

## Chapter 9

# Principal Component Analysis



### 9.1. Introduction

We will adopt the same notations as in the previous chapters. Lower-case letters  $x, y, \dots$  will denote real scalar variables, whether mathematical or random. Capital letters  $X, Y, \dots$  will be used to denote real matrix-variate mathematical or random variables, whether square or rectangular matrices are involved. A tilde will be placed on top of letters such as  $\tilde{x}, \tilde{y}, \tilde{X}, \tilde{Y}$  to denote variables in the complex domain. Constant matrices will for instance be denoted by  $A, B, C$ . A tilde will not be used on constant matrices unless the point is to be stressed that the matrix is in the complex domain. The determinant of a square matrix  $A$  will be denoted by  $|A|$  or  $\det(A)$  and, in the complex case, the absolute value or modulus of the determinant of  $A$  will be denoted as  $|\det(A)|$ . When matrices are square, their order will be taken as  $p \times p$ , unless specified otherwise. When  $A$  is a full rank matrix in the complex domain, then  $AA^*$  is Hermitian positive definite where an asterisk designates the complex conjugate transpose of a matrix. Additionally,  $dX$  will indicate the wedge product of all the distinct differentials of the elements of the matrix  $X$ . Letting the  $p \times q$  matrix  $X = (x_{ij})$  where the  $x_{ij}$ 's are distinct real scalar variables,  $dX = \wedge_{i=1}^p \wedge_{j=1}^q dx_{ij}$ . For the complex matrix  $\tilde{X} = X_1 + iX_2$ ,  $i = \sqrt{-1}$ , where  $X_1$  and  $X_2$  are real,  $d\tilde{X} = dX_1 \wedge dX_2$ .

The requisite theory for the study of *Principal Component Analysis* has already been introduced in Chap. 1, namely, the problem of optimizing a real quadratic form that is subject to a constraint. We shall formulate the problem with respect to a practical situation consisting of selecting the most “relevant” variables in a study. Suppose that a scientist would like to devise a “good health” index in terms of certain indicators. After selecting a random sample of individuals belonging to a population that is homogeneous with respect to a variety of factors, such as age group, racial background and environmental conditions, she managed to secure measurements on  $p = 15$  variables, including for instance,  $x_1$ : weight,  $x_2$ : systolic pressure,  $x_3$ : blood sugar level, and  $x_4$ : height. She now

faces a quandary as there is an excessive number of variables and some of them may not be relevant to her investigation. A methodology is thus required for discarding the unimportant ones. As a result, the number of variables will be reduced and the interpretation of the results, possibly facilitated. So, what might be the most pertinent variables in any such study? If all the observations of a particular variable  $x_j$  are concentrated around a certain value,  $\mu_j$ , then that variable is more or less predetermined. As an example, suppose that the height of the individuals comprising a study group is neighboring 1.8 meters and that, on the other hand, it is observed that the weight measurements are comparatively spread out. On account of this, while height is not a particularly consequential variable in connection with this study, weight is. Accordingly, we can utilize the criterion: *the larger the variance of a variable, the more relevant this variable is*. Let the  $p \times 1$  vector  $X$ ,  $X' = (x_1, \dots, x_p)$ , encompass all the variables on which measurements are available. Let the covariance matrix associated with  $X$  be  $\Sigma$ , that is,  $\text{Cov}(X) = \Sigma$ . Since linear functions also contain individual variables, we may consider linear functions such as  $u = a_1x_1 + \dots + a_px_p = A'X = X'A$ , a prime designating a transpose, where

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, \quad A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \quad \text{and} \quad \Sigma = (\sigma_{ij}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}. \quad (9.1.1)$$

Then,

$$\text{Var}(u) = \text{Var}(A'X) = \text{Var}(X'A) = A'\Sigma A. \quad (9.1.2)$$

## 9.2. Principal Components

As will be explained further, the central objective in connection with the derivation of principal components consists of maximizing  $A'\Sigma A$ . Such an exercise would indeed prove meaningless unless some constraint is imposed on  $A$ , considering that, for an arbitrary vector  $A$ , the minimum of  $A'\Sigma A$  occurs at zero and the maximum, at  $+\infty$ ,  $\Sigma = E[X - E(X)][X - E(X)]'$  being either positive definite or positive semi-definite. Since  $\Sigma$  is symmetric and non-negative definite, its eigenvalues, denoted by  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ , are real. Moreover,  $\Sigma$  being symmetric, there exists an orthonormal matrix  $P$ ,  $PP' = I$ ,  $P'P = I$ , such that

$$P'\Sigma P = \text{diag}(\lambda_1, \dots, \lambda_p) \equiv \Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix} \quad (9.2.1)$$

and

$$\Sigma = P \Lambda P' = \lambda_1 P_1 P_1' + \cdots + \lambda_p P_p P_p', \quad (9.2.2)$$

where  $P_1, \dots, P_p$  constitute the columns of  $P$ ,  $P_i$  denoting a normalized eigenvector corresponding to  $\lambda_i$ ,  $i = 1, \dots, p$ ; this is expounded for instance in Mathai and Haubold (2017a). Note that all real symmetric matrices, including those having repeated eigenvalues, can be diagonalized. Since the optimization problem is pointless when  $A$  is arbitrary, the search for an optimum shall be confined to vectors  $A$  such that  $A'A = 1$ , that is, vectors lying on the unit sphere in  $\Re^p$ . Without any loss of generality, the coefficients of the linear function can be selected so that the Euclidean norm of the coefficient vector is unity, in which case a minimum and a maximum will both exist. Hence, the problem can be restated as follows:

$$\text{Maximize } A' \Sigma A \text{ subject to } A'A = 1. \quad (i)$$

We will resort to the method of Lagrangian multipliers to optimize  $A' \Sigma A$  subject to the constraint  $A'A = 1$ . Let

$$\phi_1 = A' \Sigma A - \lambda(A'A - 1) \quad (ii)$$

where  $\lambda$  is a Lagrangian multiplier. Differentiating  $\phi_1$  with respect to  $A$  and equating the result to a null vector (vector/matrix derivatives are discussed in Chap. 1, as well as in Mathai 1997), we have the following:

$$\frac{\partial \phi_1}{\partial A} = O \Rightarrow 2 \Sigma A - 2 \lambda A = O \Rightarrow \Sigma A = \lambda A. \quad (iii)$$

On premultiplying (iii) by  $A'$ , we have

$$A' \Sigma A = \lambda A'A = \lambda. \quad (9.2.3)$$

In order to obtain a non-null solution for  $A$  in (iii), the coefficient matrix  $\Sigma - \lambda I$  has to be singular or, equivalently, its determinant has to be zero, that is,  $|\Sigma - \lambda I| = 0$ , which implies that  $\lambda$  is an eigenvalue of  $\Sigma$ ,  $A$  being the corresponding eigenvector. Thus, it follows from (9.2.3) that the maximum of the quadratic form  $A' \Sigma A$ , subject to  $A'A = 1$ , is the largest eigenvalue of  $\Sigma$ :

$$\max_{A'A=1} [A' \Sigma A] = \lambda_1 = \text{the largest eigenvalue of } \Sigma.$$

Similarly,

$$\min_{A'A=1} [A' \Sigma A] = \lambda_p = \text{the smallest eigenvalue of } \Sigma. \quad (9.2.4)$$

Since  $\text{Var}(A'X) = A'\Sigma A = \lambda$ , the largest variance associated with a linear combination  $A'X$  wherein the vector  $A$  is normalized, is equal to  $\lambda_1$ , the largest eigenvalue of  $\Sigma$ , and letting  $A_1$  be the normalized ( $A_1'A_1 = 1$ ) eigenvector corresponding  $\lambda_1$ ,  $u_1 = A_1'X$  will be that linear combination of  $X$  having the maximum variance. Thus,  $u_1$  is called the first principal component which is the linear function of  $X$  having the maximum variance. Although normalized, the vector  $A_1$  is not unique, as  $(-A_1)'(-A_1)$  is also equal to one. In order to ensure the unicity, we will require that the first nonzero element of  $A_1$ —and the other  $p - 1$  normalized eigenvectors—be positive. Recall that since  $\Sigma$  is a real symmetric matrix, the  $\lambda_j$ 's are real and so are the corresponding eigenvectors. Consider the second largest eigenvalue  $\lambda_2$  and determine the associated normalized eigenvector  $A_2$ ; then  $u_2 = A_2'X$  will be the second principal component. Since the matrix  $P$  that diagonalizes  $\Sigma$  into the diagonal matrix of its eigenvalues is orthonormal, the normalized eigenvectors  $A_1, A_2, \dots, A_p$  are necessarily orthogonal to each other, which means that the corresponding principal components  $u_1 = A_1'X, u_2 = A_2'X, \dots, u_p = A_p'X$  will be uncorrelated. Let us see whether uncorrelated normalized eigenvectors could be obtained by making use of the above procedure. When constructing  $A_2$ , we can impose an additional condition to the effect that  $A'X$  should be uncorrelated with  $A_1'X$ ,  $A_1'\Sigma A_1$  being equal to  $\lambda_1$ , the largest eigenvalue of  $\Sigma$ . The covariance between  $A'X$  and  $A_1'X$  is

$$\text{Cov}(A'X, A_1'X) = A'\text{Cov}(X)A_1 = A'\Sigma A_1 = A_1'\Sigma A. \quad (9.2.5)$$

Hence, we may require that  $A'\Sigma A_1 = A_1'\Sigma A = 0$ . However,

$$0 = A'\Sigma A_1 = A'(\Sigma A_1) = A'\lambda_1 A_1 = \lambda_1 A'A_1 \Rightarrow A'A_1 = 0. \quad (iv)$$

Observe that  $\lambda_1 > 0$ , noting that  $\Sigma$  would be a null matrix if its largest eigenvalue were equal to zero, in which case no optimization problem would remain to be solved. Consider

$$\phi_2 = A'\Sigma A - 2\mu_1(A'\Sigma A_1 - 0) - \mu_2(A'A - 1) \quad (v)$$

where  $\mu_1$  and  $\mu_2$  are the Lagrangian multipliers. Now, differentiating  $\phi_2$  with respect to  $A$  and equating the result to a null vector, we have the following:

$$\frac{\partial \phi_2}{\partial A} = 0 \Rightarrow 2\Sigma A - 2\mu_1(\Sigma A_1) - 2\mu_2 A = 0. \quad (vi)$$

Premultiplying (vi) by  $A_1'$  yields

$$A_1'\Sigma A - \mu_1 A_1'\Sigma A_1 - \mu_2 A_1'A = 0 \Rightarrow 0 - \mu_1 \lambda_1 - 0 \Rightarrow \mu_1 = 0, \quad (vii)$$

which entails that the added condition of uncorrelatedness with  $u_1 = A'_1 X$  is superfluous. When determining  $A_j$ , we could require that  $A'X$  be uncorrelated with the principal components  $u_1 = A'_1 X, \dots, u_{j-1} = A'_{j-1} X$ ; however, as it turns out, these conditions become redundant when optimizing  $A'\Sigma A$  subject to  $A'A = 1$ . Thus, after determining the normalized eigenvector  $A_j$  (whose first nonzero element is positive) corresponding the  $j$ -th largest eigenvalue of  $\Sigma$ , we form the  $j$ -th principal component,  $u_j = A'_j X$ , which will necessarily be uncorrelated with the preceding principal components. The question that arises at this juncture is whether all of  $u_1, \dots, u_p$  are needed or a subset thereof would suffice? For instance, we could interrupt the computations when the variance of the principal component  $u_j = A'_j X$ , namely  $\lambda_j$ , falls below a predetermined threshold, in which case we can regard the remaining principal components,  $u_{j+1}, \dots, u_p$ , as unimportant and omit the associated calculations. In this way, a reduction in the number of variables is achieved as the original number of variables  $p$  is reduced to  $j < p$  principal components. However, this reduction in the number of variables could be viewed as a compromise since the new variables are linear functions of all the original ones and so, may not be as interpretable in a real-life situation. Other drawbacks will be considered in the next section. Observe that since  $\text{Var}(u_j) = \lambda_j$ ,  $j = 1, \dots, p$ , the fraction

$$v_1 = \frac{\lambda_1}{\sum_{j=1}^p \lambda_j} = \text{the proportion of the total variation accounted for by } u_1, \quad (9.2.6)$$

and letting  $r < p$ ,

$$v_r = \frac{\sum_{j=1}^r \lambda_j}{\sum_{j=1}^p \lambda_j} = \text{the proportion of total variation accounted for by } u_1, \dots, u_r. \quad (9.2.7)$$

If  $v_1 = 0.7$ ,  $v_1$  accounts for 70% of the total variation in the original variables or 70% of the total variation is due to the first principal component. If  $r = 3$  and  $v_3 = 0.99$ , then the sum of the first three principal components accounts for 99% of the total variation. We can also use this percentage of the total variation as a stopping rule for the determination of the principal components. For example when  $v_r$  of (9.2.7) is say, greater than or equal to 95%, we may interrupt the determination of the principal components beyond  $u_r$ .

**Example 9.2.1.** Even though reducing the number of variables is a main objective of Principal Component Analysis, for illustrative purposes, we will consider a case involving three variables, that is,  $p = 3$ . Compute the principal components associated with the following covariance matrix:

$$V = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 3 & 1 \\ 0 & 1 & 3 \end{bmatrix}.$$

**Solution 9.2.1.** Let us verify that  $V$  is positive definite. The leading minors of  $V = V'$  being

$$\begin{aligned} |(3)| &= 3 > 0, \quad \begin{vmatrix} 3 & -1 \\ -1 & 3 \end{vmatrix} = 3^2 - (-1)^2 = 8 > 0, \quad \begin{vmatrix} 3 & -1 & 0 \\ -1 & 3 & 1 \\ 0 & 1 & 3 \end{vmatrix} \\ &= 3[3^2 - (-1)^2] - (-1)[(-1)(3) - 0] = 21 > 0, \end{aligned}$$

$V > O$ . Let us compute the eigenvalues of  $V$ . Consider the equation  $|V - \lambda I| = 0 \Rightarrow (3 - \lambda)[(3 - \lambda)^2 - 1] - (3 - \lambda) = 0 \Rightarrow (3 - \lambda)[(3 - \lambda)^2 - 2] = 0 \Rightarrow (3 - \lambda)(3 - \lambda \pm \sqrt{2}) = 0$ . Hence the eigenvalues are  $\lambda_1 = 3 + \sqrt{2}$ ,  $\lambda_2 = 3$ ,  $\lambda_3 = 3 - \sqrt{2}$ . Let us compute an eigenvector corresponding to  $\lambda_1 = 3 + \sqrt{2}$ . Consider  $(V - \lambda_1 I)X = O$ , that is,

$$\begin{bmatrix} 3 - \lambda_1 & -1 & 0 \\ -1 & 3 - \lambda_1 & 1 \\ 0 & 1 & 3 - \lambda_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \quad (i)$$

There are three linear equations involving the  $x_j$ 's in (i). Since the matrix  $V - \lambda I$  is singular, we need only consider any two of these three linear equations and solve to obtain a solution. Let the first equation be  $-\sqrt{2}x_1 - x_2 = 0 \Rightarrow x_2 = -\sqrt{2}x_1$ , and the second one be  $-x_1 - \sqrt{2}x_2 + x_3 = 0 \Rightarrow -x_1 + 2x_1 + x_3 = 0 \Rightarrow x_3 = -x_1$ . Now, one solution is

$$X_1 = \begin{bmatrix} 1 \\ -\sqrt{2} \\ -1 \end{bmatrix}, \text{ which once normalized is } A_1 = \frac{1}{2} \begin{bmatrix} 1 \\ -\sqrt{2} \\ -1 \end{bmatrix}.$$

Observe that  $X_1$  also satisfies the third equation in (i) and that

$$\begin{aligned} A_1' V A_1 &= \frac{1}{4} [1, -\sqrt{2}, -1] \begin{bmatrix} 3 & -1 & 0 \\ -1 & 3 & 1 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -\sqrt{2} \\ -1 \end{bmatrix} \\ &= \frac{1}{4} [3(1)^2 + 3(-\sqrt{2})^2 + 3(-1)^2 - 2(1)(-\sqrt{2}) + 2(-\sqrt{2})(-1)] \\ &= 3 + \sqrt{2} = \lambda_1. \end{aligned}$$

Thus, the first principal component, denoted by  $u_1$ , is

$$u_1 = A_1' X = \frac{1}{2} [x_1 - \sqrt{2}x_2 - x_3]. \quad (ii)$$

Now, consider an eigenvector corresponding to  $\lambda_2 = 3$ .  $(V - \lambda_2 I)X = O$  provides the first equation:  $-x_2 = 0 \Rightarrow x_2 = 0$ , the third equation giving  $-x_1 + x_3 = 0 \Rightarrow x_1 = x_3$ . Therefore, one solution for  $X$  is

$$X_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \text{ whose normalized form is } A_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},$$

so that the second principal component is

$$u_2 = \frac{1}{\sqrt{2}}[x_1 + x_3]. \quad (iii)$$

It can readily be checked that  $A_2' V A_2 = 3 = \lambda_2$ . Let us now obtain an eigenvector corresponding to the eigenvalue  $\lambda_3 = 3 - \sqrt{2}$ . Consider the linear system  $(V - \lambda_3)X = O$ . The first equation is  $\sqrt{2}x_1 - x_2 = 0 \Rightarrow x_2 = \sqrt{2}x_1$ , and the third one gives  $x_2 + \sqrt{2}x_3 = 0 \Rightarrow x_2 = -\sqrt{2}x_3$ . One solution is

$$X_3 = \begin{bmatrix} 1 \\ \sqrt{2} \\ -1 \end{bmatrix}, \text{ its normalized form being } A_3 = \frac{1}{2} \begin{bmatrix} 1 \\ \sqrt{2} \\ -1 \end{bmatrix}.$$

The third principal component is then

$$u_3 = \frac{1}{2}[x_1 + \sqrt{2}x_2 - x_3]. \quad (iv)$$

It is easily verified that  $A_3' V A_3 = \lambda_3 = 3 - \sqrt{2}$ . As well,

$$\begin{aligned} \text{Var}(u_1) &= \frac{1}{4}[\text{Var}(x_1) + 2 \text{Var}(x_2) + \text{Var}(x_3) - 2\sqrt{2} \text{Cov}(x_1, x_2) \\ &\quad - 2 \text{Cov}(x_1, x_3) + 2\sqrt{2} \text{Cov}(x_2, x_3)] \\ &= \frac{1}{4}[3 + 2 \times 3 + 3 + 2\sqrt{2}(-1) - 2(0) + 2\sqrt{2}] \\ &= \frac{1}{4}[3 \times 4 + 4\sqrt{2}] = 3 + \sqrt{2} = \lambda_1. \end{aligned}$$

Similar calculations will confirm that  $\text{Var}(u_2) = 3 = \lambda_2$  and  $\text{Var}(u_3) = 2 - \sqrt{2} = \lambda_3$ . Now, consider the covariance between  $u_1$  and  $u_2$ :

$$\begin{aligned} \text{Cov}(u_1, u_2) &= \frac{1}{\sqrt{2}}[\text{Var}(x_1) + \text{Cov}(x_1, x_3) - \sqrt{2} \text{Cov}(x_1, x_2) - \sqrt{2} \text{Cov}(x_2, x_3) \\ &\quad - \text{Cov}(x_3, x_1) - \text{Var}(x_3)] \\ &= \frac{1}{4}[3 + 0 + \sqrt{2} - \sqrt{2} - 0 - 0 - 3] = 0. \end{aligned}$$

It can be likewise verified that  $\text{Cov}(u_1, u_3) = 0$  and  $\text{Cov}(u_2, u_3) = 0$ . Note that  $u_1$ ,  $u_2$ , and  $u_3$  respectively account for 49.1%, 33.3% and 17.6% of the total variation. As none of these proportions is negligibly small, all three principal components are deemed relevant and, in this instance, it is not indicated to reduce the number of the original variables. Although we still end up with as many variables, the  $u_j$ 's,  $j = 1, 2, 3$ , are uncorrelated, which was not the case for the original variables.

### 9.3. Issues to Be Mindful of when Constructing Principal Components

Since variances and covariances are expressed in units of measurement, principal components also depend upon the scale on which measurements on the individual variables are made. If we change the units of measurement, the principal components will differ. Suppose that  $x_j$  is multiplied by a real scalar constant  $d_j$ ,  $j = 1, \dots, p$ , where some of the  $d_j$ 's are not equal to 1. This is equivalent to changing the units of measurement of some of the variables. Let

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}, Y = \begin{bmatrix} d_1 x_1 \\ \vdots \\ d_p x_p \end{bmatrix} = DX, D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_p \end{bmatrix}, \quad (9.3.1)$$

and consider the linear functions  $A'X$  and  $A'Y$ . Then,

$$\text{Var}(A'X) = A'\Sigma A, \quad \text{Var}(A'Y) = A'\text{Var}(DX)A = A'D\Sigma DA. \quad (9.3.2)$$

Since the eigenvalues of  $\Sigma$  and  $D\Sigma D$  differ, so will the corresponding principal components, and if the original variables are measured in various units of measurements, it would be advisable to attempt to standardize them. Letting  $R$  denote the correlation matrix which is scale-invariant, observe that the covariance matrix  $\Sigma = \Sigma_1 R \Sigma_1$  where  $\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_p)$ ,  $\sigma_j^2$  being the variance of  $x_j$ ,  $j = 1, \dots, p$ . Thus, if the original variables are scaled by the inverses of their associated standard deviations, that is,  $\frac{x_j}{\sigma_j}$  or equivalently, via the transformation  $\Sigma_1^{-1}X$ , the resulting covariance matrix is  $\text{Cov}(\Sigma_1^{-1}X) = R$ , the correlation matrix. Accordingly, constructing the principal components by making use of  $R$  instead of  $\Sigma$ , will mitigate the issue stemming from the scale of measurement.

If  $\lambda_1, \dots, \lambda_p$  are the eigenvalues of  $\Sigma$ , then  $\lambda_1^k, \dots, \lambda_p^k$  will be the eigenvalues of  $\Sigma^k$ . Moreover,  $\lambda$  and  $\lambda^k$  will share the same eigenvectors. Note that the collection  $(\lambda_1^k, \dots, \lambda_p^k)$  will be well separated compared to the set  $(\lambda_1, \dots, \lambda_p)$  when the  $\lambda_i$ 's are distinct and greater than one. Hence, in some instances, it might be preferable to construct principal



components by making use of  $\Sigma^k$  for an appropriate value of  $k$  instead of  $\Sigma$ . Observe that, in certain situations, it is not feasible to provide a physical interpretation of the principal components which are a linear function of the original  $x_1, \dots, x_p$ . Nonetheless, they can at times be informative by pointing to the average of certain variables (for example,  $u_2$  in the previous numerical example) or by eliciting contrasts between two sets of variables (for example,  $u_3$  in the previous numerical example, which opposes  $x_3$  to  $x_1$  and  $x_2$ ).

To illustrate how the eigenvalues of a correlation matrix are determined, we revisit Example 9.2.1 wherein the variances of the  $x_j$ 's or the diagonal elements of the covariance matrix are all equal to 3. Thus, in this case, the correlation matrix is  $R = \frac{1}{3}V$ , and the eigenvalues of  $R$  will be  $\frac{1}{3}$  times the eigenvalues of  $V$ . However, the normalized eigenvectors will remain identical to those of  $V$ , and therefore the principal components will not change. We now consider an example wherein the diagonal elements of the covariance matrix are different.

**Example 9.3.1.** Let  $X' = (x_1, x_2, x_3)$  where the  $x_j$ 's are real scalar random variables. Compute the principal components resulting from  $R$ , the correlation matrix of  $X$ , where

$$R = \begin{bmatrix} 1 & 0 & \frac{1}{\sqrt{6}} \\ 0 & 1 & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & 1 \end{bmatrix}.$$

**Solution 9.3.1.** Let us compute the eigenvalues of  $R$ . Consider the equation

$$\begin{aligned} |R - \lambda I| = 0 &\Rightarrow (1 - \lambda) \left[ (1 - \lambda)^2 - \frac{1}{6} \right] - \frac{1}{\sqrt{6}} \left[ -\frac{1}{\sqrt{6}}(1 - \lambda) \right] = 0 \\ &\Rightarrow (1 - \lambda) \left[ (1 - \lambda)^2 - \frac{1}{3} \right] = 0. \end{aligned}$$

Thus, the eigenvalues are  $\lambda_1 = 1 + \frac{1}{\sqrt{3}}$ ,  $\lambda_2 = 1$ ,  $\lambda_3 = 1 - \frac{1}{\sqrt{3}}$ . Let us compute an eigenvector corresponding to  $\lambda_1 = 1 + \frac{1}{\sqrt{3}}$ . Consider the system  $(R - \lambda_1 I)X = O$ . The first equation is

$$-\frac{1}{\sqrt{3}}x_1 + \frac{1}{\sqrt{6}}x_3 = 0 \Rightarrow x_1 = \frac{1}{\sqrt{2}}x_3,$$

the second equation being

$$-\frac{1}{\sqrt{3}}x_2 - \frac{1}{\sqrt{6}}x_3 = 0 \Rightarrow x_2 = -\frac{1}{\sqrt{2}}x_3.$$

Let us take  $x_3 = \sqrt{2}$ . Then, an eigenvector corresponding to  $\lambda_1$  is

$$X_1 = \begin{bmatrix} 1 \\ -1 \\ \sqrt{2} \end{bmatrix}, \text{ and its normalized form is } A_1 = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ \sqrt{2} \end{bmatrix}.$$

Let us verify that the third equation is also satisfied by  $X_1$  or  $A_1$ . This is the case since  $\frac{1}{\sqrt{6}} + \frac{1}{\sqrt{6}} - \frac{\sqrt{2}}{\sqrt{3}} = 0$ , and the first principal component is indeed  $u_1 = \frac{1}{2}[x_1 - x_2 + \sqrt{2}x_3]$ . As well, the variance of  $u_1$  is equal to  $\lambda_1$ :

$$\begin{aligned} \text{Var}(u_1) &= \frac{1}{4}[\text{Var}(x_1) + \text{Var}(x_2) + 2\text{Var}(x_3) - 2\text{Cov}(x_1, x_2) \\ &\quad + 2\sqrt{2}\text{Cov}(x_1, x_3) - 2\sqrt{2}\text{Cov}(x_2, x_3)] \\ &= \frac{1}{4}\left[1 + 1 + 2 + \frac{2\sqrt{2}}{\sqrt{6}} + \frac{2\sqrt{2}}{\sqrt{6}}\right] = 1 + \frac{1}{\sqrt{3}} = \lambda_1. \end{aligned}$$

Let us compute an eigenvector corresponding to the eigenvalue  $\lambda_2 = 1$ . Consider the linear system  $(R - \lambda_2 I)X = O$ . The first equation,  $\frac{1}{\sqrt{6}}x_3 = 0$ , gives  $x_3 = 0$  and the second one also yields  $x_3 = 0$ ; as for the third one,  $\frac{x_1}{\sqrt{6}} - \frac{x_2}{\sqrt{6}} = 0 \Rightarrow x_1 = x_2$ . Letting  $x_1 = 1$ , an eigenvector is

$$X_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \text{ its normalized form being given by } A_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Thus, the second principal component is  $u_2 = \frac{1}{\sqrt{2}}[x_1 + x_2]$ . In the case of  $\lambda_3$ , we consider the system  $(R - \lambda_3 I)X = O$  whose first equation is

$$\frac{1}{\sqrt{3}}x_1 + \frac{1}{\sqrt{6}}x_3 = 0 \Rightarrow x_1 = -\frac{1}{\sqrt{2}}x_3,$$

the second one being

$$\frac{x_2}{\sqrt{3}} + \frac{x_3}{\sqrt{6}} = 0 \Rightarrow x_2 = \frac{1}{\sqrt{2}}x_3.$$

Letting  $x_3 = \sqrt{2}$ , an eigenvector corresponding to  $\lambda_3$  is

$$X_3 = \begin{bmatrix} -1 \\ 1 \\ \sqrt{2} \end{bmatrix}, \text{ its normalized form being } A_3 = \frac{1}{2} \begin{bmatrix} -1 \\ 1 \\ \sqrt{2} \end{bmatrix}.$$

For the matrix  $[A_1, A_2, A_3]$  to be uniquely determined, we may multiply  $A_3$  by  $-1$  so that the first nonzero element is positive. Hence, the third principal component is  $u_3 = \frac{1}{2}[x_1 - x_2 - \sqrt{2}x_3]$ . As was shown in the case of  $u_1$ , it can also be verified that  $\text{Var}(u_2) = 1 = \lambda_2$ ,  $\text{Var}(u_3) = \lambda_3 = 1 - \frac{1}{\sqrt{3}}$ ,  $\text{Cov}(u_1, u_2) = 0$ ,  $\text{Cov}(u_1, u_3) = 0$  and  $\text{Cov}(u_2, u_3) = 0$ . We note that

$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{1 + 1/\sqrt{3}}{3} \approx 0.525,$$

$$\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{2 + 1/\sqrt{3}}{3} \approx 0.859.$$

Thus, almost 53% of the total variation is accounted for by the first principal component and nearly 86% of the total variation is due to the first two principal components.

#### 9.4. The Vector of Principal Components

Observe that the determinant of a matrix is the product of its eigenvalues and that its trace is the sum of its eigenvalues, that is,  $|\Sigma| = \lambda_1 \cdots \lambda_p$  and  $\text{tr}(\Sigma) = \lambda_1 + \cdots + \lambda_p$ . As previously pointed out, the determinant of a covariance matrix corresponds to Wilks' concept of *generalized variance*. Let us consider the vector of principal components. The principal components are  $u_j = A'_j X$ , with  $\text{Var}(u_j) = A'_j \Sigma A_j = \lambda_j$ ,  $j = 1, \dots, p$ , and  $\text{Cov}(u_i, u_j) = 0$  for all  $i \neq j$ . Thus,

$$U = \begin{bmatrix} u_1 \\ \vdots \\ u_p \end{bmatrix} = \begin{bmatrix} A'_1 X \\ \vdots \\ A'_p X \end{bmatrix} \Rightarrow \text{Cov}(U) = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}. \quad (9.4.1)$$

The determinant of the covariance matrix is the product  $\lambda_1 \cdots \lambda_p$  and its trace, the sum  $\lambda_1 + \cdots + \lambda_p$ . Hence the following result:

**Theorem 9.4.1.** *Let  $X$  be a  $p \times 1$  real vector whose associated covariance matrix is  $\Sigma$ . Let the principal components of  $X$  be denoted by  $u_j = A'_j X$  with  $\text{Var}(u_j) = A'_j \Sigma A_j = \lambda_j = j$ -th largest eigenvalue of  $\Sigma$ , and  $U' = (u_1, \dots, u_p)$  with  $\text{Cov}(U) \equiv \Sigma_u$ . Then,  $|\Sigma_u| = |\Sigma| =$  product of the eigenvalues,  $\lambda_1 \cdots \lambda_p$ , and  $\text{tr}(\Sigma_u) = \text{tr}(\Sigma) =$  sum of the eigenvalues,  $\lambda_1 + \cdots + \lambda_p$ . Observe that the determinant as well as the eigenvalues and the trace are invariant with respect to orthonormal transformations or rotations of the coordinate axes.*

Let  $A = [A_1, A_2, \dots, A_p]$ . Then

$$\Sigma A = A\Lambda, \quad A'A = I, \quad A'\Sigma A = \Lambda. \quad (9.4.2)$$

Note that  $U' = (u_1, \dots, u_p)$  where  $u_1, \dots, u_p$  are linearly independent. We have not assumed any distribution for the  $p \times 1$  vector  $X$  so far. If the  $p \times 1$  vector  $X$  has a  $p$ -variate nonsingular Gaussian distribution, that is,  $X \sim N_p(O, \Sigma)$ ,  $\Sigma > O$ , then

$$u_j \sim N_1(0, \lambda_j), \quad U \sim N_p(O, \Lambda). \quad (9.4.3)$$

Since *Principal Components* involve variances or covariances, which are free of any location parameter, we may take the mean value vector to be a null vector without any loss of generality. Then,  $E[u_j] = 0$  by assumption,  $\text{Var}(u_j) = A'_j \Sigma A_j = \lambda_j$ ,  $j = 1, \dots, p$  and  $\text{Cov}(U) = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ . Accordingly,  $\frac{\lambda_1}{\lambda_1 + \dots + \lambda_p}$  is the proportion of the total variance accounted for by the largest eigenvalue  $\lambda_1$ , where  $\lambda_1$  is equal to the variance of the first principal component. Similarly,  $\frac{\lambda_1 + \dots + \lambda_r}{\lambda_1 + \dots + \lambda_p}$  is the proportion of the total variance due to the first  $r$  principal components,  $r \leq p$ .

**Example 9.4.1.** Let  $X \sim N_3(\mu, \Sigma)$ ,  $\Sigma > O$ , where

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mu = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & -1 \\ 1 & -1 & 3 \end{bmatrix}.$$

Derive the densities of the principal components of  $X$ .

**Solution 9.4.1.** Let us determine the eigenvalues of  $\Sigma$ . Consider the equation

$$\begin{aligned} |\Sigma - \lambda I| = 0 &\Rightarrow \begin{vmatrix} 2 - \lambda & 0 & 1 \\ 0 & 2 - \lambda & -1 \\ 1 & -1 & 3 - \lambda \end{vmatrix} = 0 \\ &\Rightarrow (2 - \lambda)[(2 - \lambda)(3 - \lambda) - 1] - (2 - \lambda) = 0 \\ &\Rightarrow (2 - \lambda)[\lambda^2 - 5\lambda + 4] = 0 \Rightarrow \lambda_1 = 4, \lambda_2 = 2, \lambda_3 = 1. \end{aligned}$$

Let us compute an eigenvector corresponding to  $\lambda_1 = 4$ . Consider the linear system  $(\Sigma - \lambda_1 I)X = O$ , whose first equation gives  $-2x_1 + x_3 = 0$  or  $x_1 = \frac{1}{2}x_3$ , the second equation,  $-2x_2 - x_3 = 0$ , yielding  $x_2 = -\frac{1}{2}x_3$ . Letting  $x_3 = 2$ , one solution is

$$X_1 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \quad \text{and its normalized form is } A_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}.$$

Thus, the first principal component is  $u_1 = A'_1 X = \frac{1}{\sqrt{6}}[x_1 - x_2 + 2x_3]$  with  $E[u_1] = \frac{1}{\sqrt{6}}[1 - 0 - 2] = -\frac{1}{\sqrt{6}}$  and

$$\begin{aligned}\text{Var}(u_1) &= \frac{1}{6}[\text{Var}(x_1) + \text{Var}(x_2) + 4\text{Var}(x_3) - 2\text{Cov}(x_1, x_2) \\ &\quad + 4\text{Cov}(x_1, x_3) - 4\text{Cov}(x_2, x_3)] \\ &= \frac{1}{6}[2 + 2 + 4 \times 3 - 2(0) + 4(1) - 4(-1)] = 4 = \lambda_1.\end{aligned}$$

Since  $u_1$  is a linear function of normal variables,  $u_1$  has the following Gaussian distribution:

$$u_1 \sim N_1\left(-\frac{1}{\sqrt{6}}, 4\right).$$

Let us compute an eigenvector corresponding to  $\lambda_2 = 2$ . Consider the system  $(\Sigma - \lambda_2 I)X = O$ , whose first and second equations give  $x_3 = 0$ , the third equation  $x_1 - x_2 + x_3 = 0$  yielding  $x_1 = x_2$ . Hence, one solution is

$$X_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \text{ its normalized form being } A_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix},$$

and the second principal component is  $u_2 = A'_2 X = \frac{1}{\sqrt{2}}[x_1 + x_2]$  with  $E[u_2] = \frac{1}{\sqrt{2}}[1 + 0] = \frac{1}{\sqrt{2}}$ . Let us verify that the variance of  $u_2$  is  $\lambda_2 = 2$ :

$$\text{Var}(u_2) = \frac{1}{2}[\text{Var}(x_1) + \text{Var}(x_2) + 2\text{Cov}(x_1, x_2)] = \frac{1}{2}[2 + 2 + 2(0)] = 2 = \lambda_2.$$

Hence,  $u_2$  has the following real univariate normal distribution:

$$u_2 \sim N_1\left(\frac{1}{\sqrt{2}}, 2\right).$$

We finally construct an eigenvector associated with  $\lambda_3 = 1$ . Consider the linear system  $(\Sigma - \lambda_3 I)X = O$ . In this case, the first equation is  $x_1 + x_3 = 0 \Rightarrow x_1 = -x_3$  and the second one is  $x_2 - x_3 = 0$  or  $x_2 = x_3$ . Let  $x_3 = 1$ . One eigenvector is

$$X_3 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \text{ its normalized form being } A_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}.$$

In order to ensure the uniqueness of the matrix  $[A_1, A_2, A_3]$ , we may multiply  $A_3$  by  $-1$  to ensure that the first nonzero element is positive. Then, the third principal component is  $u_3 = \frac{1}{\sqrt{3}}[x_1 - x_2 - x_3]$  with  $E[u_3] = \frac{1}{\sqrt{3}}[1 - 0 + 1] = \frac{2}{\sqrt{3}}$ . The variance of  $u_3$  is indeed equal to  $\lambda_3$  as

$$\begin{aligned}\text{Var}(u_3) &= \frac{1}{3}[\text{Var}(x_1) + \text{Var}(x_2) + \text{Var}(x_3) - 2\text{Cov}(x_1, x_2) \\ &\quad - 2\text{Cov}(x_1, x_3) + 2\text{Cov}(x_2, x_3)] \\ &= \frac{1}{3}[2 + 2 + 3 - 2(0) - 2(1) + 2(-1)] = 1.\end{aligned}$$

Thus,

$$u_3 \sim N_1\left(\frac{2}{\sqrt{3}}, 1\right).$$

It can easily be verified that  $\text{Cov}(u_1, u_2) = 0$ ,  $\text{Cov}(u_1, u_3) = 0$  and  $\text{Cov}(u_2, u_3) = 0$ . Accordingly, letting

$$U = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} A_1' X \\ A_2' X \\ A_3' X \end{bmatrix},$$

$$\text{Cov}(U) = \begin{bmatrix} A_1' \Sigma A_1 & A_1' \Sigma A_2 & A_1' \Sigma A_3 \\ A_2' \Sigma A_1 & A_2' \Sigma A_2 & A_2' \Sigma A_3 \\ A_3' \Sigma A_1 & A_3' \Sigma A_2 & A_3' \Sigma A_3 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}.$$

As well,  $\Sigma A_j = \lambda_j A_j$ ,  $j = 1, 2, 3$ , that is,

$$\begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & -1 \\ 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This completes the computations.

#### 9.4.1. Principal components viewed from differing perspectives

Let  $X$  be a  $p \times 1$  real vector whose covariance matrix is  $\text{Cov}(X) = \Sigma > O$ . Assume that  $E(X) = O$ . Then,  $X' \Sigma^{-1} X = c > 0$  is commonly referred to as an ellipsoid of concentration, centered at the origin of the coordinate system, with  $X' X$  being the square of the distance between the origin and a point  $X$  on the surface of this ellipsoid. A principal axis of this ellipsoid is defined when this squared distance has a stationary point.

The stationary points are determined by optimizing  $X'X$  subject to  $X'\Sigma^{-1}X = c > 0$ . Consider

$$\begin{aligned} w = X'X - \lambda(X'\Sigma^{-1}X - c), \quad \frac{\partial w}{\partial X} = O &\Rightarrow 2X - 2\lambda\Sigma^{-1}X = O \\ \Rightarrow \Sigma^{-1}X = \frac{1}{\lambda}X &\Rightarrow \Sigma X = \lambda X \end{aligned} \quad (i)$$

where  $\lambda$  is a Lagrangian multiplier. It is seen from (i) that  $\lambda$  is an eigenvalue of  $\Sigma$  and  $X$  is the corresponding eigenvector. Letting  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$  be the eigenvalues and  $A_1, \dots, A_p$  be the corresponding eigenvectors,  $A_1, \dots, A_p$  give the principal axes of this ellipsoid of concentration. It follows from (i) that

$$c = A'_j \Sigma^{-1} A_j = \frac{1}{\lambda_j} A'_j A_j \Rightarrow A'_j A_j = \lambda_j c.$$

Thus, the length of the  $j$ -th principal axis is  $2\sqrt{A'_j A_j} = 2\sqrt{\lambda_j c}$ .

As another approach, consider a plane passing through the origin. The equation of this plane will be  $\beta'X = 0$  where  $\beta$  is a  $p \times 1$  constant vector and  $X$  is the  $p \times 1$  vector of the coordinate system. Without any loss of generality, let  $\beta'\beta = 1$ . Let the  $p \times 1$  vector  $Y$  be a point in the Euclidean space. The distance between this point and the plane is then  $\beta'Y$ . Letting  $Y$  be a random point such that  $E(Y) = O$  and  $\text{Cov}(Y) = \Sigma > O$ , the expected squared distance from this point to the plane is  $E[\beta'Y]^2 = E[\beta'YY'\beta] = \beta'E(YY')\beta = \beta'\Sigma\beta$ . Accordingly, the two-dimensional planar manifold of closest fit to the point  $Y$  is that plane whose coefficient vector  $\beta$  is such that  $\beta'\Sigma\beta$  is minimized subject to  $\beta'\beta = 1$ . This, once again, leads to the eigenvalue problem encountered in principal component analysis.

### 9.5. Sample Principal Components

When  $X \sim N_p(O, \Sigma)$ ,  $\Sigma > O$ , the maximum likelihood estimator of  $\Sigma$  is  $\hat{\Sigma} = \frac{1}{n}S$  where  $n > p$  is the sample size and  $S$  is the sample sum of products matrix, as defined for instance in Mathai and Haubold (2017b). In general, whether  $\Sigma$  is nonsingular or singular and  $X$  is  $p$ -variate Gaussian or not, we may take  $\hat{\Sigma}$  as an estimate of  $\Sigma$ . The sample eigenvalues and eigenvectors are then available from  $\hat{\Sigma}\beta = k\beta$  where  $\beta \neq O$  is a  $p \times 1$  eigenvector corresponding to the eigenvalue  $k$  of  $\hat{\Sigma} = \frac{1}{n}S$ . For a non-null  $\beta$ ,  $(\hat{\Sigma} - kI)\beta = O \Rightarrow \hat{\Sigma} - kI$  is singular or  $|\hat{\Sigma} - kI| = 0$  and  $k$  is a solution of

$$|\hat{\Sigma} - kI| = 0 \Rightarrow (\hat{\Sigma} - k_j I)B_j = O, \quad j = 1, \dots, p. \quad (9.5.1)$$

We may only consider the normalized eigenvectors  $B_j$  such that  $B'_j B_j = 1$ ,  $j = 1, \dots, p$ . If the eigenvalues of  $\Sigma$  are distinct, it can be shown that the eigenvalues of  $\hat{\Sigma}$  are also

distinct,  $k_1 > k_2 > \dots > k_p$  almost surely. Note that even though  $B_j' B_j = 1$ ,  $B_j$  is not unique as one also has  $(-B_j)'(-B_j) = 1$ . Thus, we require that the first nonzero element of  $B_j$  be positive to ensure uniqueness. Since  $\hat{\Sigma}$  is a real symmetric matrix, its eigenvalues  $k_1, \dots, k_p$  are real, and so are the corresponding normalized eigenvectors  $B_1, \dots, B_p$ . As well, there exist a full set of orthonormal eigenvectors  $B_1, \dots, B_p$ ,  $B_j' B_j = 1$ ,  $B_j' B_i = 0$ ,  $i \neq j$ ,  $j = 1, \dots, p$ , such that for  $B = (B_1, \dots, B_p)$ ,

$$B' \hat{\Sigma} B = K = \text{diag}(k_1, \dots, k_p), \quad \hat{\Sigma} B = B K \quad \text{and} \quad \hat{\Sigma} = k_1 B_1 B_1' + \dots + k_p B_p B_p'. \quad (9.5.2)$$

Also, observe that  $\frac{k_1}{k_1 + \dots + k_p}$  is the proportion of the total variation in the data which is accounted for by the first principal component. Similarly,  $\frac{k_1 + \dots + k_r}{k_1 + \dots + k_p}$  is the proportion of the total variation due to the first  $r$  principal components,  $r \leq p$ .

**Example 9.5.1.** Let  $X$  be a  $3 \times 1$  vector of real scalar random variables,  $X' = [x_1, x_2, x_3]$ , with  $E[X] = \mu$  and covariance matrix  $\text{Cov}(X) = \Sigma > O$  where both  $\mu$  and  $\Sigma$  are unknown. Let the following observation vectors be a simple random sample of size 5 from this population:

$$X_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad X_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad X_3 = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}, \quad X_4 = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}, \quad X_5 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}.$$

Compute the principal components of  $X$  from an estimate of  $\Sigma$  that is based on those observations.

**Solution 9.5.1.** An estimate of  $\Sigma$  is  $\hat{\Sigma} = \frac{1}{n} S$  where  $n$  is the sample size and  $S$  is the sample sum of products matrix. In this case,  $n = 5$ . We first compute  $S$ . To this end, we determine the sample average vector, the sample matrix, the deviation matrix and finally  $S$ . The sample average is  $\bar{X} = \frac{1}{n}[X_1 + X_2 + X_3 + X_4 + X_5]$  and the sample matrix is  $\mathbf{X} = [X_1, X_2, X_3, X_4, X_5]$ . The matrix of sample means is  $\bar{\mathbf{X}} = [\bar{X}, \bar{X}, \bar{X}, \bar{X}, \bar{X}]$ . The deviation matrix is  $\mathbf{X}_d = \mathbf{X} - \bar{\mathbf{X}}$  and the sample sum of products matrix  $S = \mathbf{X}_d \mathbf{X}_d' = [\mathbf{X} - \bar{\mathbf{X}}][\mathbf{X} - \bar{\mathbf{X}}]'$ . Based on the given random sample, these quantities are



$$\begin{aligned}\bar{X} &= \frac{1}{5} \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} \right\} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \\ \mathbf{X} &= [X_1, X_2, X_3, X_4, X_5] = \begin{bmatrix} 1 & -1 & 1 & -1 & 0 \\ 1 & 1 & -1 & -1 & 0 \\ 0 & 0 & -1 & -1 & 2 \end{bmatrix}, \\ \mathbf{X}_d &= [X_1 - \bar{X}, X_2 - \bar{X}, X_3 - \bar{X}, X_4 - \bar{X}, X_5 - \bar{X}] \\ &= \begin{bmatrix} 1 & -1 & 1 & -1 & 0 \\ 1 & 1 & -1 & -1 & 0 \\ 0 & 0 & -1 & -1 & 2 \end{bmatrix}, \\ S &= \begin{bmatrix} 1 & -1 & 1 & -1 & 0 \\ 1 & 1 & -1 & -1 & 0 \\ 0 & 0 & -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 & -1 & 0 \\ 1 & 1 & -1 & -1 & 0 \\ 0 & 0 & -1 & -1 & 2 \end{bmatrix}' \\ &= \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 6 \end{bmatrix}.\end{aligned}$$

An estimate of  $\Sigma$  is  $\hat{\Sigma} = \frac{1}{n}S = \frac{1}{5}S$ . However, since the eigenvalues of  $\hat{\Sigma}$  are those of  $S$  multiplied by  $\frac{1}{5}$  and the normalized eigenvectors of  $\hat{\Sigma}$  and  $S$  will then be identical, we will work with  $S$ . The eigenvalues of  $S$  are available from the equation  $|S - \lambda I| = 0$ . That is,  $(4 - \lambda)[(4 - \lambda)(6 - \lambda) - 4] = 0 \Rightarrow (4 - \lambda)[\lambda^2 - 10\lambda + 20] = 0 \Rightarrow \lambda_1 = 5 + \sqrt{5}$ ,  $\lambda_2 = 4$ ,  $\lambda_3 = 5 - \sqrt{5}$ . An eigenvector corresponding to  $\lambda_1 = 5 + \sqrt{5}$  can be determined from the system  $(S - \lambda_1 I)X = O$  wherein first equation gives  $x_1 = 0$  and the third one yields  $2x_2 + (1 - \sqrt{5})x_3 = 0$ . Taking  $x_3 = 2$ ,  $x_2 = -(1 - \sqrt{5})$ , and it is easily verified that these values also satisfy the second equation. Thus, an eigenvector, denoted by  $Y_1$ , is the following:

$$Y_1 = \begin{bmatrix} 0 \\ \sqrt{5} - 1 \\ 2 \end{bmatrix} \text{ with its normalized form being } A_1 = \frac{1}{\sqrt{10 - 2\sqrt{5}}} \begin{bmatrix} 0 \\ \sqrt{5} - 1 \\ 2 \end{bmatrix},$$

so that the first principal component is  $u_1 = \frac{1}{\sqrt{10 - 2\sqrt{5}}}[(\sqrt{5} - 1)x_2 + 2x_3]$ . Let us verify that the variance of  $u_1$  equals  $\lambda_1$ :

$$\begin{aligned}
\text{Var}(u_1) &= \frac{1}{10 - 2\sqrt{5}} [(\sqrt{5} - 1)^2 \text{Var}(x_2) + 4\text{Var}(x_3) + 2(\sqrt{5} - 1)\text{Cov}(x_2, x_3)] \\
&= \frac{1}{10 - 2\sqrt{5}} [(6 - 2\sqrt{5})(4) + 4(6) + 2(\sqrt{5} - 1)(4)] = \frac{20(5 + \sqrt{5})}{20} \\
&= 5 + \sqrt{5} = \lambda_1.
\end{aligned}$$

An eigenvector corresponding to  $\lambda_2 = 4$  is available from  $(S - \lambda_2 I)X = O$ . The first equation shows that all variables are arbitrary. The second and third equations yield  $x_3 = 0$  and  $x_2 = 0$ . Taking  $x_1 = 1$ , an eigenvector corresponding to  $\lambda_2 = 4$ , denoted by  $Y_2$ , is

$$Y_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ which is already normalized so that } A_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Thus, the second principal component is  $u_2 = x_1$  and its variance is  $\text{Var}(x_1) = 4$ . An eigenvector corresponding to  $\lambda_3 = 5 - \sqrt{5}$  can be determined from the linear system  $(S - \lambda_3 I)X = O$  whose first equation yields  $x_1 = 0$ , the third one giving  $2x_2 + (1 + \sqrt{5})x_3 = 0$ . Taking  $x_3 = 2$ ,  $x_2 = -(1 + \sqrt{5})$ , and it is readily verified that these values satisfy the second equation as well. Then, an eigenvector, denoted by  $Y_3$ , is

$$Y_3 = \begin{bmatrix} 0 \\ -(\sqrt{5} + 1) \\ 2 \end{bmatrix}, \text{ its normalized form being } A_3 = \frac{1}{\sqrt{10 + 2\sqrt{5}}} \begin{bmatrix} 0 \\ -(\sqrt{5} + 1) \\ 2 \end{bmatrix}.$$

In order to select the matrix  $[A_1, A_2, A_3]$  uniquely, we may multiply  $A_3$  by  $-1$  so that the first nonzero element is positive. Thus, the third principal component is  $u_3 = \frac{1}{\sqrt{10 + 2\sqrt{5}}} [(1 + \sqrt{5})x_2 - 2x_3]$  and, as the following calculations corroborate, its variance is indeed equal to  $\lambda_3$ :

$$\begin{aligned}
\text{Var}(u_3) &= \frac{1}{\sqrt{10 + 2\sqrt{5}}} [(1 + \sqrt{5})^2 \text{Var}(x_2) + 4\text{Var}(x_3) - 4(1 + \sqrt{5})\text{Cov}(x_2, x_3)] \\
&= \frac{1}{2(5 + \sqrt{5})} [4(6 + 2\sqrt{5}) + 4(6) - 4(1 + \sqrt{5})(2)] = \frac{20}{5 + \sqrt{5}} = 5 - \sqrt{5} = \lambda_3.
\end{aligned}$$

Additionally,

$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{5 + \sqrt{5}}{14} \approx 0.52,$$

that is, approximately 52% of the total variance is due to  $u_1$  or nearly 52% of the total variation in the data is accounted for by the first principal component. Also observe that

$$S[A_1, A_2, A_3] = [A_1, A_2, A_3]D, \quad D = \text{diag}(5 + \sqrt{5}, 4, 5 - \sqrt{5}) \text{ or}$$

$$S = [A_1, A_2, A_3]D \begin{bmatrix} A'_1 \\ A'_2 \\ A'_3 \end{bmatrix}.$$

That is,

$$\begin{aligned} S &= \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 6 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & 0 \\ \frac{\sqrt{5}-1}{\sqrt{10-2\sqrt{5}}} & 0 & -\frac{(1+\sqrt{5})}{\sqrt{10+2\sqrt{5}}} \\ \frac{2}{\sqrt{10-2\sqrt{5}}} & 0 & \frac{2}{\sqrt{10+2\sqrt{5}}} \end{bmatrix} \begin{bmatrix} 5 + \sqrt{5} & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 5 - \sqrt{5} \end{bmatrix} \begin{bmatrix} 0 & \frac{\sqrt{5}-1}{\sqrt{10-2\sqrt{5}}} & \frac{2}{\sqrt{10-2\sqrt{5}}} \\ 1 & 0 & 0 \\ 0 & -\frac{(1+\sqrt{5})}{\sqrt{10+2\sqrt{5}}} & \frac{2}{\sqrt{10+2\sqrt{5}}} \end{bmatrix}, \end{aligned}$$

which completes the computations.

### 9.5.1. Estimation and evaluation of the principal components

If  $X \sim N_p(O, \Sigma)$ ,  $\Sigma > O$ , then the maximum likelihood estimator of  $\Sigma$ , denoted by  $\hat{\Sigma} = \frac{1}{n}S$  where  $n$  is the sample size and  $S$  is the sample sum of products matrix. How can one evaluate the eigenvalues and eigenvectors in order to construct the principal components? One method consists of solving the polynomial equation  $|\Sigma - \lambda I| = 0$  for the population eigenvalues  $\lambda_1, \dots, \lambda_p$ , or  $|\hat{\Sigma} - kI| = 0$  for obtaining the sample eigenvalues  $k_1 \geq k_2 \geq \dots \geq k_p$ . Direct evaluation of  $k$  by solving the determinantal equation is not difficult when  $p$  is small. However, for large values of  $p$ , one has to resort to mathematical software or some iterative process. We will illustrate such an iterative process for the population values  $\lambda_1 > \lambda_2 > \dots > \lambda_p$  and the corresponding normalized eigenvectors  $A_1, \dots, A_p$  (using our notations). Let us assume that the eigenvalues are distinct. Consider the following equation for determining the eigenvalues and eigenvectors:

$$\Sigma A_j = \lambda_j A_j, \quad j = 1, \dots, k. \quad (9.5.3)$$

Take any initial  $p$ -vector  $W_o$  that is not orthogonal to  $A_1$ , the eigenvector corresponding to the largest eigenvalue  $\lambda_1$ . If  $W_o$  is orthogonal to  $A_1$ , then, in the iterative process,  $W_j$  will not reach  $A_1$ . Let  $Y_o = \frac{1}{\sqrt{W'_o W_o}} W_o$  be the normalized  $W_o$ . Consider the equation

$$\Sigma Y_{j-1} = W_j, \quad \text{that is, } \Sigma Y_0 = W_1, \quad \Sigma Y_1 = W_2, \dots, \quad Y_j = \frac{1}{\sqrt{(W'_j W_j)}} W_j, \quad j = 1, \dots. \quad (9.5.4)$$

Halt the iteration when  $W_j$  approximately agrees with  $W_{j-1}$ , that is,  $W_j$  converges to some  $W$  which will then be  $\lambda_1 A_1$  or  $Y_j$  converges to  $Y_1$ . At each stage, compute the quadratic form  $\delta_j = Y_j' \Sigma Y_j$  and make sure that  $\delta_j$  is increasing. Suppose that  $W_j$  converges to some  $W$  for certain values of  $j$  or as  $j \rightarrow \infty$ . At that stage, the equation is  $\Sigma Y = W$  where  $Y = \frac{1}{\sqrt{W'W}} W$ , the normalized  $W$ . Then, the equation is  $\Sigma Y = \sqrt{W'W} Y$ . In other words,  $\sqrt{W'W} = \lambda_1$  and  $Y = A_1$ , which are the largest eigenvalue of  $\Sigma$  and the corresponding eigenvector  $A_1$ . That is,

$$\lim_{j \rightarrow \infty} \sqrt{W_j' W_j} = \lambda_1$$

$$\lim_{j \rightarrow \infty} \left[ \frac{1}{\sqrt{(W_j' W_j)}} W_j \right] = \lim_{j \rightarrow \infty} Y_j = A_1.$$

The rate of convergence in (9.5.4) depends upon the ratio  $\frac{\lambda_2}{\lambda_1}$ . If  $\lambda_2$  is close to  $\lambda_1$ , then the convergence will be very slow. Hence, the larger the difference between  $\lambda_1$  and  $\lambda_2$ , the faster the convergence. It is thus indicated to raise  $\Sigma$  to a suitable power  $m$  and initiate the iteration on this  $\Sigma^m$  so that the difference between the resulting  $\lambda_1$  and  $\lambda_2$  be magnified accordingly,  $\lambda_j^m$ ,  $j = 1, \dots, p$ , being the eigenvalues of  $\Sigma^m$ . If an eigenvalue is equal to one, then  $\Sigma$  must first be multiplied by a constant so that all the resulting eigenvalues of  $\Sigma$  are well separated, which will not affect the eigenvectors. As well, observe that  $\Sigma$  and  $\Sigma^m$ ,  $m = 1, 2, \dots$ , share the same eigenvectors even though the eigenvalues are  $\lambda_j$  and  $\lambda_j^m$ ,  $j = 1, \dots, p$ , respectively; in other words, the normalized eigenvectors remain the same. After obtaining the largest eigenvalue  $\lambda_1$  and the corresponding normalized eigenvector  $A_1$ , consider

$$\Sigma_2 = \Sigma - \lambda_1 A_1 A_1', \quad \Sigma = \lambda_1 A_1 A_1' + \lambda_2 A_2 A_2' + \dots + \lambda_p A_p A_p',$$

where  $A_j$  is the column eigenvector corresponding to  $\lambda_j$  so that  $A_j A_j'$  is a  $p \times p$  matrix for  $j = 1, 2, \dots, p$ . Now, carry out the iteration on  $\Sigma_2$ , as was previously done on  $\Sigma$ . This will produce  $\lambda_2$  and the corresponding normalized eigenvector  $A_2$ . Note that  $\lambda_2$  is the largest eigenvalue of  $\Sigma_2$ . Next, consider

$$\Sigma_3 = \Sigma_2 - \lambda_2 A_2 A_2'$$

and continue this process until all the required eigenvectors are obtained. Similarly, for small  $p$ , the sample eigenvalues  $k_1 \geq k_2 \geq \dots \geq k_p$  are available by solving the equation  $|\hat{\Sigma} - kI| = 0$ ; otherwise, the sample eigenvalues and eigenvectors can be computed by applying the iterative process described above with  $\Sigma$  replaced by  $\hat{\Sigma}$ ,  $\lambda_j$  interchanged with  $k_j$  and  $B_j$  substituted to the eigenvector  $A_j$ ,  $j = 1, \dots, p$ .

**Example 9.5.2.** Principal component analysis is especially called for when the number of variables involved is large. For the sake of illustration, we will take  $p = 2$ . Consider the following  $2 \times 2$  covariance matrix

$$\Sigma = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} = \text{Cov}(X), \quad X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Construct the principal components associated with this matrix while working with the symbolic representations of the eigenvalues  $\lambda_1$  and  $\lambda_2$  rather than their actual values.

**Solution 9.5.2.** Let us first evaluate the eigenvalues of  $\Sigma$  in the customary fashion. Consider

$$\begin{aligned} |\Sigma - \lambda I| = 0 &\Rightarrow (2 - \lambda)(1 - \lambda) - 1 = 0 \Rightarrow \lambda^2 - 3\lambda + 1 = 0 & (i) \\ &\Rightarrow \lambda_1 = \frac{1}{2}[3 + \sqrt{5}], \quad \lambda_2 = \frac{1}{2}[3 - \sqrt{5}]. \end{aligned}$$

Let us compute an eigenvector corresponding to the largest eigenvalue  $\lambda_1$ . Then,

$$\begin{aligned} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \lambda_1 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Rightarrow \\ (2 - \lambda_1)x_1 - x_2 &= 0 \\ -x_1 + (1 - \lambda_1)x_2 &= 0. & (ii) \end{aligned}$$

Since the system is singular, we need only solve one of these equations. Letting  $x_2 = 1$  in (ii),  $x_1 = (1 - \lambda_1)$ . For illustrative purposes, we will complete the remaining steps with the general parameters  $\lambda_1$  and  $\lambda_2$  rather than their numerical values. Hence, one eigenvector is

$$C = \begin{bmatrix} 1 - \lambda_1 \\ 1 \end{bmatrix}, \quad \|C\| = \sqrt{1 + (1 - \lambda_1)^2} \Rightarrow A_1 = \frac{1}{\|C\|}C = \frac{1}{\sqrt{1 + (1 - \lambda_1)^2}} \begin{bmatrix} 1 - \lambda_1 \\ 1 \end{bmatrix}.$$

Thus, the principal components are the following:

$$\begin{aligned} u_1 &= \frac{1}{\|C\|}C'X = \frac{1}{\sqrt{1 + (1 - \lambda_1)^2}}\{(1 - \lambda_1)x_1 + x_2\} \\ u_2 &= \frac{1}{\sqrt{1 + (1 - \lambda_2)^2}}\{(1 - \lambda_2)x_1 + x_2\}. \end{aligned}$$

Let us verify that  $\text{Var}(u_1) = \lambda_1$  and  $\text{Var}(u_2) = \lambda_2$ . Note that

$$\|C\|^2 = 1 + (1 - \lambda_1)^2 = \lambda_1^2 - 2\lambda_1 + 2 = (\lambda_1^2 - 3\lambda_1 + 1) + \lambda_1 + 1 = \lambda_1 + 1, \quad (iii)$$

given the characteristic equation (i). However,

$$\begin{aligned}
 \text{Var}(C'X) &= \text{Var}((1 - \lambda_1)x_1 + x_2) \\
 &= (1 - \lambda_1)^2 \text{Var}(x_1) + \text{Var}(x_2) + 2(1 - \lambda_1)\text{Cov}(x_1, x_2) \\
 &= 2(1 - \lambda_1)^2 + 1 - 2(1 - \lambda_1) \\
 &= 2\lambda_1^2 - 2\lambda_1 + 1 = 2(\lambda_1^2 - 3\lambda_1 + 1) + 4\lambda_1 - 1 = 4\lambda_1 - 1, \quad (iv)
 \end{aligned}$$

given (i), the characteristic (or starting) equation. We now have  $\text{Var}(C'X) = \text{Var}((1 - \lambda_1)x_1 + x_2) = 4\lambda_1 - 1$  from (iv) and  $\|C\| = \lambda_1 + 1$  from (iii). Hence, we must have  $\text{Var}(C'X) = \|C\|^2\lambda_1 = [1 + (1 - \lambda_1)^2]\lambda_1 = (\lambda_1 + 1)\lambda_1$  from (iii). Since

$$\lambda_1(\lambda_1 + 1) = \lambda_1^2 + \lambda_1 = (\lambda_1^2 - 3\lambda_1 + 1) + 4\lambda_1 - 1 = 4\lambda_1 - 1,$$

agrees with (iv), the result is verified for  $u_1$ . Moreover, on replacing  $\lambda_1$  by  $\lambda_2$ , we have  $\text{Var}(u_2) = \lambda_2$ .

### 9.5.2. L1 and L2 norm principal components

For any  $p \times 1$  vector  $Y$ ,  $Y' = (y_1, \dots, y_p)$ , where  $y_1, \dots, y_p$  are real quantities, the L2 and L1 norms are respectively defined as follows:

$$\|Y\|_2 = (y_1^2 + \dots + y_p^2)^{\frac{1}{2}} = [Y'Y]^{\frac{1}{2}} \Rightarrow \|Y\|_2^2 = Y'Y \quad (9.5.5)$$

$$\|Y\|_1 = |y_1| + |y_2| + \dots + |y_p| = \sum_{j=1}^p |y_j|. \quad (9.5.6)$$

In Sect. 9.5, we set up the principal component analysis on the basis of the sample sum of products matrix, assuming a  $p$ -variate real population, not necessarily Gaussian, having  $\mu$  as its mean value vector and  $\Sigma > O$  as its covariance matrix. Then, we considered  $X_j$ ,  $j = 1, \dots, n$ , iid vector random variables, that is, a simple random sample of size  $n$  from this population such that  $E[X_j] = \mu$  and  $\text{Cov}(X_j) = \Sigma > O$ ,  $j = 1, \dots, n$ . We denoted the sample matrix by  $\mathbf{X} = [X_1, \dots, X_n]$ , the sample average by  $\bar{X} = \frac{1}{n}[X_1 + \dots + X_n]$ , the matrix of sample means by  $\bar{\mathbf{X}} = [\bar{X}, \dots, \bar{X}]$  and the sample sum of products matrix by  $S = [\mathbf{X} - \bar{\mathbf{X}}][\mathbf{X} - \bar{\mathbf{X}}]'$ . If the population mean value vector is the null vector, that is,  $\mu = O$ , then we can take  $S = \mathbf{X}\mathbf{X}'$ . For convenience, we will consider this case. For determining the sample principal components, we then maximized

$$A'\mathbf{X}\mathbf{X}'A \text{ subject to } A'A = 1$$

where  $A$  is an arbitrary constant vector that will result in the coefficient vector of the principal components. This can also be stated in terms of maximizing the square of an  $L2$  norm subject to  $A'A = 1$ , that is,

$$\max_{A'A=1} \|A'X\|_2^2. \quad (9.5.7)$$

When carrying out statistical analyses, it turns out that the  $L2$  norm is more sensitive to outliers than the  $L1$  norm. If we utilize the  $L1$  norm, the problem corresponding to (9.5.7) is

$$\max_{A'A=1} \|A'X\|_1. \quad (9.5.8)$$

Observe that when  $\mu = O$ , the sum of products matrix can be expressed as

$$S = \mathbf{X}\mathbf{X}' = \sum_{j=1}^p X_j X_j', \quad \mathbf{X} = [X_1, \dots, X_n],$$

where  $X_j$  is the  $j$ -th column of  $\mathbf{X}$  or  $j$ -th sample vector. Then, the initial optimization problem can be restated as follows:

$$\max_{A'A=1} \|A'X\|_2^2 = \max_{A'A=1} \sum_{j=1}^n A'X_j X_j' A = \max_{A'A=1} \sum_{j=1}^n \|A'X_j\|_2^2, \quad (9.5.9)$$

and the corresponding  $L1$  norm optimization problem can be formulated as follows:

$$\max_{A'A=1} \|A'X\|_1 = \max_{A'A=1} \sum_{j=1}^n \|A'X_j\|_1. \quad (9.5.10)$$

We have obtained exact analytical solutions for the coefficient vector  $A$  in (9.5.9); however, this is not possible when attempting to optimize (9.5.10) subject to  $A'A = 1$ . Thankfully, iterative procedures are available in this case.

Let us consider a generalization of the basic principal component analysis. Let  $W$  be a  $p \times m$ ,  $m < p$ , matrix of full rank  $m$ . Then, the general problem in  $L2$  norm optimization is the following:

$$\max_{W'W=I} \text{tr}(W'SW) = \max_{W'W=I} \sum_{j=1}^n \|W'X_j\|_2^2 \quad (9.5.11)$$

where  $I$  is the  $m \times m$  identity matrix, the corresponding  $L1$  norm optimization problem being

$$\max_{W'W=I} \sum_{j=1}^n \|W'X_j\|_1. \quad (9.5.12)$$

In Sect. 9.4.1, we have considered a dual problem of minimization for constructing principal components. The dual problems of minimization corresponding to (9.5.11) and (9.5.12) can be stated as follows:

$$\min_{W'W=I} \sum_{j=1}^n \|X_j - WW'X_j\|_2^2 \quad (9.5.13)$$

with respect to the  $L2$  norm, and

$$\min_{W'W=I} \sum_{j=1}^n \|X_j - WW'X_j\|_1 \quad (9.5.14)$$

with respect to the  $L1$  norm. The form appearing in (9.5.13) suggests that the construction of principal components by making use of the  $L2$  norm is also connected to general model building, Factor Analysis and related topics. The general mathematical problem pertaining to the basic structure in (9.5.13) is referred to as low-rank matrix factorization. Readers interested in such statistical or mathematical problems may refer to the survey article, Shi et al. (2017).  $L1$  norm optimization problems are for instance discussed in Kwak (2008) and Nie and Huang (2016).

## 9.6. Distributional Aspects of Eigenvalues and Eigenvectors

Let us examine the distributions of the variances of the sample principal components and the coefficient vectors in the sample principal components. Let  $A$  be a  $p \times p$  orthonormal matrix whose columns are denoted by  $A_1, \dots, A_p$ , so that  $A_j'A_j = 1$ ,  $A_i'A_j = 0$ ,  $i \neq j$ , or  $AA' = I$ ,  $A'A = I$ . Let  $u_j = A_j'X$  be the sample principal components for  $j = 1, \dots, p$ . Let the  $p \times 1$  vector  $X$  have a nonsingular normal density with the null vector as its mean value vector, that is,  $X \sim N_p(O, \Sigma)$ ,  $\Sigma > O$ . We can assume that the Gaussian distribution has a null mean value vector without any loss of generality since we are dealing with variances and covariances, which are free of any location parameter. Consider a simple random sample of size  $n$  from this normal population. Letting  $S$  denote the sample sum of products matrix, the maximum likelihood estimate of  $\Sigma$  is  $\hat{\Sigma} = \frac{1}{n}S$ . Given that  $S$  has a Wishart distribution having  $n - 1$  degrees of freedom, the density of  $\hat{\Sigma}$ , denoted by  $f(\hat{\Sigma})$ , is given by

$$f(\hat{\Sigma}) = \frac{n^{\frac{mp}{2}}}{|2\Sigma|^{\frac{m}{2}} \Gamma_p(\frac{m}{2})} |\hat{\Sigma}|^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{n}{2}\text{tr}(\Sigma^{-1}\hat{\Sigma})},$$

where  $\Sigma$  is the population covariance matrix and  $m = n - 1$ ,  $n$  being the sample size. Let  $k_1, \dots, k_p$  be the eigenvalues of  $\hat{\Sigma}$ ; it was shown that  $k_1 \geq k_2 \geq \dots \geq k_p > 0$  are



actually the variances of the sample principal components. Letting  $B_j$ ,  $j = 1, \dots, p$ , denote the coefficient vectors of the sample principal components and  $B = (B_1, \dots, B_p)$ , it was established that  $B' \hat{\Sigma} B = \text{diag}(k_1, \dots, k_p) \equiv D$ . The first nonzero component of  $B_j$  is required to be positive for  $j = 1, \dots, p$ , so that  $B$  be uniquely determined. Then, the joint density of  $B$  and  $D$  is available by transforming  $\hat{\Sigma}$  in terms of  $B$  and  $D$ . Let the joint density of  $B$  and  $D$  and the marginal densities of  $D$  and  $B$  be respectively denoted by  $g(D, B)$ ,  $g_1(D)$  and  $g_2(B)$ . In light of the procedures and results presented in Chap. 8 or in Mathai (1997), we have the following joint density:

$$g(D, B)dD \wedge dB = \frac{n^{\frac{mp}{2}}}{|2\Sigma|^{\frac{m}{2}} \Gamma_p(\frac{m}{2})} \left\{ \prod_{j=1}^p k_j \right\}^{\frac{m}{2} - \frac{p+1}{2}} \left\{ \prod_{i < j} (k_i - k_j) \right\} \\ \times e^{-\frac{n}{2} \text{tr}((B' \Sigma^{-1} B) D)} h(B) dD \quad (9.6.1)$$

where  $h(B)$  is the differential element corresponding to  $B$ , which is given in Chap. 8 and in Mathai (1997). The marginal densities of  $D$  and  $B$  are not explicitly available in a convenient form for a general  $\Sigma$ . In that case, they can only be expressed in terms of hypergeometric functions of matrix argument and zonal polynomials. See, for example, Mathai et al. (1995) for a discussion of zonal polynomials. However, when  $\Sigma = I$ , the joint and marginal densities are available in closed forms. They are

$$g(D, B)dD \wedge dB = \frac{n^{\frac{mp}{2}}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} \left\{ \prod_{j=1}^p k_j \right\}^{\frac{m}{2} - \frac{p+1}{2}} \left\{ \prod_{i < j} (k_i - k_j) \right\} e^{-\frac{n}{2}(k_1 + \dots + k_p)} h(B) dD, \quad (9.6.2)$$

$$g_1(D) = \frac{n^{\frac{mp}{2}}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \left\{ \prod_{j=1}^p k_j \right\}^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{n}{2}(k_1 + \dots + k_p)} \left\{ \prod_{i < j} (k_i - k_j) \right\}, \quad (9.6.3)$$

$$g_2(B)dB = \frac{\Gamma_p(\frac{p}{2})}{\pi^{\frac{p^2}{2}}} h(B), \quad BB' = I = B'B, \quad k_1 > k_2 > \dots > k_p > 0. \quad (9.6.4)$$

Given the representation of the joint density  $g(D, B)$  appearing in (9.6.2), it is readily seen that  $D$  and  $B$  are independently distributed. Similar results are obtainable for  $\Sigma = \sigma^2 I$  where  $\sigma^2 > 0$  is a real scalar quantity and  $I$  is the identity matrix.

**Example 9.6.1.** Show that the function  $g_1(D)$  given in (9.6.3) is a density for  $p = 2$ ,  $n = 6$ ,  $m = 5$ , and derive the density of  $k_1$  for this special case.

**Solution 9.6.1.** We have  $\frac{p+1}{2} = \frac{3}{2}$ ,  $\frac{m}{2} - \frac{3}{2} = 1$ . As  $g_1(D)$  is manifestly nonnegative, it suffices to show that the total integral equals 1. When  $p = 2$  and  $n = 6$ , the constant part of (9.6.3) is the following:

$$\begin{aligned} \frac{n^{\frac{mp}{2}}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} &= \frac{6^5}{2^5 \Gamma_p(\frac{5}{2})} \frac{\pi^2}{\Gamma_2(1)} \\ &= \frac{3^5}{\sqrt{\pi} \Gamma(\frac{5}{2}) \Gamma(\frac{2}{2})} \frac{\pi^2}{\sqrt{\pi} \Gamma(1) \sqrt{\pi}} \\ &= \frac{3^5}{\sqrt{\pi} (\frac{3}{2}) (\frac{1}{2}) \sqrt{\pi}} \frac{\pi^2}{\sqrt{\pi} \sqrt{\pi}} = 4(3^4). \end{aligned} \quad (i)$$

As for the integral part, we have

$$\begin{aligned} &\int_{k_1=0}^{\infty} \int_{k_2=0}^{k_1} k_1 k_2 (k_1 - k_2) e^{-3(k_1+k_2)} dk_1 \wedge dk_2 \\ &= \int_{k_1=0}^{\infty} k_1^2 e^{-3k_1} \left[ \int_{k_2=0}^{k_1} k_2 e^{-3k_2} dk_2 \right] dk_1 - \int_{k_1=0}^{\infty} k_1 e^{-3k_1} \left[ \int_{k_2=0}^{k_1} k_2^2 e^{-3k_2} dk_2 \right] dk_1 \\ &= \int_{k_1=0}^{\infty} \left( k_1^2 e^{-3k_1} \left[ -\frac{k_1}{3} e^{-3k_1} + \frac{1}{3^2} (1 - e^{-3k_1}) \right] \right. \\ &\quad \left. - k_1 e^{-3k_1} \left[ -\frac{k_1^2}{3} e^{-3k_1} - \frac{2k_1}{3^2} e^{-3k_1} + \frac{2}{3^3} (1 - e^{-3k_1}) \right] \right) dk_1 \quad (ii) \\ &= -\frac{1}{3} [\Gamma(4)6^{-4}] + \frac{1}{9} [\Gamma(3)3^{-3}] - \frac{1}{3^2} [\Gamma(3)6^{-3}] \\ &\quad + \frac{1}{3} [\Gamma(4)6^{-4}] + \frac{2}{3^2} [\Gamma(3)6^{-3}] - \frac{2}{3^3} [\Gamma(2)3^{-2}] + \frac{2}{3^2} [\Gamma(2)6^{-2}] \\ &= \frac{1}{3^2} [\Gamma(3)6^{-3}] + \frac{2}{3^3} [\Gamma(2)6^{-2}] + \frac{1}{3^2} [\Gamma(3)3^{-3}] - \frac{2}{3^3} [\Gamma(2)3^{-2}] \\ &= \left[ \frac{1}{4} + \frac{2}{4} + 2 - 2 \right] \frac{1}{3^5} = \frac{1}{4(3^4)}. \quad (iii) \end{aligned}$$

The product of (i) and (iii) being equal to 1, this establishes that  $g_1(D)$  is indeed a density function for  $p = 2$  and  $n = 6$ . On integrating out  $k_2$ , the resulting integrand appearing in (ii) yields the marginal density of  $k_1$ . Denoting the marginal density of  $k_1$  by  $g_{11}(k_1)$ , after some simplifications, we have

$$g_{11}(k_1) = 4(3^4) \left\{ \left( \frac{k_1^2}{3^2} + \frac{2k_1}{3^3} \right) e^{-6k_1} + \left( \frac{k_1^2}{3^2} - \frac{2k_1}{3^3} \right) e^{-3k_1} \right\}, \quad 0 < k_1 < \infty,$$

and zero elsewhere. Similarly, given  $g_1(D)$ , the marginal density of  $k_2$  is obtained by integrating out  $k_1$  from  $k_2$  to  $\infty$ . This completes the computations.

### 9.6.1. The distributions of the largest and smallest eigenvalues

One can determine the distribution of the  $j$ -th largest eigenvalue  $k_j$  and thereby, the distribution of the largest one,  $k_1$ , as well as that of the smallest one,  $k_p$ . Actually, many authors have been investigating the problem of deriving the distributions of the eigenvalues of a central Wishart matrix and matrix-variate type-1 beta and type-2 beta distributions. Some of the test statistics discussed in Chap. 6 can also be treated as eigenvalue problems involving real and complex type-1 beta matrices. Let  $k_1 > k_2 > \cdots > k_p > 0$  be the eigenvalues of the real sampling distribution of a Wishart matrix generated from a  $p$ -variate real Gaussian sample. Then, as given in (9.6.3), the joint density of  $k_1, \dots, k_p$ , denoted by  $g_1(D)$ , is the following:

$$g_1(D)dD = \frac{n^{\frac{mp}{2}}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \left\{ \prod_{j=1}^p k_j^{\frac{m}{2} - \frac{p+1}{2}} \right\} e^{-\frac{n}{2}(k_1 + \cdots + k_p)} \left\{ \prod_{i < j} (k_i - k_j) \right\} dD \quad (9.6.5)$$

where  $m = n - 1$  is the number of degrees of freedom,  $n$  being the sample size. More generally, we may consider a  $p \times p$  real matrix-variate gamma distribution having  $\alpha$  as its shape parameter and  $aI$ ,  $a > 0$ , as its scale parameter matrix,  $I$  denoting the identity matrix. Noting that the eigenvalues of  $aS$  are equal to the eigenvalues of  $S$  multiplied by the scalar quantity  $a$ , we may take  $a = 1$  without any loss of generality. Let  $g(S)dS$  denote the density of the resulting gamma matrix and  $g(D)$  be  $g(S)$  expressed in terms of  $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$ , the eigenvalues of  $S$ . Then, given the Jacobian of the transformation  $S \rightarrow D$  specified in (9.6.7), the joint density of  $\lambda_1, \dots, \lambda_p$ , denoted by  $g_1(D)$  is obtained as

$$g(D)dS = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\alpha) \Gamma_p(\frac{p}{2})} \left\{ \prod_{j=1}^p \lambda_j^{\alpha - \frac{p+1}{2}} \right\} e^{-(\lambda_1 + \cdots + \lambda_p)} \left\{ \prod_{i < j} (\lambda_i - \lambda_j) \right\} dD \equiv g_1(D)dD \quad (9.6.6)$$

where  $dD = d\lambda_1 \wedge \cdots \wedge d\lambda_p$ . The corresponding  $p \times p$  complex matrix-variate gamma distribution will have real eigenvalues, also denoted by  $\lambda_1 > \cdots > \lambda_p > 0$ , the matrix  $\tilde{S}$  being Hermitian positive definite in this case. Then, in light of the relationship (9.6a.2) between the differential elements of  $\tilde{S}$  and  $D$ , the joint density of  $\lambda_1, \dots, \lambda_p$ , denoted by  $\tilde{g}_1(D)$ , is given by

$$\tilde{g}(D)d\tilde{S} = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(\alpha)\tilde{\Gamma}_p(p)} \left\{ \prod_{j=1}^p \lambda_j^{\alpha-p} \right\} e^{-(\lambda_1+\dots+\lambda_p)} \left\{ \prod_{i<j} (\lambda_i - \lambda_j)^2 \right\} dD \equiv \tilde{g}_1(D)dD \quad (9.6a.1)$$

where  $\tilde{g}(D)$  is  $\tilde{g}(\tilde{S})$  expressed in terms of  $D$ ,  $\tilde{g}(\tilde{S})$  denoting the density function of a matrix-variate gamma random variable whose shape parameter and scale parameter matrix are  $\alpha$  and  $I$ , respectively.

As explained for instance in Mathai (1997), when  $S$  and  $\tilde{S}$  are the  $p \times p$  gamma matrices in the real and complex domains, the integration over the Stiefel manifold yields

$$dS = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \left\{ \prod_{i<j} (\lambda_i - \lambda_j) \right\} dD \quad (9.6.7)$$

and

$$d\tilde{S} = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)} \left\{ \prod_{i<j} (\lambda_i - \lambda_j)^2 \right\} dD, \quad (9.6a.2)$$

respectively. When endeavoring to derive the marginal density of  $\lambda_j$  for any fixed  $j$ , the difficulty arises from the factor  $\prod_{i<j} (\lambda_i - \lambda_j)$ . So, let us first attempt to simplify this factor.

### 9.6.2. Simplification of the factor $\prod_{i<j} (\lambda_i - \lambda_j)$

It is well known that one can write  $\prod_{i<j} (\lambda_i - \lambda_j)$  as a Vandermonde determinant which, incidentally, has been utilized in connection with nonlinear transformations in Mathai (1997). That is,

$$\prod_{i<j} (\lambda_i - \lambda_j) = \begin{vmatrix} \lambda_1^{p-1} & \lambda_1^{p-2} & \dots & \lambda_1 & 1 \\ \lambda_2^{p-1} & \lambda_2^{p-2} & \dots & \lambda_2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_p^{p-1} & \lambda_p^{p-2} & \dots & \lambda_p & 1 \end{vmatrix} \equiv |A| = |(a_{ij})|, \quad (9.6.8)$$

where the  $(i, j)$ -th element  $a_{ij} = \lambda_i^{p-j}$  for all  $i$  and  $j$ . We could consider a cofactor expansion of the determinant,  $|A|$ , consisting of expanding it in terms of the elements and their cofactors along any row. In this case, it would be advantageous to do so along the  $i$ -th row as the cofactors would then be free of  $\lambda_i$  and the coefficients of the cofactors would only be powers of  $\lambda_i$ . However, we would then lose the symmetry with respect to the elements of the cofactors in this instance. Since symmetry is required for the procedure to be discussed, we will consider the general expansion of a determinant, that is,

$$|A| = \sum_K (-1)^{\rho_K} a_{1k_1} a_{2k_2} \cdots a_{pk_p} = \sum_K (-1)^{\rho_K} \lambda_1^{p-k_1} \lambda_2^{p-k_2} \cdots \lambda_p^{p-k_p} \quad (9.6.9)$$

where  $K = (k_1, \dots, k_p)$  and  $(k_1, \dots, k_p)$  is a given permutation of the numbers  $(1, 2, \dots, p)$ . Defining  $\rho_K$  as the number of transpositions needed to bring  $(k_1, \dots, k_p)$  into the natural order  $(1, 2, \dots, p)$ ,  $(-1)^{\rho_K}$  will correspond to a  $-$  sign for the corresponding term if  $\rho_K$  is odd, the sign being otherwise positive. For example, for  $p = 3$ , the possible permutations are  $(1, 2, 3)$ ,  $(1, 3, 2)$ ,  $(2, 3, 1)$ ,  $(2, 1, 3)$ ,  $(3, 1, 2)$ ,  $(3, 2, 1)$ , so that there are  $3! = 6$  terms. For the sequence  $(1, 2, 3)$ ,  $k_1 = 1$ ,  $k_2 = 2$  and  $k_3 = 3$ ; for the sequence  $(1, 3, 2)$ ,  $k_1 = 1$ ,  $k_2 = 3$  and  $k_3 = 2$ , and so on,  $\sum_K$  representing the sum of all such terms multiplied by  $(-1)^{\rho_K}$ . For  $(1, 2, 3)$ ,  $\rho_K = 0$  corresponding to a plus sign; for  $(1, 3, 2)$ ,  $\rho_K = 1$  corresponding to a minus sign, and so on. Other types of expansions of  $|A|$  could also be utilized. As it turns out, the general expansion given in (9.6.9) is the most convenient one for deriving the marginal densities.

In the complex case,

$$\prod_{i < j} (\lambda_i - \lambda_j)^2 = |A|^2 = |AA'| = |A'A|$$

where

$$A' = \begin{vmatrix} \lambda_1^{p-1} & \lambda_2^{p-1} & \cdots & \lambda_p^{p-1} \\ \lambda_1^{p-2} & \lambda_2^{p-2} & \cdots & \lambda_p^{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{vmatrix} = [\alpha_1, \alpha_2, \dots, \alpha_p], \quad \alpha_j = \begin{bmatrix} \lambda_j^{p-1} \\ \lambda_j^{p-2} \\ \vdots \\ 1 \end{bmatrix}. \quad (i)$$

Observe that  $\alpha_j$  only contains  $\lambda_j$  and that  $A'A = \alpha_1 \alpha_1' + \cdots + \alpha_p \alpha_p'$ , so that for instance the  $(i, j)$ -th element in  $\alpha_1 \alpha_1'$  is  $\lambda_1^{2p-(i+j)}$ . Accordingly, the  $(i, j)$ -th element of  $\alpha_1 \alpha_1' + \cdots + \alpha_p \alpha_p'$  is  $\sum_{r=1}^p \lambda_r^{2p-(i+j)}$ . Thus, letting  $B = A'A = (b_{ij})$ ,  $b_{ij} = \sum_{r=1}^p \lambda_r^{2p-(i+j)}$ . Now, consider the expansion (9.6.9) of  $|B|$ , that is,

$$\prod_{i < j} (\lambda_i - \lambda_j)^2 = |B| = |A'A| = \sum_K (-1)^{\rho_K} b_{1k_1} b_{2k_2} \cdots b_{pk_p} \quad (ii)$$

where  $K = (k_1, \dots, k_p)$  and  $(k_1, \dots, k_p)$  is a permutation of the sequence  $(1, 2, \dots, p)$ . Note that

$$\begin{aligned}
b_{1k_1} &= \lambda_1^{2p-(1+k_1)} + \lambda_2^{2p-(1+k_1)} + \dots + \lambda_p^{2p-(1+k_1)} \\
b_{2k_2} &= \lambda_1^{2p-(2+k_2)} + \lambda_2^{2p-(2+k_2)} + \dots + \lambda_p^{2p-(2+k_2)} \\
&\vdots \\
b_{pk_p} &= \lambda_1^{2p-(p+k_p)} + \lambda_2^{2p-(p+k_p)} + \dots + \lambda_p^{2p-(p+k_p)}. \tag{iii}
\end{aligned}$$

Let us write

$$b_{1k_1} b_{2k_2} \cdots b_{pk_p} = \sum_{r_1, \dots, r_p} \lambda_1^{r_1} \cdots \lambda_p^{r_p}. \tag{iv}$$

Then,

$$\prod_{i < j} (\lambda_i - \lambda_j)^2 = |B| = \sum_K (-1)^{\rho_K} \left[ \sum_{r_1, \dots, r_p} \lambda_1^{r_1} \cdots \lambda_p^{r_p} \right] \tag{9.6a.3}$$

where the  $r_j$ 's,  $j = 1, \dots, p$ , are nonnegative integers. We may now express the joint density of the eigenvalues in a systematic way.

### 9.6.3. The distributions of the eigenvalues

The joint density of  $\lambda_1, \dots, \lambda_p$  as specified in (9.6.6) and (9.6a.1) can be expressed as follows. In the real case,

$$\begin{aligned}
f(D)dD &= \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})\Gamma_p(\alpha)} \left( \prod_{j=1}^p \lambda_j^{\alpha - \frac{p+1}{2}} \right) e^{-(\lambda_1 + \dots + \lambda_p)} \left( \sum_K (-1)^{\rho_K} \lambda_1^{p-k_1} \cdots \lambda_p^{p-k_p} \right) dD \\
&= \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})\Gamma_p(\alpha)} \sum_K (-1)^{\rho_K} (\lambda_1^{m_1} \cdots \lambda_p^{m_p}) e^{-(\lambda_1 + \dots + \lambda_p)} dD \tag{9.6.10}
\end{aligned}$$

with

$$m_j = \alpha - \frac{p+1}{2} + p - k_j. \tag{i}$$

In the complex case, the joint density is

$$\begin{aligned}
\tilde{f}(D)dD &= \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)\tilde{\Gamma}_p(\alpha)} \sum_K (-1)^{\rho_K} \sum_{r_1, \dots, r_p} (\lambda_1^{\alpha-p+r_1} \cdots \lambda_p^{\alpha-p+r_p}) e^{-(\lambda_1 + \dots + \lambda_p)} dD \\
&= \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)\tilde{\Gamma}_p(\alpha)} \sum_K (-1)^{\rho_K} \sum_{r_1, \dots, r_p} (\lambda_1^{m_1} \cdots \lambda_p^{m_p}) e^{-(\lambda_1 + \dots + \lambda_p)} dD \tag{9.6a.4}
\end{aligned}$$

with  $m_j = \alpha - p + r_j$  and  $r_j$  as defined in (9.6a.3). For convenience, we will use the same symbol  $m_j$  for both the real and complex cases; however, in the real case,  $m_j = \alpha - \frac{p+1}{2} + p - k_j$  and, in the complex case,  $m_j = \alpha - p + r_j$ .

The distributions of the largest, smallest and  $j$ -th largest eigenvalues were considered by many authors. Earlier works mostly dealt with eigenvalue problems associated with testing hypotheses on the parameters of one or more real Gaussian populations. In such situations, a one-to-one function of the likelihood ratio statistics could be explored in terms of the eigenvalues of a real type-1 beta distributed matrix-variate random variable. In a series of papers, Pillai constructed the distributions of the seven largest eigenvalues in the type-1 beta distribution and produced percentage points as well; the reader may refer for example to Pillai (1964) and the references therein. In a series of papers including Khatri (1964), Khatri addressed the distributions of eigenvalues in the real and complex domains. In a series of papers, Krishnaiah and his co-authors dealt with various distributional aspects of eigenvalues, see for instance Krishnaiah et al. (1973). Clemm et al. (1973) computed upper percentage points of the distribution of the eigenvalues of the Wishart matrix. James (1964) considered the eigenvalue problem of different types of matrix-variate random variables and determined their distributions in terms of functions of matrix arguments and zonal polynomials. In a series of papers Davis, dealt with the distributions of eigenvalues by creating and solving systems of differential equations, see for example Davis (1972). Edelman (1991) discussed the distributions and moments of the smallest eigenvalue of Wishart type matrices. Johnstone (2001) examined the distribution of the largest eigenvalue in Principal Component Analysis. Recently, Chiani (2014) and James and Lee (2021) discussed the distributions of the eigenvalues of Wishart matrices. The methods employed in these papers lead to representations of the distributions of eigenvalues in terms of Pfaffians of skew symmetric matrices, incomplete gamma functions, multiple integrals, functions of matrix argument and zonal polynomials, ratios of determinants, solutions of differential equations, and so on. None of those methods yield tractable forms for the distribution or density functions of eigenvalues.

In the next subsections, we provide, in explicit forms, the exact distributions of any of the  $j$ -th largest eigenvalue of a general real or complex matrix-variate gamma type matrix, either as finite sums when a certain quantity is a positive integer or as a product of infinite series in the general non-integer case. These include, for instance, the distributions of the largest and smallest eigenvalue as well as the joint distributions of several of the largest or smallest eigenvalues, and readily apply to the real and complex Wishart distributions.

#### 9.6.4. Density of the smallest eigenvalue $\lambda_p$ in the real matrix-variate gamma case

We will initially examine the situation where  $m_j = \alpha - \frac{p+1}{2} + p - k_j$  is an integer, so that  $m_j$  in the real matrix-variate gamma case is a positive integer. We will integrate out  $\lambda_1, \dots, \lambda_{p-1}$  to obtain the marginal density of  $\lambda_p$ . Since  $m_j$  is a positive integer, we can integrate by parts. For instance,

$$\begin{aligned} \int_{\lambda_1=\lambda_2}^{\infty} \lambda_1^{m_1} e^{-\lambda_1} d\lambda_1 &= [-\lambda_1^{m_1} e^{-\lambda_1}]_{\lambda_2}^{\infty} + [-m_1 \lambda_1^{m_1-1} e^{-\lambda_1}]_{\lambda_2}^{\infty} + \dots + [-e^{-\lambda_1}]_{\lambda_2}^{\infty} \\ &= \sum_{\mu_1=0}^{m_1} \frac{m_1!}{(m_1 - \mu_1)!} \lambda_2^{m_1 - \mu_1} e^{-\lambda_2}, \end{aligned} \quad (i)$$

and integrating  $\lambda_1, \dots, \lambda_{j-1}$  gives the following:

$$\begin{aligned} &\int_{\lambda_1} \int_{\lambda_2} \dots \int_{\lambda_{j-1}} \lambda_1^{m_1} \dots \lambda_{j-1}^{m_{j-1}} e^{-(\lambda_1 + \dots + \lambda_{j-1})} d\lambda_1 \wedge \dots \wedge d\lambda_{j-1} \\ &= \sum_{\mu_1=0}^{m_1} \frac{m_1!}{(m_1 - \mu_1)!} \sum_{\mu_2=0}^{m_1 - \mu_1 + m_2} \frac{(m_1 - \mu_1 + m_2)!}{2^{\mu_2+1} (m_1 - \mu_1 + m_2 - \mu_2)!} \dots \\ &\quad \times \sum_{\mu_{j-1}=0}^{m_1 - \mu_1 + \dots + m_{j-1}} \frac{(m_1 - \mu_1 + \dots + m_{j-1})!}{(j-1)^{\mu_{j-1}+1} (m_1 - \mu_1 + \dots + m_{j-1} - \mu_{j-1})!} \lambda_j^{m_1 - \mu_1 + \dots + m_{j-1} - \mu_{j-1}} e^{-j\lambda_j} \\ &\equiv \phi_{j-1}(\lambda_j). \end{aligned} \quad (ii)$$

Hence, the following result:

**Theorem 9.6.1.** *When  $m_j = \alpha - \frac{p+1}{2} + p - k_j$  is a positive integer, where  $m_j$  is defined in (9.6.10), the marginal density of the smallest eigenvalue  $\lambda_p$  of the  $p \times p$  real gamma distributed matrix with parameters  $(\alpha, I)$ , denoted by  $f_{1p}(\lambda_p)$ , is the following:*

$$\begin{aligned} &f_{1p}(\lambda_p) d\lambda_p \\ &= c_K \phi_{p-1}(\lambda_p) \lambda_p^{m_p} e^{-\lambda_p} \\ &= c_K \sum_{\mu_1=0}^{m_1} \frac{m_1!}{(m_1 - \mu_1)!} \sum_{\mu_2=0}^{m_1 - \mu_1 + m_2} \frac{(m_1 - \mu_1 + m_2)!}{2^{\mu_2+1} (m_1 - \mu_1 + m_2 - \mu_2)!} \dots \sum_{\mu_{p-1}=0}^{m_1 - \mu_1 + \dots + m_{p-1}} \\ &\quad \times \frac{(m_1 - \mu_1 + \dots + m_{p-1})!}{(p-1)^{\mu_{p-1}+1} (m_1 - \mu_1 + \dots + m_{p-1} - \mu_{p-1})!} \lambda_p^{m_1 - \mu_1 + \dots + m_{p-1} - \mu_{p-1} + m_p} e^{-p\lambda_p} d\lambda_p \end{aligned} \quad (9.6.11)$$



for  $0 < \lambda_p < \infty$ , where

$$c_K = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})\Gamma_p(\alpha)} \sum_K (-1)^{\rho_K},$$

$\phi_{j-1}(\lambda_j)$  is specified in (ii) and  $K = (k_1, \dots, k_p)$  is defined in (9.6.9).

In the complex  $p \times p$  matrix-variate gamma case,  $r_j$  is as defined in (9.6a.3) and the expression for  $\phi_{j-1}(\lambda_j)$  given in (ii) remains the same with the exception that  $m_j$  in the complex case is  $m_j = \alpha - p + r_j$ . Then, in the complex domain, the density of  $\lambda_p$ , denoted by  $\tilde{f}_{1p}(\lambda_p)$ , is the following:

**Theorem 9.6a.1.** When  $m_j = \alpha - p + r_j$  is a positive integer, where  $r_j$  is defined in (9.6a.3), the density of the smallest eigenvalue of the complex matrix-variate gamma distribution is the following:

$$\begin{aligned} & \tilde{f}_{1p}(\lambda_p) d\lambda_p \\ &= \tilde{c}_K \phi_{p-1}(\lambda_p) \lambda_p^{m_p} e^{-p\lambda_p} \\ &= \tilde{c}_K \sum_{\mu_1=0}^{m_1} \frac{m_1!}{(m_1 - \mu_1)!} \sum_{\mu_2=0}^{m_1 - \mu_1 + m_2} \frac{(m_1 - \mu_1 + m_2)!}{2^{\mu_2+1} (m_1 - \mu_1 + m_2 - \mu_2)!} \cdots \sum_{\mu_{p-1}=0}^{m_1 - \mu_1 + \cdots + m_{p-1}} \\ & \times \frac{(m_1 - \mu_1 + \cdots + m_{p-1})!}{(p-1)^{\mu_{p-1}+1} (m_1 - \mu_1 + \cdots + m_{p-1} - \mu_{p-1})!} \lambda_p^{m_1 - \mu_1 + \cdots + m_{p-1} - \mu_{p-1} + m_p} e^{-p\lambda_p} d\lambda_p \end{aligned} \quad (9.6a.5)$$

for  $0 < \lambda_p < \infty$ , where

$$\tilde{c}_K = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)\tilde{\Gamma}_p(\alpha)} \sum_K (-1)^{\rho_K} \sum_{r_1, \dots, r_k}.$$

**Note 9.6.1.** In the complex Wishart case,  $\alpha = m$  where  $m$  is the number of degrees of freedom, which is a positive integer. Hence, Theorem 9.6a.1 gives the final result in the general case for that distribution. One can obtain the joint density of the  $p - j + 1$  smallest eigenvalues from  $\phi_{j-1}(\lambda_j)$  as defined in (ii), both in the real and complex cases. If the scale parameter matrix of the gamma distribution is of the form  $aI$  where  $a > 0$  is a real scalar and  $I$  is the identity matrix, then the distributions of the eigenvalues can also be obtained from the proposed procedure since for any square matrix  $B$ , the eigenvalues of  $aB$  are  $a v_j$ 's where the  $v_j$ 's are the eigenvalues of  $B$ . In the case of real Wishart distributions originating from a sample of size  $n$  from a  $p$ -variate Gaussian population

whose covariance matrix is the identity matrix, the  $\lambda_j$ 's are multiplied by the constant  $a = \frac{n}{2}$ . If the Wishart matrix is not an estimator obtained from the sample values, they should only be multiplied by  $a = \frac{1}{2}$  in the real case, no multiplicative constant being necessary in the complex Wishart case.

### 9.6.5. Density of the largest eigenvalue $\lambda_1$ in the real matrix-variate gamma case

Consider the case of  $m_j = \alpha - \frac{p+1}{2} + p - k_j$  being an integer first. Then, in the real case,  $m_j$  is as defined in (9.6.10). One has to integrate out  $\lambda_p, \dots, \lambda_2$  in order to obtain the marginal density of  $\lambda_1$ . As the initial step, consider the integration of  $\lambda_p$ , that is,

$$\begin{aligned} \text{Step 1 integral : } & \int_{\lambda_p=0}^{\lambda_{p-1}} \lambda_p^{m_p} e^{-\lambda_p} d\lambda_p \\ & = [-\lambda_p^{m_p} e^{-\lambda_p}]_0^{\lambda_{p-1}} + \dots + [-m_p! e^{-\lambda_p}]_0^{\lambda_{p-1}} \\ & = m_p! - \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \lambda_{p-1}^{m_p - v_p} e^{-\lambda_{p-1}}. \end{aligned} \quad (i)$$

Now, multiplying each term by  $\lambda_{p-1}^{m_{p-1}} e^{-\lambda_{p-1}}$  and integrating by parts, we have the second step integral:

Step 2 integral

$$\begin{aligned} & = m_p! \int_{\lambda_{p-1}=0}^{\lambda_{p-2}} \lambda_{p-1}^{m_{p-1}} e^{-\lambda_{p-1}} d\lambda_{p-1} - \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \int_{\lambda_{p-1}=0}^{\lambda_{p-2}} \lambda_{p-1}^{m_p - v_p + m_{p-1}} e^{-2\lambda_{p-1}} d\lambda_{p-1} \\ & = m_p! m_{p-1}! - m_p! \sum_{v_{p-1}=0}^{m_{p-1}} \frac{m_{p-1}!}{(m_{p-1} - v_{p-1})!} \lambda_{p-2}^{m_{p-1} - v_{p-1}} e^{-\lambda_{p-2}} \\ & \quad - \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \frac{(m_p - v_p + m_{p-1})!}{2^{m_p - v_p + m_{p-1}}} \\ & \quad + \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \sum_{v_{p-1}=0}^{m_{p-1}} \frac{(m_p - v_p + m_{p-1})!}{2^{v_{p-1}+1} (m_p - v_p + m_{p-1} - v_{p-1})!} \lambda_{p-2}^{m_p - v_p + m_{p-1} - v_{p-1}} e^{-2\lambda_{p-2}}. \end{aligned} \quad (ii)$$

At the  $j$ -th step of integration, there will be  $2^j$  terms of which  $\frac{2^j}{2} = 2^{j-1}$  will be positive and  $2^{j-1}$  will be negative. All the terms at the  $j$ -th step can be generated by  $2^j$  sequences

of zeros and ones. Each sequence (each row) consists of  $j$  positions with a zero or one occupying each position. Depending on whether the number of ones in a sequence is odd or even, the corresponding term will start with a minus or a plus sign, respectively. The following are the sequences, where the last column indicates the sign of the term:

$$\begin{array}{l}
 \text{Step 1: } \begin{array}{c} 0 \quad + \\ 1 \quad - \end{array}, \quad \text{Step 2: } \begin{array}{c} 0 \quad 0 \quad + \\ 0 \quad 1 \quad - \\ 1 \quad 0 \quad - \\ 1 \quad 1 \quad + \end{array}, \quad \text{Step 3: } \begin{array}{c} 0 \quad 0 \quad 0 \quad + \\ 0 \quad 0 \quad 1 \quad - \\ 0 \quad 1 \quad 0 \quad - \\ 0 \quad 1 \quad 1 \quad + \end{array}, \quad \text{Step 3: } \begin{array}{c} 1 \quad 0 \quad 0 \quad - \\ 1 \quad 0 \quad 1 \quad + \\ 1 \quad 1 \quad 0 \quad + \\ 1 \quad 1 \quad 1 \quad - \end{array} \\
 \\
 \text{Step 4: } \begin{array}{c} 0 \quad 0 \quad 0 \quad 0 \quad + \\ 0 \quad 0 \quad 0 \quad 1 \quad - \\ 0 \quad 0 \quad 1 \quad 0 \quad - \\ 0 \quad 0 \quad 1 \quad 1 \quad + \end{array}, \quad \begin{array}{c} 0 \quad 1 \quad 0 \quad 0 \quad - \\ 0 \quad 1 \quad 0 \quad 1 \quad + \\ 0 \quad 1 \quad 1 \quad 0 \quad + \\ 0 \quad 1 \quad 1 \quad 1 \quad - \end{array}, \quad \begin{array}{c} 1 \quad 0 \quad 0 \quad 0 \quad - \\ 1 \quad 0 \quad 0 \quad 1 \quad + \\ 1 \quad 0 \quad 1 \quad 0 \quad + \\ 1 \quad 0 \quad 1 \quad 1 \quad - \end{array}, \quad \begin{array}{c} 1 \quad 1 \quad 0 \quad 0 \quad + \\ 1 \quad 1 \quad 0 \quad 1 \quad - \\ 1 \quad 1 \quad 1 \quad 0 \quad - \\ 1 \quad 1 \quad 1 \quad 1 \quad + \end{array}.
 \end{array}$$

All the terms at the  $j$ -th step can be written down by using the following rules:

- (1): If the first entry in a sequence (row) is zero, then the corresponding factor in the term is  $m_p!$ ;
- (2): If the first entry in a sequence is 1, then the corresponding factor in the term is  $\sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p-v_p)!}$  or this sum multiplied by  $\lambda_{p-1}^{m_p-v_p} e^{-\lambda_{p-1}}$  if this 1 is the last entry in the sequence;
- (3): If the  $r$ -th and  $(r-1)$ -th entries in the sequence both equal zero, then the corresponding factor in the term is  $m_{p-r+1}!$ ;
- (4): If the  $r$ -th entry in the sequence is zero and the  $(r-1)$ -th entry is 1, then the corresponding factor in the term is  $\frac{(n_{r-1}+m_{p-r+1})!}{(\eta_{r-1}+1)^{n_{r-1}+m_{p-r+1}+1}}$ , where  $n_{r-1}$  is the argument of the denominator factorial and  $(\eta_{r-1})^{v_{p-r+2}+1}$  is the factor in the denominator corresponding to the  $(r-1)$ -th entry;
- (5): If the  $r$ -th entry in the sequence is 1 and the  $(r-1)$ -th entry is zero, then the corresponding factor in the term is  $\sum_{v_{p-r+1}=0}^{m_{p-r+1}} \frac{m_{p-r+1}!}{(m_{p-r+1}-v_{p-r+1})!}$  or this sum multiplied by  $\lambda_{p-r}^{m_{p-r+1}-v_{p-r+1}} e^{-\lambda_{p-r}}$  if this 1 is the last entry in the sequence;
- (6): If the  $r$ -th and  $(r-1)$ -th entries in the sequence are both equal to 1, then the corresponding factor in the term is  $\sum_{v_{p-r+1}=0}^{n_{r-1}+m_{p-r+1}} \frac{(n_{r-1}+m_{p-r+1})!}{(\eta_{r-1}+1)^{v_{p-r+1}+1} (n_{r-1}+m_{p-r+1}-v_{p-r+1})!}$  or this sum multiplied by  $\lambda_{p-r}^{n_{r-1}+m_{p-r+1}-v_{p-r+1}} e^{-(\eta_{r-1}+1)\lambda_{p-r}}$  if this 1 is the last entry in the sequence, where  $n_{r-1}$  and  $\eta_{r-1}$  are defined in rule (4). By applying the above rules, let us write down the terms at step 3, that is,  $j = 3$ . The sequences are then the following:

$$\text{Step 3 sequences: } \begin{array}{cccc} 0 & 0 & 0 & + \\ 0 & 0 & 1 & - \\ 0 & 1 & 0 & - \\ 0 & 1 & 1 & + \end{array}, \begin{array}{cccc} 1 & 0 & 0 & - \\ 1 & 0 & 1 & + \\ 1 & 1 & 0 & + \\ 1 & 1 & 1 & - \end{array}. \quad (iii)$$

The corresponding terms in the order of the sequences are the following:

Step 3 integral

$$\begin{aligned} &= m_p! m_{p-1}! m_{p-2}! - m_p! m_{p-1}! \sum_{v_{p-2}=0}^{m_{p-2}} \frac{m_{p-2}!}{(m_{p-2} - v_{p-2})!} \lambda_{p-3}^{m_{p-2}-v_{p-2}} e^{-\lambda_{p-3}} \\ &\quad - m_p! \sum_{v_{p-1}=0}^{m_{p-1}} \frac{m_{p-1}!}{(m_{p-1} - v_{p-1})!} \frac{(m_{p-1} - v_{p-1} + m_{p-2})!}{2^{m_{p-1}-v_{p-1}+m_{p-2}+1}} \\ &\quad + m_p! \sum_{v_{p-1}=0}^{m_{p-1}} \frac{(m_{p-1})!}{(m_{p-1} - v_{p-1})!} \sum_{v_{p-2}=0}^{m_{p-1}-v_{p-1}+m_{p-2}} \frac{(m_{p-1} - v_{p-1} + m_{p-2})!}{2^{v_{p-2}+1}(m_{p-1} - v_{p-1} + m_{p-2} - v_{p-2})!} \\ &\quad \quad \quad \times \lambda_{p-3}^{m_{p-1}-v_{p-1}+m_{p-2}-v_{p-2}} e^{-2\lambda_{p-3}} \\ &\quad - \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \frac{(m_p - v_p + m_{p-1})!}{2^{m_p-v_p+m_{p-1}}} m_{p-2}! \\ &\quad + \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \frac{(m_p - v_p + m_{p-1})!}{2^{m_p-v_p+m_{p-1}+1}} \sum_{v_{p-2}=0}^{m_{p-2}} \frac{m_{p-2}!}{(m_{p-2} - v_{p-2})!} \lambda_{p-3}^{m_{p-2}-v_{p-2}} e^{-\lambda_{p-3}} \\ &\quad + \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \sum_{v_{p-1}=0}^{m_p-v_p+m_{p-1}} \frac{(m_p - v_p + m_{p-1})!}{(m_p - v_p + m_{p-1} - v_{p-1})!} \\ &\quad \quad \quad \times \frac{(m_p - v_p + m_{p-1} - v_{p-1} + m_{p-2})!}{3^{m_p-v_p+m_{p-1}-v_{p-1}+m_{p-2}}} \\ &\quad - \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \sum_{v_{p-1}=0}^{m_p-v_p+m_{p-1}} \frac{(m_p - v_p + m_{p-1})!}{2^{v_{p-1}+1}(m_p - v_p + m_{p-1} - v_{p-1})!} \\ &\quad \quad \quad \times \sum_{v_{p-2}=0}^{m_p-v_p+m_{p-1}-v_{p-1}+m_{p-2}} \frac{(m_p - v_p + m_{p-1} - v_{p-1} + m_{p-2})!}{3^{v_{p-2}+1}(m_p - v_p + m_{p-1} - v_{p-1} + m_{p-2} - v_{p-2})!} \\ &\quad \quad \quad \times \lambda_{p-3}^{m_p-v_p+m_{p-1}-v_{p-1}+m_{p-2}-v_{p-2}} e^{-3\lambda_{p-3}}. \quad (iv) \end{aligned}$$

The terms in (iv) are to be multiplied by  $\lambda_{p-3}^{m_{p-3}} e^{-\lambda_{p-3}}$  to obtain the final result if we are stopping, that is, if  $p = 4$ . The terms in (iv) can be verified by multiplying the step 2

integral by  $\lambda_{p-2}^{m_{p-2}} e^{-\lambda_{p-2}}$  and then integrating (ii), term by term. Denoting the sum of the terms at the  $j$ -th step by  $\psi_j(\lambda_{p-j})$ , the density of the largest eigenvalue  $\lambda_1$ , denoted by  $f_{11}(\lambda_1)$ , is the following:

**Theorem 9.6.2.** *When  $m_j = \alpha - \frac{p+1}{2} + p - k_j$  is a positive integer and  $\psi_j(\lambda_{p-j})$  is as defined in the preceding paragraph, where  $k_j$  is given in (9.6.10), the density of  $\lambda_1$  in the real matrix-variate gamma case, denoted by  $f_{11}(\lambda_1)$ , is the following:*

$$f_{11}(\lambda_1) d\lambda_1 = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\alpha) \Gamma_p(\frac{p}{2})} \sum_K (-1)^{\rho_k} \psi_{p-1}(\lambda_1) \lambda_1^{m_1} e^{-\lambda_1} d\lambda_1, \quad 0 < \lambda_1 < \infty, \quad (9.6.12)$$

where it is assumed that  $m_j$  is a positive integer.

In the corresponding complex case, the procedure is parallel and the expression for  $\psi_j(\lambda_{p-j})$  remains the same, except that  $m_j$  will then be equal to  $\alpha - p + r_j$ , where  $r_j$  is defined in (9.6a.3). Assuming that  $m_j$  is a positive integer and letting the density in the complex case be denoted by  $\tilde{f}_{11}(\lambda_1)$ , we have the following result:

**Theorem 9.6a.2.** *Letting  $m_j = \alpha - p + r_j$  be a positive integer and  $\psi_j(\lambda_{p-j})$  have the same representation as in the real case except that  $m_j = \alpha - p + r_j$ , in the complex case, the density of  $\lambda_1$ , denoted by  $\tilde{f}_{11}(\lambda_1)$ , is the following:*

$$\tilde{f}_{11}(\lambda_1) d\lambda_1 = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p) \tilde{\Gamma}_p(\alpha)} \sum_K (-1)^{\rho_K} \sum_{r_1, \dots, r_k} \psi_{p-1}(\lambda_1) \lambda_1^{m_1} e^{-\lambda_1} d\lambda_1, \quad 0 < \lambda_1 < \infty. \quad (9.6a.6)$$

**Note 9.6.2.** One can also compute the density of the  $j$ -th eigenvalue  $\lambda_j$  from Theorems 9.6.1 and 9.6.2. For obtaining the density of  $\lambda_j$ , one has to integrate out  $\lambda_1, \dots, \lambda_{j-1}$  and  $\lambda_p, \lambda_{p-1}, \dots, \lambda_{j+1}$ , the resulting expressions being available from the  $(j-1)$ -th step when integrating  $\lambda_1, \dots, \lambda_{j-1}$  and from the  $(p-j)$ -th step when integrating  $\lambda_p, \lambda_{p-1}, \dots, \lambda_{j+1}$ .

### 9.6.6. Density of the largest eigenvalue $\lambda_1$ in the general real case

By general case, it is meant that  $m_j = \alpha - \frac{p+1}{2} + p - k_j$  is not a positive integer. In the real Wishart case,  $m_j$  will then be a half-integer; however in the general gamma case  $\alpha$  can be any real number greater than  $\frac{p-1}{2}$ . In this general case, we will expand the exponential part and then integrate term by term. That is,

$$\begin{aligned}
\text{Step 1 integral: } \int_{\lambda_p=0}^{\lambda_{p-1}} \lambda_p^{m_p} e^{-\lambda_p} d\lambda_p &= \sum_{v_p=0}^{\infty} \frac{(-1)^{v_p}}{v_p!} \int_{\lambda_p=0}^{\lambda_{p-1}} \lambda_p^{m_p+v_p} d\lambda_p \\
&= \sum_{v_p=0}^{\infty} \frac{(-1)^{v_p}}{v_p!} \frac{1}{(m_p + v_p + 1)} \lambda_{p-1}^{m_p+v_p+1}. \tag{i}
\end{aligned}$$

Continuing this process, we have

$$\begin{aligned}
\text{Step } j \text{ integral} &= \sum_{v_p=0}^{\infty} \frac{(-1)^{v_p}}{v_p!} \frac{1}{m_p + v_p + 1} \sum_{v_{p-1}=0}^{\infty} \frac{(-1)^{v_{p-1}}}{v_{p-1}!} \frac{1}{m_p + v_p + m_{p-1} + v_{p-1} + 2} \\
&\quad \cdots \sum_{v_{p-j+1}=0}^{\infty} \frac{(-1)^{v_{p-j+1}}}{v_{p-j+1}!} \frac{1}{m_p + v_p + \cdots + m_{p-j+1} + v_{p-j+1} + j} \\
&\equiv \Delta_j(\lambda_{p-j}). \tag{ii}
\end{aligned}$$

Then, in the general real case, the density of  $\lambda_1$ , denoted by  $f_{21}(\lambda_1)$ , is the following:

**Theorem 9.6.3.** When  $m_j = \alpha - \frac{p+1}{2} + p - k_j$  is not a positive integer, where  $k_j$  is as specified in (9.6.10), the density of the largest eigenvalue  $\lambda_1$  in the general real matrix-variate gamma case, denoted by  $f_{21}(\lambda_1)$ , is given by

$$f_{21}(\lambda_1)d\lambda_1 = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})\Gamma_p(\alpha)} \sum_K (-1)^{\rho_K} \Delta_{p-1}(\lambda_1) \lambda_1^{m_1} e^{-\lambda_1} d\lambda_1, \quad 0 < \lambda_1 < \infty, \tag{9.6.13}$$

where  $\Delta_j(\lambda_{p-j})$  is defined in (ii).

The corresponding density of  $\lambda_1$  in the general situation of the complex matrix-variate gamma distribution is given in the next theorem. Observe that in the complex Wishart case,  $m_j$  is an integer and hence there is no general case to consider.

**Theorem 9.6a.3.** When  $m_j = \alpha - p + r_j$  is not a positive integer, where  $r_j$  is as defined in (9.6a.3), the density of  $\lambda_1$  in the complex case, denoted by  $\tilde{f}_{21}(\lambda_1)$ , is given by

$$\tilde{f}_{21}(\lambda_1)d\lambda_1 = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)\tilde{\Gamma}_p(\alpha)} \sum_K (-1)^{\rho_K} \sum_{r_1, \dots, r_p} \Delta_{p-1}(\lambda_1) \lambda_1^{m_1} e^{-\lambda_1} d\lambda_1, \quad 0 < \lambda_1 < \infty, \tag{9.6a.7}$$

where the  $\Delta_j(\lambda_{p-j})$  has the representation specified in (ii) except that  $m_j = \alpha - p + r_j$ .

### 9.6.7. Density of the smallest eigenvalue $\lambda_p$ in the general real case

Once again, ‘general case’ is understood to mean that  $m_j = \alpha - \frac{p+1}{2} + p - k_j$  is not a positive integer, where  $k_j$  is defined in (9.6.10). For the real Wishart distribution, ‘general case’ corresponds to  $m_j$  being a half-integer. In order to determine the density of the smallest eigenvalue, we will integrate out  $\lambda_1, \dots, \lambda_{p-1}$ . We initially evaluate the following integral:

$$\begin{aligned} \text{Step 1 integral: } \int_{\lambda_1=\lambda_2}^{\infty} \lambda_1^{m_1} e^{-\lambda_1} d\lambda_1 &= \Gamma(m_1 + 1) - \int_{\lambda_1=0}^{\lambda_2} \lambda_1^{m_1} e^{-\lambda_1} d\lambda_1 \\ &= \Gamma(m_1 + 1) - \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!} \frac{1}{m_1 + \mu_1 + 1} \lambda_2^{m_1 + \mu_1 + 1}. \end{aligned} \quad (i)$$

The second step consists of integrating out  $\lambda_2$  from the expression obtained in (i) multiplied by  $\lambda_2^{m_2} e^{-\lambda_2}$ :

Step 2 integral:

$$\begin{aligned} \Gamma(m_1 + 1) \int_{\lambda_2=\lambda_3}^{\infty} \lambda_2^{m_2} e^{-\lambda_2} d\lambda_2 - \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!} \frac{1}{m_1 + \mu_1 + 1} \int_{\lambda_2=\lambda_3}^{\infty} \lambda_2^{m_1 + \mu_1 + m_2 + 1} e^{-\lambda_2} d\lambda_2 \\ = \Gamma(m_1 + 1) \Gamma(m_2 + 1) - \Gamma(m_1 + 1) \sum_{\mu_2=0}^{\infty} \frac{(-1)^{\mu_2}}{\mu_2!} \frac{1}{m_2 + \mu_2 + 1} \lambda_3^{m_2 + \mu_2 + 1} \\ - \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1! (m_1 + \mu_1 + 1)} \Gamma(m_1 + \mu_1 + m_2 + 2) \\ + \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!} \frac{1}{m_1 + \mu_1 + 1} \sum_{\mu_2=0}^{\infty} \frac{(-1)^{\mu_2}}{\mu_2!} \frac{\lambda_3^{m_1 + \mu_1 + m_2 + \mu_2 + 2}}{m_1 + \mu_1 + m_2 + \mu_2 + 2}. \end{aligned} \quad (ii)$$

A pattern is now seen to emerge. At step  $j$ , there will be  $2^j$  terms, of which  $2^{j-1}$  will start with a plus sign and  $2^{j-1}$  will start with a minus sign. All the terms at the  $j$ -th step are available from the  $2^j$  sequences of zeros and ones provided in Sect. 9.6.5. The terms can be written down by utilizing the following rules:

- (1): If the sequence starts with a zero, then the corresponding factor in the term is  $\Gamma(m_1 + 1)$ ;
- (2): If the sequence starts with a 1, then the corresponding factor in the term is  $\sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!} \frac{1}{m_1 + \mu_1 + 1}$  or this series multiplied by  $\lambda_2^{m_1 + \mu_1 + 1}$  if this 1 is the last entry in the sequence;
- (3): If the  $r$ -th entry in the sequence is a zero and the  $(r - 1)$ -th entry in the sequence is also zero, then the corresponding factor in the term is  $\Gamma(m_r + 1)$ ;

(4): If the  $r$ -th entry in the sequence is a zero and the  $(r - 1)$ -th entry in the sequence is a 1, then the corresponding factor in the term is  $\Gamma(n_{r-1} + m_r + 1)$  where  $n_{r-1}$  is the denominator factor in the  $(r - 1)$ -th factor excluding the factorial;

(5): If the  $r$ -th entry in the sequence is 1 and the  $(r - 1)$ -th entry is zero, then the corresponding factor in the term is  $\sum_{\mu_r=0}^{\infty} \frac{(-1)^{\mu_r}}{\mu_r!} \frac{1}{m_r + \mu_r + 1}$  or this series multiplied by  $\lambda_{r+1}^{m_r + \mu_r + 1}$  if this 1 happens to be the last entry in the sequence;

(6): If the  $r$ -th entry in the sequence is 1 and the  $(r - 1)$ -th entry is also 1, then the corresponding factor in the term is  $\sum_{\mu_r=0}^{\infty} \frac{(-1)^{\mu_r}}{\mu_r!} \frac{1}{n_{r-1} + m_r + \mu_r + 1}$  or this series multiplied by  $\lambda_{r+1}^{n_{r-1} + m_r + \mu_r + 1}$  if this 1 is the last entry in the sequence, where  $n_{r-1}$  is the factor appearing in the denominator of the  $(r - 1)$ -th factor excluding the factorial.

These rules enable one to write down all the terms at any step. For example for  $j = 3$ , that is, at the third step, the terms are available from the following step 3 sequences:

$$\begin{array}{cccc} 0 & 0 & 0 & + \\ 0 & 0 & 1 & - \\ 0 & 1 & 0 & - \\ 0 & 1 & 1 & + \end{array} , \quad \begin{array}{cccc} 1 & 0 & 0 & - \\ 1 & 0 & 1 & + \\ 1 & 1 & 0 & + \\ 1 & 1 & 1 & - \end{array} .$$

The terms corresponding to the sequences in the order are the following:

Step 3 integral

$$\begin{aligned} &= \Gamma(m_1 + 1)\Gamma(m_2 + 1)\Gamma(m_3 + 1) \\ &- \Gamma(m_1 + 1)\Gamma(m_2 + 1) \sum_{\mu_3=0}^{\infty} \frac{(-1)^{\mu_3}}{\mu_3!} \frac{1}{m_3 + \mu_3 + 1} \lambda_4^{m_3 + \mu_3 + 1} \\ &- \Gamma(m_1 + 1) \sum_{\mu_2=0}^{\infty} \frac{(-1)^{\mu_2}}{\mu_2!(m_2 + \mu_2 + 1)} \Gamma(m_2 + \mu_2 + m_3 + 2) \\ &+ \Gamma(m_1 + 1) \sum_{\mu_2=0}^{\infty} \frac{(-1)^{\mu_2}}{\mu_2!} \frac{1}{m_2 + \mu_2 + 1} \sum_{\mu_3=0}^{\infty} \frac{(-1)^{\mu_3}}{\mu_3!} \frac{\lambda_4^{m_2 + \mu_2 + m_3 + \mu_3 + 2}}{m_2 + \mu_2 + m_3 + \mu_3 + 2} \\ &- \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!(m_1 + \mu_1 + 1)} \Gamma(m_1 + \mu_1 + m_2 + 2) \Gamma(m_3 + 1) \\ &+ \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!(m_1 + \mu_1 + 1)} \Gamma(m_1 + \mu_1 + m_2 + 2) \sum_{\mu_3=0}^{\infty} \frac{(-1)^{\mu_3}}{\mu_3!} \frac{\lambda_4^{m_3 + \mu_3 + 1}}{(m_3 + \mu_3 + 1)} \end{aligned}$$



$$\begin{aligned}
 & + \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!(m_1 + \mu_1 + 1)} \sum_{\mu_2=0}^{\infty} \frac{(-1)^{\mu_2}}{\mu_2!(m_1 + \mu_1 + m_2 + \mu_2 + 2)} \\
 & \quad \times \Gamma(m_1 + \mu_1 + m_2 + \mu_2 + m_3 + 3) \\
 & - \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!(m_1 + \mu_1 + 1)} \sum_{\mu_2=0}^{\infty} \frac{(-1)^{\mu_2}}{\mu_2!(m_1 + \mu_1 + m_2 + \mu_2 + 2)} \\
 & \quad \times \sum_{\mu_3=0}^{\infty} \frac{(-1)^{\mu_3}}{\mu_3!} \frac{\lambda_4^{m_1+\mu_1+m_2+\mu_2+m_3+\mu_3+3}}{(m_1 + \mu_1 + m_2 + \mu_2 + m_3 + \mu_3 + 3)}. \tag{iii}
 \end{aligned}$$

Then, the step  $j$  will be the following, denoted by  $w_j(\lambda_{j+1})$ :

$$\begin{aligned}
 w_j(\lambda_{j+1}) & = \Gamma(m_1 + 1) \cdots \Gamma(m_j + 1) - \cdots + (-1)^j \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!(m_1 + \mu_1 + 1)} \cdots \\
 & \quad \times \sum_{\mu_j=0}^{\infty} \frac{(-1)^{\mu_j}}{\mu_j!} \frac{\lambda_{j+1}^{m_1+\mu_1+\cdots+m_j+\mu_j+j}}{(m_1 + \mu_1 + \cdots + m_j + \mu_j + j)}. \tag{iv}
 \end{aligned}$$

**Theorem 9.6.4.** *The density of  $\lambda_p$  for the general real matrix-variate gamma distribution, denoted by  $f_{2p}(\lambda_p)$ , is the following:*

$$f_{2p}(\lambda_p)d\lambda_p = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})\Gamma_p(\alpha)} \sum_K (-1)^{\rho_K} w_{p-1}(\lambda_p) \lambda_p^{m_p} e^{-\lambda_p} d\lambda_p, \quad 0 < \lambda_p < \infty, \tag{9.6.14}$$

where the  $w_j(\lambda_{j+1})$  is defined in (iv).

The corresponding distribution of  $\lambda_p$  for a general complex matrix-variate gamma distribution, denoted by  $\tilde{f}_{2p}(\lambda_p)$ , is the following:

**Theorem 9.6a.4.** *In the general complex case, in which instance  $m_j = \alpha - p + r_j$  is not a positive integer,  $r_j$  being as defined in (9.6a.3), the density of the smallest eigenvalue  $\lambda_p$ , denoted by  $\tilde{f}_{2p}(\lambda_p)$ , is given by*

$$\tilde{f}_{2p}(\lambda_p)d\lambda_p = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)\tilde{\Gamma}_p(\alpha)} \sum_K (-1)^{\rho_K} \sum_{r_1, \dots, r_p} w_{p-1}(\lambda_p) \lambda_p^{m_p} e^{-\lambda_p} d\lambda_p, \quad 0 < \lambda_p < \infty, \tag{9.6a.8}$$

where  $w_j(\lambda_{j+1})$  has the representation given in (iv) above for the real case, except that  $m_j = \alpha - p + r_j$ .

**Note 9.6.3.** In the complex Wishart case,  $\alpha = m$  where  $m > p - 1$  is the number of degrees of freedom, which is a positive integer. Hence, in this instance, one would simply apply Theorem 9.6a.1. It should also be observed that one can integrate out  $\lambda_1, \dots, \lambda_{j-1}$  by using the procedure described in Theorem 9.6.4 and integrate out  $\lambda_p, \dots, \lambda_{j+1}$  by employing the procedure provided in Theorem 9.6.3, and thus derive the density of  $\lambda_j$  or the joint density of any set of successive  $\lambda_j$ 's. In a similar manner, one can obtain the density of  $\lambda_j$  or the joint density of any set of successive  $\lambda_j$ 's in the complex domain by making use of the procedures outlined in Theorems 9.6a.3 and 9.6a.4.

### References

- M. Chiani (2014): Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the Tracy-Widom distribution, *Journal of Multivariate Analysis*, **129**, 69–81.
- D. S. Clemm, A. K. Chattopadhyay and P. R. Krishnaiah (1973): Upper percentage points of the individual roots of the Wishart matrix, *Sankhya, Series B*, **35**(3), 325–338.
- A. W. Davis (1972): On the marginal distributions of the latent roots of the multivariate beta matrix, *The Annals of Mathematical Statistics*, **43**(5), 1664–1670.
- A. Edelman (1991): The distribution and moments of the smallest eigenvalue of a random matrix of Wishart type, *Linear Algebra and its Applications*, **159**, 55–80.
- A. T. James (1964): Distributions of matrix variates and latent roots derived from normal samples, *The Annals of Mathematical Statistics*, **35**, 475–501.
- O. James and H.-N. Lee (2021): Concise probability distributions of eigenvalues of real-valued Wishart matrices, <https://arxiv.org/ftp/arxiv/paper/1402.6757.pdf>
- I. M. Johnstone (2001): On the distribution of the largest eigenvalue in Principal Components Analysis, *The Annals of Statistics*, **29**(2), 295–327.
- C. G. Khatri (1964): Distribution of the largest or smallest characteristic root under null hypothesis concerning complex multivariate normal populations. *The Annals of Mathematical Statistics*, **35**, 1807–1810
- P. R. Krishnaiah, F.J. Schuurmannan and V.B. Waikar (1973): Upper percentage points of the intermediate roots of the manova matrix, *Sankhya, Ser.B*, **35**(3), 339–358
- N. Kwak (2008): Principal component analysis based on L1-norm maximization, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **30**(9), 1672–1680.
- A. M. Mathai (1997): *Jacobians of Matrix Transformations and Functions of Matrix Argument*, World Scientific Publishing, New York.

A. M. Mathai and H. J. Haubold (2017a): *Linear Algebra for Physicists and Engineers*, De Gruyter, Germany.

A. M. Mathai and H. J. Haubold (2017b): *Probability and Statistics for Physicists and Engineers*, De Gruyter, Germany.

A. M. Mathai, S. B. Provost and T. Hayakawa (1995): *Bilinear Forms and Zonal Polynomials*, Springer Lecture Notes, New York.

F. Nie and H. Huang (2016): Non-greedy L21-norm maximization for principal component analysis, *arXiv:1603.08293v1[cs,LG]* 28 March 2016.

K. C. S. Pillai (1964): On the distribution of the largest seven roots of a matrix in multivariate analysis, *Biometrika*, **51**(1/2), 270–275.

J. Shi, X. Zheng and W. Yang (2017): Survey on probabilistic models of low-rank matrix factorization, *Entropy*, **19**, 424, doi:10.3390/e19080424.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

