much of Chapters 2 and 3 and concentrate their attention on later chapters, although Sections 2.3, 2.4, 3.3, 3.4, 3.8, and to a lesser extent 3.5, are likely to be of interest to most readers.

To derive the form of the PCs, consider first $\boldsymbol{\alpha}_1'\mathbf{x}$; the vector $\boldsymbol{\alpha}_1$ maximizes $\mathrm{var}[\boldsymbol{\alpha}_1'\mathbf{x}] = \boldsymbol{\alpha}_1'\boldsymbol{\Sigma}\boldsymbol{\alpha}_1$. It is clear that, as it stands, the maximum will not be achieved for finite $\boldsymbol{\alpha}_1$ so a normalization constraint must be imposed. The constraint used in the derivation is $\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_1 = 1$, that is, the sum of squares of elements of $\boldsymbol{\alpha}_1$ equals 1. Other constraints, for example $\mathrm{Max}_j|\alpha_{1j}| = 1$, may more useful in other circumstances, and can easily be substituted later on. However, the use of constraints other than $\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_1 =$ *constant* in the derivation leads to a more difficult optimization problem, and it will produce a set of derived variables different from the PCs.

To maximize $\boldsymbol{\alpha}_1'\boldsymbol{\Sigma}\boldsymbol{\alpha}_1$ subject to $\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_1 = 1$, the standard approach is to use the technique of Lagrange multipliers. Maximize

$$\boldsymbol{\alpha}_1'\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 - \lambda(\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_1 - 1),$$

where λ is a Lagrange multiplier. Differentiation with respect to $\boldsymbol{\alpha}_1$ gives

$$\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 - \lambda\boldsymbol{\alpha}_1 = \mathbf{0},$$

or

$$(\boldsymbol{\Sigma} - \lambda\mathbf{I}_p)\boldsymbol{\alpha}_1 = \mathbf{0},$$

where $\mathbf{I}_p$ is the $(p \times p)$ identity matrix. Thus, λ is an eigenvalue of $\boldsymbol{\Sigma}$ and $\boldsymbol{\alpha}_1$ is the corresponding eigenvector. To decide which of the p eigenvectors gives $\boldsymbol{\alpha}_1'\mathbf{x}$ with maximum variance, note that the quantity to be maximized is

$$\boldsymbol{\alpha}_1'\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_1'\lambda\boldsymbol{\alpha}_1 = \lambda\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_1 = \lambda,$$

so λ must be as large as possible. Thus, $\boldsymbol{\alpha}_1$ is the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{\Sigma}$, and $\mathrm{var}(\boldsymbol{\alpha}_1'\mathbf{x}) = \boldsymbol{\alpha}_1'\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 = \lambda_1$, the largest eigenvalue.

In general, the kth PC of $\mathbf{x}$ is $\boldsymbol{\alpha}_k'\mathbf{x}$ and $\mathrm{var}(\boldsymbol{\alpha}_k'\mathbf{x}) = \lambda_k$, where λ_k is the kth largest eigenvalue of $\boldsymbol{\Sigma}$, and $\boldsymbol{\alpha}_k$ is the corresponding eigenvector. This will now be proved for $k = 2$; the proof for $k \geq 3$ is slightly more complicated, but very similar.

The second PC, $\boldsymbol{\alpha}_2'\mathbf{x}$, maximizes $\boldsymbol{\alpha}_2'\boldsymbol{\Sigma}\boldsymbol{\alpha}_2$ subject to being uncorrelated with $\boldsymbol{\alpha}_1'\mathbf{x}$, or equivalently subject to $\mathrm{cov}[\boldsymbol{\alpha}_1'\mathbf{x}, \boldsymbol{\alpha}_2'\mathbf{x}] = 0$, where $\mathrm{cov}(x,\ y)$ denotes the covariance between the random variables x and y . But

$$\mathrm{cov}\ [\boldsymbol{\alpha}_1'\mathbf{x}, \boldsymbol{\alpha}_2'\mathbf{x}] = \boldsymbol{\alpha}_1'\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_2'\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2'\lambda_1\boldsymbol{\alpha}_1' = \lambda_1\boldsymbol{\alpha}_2'\boldsymbol{\alpha}_1 = \lambda_1\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_2.$$

Thus, any of the equations

$$\begin{aligned} \boldsymbol{\alpha}_1'\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 &= 0, & \boldsymbol{\alpha}_2'\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 &= 0, \\ \boldsymbol{\alpha}_1'\boldsymbol{\alpha}_2 &= 0, & \boldsymbol{\alpha}_2'\boldsymbol{\alpha}_1 &= 0 \end{aligned}$$