a straight line, we might not notice that the fit is not very good unless we have already had experience fitting data. Finally, the visceral experience of fitting the data manually gives us some feeling for the nature of the data that might otherwise be missed. It is a good idea to plot some data in this way, even though a computer can do it much faster.

Although the visual approach is simple, it does not yield precise results; we need to use more systematic fitting methods. The most common method for finding the best straight line fit to a series of measured points is called *linear regression* or *least squares*. Suppose we have $n$ pairs of measurements $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ and that the errors are entirely in the values of $y$. For simplicity, we assume that the uncertainties in $\{y_i\}$ all have the same magnitude. Our goal is to obtain the best fit to the linear function

$$y = mx + b. \tag{7.32}$$

The problem is to calculate the values of the parameters $m$ and $b$ for the best straight line through the $n$ data points. The difference

$$d_i = y_i - mx_i - b, \tag{7.33}$$

is a measure of the discrepancy in $y_i$. It is reasonable to assume that the best pair of values of $m$ and $b$ are those that minimize the quantity

$$\chi^2 = \sum_{i=1}^{n} (y_i - mx_i - b)^2. \tag{7.34}$$

Why should we minimize the sum of the squared differences between the experimental values $y_i$ and the analytical values $mx_i + b$, and not some other function of the differences? The justification is based on the assumption that if we did many simulations or measurements, then the values of $d_i$ would be distributed according to the Gaussian distribution (see Problems 7.5 and 7.15). Based on this assumption, it can be shown that the values of $m$ and $b$ that minimize $\chi$ yield a set of values of $mx_i + b$ that are the *most probable* set of measurements that we would find based on the available information. This link to probability is the reason we have discussed least squares fits in this chapter, even though we will not explicitly show that the difference $d_i$ is distributed according to a Gaussian distribution.

To minimize $\chi$, we take the partial derivative of $S$ with respect to $b$ and $m$:

$$\frac{\partial \chi}{\partial m} = -2 \sum_{i=1}^{n} x_i (y_i - mx_i - b) = 0 \tag{7.35a}$$

$$\frac{\partial \chi}{\partial b} = -2 \sum_{i=1}^{n} (y_i - mx_i - b) = 0. \tag{7.35b}$$

From (7.35) we obtain two simultaneous equations:

$$m \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \tag{7.36a}$$

$$m \sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i. \tag{7.36b}$$

It is convenient to define the quantities

$$\langle c \rangle = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{7.37a}$$

$$\langle y \rangle = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{7.37b}$$

$$\langle xy \rangle = \frac{1}{n} \sum_{i=1}^{n} x_i y_i, \tag{7.37c}$$

and rewrite (7.36) as

$$m\langle x^2 \rangle + b\langle x \rangle = \langle xy \rangle \tag{7.38a}$$

$$m\langle x \rangle + b = \langle y \rangle. \tag{7.38b}$$

The solution of (7.38) can be expressed as

$$m = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\sigma_x^2} \tag{7.39a}$$

$$b = \langle y \rangle - m \langle x \rangle, \tag{7.39b}$$

where

$$\sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2. \tag{7.39c}$$

Equation (7.39) determines the slope $m$ and the intercept $b$ of the best straight line through the $n$ data points. (Note that the average for the coefficients $m$ and $b$ are over the data points.)

As an example, consider the data shown in Table 7.2 for a one-dimensional random walk. To make the example more interesting, suppose that the walker takes steps of length 1 or 2 with equal probability. The direction of the step is random and $p = 1/2$. As in Section 7.2, we assume that the mean square displacement $\Delta x^2$ obeys the general relation

$$\Delta x^2 = aN^{2\nu}, \tag{7.40}$$

with an unknown exponent $\nu$ and $a$ a constant. Note that the fitting problem in (7.40) is nonlinear; that is, $\langle x^2 \rangle - \langle x \rangle^2$ depends on $N^\nu$ rather than $N$. Often a problem that looks

**Table 7.2** Computed values of the mean square displacement $\Delta x^2$ as a function of the total number of steps $N$. The mean square displacement was averaged over 1000 trials. The one-dimensional random walker takes steps of length 1 or 2 with equal probability, and the direction of the step is random with $p = 1/2$.

| $N$ | $\Delta x^2$ |
|---|---|
| 8 | 19.43 |
| 16 | 37.65 |
| 32 | 76.98 |
| 64 | 160.38 |