

where $y_j = \mathbf{u}_j^T \mathbf{z}$, which gives

$$\begin{aligned}
 & \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \int \exp \left\{ -\frac{1}{2} \mathbf{z}^T \Sigma^{-1} \mathbf{z} \right\} \mathbf{z} \mathbf{z}^T d\mathbf{z} \\
 &= \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \sum_{i=1}^D \sum_{j=1}^D \mathbf{u}_i \mathbf{u}_j^T \int \exp \left\{ -\sum_{k=1}^D \frac{y_k^2}{2\lambda_k} \right\} y_i y_j d\mathbf{y} \\
 &= \sum_{i=1}^D \mathbf{u}_i \mathbf{u}_i^T \lambda_i = \Sigma
 \end{aligned} \tag{3.45}$$

where we have made use of the eigenvector equation (3.28), together with the fact that the integral on the middle line vanishes by symmetry unless $i = j$. In the final line we have made use of the results (2.53) and (3.38), together with (3.31). Thus, we have

$$\mathbb{E}[\mathbf{x}\mathbf{x}^T] = \boldsymbol{\mu}\boldsymbol{\mu}^T + \Sigma. \tag{3.46}$$

When defining the variance for a single random variable, we subtracted the mean before taking the second moment. Similarly, in the multivariate case it is again convenient to subtract off the mean, giving rise to the *covariance* of a random vector \mathbf{x} defined by

$$\text{cov}[\mathbf{x}] = \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^T]. \tag{3.47}$$

For the specific case of a Gaussian distribution, we can make use of $\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu}$, together with the result (3.46), to give

$$\text{cov}[\mathbf{x}] = \Sigma. \tag{3.48}$$

Because the parameter matrix Σ governs the covariance of \mathbf{x} under the Gaussian distribution, it is called the covariance matrix.

3.2.3 Limitations

Although the Gaussian distribution (3.26) is often used as a simple density model, it suffers from some significant limitations. Consider the number of free parameters in the distribution. A general symmetric covariance matrix Σ will have $D(D+1)/2$ independent parameters, and there are another D independent parameters in $\boldsymbol{\mu}$, giving $D(D+3)/2$ parameters in total. For large D , the total number of parameters therefore grows quadratically with D , and the computational task of manipulating and inverting the large matrices can become prohibitive. One way to address this problem is to use restricted forms of the covariance matrix. If we consider covariance matrices that are *diagonal*, so that $\Sigma = \text{diag}(\sigma_i^2)$, we then have a total of $2D$ independent parameters in the density model. The corresponding contours of constant density are given by axis-aligned ellipsoids. We could further restrict the covariance matrix to be proportional to the identity matrix, $\Sigma = \sigma^2 \mathbf{I}$, known as an *isotropic* covariance, giving $D+1$ independent parameters in the model together with spherical surfaces of constant density. The three possibilities of general, diagonal, and isotropic covariance matrices are illustrated in Figure 3.4. Unfortunately,

Exercise 3.15