int_{-\infty}^{\infty} p_x(x) p_y\left(\frac{z - ax - c}{b}\right) dx = \int_{-\infty}^{\infty} p_x\left(\frac{z - by - c}{a}\right) p_y(y) dy \tag{6.2}$$

*Mean*

We can calculate the mean, $\bar{z}$, of $z$ from its definition [Eq (5.16)] and Eq (6.2):

$$\begin{aligned} \bar{z} &= \int_{-\infty}^{\infty} z p_z(z) dz = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (ax + by + c) p_x(x) p_y(y) dx dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ax p_x(x) p_y(y) dx dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} by p_x(x) p_y(y) dx dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c p_x(x) p_y(y) dx dy \end{aligned} \tag{6.3}$$

$$\bar{z} = a\bar{x} + b\bar{y} + c \tag{6.4}$$

  

*Variance and Standard Deviation*

We calculate the variance and standard deviation with the same approach used above to calculate the mean:

$$\sigma^2(z) = \int_{-\infty}^{\infty} (z - \bar{z})^2 \, p_z(z) dz = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[(ax + by + c) - (a\bar{x} + b\bar{y} + c)\right]^2 p_x(x) p_y(y) dx dy \qquad (6.5)$$

The missing steps can be found by expanding the expression within the integral and breaking up the sums into separate integrals.

$$\sigma^2(z) = a^2\sigma^2(x) + b^2\sigma^2(y) \qquad (6.6)$$

Even if $z$ is not a linear function of $x$ and $y$, we can get an approximate value for the variance of $z$ by using a Taylor expansion around the point $z_0 = z(x_0, y_0)$ and then approximating it as a linear function, using the derivatives at $z_0$ as the linear coefficients for $x$ and $y$.

$$\sigma^2(z_0) \approx \left(\frac{\partial z}{\partial x}\right)^2_{x=x_0} \sigma^2(x) + \left(\frac{\partial z}{\partial y}\right)^2_{y=y_0} \sigma^2(y) \qquad (6.7)$$

Eq (6.7) holds when $x$ and $y$ are independent, and it will be a good approximation to the extent that $z$ is a linear function of $x$ and $y$ over intervals in $x$ and $y$ of a few standard deviations [$\sigma(x)$, $\sigma(y)$] surrounding the point $(x_0, y_0)$. If $z$ is a linear function of $x$ and $y$, then Eq (6.7) is no longer an approximation, but an exact equation.

**Random Variable Arithmetic**

We can extend the results of the previous section to include arithmetic operations.

*Addition*

If $z = x + y$, then:

$$\sigma^2(z) = 1^2\sigma^2(x) + 1^2\sigma^2(y) = \sigma^2(x) + \sigma^2(y) \qquad (6.8)$$

*Subtraction*

If $z = x - y$, then:

$$\sigma^2(z) = 1^2\sigma^2(x) + (-1)^2\sigma^2(y) = \sigma^2(x) + \sigma^2(y) \qquad (6.9)$$

*Multiplication*

If $z = xy$, then:

  

$$\sigma^2(z_0) \approx \left(\frac{\partial z}{\partial x}\right)^2_{x=x_0} \sigma^2(x) + \left(\frac{\partial z}{\partial y}\right)^2_{y=y_0} \sigma^2(y) = y_0^2\sigma^2(x) + x_0^2\sigma^2(y) \qquad (6.10)$$

For multiplication (and division), a useful parameter is the fractional error of $z$, defined to be $\varepsilon(z) = \sigma(z)/z$. This gives the following relationship between the fractional error for $z$ in terms of the fractional errors for $x$ and $y$:

$$\varepsilon^2(z) \equiv \frac{\sigma^2(z)}{z^2} \approx \frac{y^2\sigma^2(x) + x^2\sigma^2(y)}{z^2} = \frac{\sigma^2(x)}{x^2} + \frac{\sigma^2(y)}{y^2} = \varepsilon^2(x) + \varepsilon^2(y) \qquad (6.11)$$

*Division*

If $z = x/y$, then (in terms of $\varepsilon$ as for multiplication):

$$\varepsilon^2(z) \equiv \frac{\sigma^2(z)}{z^2} \approx \frac{\left(\frac{1}{y}\right)^2 \sigma^2(x) + \left(\frac{-x}{y^2}\right)^2 \sigma^2(y)}{z^2} = \frac{\sigma^2(x)}{x^2} + \frac{\sigma^2(y)}{y^2} = \varepsilon^2(x) + \varepsilon^2(y) \qquad (6.12)$$

*Summary*

The standard deviations for the sum and difference of two random variables are equal; for the their product and quotient, the fractional errors are also equal. The latter result holds when the appropriate partial derivatives are linear over the range of a few standard deviations of the two random variables involved in the multiplication or division.

**Correlated Random Variables**

Until now we have worked with independent random variables; we now include *correlated random variables* in our discussion. If two random variables are correlated, it implies that they have the same randomness, and knowledge of one of them gives us information about the other.

If we have random variables $x$, $y$, and $z$, with $z = z(x, y)$, we define the *correlation coefficient*, $\rho$, with the equation:

$$\rho(x,y) = \frac{\overline{(x-\bar{x})(y-\bar{y})}}{\sigma(x)\sigma(y)} \qquad (6.13)$$

The correlation coefficient satisfies the condition $-1.0 \le \rho \le 1.0$. When $x$ and $y$ are not correlated, $\rho = 0.0$. For perfect correlation or anti-correlation (*i.e.* $y = x$ or $y = -x$), $\rho = 1.0$ and $-1.0$ respectively. Intermediate values of $\rho$ occur for intermediate correlations.

If the variables $x$ and $y$ have non-zero correlation, then we must modify Eq (6.7) for the standard deviation of the derived quantity $z$ to include the effect of correlation:

  

$$\sigma^2(z_0) \approx \left(\frac{\partial z}{\partial x}\right)^2_{x=x_0} \sigma^2(x) + 2\rho(x,y)\left(\frac{\partial z}{\partial x}\right)_{x=x_0}\left(\frac{\partial z}{\partial y}\right)_{y=y_0} \sigma(x)\sigma(y) + \left(\frac{\partial z}{\partial y}\right)^2_{y=y_0} \sigma^2(y) \qquad (6.14)$$

The contours of constant error in the $x$-$y$ plane are ellipses—known as *error ellipses*—which have axes parallel to the $x$ and $y$ axes for uncorrelated random variables, but not parallel for correlated random variables.

**Optimum Signal-to-Noise Ratio**

Suppose we have a set of measurements $\{d_i\}$ of a quantity $x$. The set $\{d_i\}$ may be a series of raw data, for which we may have no other knowledge of their standard deviations. In this case we assume that all the standard deviations are equal and derive the standard deviation from the data series as previously described. However, the set $\{d_i\}$ may be a list of actual measurements (such as instrumental magnitudes for a standard star). For this case each member of the series may have its own standard deviation, $\sigma_i$ (not necessarily equal to each other).

The question arises of how to best use all the input values to arrive at the best estimate for $x$. For us, the "best" estimate will be the one with the lowest error. Since the errors in the measurements are not equal, a simple average of the $\{d_i\}$ will not, in general, yield the minimum error (standard deviation) of the final result. The problem we want to solve in this section is to determine how to average the measurements in order to maximize $S/N = \bar{d} / \sigma(\bar{d})$, which is equivalent to minimizing $\sigma(\bar{d})$. Then our best estimate for $x$ (the one with the lowest error) is $\tilde{x} = \bar{d}$.

Here we shall use the "real" $\sigma_i$, rather than the measured value, $\tilde{\sigma}_i$. (When working with data, of course, we have to use $\tilde{\sigma}_i$, since that is all we know. In that case, the results of this section will be approximate, to the extent that standard deviations determined from data themselves contain errors.) It will be instructive to proceed to the desired result in several steps.

*Two Measurements with Equal Errors*

We have two measurements, $(d_1, \sigma_1)$ and $(d_2, \sigma_2)$, with $\sigma_1 = \sigma_2 = \sigma$. Since the standard deviations are equal, we *weight* them equally with weighting constants, $w_1$ and $w_2$, whose sum is 1: in this case, $w_1 = w_2 = 1/2$ (*i.e.* we average the two values).

$$\bar{d} = w_1 d_1 + w_2 d_2 \qquad (6.15)$$

$$\sigma^2(\bar{d}) \approx \left(\frac{\partial \bar{d}}{\partial d_1}\right)^2_0 \sigma_1^2 + \left(\frac{\partial \bar{d}}{\partial d_2}\right)^2_0 \sigma_2^2 = (w_1^2 + w_2^2)\sigma^2 = \frac{\sigma^2}{2} \qquad (6.16)$$

  

$$\sigma(\bar{d}) = \frac{\sigma}{\sqrt{2}} \tag{6.17}$$

*N Measurements with Equal Errors*

Again, we have no reason to weight any one measurement more than any other, so we select $1/N$ as the weighting factor for all:

$$\bar{d} = \sum_{i=1}^{N} w_i d_i = \sum_{i=1}^{N} \frac{1}{N} d_i = \frac{1}{N} \sum_{i=1}^{N} d_i \tag{6.18}$$

$$\sigma^2(\bar{d}) \approx \sum_{i=1}^{N} \left( \frac{\partial \bar{d}}{\partial d_i} \right)_0^2 \sigma_i^2 = \sum_{i=1}^{N} w_i^2 \sigma^2 = \sum_{i=1}^{N} \frac{\sigma^2}{N^2} = \frac{\sigma^2}{N} \tag{6.19}$$

$$\sigma(\bar{d}) \approx \frac{\sigma}{\sqrt{N}} \tag{6.20}$$

*Two Measurements with Unequal Errors*

If we have two measurements with unequal errors, we can choose two different weighting constants, $w_1$ and $w_2$ whose sum is 1, and write the resulting weighted mean as follows:

$$\bar{d} = w_1 d_1 + w_2 d_2 \tag{6.21}$$

$$\sigma^2(\bar{d}) = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 = w_1^2 \sigma_1^2 + (1 - w_1)^2 \sigma_2^2 \tag{6.22}$$

Now we want to choose $w_1$ to minimize the standard deviation in the above equation. At one extreme $w_1 = 1$ and $\sigma(\bar{d}) = \sigma_1$; at the other extreme, $w_1 = 0$ and $\sigma(\bar{d}) = \sigma_2$. We can do better than either extreme. To find the minimum value of $\sigma(\bar{d})$ as a function of $w_1$, we take the derivative with respect to $w_1$, set it to zero, and then solve for $w_1$:

$$\frac{d\sigma^2(\bar{d})}{dw_1} = 2w_1\sigma_1^2 - 2(1 - w_1)\sigma_2^2 = 0 \tag{6.23}$$

$$w_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{\dfrac{1}{\sigma_1^2}}{\dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2}} \tag{6.24}$$

  

and

$$w_2 = \frac{\frac{1}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \tag{6.25}$$

Note that the denominators for $w_1$ and $w_2$ are the same. It is common practice to formally define the weight as the reciprocal of the variance, and then to use the sum to normalize the average (in this case the normalization is necessary because the sum of the weights does not equal 1). Hence, we set $w_1 = 1/\sigma_1^2$ and $w2 = 1/\sigma_2^2$; then we can rewrite the above equations as:

$$\bar{d} = \frac{w_1 d_1 + w_2 d_2}{w_1 + w_2} \tag{6.26}$$

$$\sigma^2(\tilde{\bar{d}}) = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 = \frac{w_1^2\sigma_1^2 + w_2^2\sigma_2^2}{w_1 + w_2} = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{\frac{1}{w_1}\frac{1}{w_2}}{\frac{1}{w_1} + \frac{1}{w_2}} = \frac{1}{w_1 + w_2} \tag{6.27}$$

(For those familiar with electronics, this equation is analogous to the equation for the total resistance of two parallel resistors.)

*N Measurements with Unequal Errors*

For this case we want to find the set of weights, $\{w_i\}$ $i = 1, ..., N$.

$$\bar{d} = \frac{\sum_{i=1}^{N} w_i d_i}{\sum_{i=1}^{N} w_i} \tag{6.28}$$

$$\sigma^2(\bar{d}) = \sum_{i=1}^{N}\left(\frac{\partial \bar{d}}{\partial d_i}\right)_0^2 \sigma_i^2 = \frac{\sum_{i=1}^{N} w_i^2\sigma_i^2}{\left(\sum_{i=1}^{N} w_i\right)^2} \tag{6.29}$$

Now we search for the weights that minimize the variance by taking the derivative of the above equation with respect to $w_i$:

  

$$\frac{d\sigma^2(\bar{d})}{dw_i} = \frac{2w_i\sigma_i^2\left(\sum_{i=1}^{N} w_i\right)^2 - 2\sum_{i=1}^{N} w_i \sum_{i=1}^{N} w_i^2\sigma_i^2}{\left(\sum_{i=1}^{N} w_i\right)^4} \tag{6.30}$$

A minimum occurs when the numerator in the above equation equals zero, and we have the condition:

$$w_i = \frac{\sum_{i=1}^{N} w_i^2\sigma_i^2}{\sigma_i^2 \sum_{i=1}^{N} w_i} \tag{6.31}$$

A solution to the above equation is $w_i = c / \sigma_i^2$, where $c$ is an arbitrary constant. This yields the following value for $\sigma^2(\bar{d})$.

$$\sigma^2(\bar{d}) = \frac{\sum_{i=1}^{N} w_i^2\sigma_i^2}{\left(\sum_{i=1}^{N} w_i\right)^2} = \frac{\sum_{i=1}^{N} cw_i}{\left(\sum_{i=1}^{N} w_i\right)^2} = \frac{c}{\sum_{i=1}^{N} w_i} = \frac{1}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \tag{6.32}$$

In the next section we shall learn how weights are fundamentally connected to using the data to achieve the most accurate estimate of $\bar{d}$.

**Measurements with Multiple Noise Sources**

For measurements with multiple noise sources that are not correlated with one another, such as photon noise and scintillation, we express the measured quantity as the sum of its mean and noise sources, with each of the noise sources having the appropriate variance [*e.g.* the squared, un-normalized version of Eq (3.44) for scintillation and Eq (5.8) for photon noise]. The sum of the variances of noise sources yields the variance of the measurement.

  

## 6.2 MODEL FITTING WITH LEAST SQUARES

In some cases, we know that a given set of data "should" be described by a certain type of function—a constant, straight line or a sine curve, for example—and the only reason for deviation from a perfect match to this function is the noise (or random error) in the data. For these cases in which the data should fit a known functional form, we can use the method of *least squares* to find the parameters of the model that best fits the data. We can also find the rms errors in the best fitting parameters.

### Describing Data with a Model

Models usually will be described by a set of "coordinates," **x**, which can vary with each measurement, and a set of parameters, **p**, that are constant for a given model. Model fitting involves finding the set of the model parameters that best fit the data.

For example, after reducing data taken at the telescope to a set of magnitudes and their errors, we are often faced with the task of finding a model that "best" fits the results. Sometimes we already know that the results should fit a model of a given type (*i.e.*, instrumental magnitudes should be a linear function of airmass, if conditions were constant during the data recording). The set of instrumental magnitudes of a standard star and the airmasses at which the instrumental magnitudes were measured constitute the data set. The airmasses are the coordinates, and the parameters of the model are (i) the instrumental magnitude of the standard star at zero airmass, and (ii) the extinction coefficient ($k$, in magnitudes per airmass).

For each measurement of the set, a given model will yield a mean of the measurements, which can be calculated from the coordinates and parameters, using the prescription of the model. From here on, we shall refer to the data values as $v_i$ instead of $d_i$, in order to avoid confusion with $d$ as standard notation in differential calculus. So the mean of the $i$th measurement is a function of the model, **m**, at the $i$th coordinate:

$$\bar{v}_i = \mathbf{m}_i \tag{6.33}$$

$$\mathbf{m}_i = \mathbf{m}(\mathbf{x}_i, \mathbf{p}) \tag{6.34}$$

where the $i$th coordinate may be composed of several components:

$$\mathbf{x}_i = \{(x, y, z, \ldots)_i\} \tag{6.35}$$

We use the index $j$ for the set of parameters:

$$\mathbf{p} = \{p_j\} \tag{6.36}$$

### Adjusting the Model with "Least Squares"

How do we choose the best set of parameters, **p**? What are the errors in the best set of parameters? We can answer these questions under some fairly general assumptions:

  

1. The noise in the measurements follows a Gaussian distribution.

2. The mean of the distribution is a function of a model, chosen from a class of models that are described by "coordinates" and "parameters."

3. All measurements have a standard deviation $\sigma$ about their local mean. We can later relax this assumption to allow the standard deviation of each measurement to be different.

The main goal of the following derivation is to find the set of model parameters that gives the maximum probability for the particular set of measurements obtained. Another way of thinking of this is that we are trying to find the model that best fits the measurements.

For the $i$th measurement, $v_i$, we assume that the measurements follow a Gaussian distribution, with a mean

$$p(v_i)dv_i = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(v_i-\bar{v}_i)^2}{2\sigma^2}\right]dv_i \tag{6.37}$$

We define $\mathbf{v}$ to be the entire set $\{v_i\}$, $p_v(\mathbf{v})$ to be the product of all the density functions, and d$\mathbf{v}$ to be the product of all the differentials, $dv_i$. Then the probability density for the entire set is the product of the probability densities for each measurement. In the equation below, the symbol $\prod_{i=1}^{N} f_i$ denotes the product $f_1 f_2 \ldots f_N$.

$$p_{\mathbf{v}}(\mathbf{v})d\mathbf{v} = \prod_{i=1}^{N} p(v_i)dv_i = \prod_{i=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(v_i-\bar{v}_i)^2}{2\sigma^2}\right]dv_i \tag{6.38}$$

$$p_{\mathbf{v}}(\mathbf{v})d\mathbf{v} = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\prod_{i=1}^{N}\exp\left[-\frac{(v_i-\bar{v}_i)^2}{2\sigma^2}\right]dv_i \tag{6.39}$$

$$p_{\mathbf{v}}(\mathbf{v})d\mathbf{v} = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left[\frac{-\sum_{i=1}^{N}(v_i-\bar{v}_i)^2}{2\sigma^2}\right]d\mathbf{v} \tag{6.40}$$

Since we do not know the mean of each measurement, we have to use our estimates from the measurements, $\tilde{\bar{v}}_i$. Defining the *residual*, $r_i$, as the difference between the data and the model, $(v_i - \tilde{\bar{v}}_i)$, we can reformulate Eq (6.40) in terms of measured quantities (except for the standard deviation):

  

$$\tilde{p}_{\mathbf{v}}(\mathbf{v})d\mathbf{v} = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left[\frac{-\sum_{i=1}^{N}(v_i-\tilde{\bar{v}}_i)^2}{2\sigma^2}\right]d\mathbf{v}$$
$$= \frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left[\frac{-\sum_{i=1}^{N}r_i^2}{2\sigma^2}\right]d\mathbf{v} \qquad (6.41)$$

We maximize the probability density by minimizing the sum in the exponent of the last equation—which minimizes the sum of the squared differences between the data and the model. So the most probable model is that which has the minimum sum of *squared residuals*.

We can generalize Eq (6.41) for data that have unequal standard deviations. If the *i*th data value has a standard deviation $\sigma_i$, then we return to Eq (6.38) and proceed by keeping the $\sigma_i$ inside the product. Following the steps used to arrive at Eq (6.41) we find:

$$\tilde{p}_{\mathbf{v}}(\mathbf{v})d\mathbf{v} = \frac{1}{(2\pi)^{N/2}}\exp\left[-\frac{1}{2}\sum_{i=1}^{N}\frac{r_i^2}{\sigma_i^2}\right]\prod_{i=1}^{N}\frac{1}{\sigma_i^2}d\mathbf{v} \qquad (6.42)$$

Note that the exponential is outside the product term. For economy of notation we define the *weight* $w_i = 1/\sigma_i^2$ for each data value, and rewrite Eq (6.42) as:

$$\tilde{p}_{\mathbf{v}}(\mathbf{v})d\mathbf{v} = \frac{1}{(2\pi)^{N/2}}\left(\prod_{i=1}^{N}w_i\right)\exp\left[-\frac{1}{2}\sum_{i=1}^{N}w_i r_i^2\right]d\mathbf{v} \qquad (6.43)$$

So for the more general case for data with unequal errors, the most probable model is that which minimizes the sum of the *weighted squared residuals*.

**Modeling Data as a Constant**

If we think that the variations of all the measurements in a data set are just random variations about a constant mean, then we model the data as a constant, $a_0$. In this model, the mean of each measurement is $\bar{v}_i = a_0$. To find our best estimate of the constant, $\tilde{a}_0$, we want to minimize the sum of squared residuals between the measurements, $v_i$., and $\tilde{a}_0$. In this case each residual is the difference, $(v_i - \tilde{a}_0)$, so the desired sum of weighted squared residuals is given by

$$\sum_{i=1}^{N}w_i(v_i-\tilde{a}_0)^2 \qquad (6.44)$$

We minimize the sum by taking the derivative with respect to $\tilde{a}_0$, setting it to zero, and then solving for $\tilde{a}_0$:

  

$$\frac{\partial \sum_{i=1}^{N} w_i (v_i - \tilde{a}_0)^2}{\partial \tilde{a}_0} = 0$$

$$= \sum_{i=1}^{N} w_i \frac{\partial (v_i - \tilde{a}_0)^2}{\partial \tilde{a}_0} \quad . \tag{6.45}$$

$$= \sum_{i=1}^{N} w_i \left(-2v_i + 2\tilde{a}_0\right)$$

This leads to

$$\sum_{i=1}^{N} 2w_i \tilde{a}_0 = \sum_{i=1}^{N} 2w_i v_i \tag{6.46}$$

Solving for $\tilde{a}_0$,

$$\tilde{a}_0 = \frac{\sum_{i=1}^{N} w_i v_i}{\sum_{i=1}^{N} w_i} \tag{6.47}$$

If all the standard deviations are equal, this implies all the weights are equal and Eq (6.47) simplifies to:

$$\tilde{a}_0 = \frac{1}{N} \sum_{i=1}^{N} v_i \tag{6.48}$$

Hence the least squares fit of data modeled as a constant is just the (weighted) average of the data points. Stated another way, the sum of squared residuals between a set of data and a constant is minimized of the constant is chosen to be the (weighted) average of the data values. If all the standard deviations are equal, then the weighted average becomes the average.

For the case when $\tilde{a}_0$ is an average of N measurements of equal errors, Eq (6.20), then:

$$\sigma(\tilde{a}_0) = \frac{\sigma}{\sqrt{N}} \tag{6.49}$$

and, according to Eq (5.25),

$$\tilde{\sigma}(\tilde{a}_0) = \frac{1}{\sqrt{N}} \left[ \frac{1}{N-1} \sum_{i=1}^{N} (v_i - \tilde{a}_0)^2 \right]^{1/2} \tag{6.50}$$

  

We use Eq (6.49) when the standard deviation of each data point is known from other information, but we use Eq (6.50) when we are determining the standard deviation from the scatter of the data from their average.

**Modeling Data as a Line**

Next, we consider the least squares solution of a line with constant term $a_0$ and slope $a_1$, described by the equation:

$$\bar{v}_i = a_0 + a_1(x_i - x_0) \tag{6.51}$$

Our goal is to find the "best" values of $a_0$ and $a_1$. Objectively, this means finding the values that minimize the sum of the squared residuals between the line and the data. We define $w_i$, the weight of the $i$th point as the inverse of the variance for that point:

$$w_i = \frac{1}{\sigma_i^2} \tag{6.52}$$

Using our previous definition of $r_i$ as the difference between the model and each data point, the sum of squared residuals can be written:

$$\sum_{i=1}^{N} w_i r_i^2 = \sum_{i=1}^{N} w_i \left\{ v_i - \left[ \tilde{a}_0 + \tilde{a}_1(x_i - x_0) \right] \right\}^2 \tag{6.53}$$

To minimize the right-hand side of Eq (6.53), we take the partial derivatives of the sum with respect to the parameters $a_0$ and $a_1$ and set them to zero:

$$\frac{\partial \sum_{i=1}^{N} w_i \left\{ v_i - \left[ \tilde{a}_0 + \tilde{a}_1(x_i - x_0) \right] \right\}^2}{\partial \tilde{a}_0} = \sum_{i=1}^{N} -2w_i \left\{ v_i - \left[ \tilde{a}_0 + \tilde{a}_1(x_i - x_0) \right] \right\} = 0 \tag{6.54}$$

$$\frac{\partial \sum_{i=1}^{N} w_i \left\{ v_i - \left[ \tilde{a}_0 + \tilde{a}_1(x_i - x_0) \right] \right\}^2}{\partial \tilde{a}_1} = \sum_{i=1}^{N} -2w_i (x_i - x_0) \left\{ v_i - \left[ \tilde{a}_0 + \tilde{a}_1(x_i - x_0) \right] \right\} = 0 \tag{6.55}$$

Setting the above two equations equal to zero and rewriting them to collect terms, we have the following pair of linear equations for $\tilde{a}_0$ and $\tilde{a}_1$:

$$\begin{aligned} \tilde{a}_0 \sum_{i=1}^{N} w_i + \tilde{a}_1 \sum_{i=1}^{N} w_i (x_i - x_0) &= \sum_{i=1}^{N} w_i v_i \\ \tilde{a}_0 \sum_{i=1}^{N} w_i (x_i - x_0) + \tilde{a}_1 \sum_{i=1}^{N} w_i (x_i - x_0)^2 &= \sum_{i=1}^{N} w_i v_i (x_i - x_0) \end{aligned} \tag{6.56}$$

  

Before writing the solution to these linear equations, we by adopt more compact notation, as defined below:

$$\begin{aligned} \Sigma_w &\equiv \sum_{i=1}^{N} w_i & \Sigma_{wx} &\equiv \sum_{i=1}^{N} w_i (x_i - x_0) & \Sigma_{wv} &\equiv \sum_{i=1}^{N} w_i v_i \\ \Sigma_{wxx} &\equiv \sum_{i=1}^{N} w_i (x_i - x_0)^2 & \Sigma_{wvx} &\equiv \sum_{i=1}^{N} w_i v_i (x_i - x_0) & \Lambda &\equiv \Sigma_{wxx}\Sigma_w - \Sigma_{wx}^2 \end{aligned} \tag{6.57}$$

Solving Eq (6.56) for $\tilde{a}_0$ and $\tilde{a}_1$ with the definitions of Eq (6.57) we find:

$$\tilde{a}_0 = \frac{\Sigma_{wxx}\Sigma_{wv} - \Sigma_{wx}\Sigma_{wvx}}{\Lambda} \tag{6.58}$$

$$\tilde{a}_1 = \frac{\Sigma_{wvx}\Sigma_{w} - \Sigma_{wx}\Sigma_{wv}}{\Lambda} \tag{6.59}$$

Eqs (6.57) and (6.58) will be useful in a variety of applications, such as in the analysis of extinction data, where the slope ($\tilde{a}_1$) is the extinction coefficient, and the intercept ($\tilde{a}_0$) is the zero point correction for the instrumental magnitude.

In order to write expressions for the errors in the slope and intercept, we shall need the weighted sum of the squared residuals, which is as follows:

$$\Sigma_{wrr} \equiv \sum_{i=1}^{N} w_i r_i^2 = \sum_{i=1}^{N} w_i \left\{ v_i - \left[ \tilde{a}_0 + \tilde{a}_1 (x_i - x_0) \right] \right\}^2 \tag{6.60}$$

The errors in the intercept and slope are given by:

$$\tilde{\sigma}^2(\tilde{a}_0) = \frac{\Sigma_{wrr}\Sigma_{wxx}}{(N-2)\Lambda} \tag{6.61}$$

$$\tilde{\sigma}^2(\tilde{a}_1) = \frac{\Sigma_{wrr}\Sigma_{w}}{(N-2)\Lambda} \tag{6.62}$$

The division by $N - 2$ (not derived here) occurs because of the two free parameters used in fitting the line, analogous to the division by $N - 1$ for fitting a constant [Eq (5.A.9)].

An assumption of the preceding derivation is that all variances, $\sigma_i$, are known, and these were used to establish the weights, $w_i$, through Eq. (6.52). Another common situation is that the variances are not known, but they can be assumed to be equal for each measurement. For this case, one sets all weights equal to 1 and calculates the common variance, $\tilde{\sigma}^2$, with the equation:

  

$$\tilde{\sigma}^2 = \frac{\Sigma_{wrr}}{(N-2)} \tag{6.63}$$

Eq. (6.63) can also be used in the case where the relative weights are known, but there is an unknown common scale factor among them.

The correlation coefficient [Eq (6.13)], $\rho(\tilde{a}_0, \tilde{a}_1)$, (where $-1 \le \rho(\tilde{a}_0, \tilde{a}_1) \le 1$) is given by the equation:

$$\rho(\tilde{a}_0, \tilde{a}_1) = -\frac{\Sigma_{wx}}{\sqrt{\Sigma_w \Sigma_{wxx}}} \tag{6.64}$$

The *covariance*, $\text{cov}(\tilde{a}_0, \tilde{a}_1)$, is the product of the correlation coefficient of these two quantities and their standard deviations:

$$\text{cov}(\tilde{a}_0, \tilde{a}_1) = \rho(\tilde{a}_0, \tilde{a}_1)\sigma(\tilde{a}_0)\sigma(\tilde{a}_1) = -\frac{\Sigma_{wrr}\Sigma_{wx}}{(N-2)\Lambda} \tag{6.65}$$

As we shall see later, it will sometimes be convenient to select the parameter $x_0$ so that their correlation coefficient, $\rho(\tilde{a}_0, \tilde{a}_1)$, will be zero. We can do this by substituting $\Sigma_{wx} = 0$ in its definition in Eq (6.57) and then solving for $x_0$:

$$x_0 = \frac{\sum_{i=1}^{N} w_i x_i}{\Sigma_w} \tag{6.66}$$

Having values for $\tilde{a}_0$ and $\tilde{a}_1$, we can now generalize Eq (6.48) to apply for any $x$, not just an $x_i$ for which we have data:

$$\tilde{\bar{v}} = \tilde{a}_0 + \tilde{a}_1(x - x_0) \tag{6.67}$$

The error for any $\tilde{\bar{v}}$ will depend on the errors in $\tilde{a}_0$ and $\tilde{a}_1$, and we can write an equation for the error in the value of $\tilde{\bar{v}}$ for any $x$:

$$\tilde{\sigma}^2(\tilde{\bar{v}}) = \left(\frac{\partial \tilde{\bar{v}}}{\partial \tilde{a}_0}\right)^2 \tilde{\sigma}^2(\tilde{a}_0) + 2\left(\frac{\partial \tilde{\bar{v}}}{\partial \tilde{a}_0}\right)\left(\frac{\partial \tilde{\bar{v}}}{\partial \tilde{a}_1}\right)\text{cov}(\tilde{a}_0, \tilde{a}_1) + \left(\frac{\partial \tilde{\bar{v}}}{\partial \tilde{a}_1}\right)^2 \tilde{\sigma}^2(\tilde{a}_1) \tag{6.68}$$

Substituting the values for the partial derivatives:

$$\tilde{\sigma}^2(\tilde{\bar{v}}) = \tilde{\sigma}^2(\tilde{a}_0) + 2(x - x_0)\text{cov}(\tilde{a}_0, \tilde{a}_1) + (x - x_0)^2 \tilde{\sigma}^2(\tilde{a}_1) \tag{6.69}$$

  

This can be written in another form by substituting values for the variances and covariances:

$$\tilde{\sigma}^2(\tilde{\bar{v}}) = \frac{\Sigma_{wrr}}{(N-2)\Lambda}\left[\Sigma_{wxx}^2 - 2(x-x_0)\Sigma_{wx} + (x-x_0)^2\Sigma_w\right] \tag{6.70}$$

If $x_0$ has been chosen to satisfy Eq (6.66), then $\text{cov}(\tilde{a}_0, \tilde{a}_1)$ will be zero and we have the simpler form:

$$\tilde{\sigma}^2(\tilde{\bar{v}}) = \tilde{\sigma}^2(\tilde{a}_0) + (x-x_0)^2\tilde{\sigma}^2(\tilde{a}_1) = \frac{\Sigma_{wrr}}{(N-2)\Lambda}\left[\Sigma_{wxx}^2 + (x-x_0)^2\Sigma_w\right] \tag{6.71}$$

Inspection of Eq(6.71) shows us that the error in our estimate of a value, $\tilde{\bar{v}}(x)$, from our fitted line is a minimum for $x = x_0$, and grows quadratically as the values of $x$ depart from $x_0$. In other words, we have our best knowledge at the (weighted) average $x$-coordinate of our data and the errors get larger as we depart from it (or even extrapolate the fitted line well beyond the range of the data).

*Other Models*

Other common models that can easily be fitted to data with the least-squares method are higher order polynomials and Fourier series. If the variable parameters of the model are the coefficients of these series, then all derivatives of the model function with respect to the model parameters are constants, and the problem has an exact solution that can be found by extending the method that we just used for fitting a straight line. Such methods are examples of *linear least-squares* methods. We also can extend the least-squares method to *non-linear* models, such as a sine wave in which we are fitting not only the amplitude but the frequency as well. The approach is to approximate these functions with a Taylor expansion in the region of interest, which reduces the problem to a linear one. Non-linear least-squares methods are commonly used for modeling data, but are beyond the scope of these notes.

**Internal Errors and External Errors**

The *internal error* of a quantity derived from a data set is calculated by propagating the errors derived from the scatter of the fundamental data points. The *external error* is calculated from the deviations of the data points from a model. If the two "agree," then we believe that we understand our data. If there is doubt, one can perform statistical tests, such as the *chi-square* test, or acquire more data. Usually the latter leads to more firm conclusions.

  