**Figure 8.2** The red curve shows a plot of the error between the numerical evaluation of a gradient using finite differences (8.24) and the analytical result, as a function of ϵ. As ϵ decreases, the plot initially shows a linear decrease in error, and this represents a power law behaviour since the axes are logarithmic. The slope of this line is 1 which shows that this error behaves like $\mathcal{O}(\epsilon)$. At some point the evaluated gradient reaches the limit of numerical round-off and further reduction in ϵ leads to a noisy line, which again follows a power law but where the error now increases with decreasing ϵ. The blue curve shows the corresponding result for central differences (8.25). We see a much smaller error compared to finite differences, and the slope of the line is 2 which shows that the error is $\mathcal{O}(\epsilon^2)$.
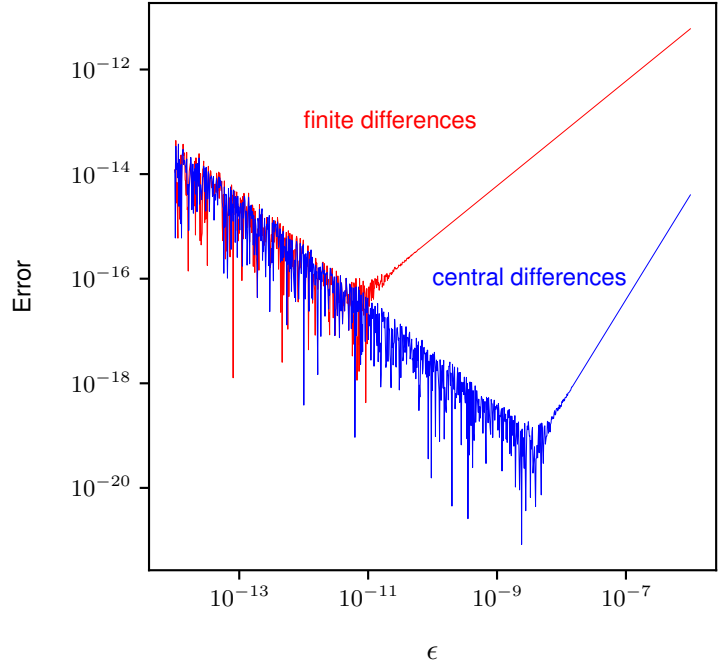
8.1.5 The Jacobian matrix

We have seen how the derivatives of an error function with respect to the weights can be obtained by propagating errors backwards through the network. Backpropagation can also be used to calculate other derivatives. Here we consider the evaluation of the *Jacobian* matrix, whose elements are given by the derivatives of the network outputs with respect to the inputs:

$$J_{ki} \equiv \frac{\partial y_k}{\partial x_i} \tag{8.26}$$

where each such derivative is evaluated with all other inputs held fixed. Jacobian matrices play a useful role in systems built from a number of distinct modules, as illustrated in Figure 8.3. Each module can comprise a fixed or learnable function, which can be linear or nonlinear, so long as it is differentiable.

Suppose we wish to minimize an error function E with respect to the parameter w in Figure 8.3. The derivative of the error function is given by

$$\frac{\partial E}{\partial w} = \sum_{k,j} \frac{\partial E}{\partial y_k} \frac{\partial y_k}{\partial z_j} \frac{\partial z_j}{\partial w} \tag{8.27}$$

in which the Jacobian matrix for the red module in Figure 8.3 appears as the middle term on the right-hand side.

Because the Jacobian matrix provides a measure of the local sensitivity of the outputs to changes in each of the input variables, it also allows any known errors Δx_i