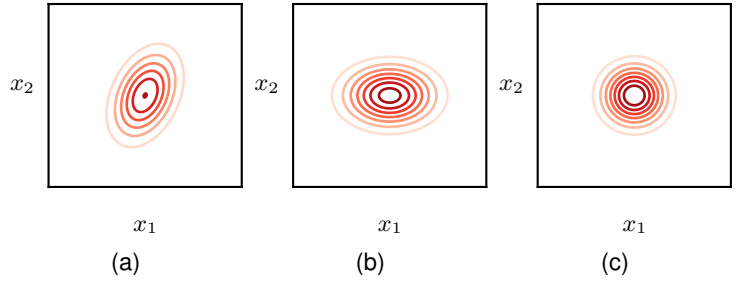


**Figure 3.4** Contours of constant probability density for a Gaussian distribution in two dimensions in which the covariance matrix is (a) of general form, (b) diagonal, in which case the elliptical contours are aligned with the coordinate axes, and (c) proportional to the identity matrix, in which case the contours are concentric circles.



whereas such approaches limit the number of degrees of freedom in the distribution and make inversion of the covariance matrix a much faster operation, they also greatly restrict the form of the probability density and limit its ability to capture interesting correlations in the data.

A further limitation of the Gaussian distribution is that it is intrinsically unimodal (i.e., has a single maximum) and so is unable to provide a good approximation to multimodal distributions. Thus, the Gaussian distribution can be both too flexible, in the sense of having too many parameters, and too limited in the range of distributions that it can adequately represent. We will see later that the introduction of *latent* variables, also called *hidden* variables or *unobserved* variables, allows both of these problems to be addressed. In particular, a rich family of multimodal distributions is obtained by introducing discrete latent variables leading to mixtures of Gaussians. Similarly, the introduction of continuous latent variables leads to models in which the number of free parameters can be controlled independently of the dimensionality  $D$  of the data space while still allowing the model to capture the dominant correlations in the data set.

Section 3.2.9

Chapter 16

### 3.2.4 Conditional distribution

An important property of a multivariate Gaussian distribution is that if two sets of variables are jointly Gaussian, then the conditional distribution of one set conditioned on the other is again Gaussian. Similarly, the marginal distribution of either set is also Gaussian.

First, consider the case of conditional distributions. Suppose that  $\mathbf{x}$  is a  $D$ -dimensional vector with Gaussian distribution  $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$  and that we partition  $\mathbf{x}$  into two disjoint subsets  $\mathbf{x}_a$  and  $\mathbf{x}_b$ . Without loss of generality, we can take  $\mathbf{x}_a$  to form the first  $M$  components of  $\mathbf{x}$ , with  $\mathbf{x}_b$  comprising the remaining  $D - M$  components, so that

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix}. \quad (3.49)$$

We also define corresponding partitions of the mean vector  $\boldsymbol{\mu}$  given by

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix} \quad (3.50)$$