

## Chapter 10

# Canonical Correlation Analysis



### 10.1. Introduction

We will keep utilizing the same notations in this chapter. More specifically, 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 above 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 necessary theory for the study of *Canonical Correlation Analysis* has already been introduced in Chap. 1, including the problem of optimizing a real bilinear form subject to two quadratic form constraints. This topic happens to be connected to the prediction problem. In regression analysis, the objective consists of seeking the best prediction function of a real scalar variable  $y$  based on a collection of preassigned real scalar variables  $x_1, \dots, x_k$ . It was previously determined that the regression of  $y$  on  $x_1, \dots, x_k$ , or the *best* predictor of  $y$  at preassigned values of  $x_1, \dots, x_k$ , is the conditional expectation of  $y$  at the specified values of  $x_1, \dots, x_k$ , that is,  $E[y|x_1, \dots, x_k]$  where  $E$  denotes the expected value. In this case, *best* is understood to mean 'in the minimum mean square' sense. Now,

consider the following generalization of this problem. Suppose that we wish to determine the best prediction function for a set of real scalar variables  $y_1, \dots, y_q$ , on the basis of a collection of real scalar variables  $x_1, \dots, x_p$ , where  $p$  needs not be equal to  $q$ . Since individual variables are available from linear functions of those variables, we will convert the problem into one of predicting a linear function of  $y_1, \dots, y_q$  from an arbitrary linear function of  $x_1, \dots, x_p$ , and vice versa if we are interested in determining the association between two sets of variables. Let the linear functions be  $u = \alpha_1 x_1 + \dots + \alpha_p x_p = \alpha' X$  with  $\alpha' = (\alpha_1, \dots, \alpha_p)$  and  $X' = (x_1, \dots, x_p)$  and  $v = \beta_1 y_1 + \dots + \beta_q y_q = \beta' Y$  with  $\beta' = (\beta_1, \dots, \beta_q)$  and  $Y' = (y_1, \dots, y_q)$ , where the coefficient vectors  $\alpha$  and  $\beta$  are arbitrary. Let us provide an interpretation of *best predictor* in the case of two linear functions. As a criterion, we may make use of the maximum joint scatter, that is, the joint variation in  $u$  and  $v$  as measured by the covariance between  $u$  and  $v$  or, equivalently, the maximum scale-free covariance, namely, the correlation between  $u$  and  $v$ , and optimize this joint variation. Given the properties of linear functions of real scalar variables, we obtain the variances of linear functions and covariance between linear functions as follows:  $\text{Var}(u) = \alpha' \Sigma_{11} \alpha$ ,  $\text{Var}(v) = \beta' \Sigma_{22} \beta$ ,  $\text{Cov}(u, v) = \alpha' \Sigma_{12} \beta = \beta' \Sigma_{21} \alpha$ ,  $\Sigma'_{12} = \Sigma_{21}$ , where  $\Sigma_{11} > O$  and  $\Sigma_{22} > O$  are the variance-covariance matrices of  $X$  and  $Y$ , respectively, and  $\Sigma_{12} = \Sigma'_{21}$  accounts for the covariance between  $X$  and  $Y$ . Letting the augmented vector  $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$  and its associated covariance matrix be  $\Sigma$ , we have

$$\Sigma = \text{Cov} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \text{Cov}(X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \text{Cov}(Y) \end{bmatrix} \equiv \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Our aim is to maximize  $\alpha' \Sigma_{12} \beta = \beta' \Sigma_{21} \alpha$ . When the coefficient vectors  $\alpha$  and  $\beta$  are unrestricted, the optimization of  $\alpha' \Sigma_{12} \beta$  proves meaningless since the quantity  $\alpha' \Sigma_{12} \beta$  can vary from  $-\infty$  to  $\infty$ . Consequently, we impose the constraints,  $\alpha' \Sigma_{11} \alpha = 1$  and  $\beta' \Sigma_{22} \beta = 1$ , to the coefficient vectors  $\alpha$  and  $\beta$ . Accordingly, the mathematical problem consists of optimizing  $\alpha' \Sigma_{12} \beta$  subject to  $\alpha' \Sigma_{11} \alpha = 1$  and  $\beta' \Sigma_{22} \beta = 1$ .

Letting

$$w = \alpha' \Sigma_{12} \beta - \frac{\rho_1}{2} (\alpha' \Sigma_{11} \alpha - 1) - \frac{\rho_2}{2} (\beta' \Sigma_{22} \beta - 1) \quad (i)$$

where  $\rho_1$  and  $\rho_2$  are the Lagrangian multipliers, we differentiate  $w$  with respect to  $\alpha$  and  $\beta$  and equate the resulting functions to null vectors. When differentiating with respect to  $\beta$ , we may utilize the equivalent form  $\beta' \Sigma_{21} \alpha = \alpha' \Sigma_{12} \beta$ . We then obtain the following equations:

$$\frac{\partial}{\partial \alpha} w = O \Rightarrow \Sigma_{12}\beta - \rho_1 \Sigma_{11}\alpha = O \quad (ii)$$

$$\frac{\partial}{\partial \beta} w = O \Rightarrow \Sigma_{21}\alpha - \rho_2 \Sigma_{22}\beta = O. \quad (iii)$$

On pre-multiplying (ii) by  $\alpha'$  and (iii), by  $\beta'$ , and using the fact that  $\alpha' \Sigma_{11} \alpha = 1$  and  $\beta' \Sigma_{22} \beta = 1$ , one has  $\rho_1 = \rho_2 \equiv \rho$  and  $\alpha' \Sigma_{12} \beta = \rho$ . Thus,

$$\begin{bmatrix} -\rho \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\rho \Sigma_{22} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} O \\ O \end{bmatrix} \quad (10.1.1)$$

and

$$\text{Cov}(\alpha' X, \beta' Y) = \alpha' \Sigma_{12} \beta = \beta' \Sigma_{21} \alpha = \rho. \quad (10.1.2)$$

Hence, the maximum value of  $\text{Cov}(\alpha' X, \beta' Y)$  yields the largest  $\rho$ . It follows from (ii) that  $\alpha = \frac{1}{\rho} \Sigma_{11}^{-1} \Sigma_{12} \beta$  which, once substituted in (iii) yields

$$[\Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} - \rho^2 \Sigma_{22}] \beta = O \Rightarrow [\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} - \rho^2 I] \beta = O.$$

This entails that  $\rho^2 = \lambda$ , an eigenvalue of  $B = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$  or its symmetrized form  $\Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}$ , and that  $\beta$  is a corresponding eigenvector. Similarly, by obtaining a representation of  $\beta$  from (iii), substituting it in (ii) and proceeding as above, it is seen that  $\rho^2 = \lambda$  is an eigenvalue of  $A = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$  or its symmetrized form  $\Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}$ , and that  $\alpha$  is a corresponding eigenvector. Hence manifestly, all the nonzero eigenvalues of  $A$  coincide with those of  $B$ . If  $p \leq q$  and  $\Sigma_{12}$  is of full rank  $p$ , then  $A > O$  (real positive definite) and  $B \geq O$  (real positive semi-definite), whereas if  $q \leq p$  and  $\Sigma_{21}$  is of full rank  $p$ , then  $A \geq O$  (real positive semi-definite) and  $B > O$  (real positive definite). If  $p = q$  and  $\Sigma_{12}$  is of full rank  $p$ , then  $A$  and  $B$  are both positive definite. If  $p \leq q$  and  $\Sigma_{12}$  is of full rank  $p$ , then one should start with  $A$  and compute all the  $p$  nonzero eigenvalues of  $A$  since  $A$  will be of lower order; on the other hand, if  $q \leq p$  and  $\Sigma_{21}$  is of full rank  $q$ , then one ought to begin with  $B$  and determine all the nonzero eigenvalues of  $B$ . Thus, one can obtain the common nonzero eigenvalues of  $A$  and  $B$  or their symmetrized forms by making use of one of these sets of steps. Let us denote the largest value of these common eigenvalues  $\lambda = \rho^2$  by  $\lambda_{(1)}$  and the corresponding eigenvectors with respect to  $A$  and  $B$ , by  $\alpha_{(1)}$  and  $\beta_{(1)}$ , where the eigenvectors are normalized via the constraints  $\alpha'_{(1)} \Sigma_{11} \alpha_{(1)} = 1$  and  $\beta'_{(1)} \Sigma_{22} \beta_{(1)} = 1$ . Then,  $(u_1, v_1) \equiv (\alpha'_{(1)} X, \beta'_{(1)} Y)$  is the first pair of canonical variables in the sense that  $u_1$  is the best predictor of  $v_1$  and  $v_1$  is the best predictor of  $u_1$ . Similarly, letting  $\rho^2_{(i)} = \lambda_{(i)}$  be the  $i$ -th largest common

eigenvalue of  $A$  and  $B$  and the corresponding eigenvectors such that  $\alpha'_{(i)} \Sigma_{11} \alpha_{(i)} = 1$  and  $\beta'_{(i)} \Sigma_{22} \beta_{(i)} = 1$ , be denoted by  $\alpha_{(i)}$  and  $\beta_{(i)}$ , the  $i$ -th largest correlation between  $u = \alpha'X$  and  $v = \beta'Y$  will be equal to  $\alpha'_{(i)} \Sigma_{12} \beta_{(i)} = \rho_{(i)} = \sqrt{\lambda_{(i)}}$ ,  $i = 1, \dots, p$ ,  $p$  denoting the common number of nonzero eigenvalues of  $A$  and  $B$ , and occur when  $u = u_i = \alpha'_{(i)}X$  and  $v = v_i = \beta'_{(i)}Y$ ,  $u_i$  and  $v_i$  being the  $i$ -th pair of canonical variables. Clearly,  $\text{Var}(u_i)$  and  $\text{Var}(v_i)$ ,  $i = 1, \dots, p$ , are both equal to one. Once again, *best* is taken to mean 'in the minimum mean square' sense. Hence, the following results:

**Theorem 10.1.1.** *Letting  $\Sigma$ ,  $A$ ,  $B$ ,  $\rho$ ,  $\alpha_{(i)}$ ,  $\beta_{(i)}$ ,  $u_i$  and  $v_i$  be as previously defined,*

$$\max_{\alpha' \Sigma_{11} \alpha = 1, \beta' \Sigma_{22} \beta = 1} [\alpha' \Sigma_{12} \beta] = \alpha'_{(1)} \Sigma_{12} \beta_{(1)} = \rho_{(1)} \quad (10.1.3)$$

where  $\rho_{(1)}$  is the largest  $\rho$  or the largest canonical correlation, that is, the largest correlation between the first pair of canonical variables,  $u = \alpha'X$  and  $v = \beta'Y$ , which is equal to the correlation between  $u_1$  and  $v_1$ , with  $\rho_{(1)}^2 = \lambda_{(1)}$ , the common largest eigenvalue of  $A$  and  $B$ . Similarly, we have

$$\min_{\alpha' \Sigma_{11} \alpha = 1, \beta' \Sigma_{22} \beta = 1} [\alpha' \Sigma_{12} \beta] = \alpha'_{(p)} \Sigma_{12} \beta_{(p)} = \rho_{(p)} \quad (10.1.4)$$

where  $\rho_{(p)}$ , which is the smallest nonzero value of  $\rho$  with  $\rho_{(p)}^2 = \lambda_{(p)}$ , the common smallest nonzero eigenvalue of  $A$  and  $B$ , represents the smallest canonical correlation between  $u$  and  $v$  or the correlation between  $u_p$  and  $v_p$

This maximum correlation between the linear functions  $\alpha'X$  and  $\beta'Y$  or the correlation between the best predictors  $u_1$  and  $v_1$  or the maximum value of  $\rho$  is called the *first canonical correlation* between the sets  $X$  and  $Y$  in the sense the correlation between  $u_1$  and  $v_1$  attains its maximum value. When  $p = 1$  or  $q = 1$ , the canonical correlation becomes the *multiple correlation*, and when  $p = 1$  and  $q = 1$ , it is simply the correlation between two real scalar random variables. The matrix of the nonzero eigenvalues of  $A$  and  $B$ , denoted by  $\Lambda$ , is  $\Lambda = \text{diag}(\lambda_{(1)}, \dots, \lambda_{(p)})$  when  $p \leq q$  and  $\Sigma_{12}$  is of full rank  $p$ ; otherwise,  $p$  is replaced by  $q$  in  $\Lambda$ .

It should be noted that, for instance, the canonical variable  $\beta'Y$  such that  $\beta$  satisfies the constraint  $\beta' \Sigma_{22} \beta = 1$  is identical to  $b' \Sigma_{22}^{-1/2} Y$  such that  $b'b = 1$  since  $\beta' \Sigma_{22} \beta = b' \Sigma_{22}^{-1/2} \Sigma_{22} \Sigma_{22}^{-1/2} b = b'b$ . Accordingly, letting the  $\lambda_{(i)}$ 's as well as  $A$  and  $B$  be as previously defined, our definition of a canonical variable, that is,  $u_i = \alpha'_{(i)}X$  and  $v_i = \beta'_{(i)}Y$ , coincides with the customary one, that is,  $u_i^* = a'_i \Sigma_{11}^{-1/2} X$  where  $a_i$  is the eigenvector with respect to  $B$  which is associated with  $\lambda_{(i)}$  and normalized by requiring that  $a'_i a_i = 1$ ,

and  $v_i^* = b_i' \Sigma_{22}^{-1/2} Y$  where  $b_i$  is an eigenvector with respect to  $B$  corresponding to  $\lambda_{(i)}$  and such that  $b_i' b_i = 1$ . It can be readily proved that the canonical variables  $u_1^*, \dots, u_p^*$  (or equivalently the  $u_i$ 's) are uncorrelated, as  $\text{Cov}(u_i^*, u_j^*) = a_i' \Sigma_{11}^{-1/2} \Sigma_{11} \Sigma_{11}^{-1/2} a_j = 0$  for  $i \neq j$  since the normed eigenvectors  $a_i$  are orthogonal to one another. It can be similarly established that the  $v_i^*$ 's or, equivalently, the  $v_i$ 's are uncorrelated. Clearly,  $\text{Cov}(u_i^*, u_j^*) = \text{Cov}(u_i, u_j) = 1$  and  $\text{Cov}(v_i^*, v_j^*) = \text{Cov}(v_i, v_j) = 1$ . We now demonstrate that, for  $i \neq k$ , the canonical variables,  $u_i$  and  $v_k$  are uncorrelated. First, consider the equation  $A a = \lambda a$ , that is,

$$\Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} a = \lambda a.$$

On pre-multiplying both sides by  $\Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}$ , we obtain  $B b = \lambda b$  where  $b = \Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} a$ . Thus, if  $\lambda$  and  $a$  constitute an eigenvalue-eigenvector pair for  $A$ , then  $\lambda$  and  $b$  must also form an eigenvalue-eigenvector pair for  $B$ , and vice versa with  $a = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}} b$ . By definition,  $\text{Cov}(u_i^*, v_k^*) = a_i' \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}} b_k$  where the vector  $b_k = \theta \Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} a_k$ ,  $\theta$  being a positive constant such that the Euclidean norm of  $b_k$  is one. Note that since  $b_k' b_k = \theta^2 a_k' A a_k = \theta^2 a_k' \lambda_{(k)} a_k = 1$ ,  $\theta$  must be equal to  $1/\sqrt{\lambda_{(k)}}$ . Thus,  $b_k = \Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} a_k / \sqrt{\lambda_{(k)}}$  with  $\Sigma_{11}^{-\frac{1}{2}} a_k = \alpha_{(k)}$  and  $b_k = \Sigma_{22}^{-\frac{1}{2}} \beta_{(k)}$ , which is equivalent to (iii) with  $\alpha = \alpha_{(k)}$ ,  $\beta = \beta_{(k)}$  and  $\rho = \rho_{(k)}$ , that is,  $\beta_{(k)} = \Sigma_{22}^{-1} \Sigma_{21} \alpha_{(k)} / \rho_{(k)}$ . Then,  $\text{Cov}(u_i^*, v_k^*) = a_i' \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} a_k / \sqrt{\lambda_{(k)}} =$  the  $(i, k)$ th element of  $\text{diag}(\lambda_{(1)}, \dots, \lambda_{(p)}) / \sqrt{\lambda_{(k)}}$ , which is equal to 0 whenever  $i \neq k$ . As expected,  $\text{Cov}(u_k^*, v_k^*) = \sqrt{\lambda_{(k)}} = \rho_{(k)}$ , and  $\text{Cov}(u_i, v_k) = \alpha_{(i)}' \Sigma_{12} \beta_{(k)} \stackrel{(iii)}{=} \alpha_{(i)}' \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \alpha_{(k)} / \rho_{(k)} = a_i' \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} a_k / \sqrt{\lambda_{(k)}} = \text{Cov}(u_i^*, v_k^*)$  for  $i, k = 1, \dots, p$ , assuming that  $p \leq q$  and  $\Sigma_{12}$  is of full rank; if  $p \geq q$  and  $\Sigma_{12}$  is of full rank,  $A$  and  $B$  will then share  $q$  nonzero eigenvalues.

### 10.1.1. An invariance property

An interesting property of canonical correlations is now pointed out. Consider the following nonsingular transformations of  $X$  and  $Y$ : Let  $X_1 = A_1 X$  and  $Y_1 = B_1 Y$  where  $A_1$  is a  $p \times p$  nonsingular constant matrix and  $B_1$  is a  $q \times q$  constant nonsingular matrix so that  $|A_1| \neq 0$  and  $|B_1| \neq 0$ . Now, consider the linear functions  $\alpha' X_1 = \alpha' A_1 X$  and  $\beta' Y_1 = \beta' B_1 Y$  whose variances and covariance are as follows:

$$\begin{aligned} \text{Var}(\alpha' Y_1) &= \text{Var}(\alpha' A_1 Y) = \alpha' A_1 \Sigma_{11} A_1' \alpha, \quad \text{Var}(\beta' Y_1) = \text{Var}(\beta' B_1 Y) = \beta' B_1 \Sigma_{22} B_1' \beta \\ \text{Cov}(\alpha' X_1, \beta' Y_1) &= \alpha' A_1 \Sigma_{12} B_1' \beta = \beta' B_1 \Sigma_{21} A_1' \alpha. \end{aligned}$$

On imposing the conditions  $\text{Var}(\alpha'X_1) = 1$  and  $\text{Var}(\beta'Y_1) = 1$ , and maximizing  $\text{Cov}(\alpha'X_1, \beta'Y_1)$  by means of the previously used procedure, we arrive at the equations

$$A_1 \Sigma_{12} B_1' \beta - \rho_1 A_1 \Sigma_{11} A_1' \alpha = 0 \quad (iv)$$

$$-\rho_2 B_1 \Sigma_{22} B_1' \beta + B_1 \Sigma_{21} A_1' \alpha = 0. \quad (v)$$

On pre-multiplying (iv) by  $\alpha'$  and (v) by  $\beta'$ , one has  $\rho_1 = \rho_2 \equiv \rho$ , say. Equations (iv) and (v) can then be re-expressed as

$$\begin{bmatrix} -\rho A_1 \Sigma_{11} A_1' & A_1 \Sigma_{12} B_1' \\ B_1 \Sigma_{21} A_1' & -\rho B_1 \Sigma_{22} B_1' \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \\ \begin{bmatrix} A_1 & O \\ O & B_1 \end{bmatrix} \begin{bmatrix} -\rho \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\rho \Sigma_{22} \end{bmatrix} \begin{bmatrix} A_1' & O \\ O & B_1' \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Taking the determinant of the coefficient matrix and equating it to zero to determine the roots, we have

$$\left| \begin{bmatrix} A_1 & O \\ O & B_1 \end{bmatrix} \begin{bmatrix} -\rho \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\rho \Sigma_{22} \end{bmatrix} \begin{bmatrix} A_1' & O \\ O & B_1' \end{bmatrix} \right| = 0 \Rightarrow \quad (10.1.5)$$

$$\left| \begin{bmatrix} -\rho \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\rho \Sigma_{22} \end{bmatrix} \right| = 0. \quad (10.1.6)$$

As can be seen from (10.1.6), (10.1.1) and (10.1.5) have the same roots  $\rho$ , which means that the canonical correlation  $\rho$  is invariant under nonsingular linear transformations. Observe that when  $\Sigma_{12}$  is of full rank  $p$  and  $p \leq q$ ,  $\rho_{(1)}, \dots, \rho_{(p)}$  corresponding to the nonzero roots of (10.1.1) or (10.1.6), encompasses all the canonical correlations, so that, in that case, we have a matrix of canonical correlations. Hence, the following result:

**Theorem 10.1.2.** *Let  $X$ , a  $p \times 1$  vector of real scalar random variables  $x_1, \dots, x_p$ , and  $Y$ , a  $q \times 1$  vector of real scalar random variables  $y_1, \dots, y_q$ , have a joint distribution. Then, the canonical correlations between  $X$  and  $Y$  are invariant under nonsingular linear transformations, that is, the canonical correlations between  $X$  and  $Y$  are the same as those between  $A_1 X$  and  $B_1 Y$  where  $|A_1| \neq 0$  and  $|B_1| \neq 0$ .*

## 10.2. Pairs of Canonical Variables

As previously explained,  $\lambda_{(1)}$  which denotes the largest eigenvalue of the matrix  $A = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$  or its symmetrized form  $\Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}$ , as well the largest

eigenvalue of  $B = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$  or  $\Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}$ , also turns out to be equal to  $\rho_{(1)}^2$ , the square of the largest root of equation (10.1.1). Having evaluated  $\lambda_{(1)}$ , we compute the corresponding eigenvectors  $\alpha_{(1)}$  and  $\beta_{(1)}$  and normalize them via the constraints  $\alpha'_{(1)} \Sigma_{11} \alpha_{(1)} = 1$  and  $\beta'_{(1)} \Sigma_{22} \beta_{(1)} = 1$ , which produces the first pair of canonical variables:  $(u_1, v_1) = (\alpha'_{(1)} X, \beta'_{(1)} Y)$ . We then take the second largest nonzero eigenvalue of  $A$  or  $B$ , denote it by  $\lambda_{(2)}$ , compute the corresponding eigenvectors  $\alpha_{(2)}$  and  $\beta_{(2)}$  and normalize them so that  $\alpha'_{(2)} \Sigma_{11} \alpha_{(2)} = 1$  and  $\beta'_{(2)} \Sigma_{22} \beta_{(2)} = 1$ , which yields the second pair of canonical variables:  $(u_2, v_2) = (\alpha'_{(2)} X, \beta'_{(2)} Y)$ . Continuing this process with all of the  $p$  nonzero eigenvalues if  $p \leq q$  and  $\Sigma_{12}$  is of full rank  $p$ , or with all of the  $q$  nonzero eigenvalues if  $q \leq p$  and  $\Sigma_{21}$  is of full rank  $q$ , will produce a complete set of canonical variables pairs:  $(u_i, v_i) = (\alpha'_{(i)} X, \beta'_{(i)} Y)$ ,  $i = 1, \dots, p$  or  $q$ .

Since the symmetrized forms of  $A$  and  $B$  are symmetric and nonnegative definite, all of their eigenvalues will be nonnegative and all nonzero eigenvalues will be positive. As is explained in Chapter 1 and Mathai and Haubold (2017a), all the eigenvalues of real symmetric matrices are real and for such matrices, there exists a full set of orthogonal eigenvectors whether some of the roots are repeated or not. Hence,  $\alpha'_{(j)} X$  will be uncorrelated with all the linear functions  $\alpha'_{(r)} X$ ,  $r = 1, 2, \dots, j - 1$ , and  $\beta'_{(j)} Y$  will be uncorrelated with  $\beta'_{(r)} Y$ ,  $r = 1, 2, \dots, j - 1$ . When constructing the second pair of canonical variables, we may impose the condition that the second linear functions  $\alpha' X$  and  $\beta' Y$  must be uncorrelated with the first pair  $\alpha'_{(1)} X$  and  $\beta'_{(1)} Y$ , respectively, by taking two more Lagrangian multipliers, adding the conditions  $\text{Cov}(\alpha' X, \alpha'_{(1)} X) = \alpha' \Sigma_{11} \alpha_{(1)} = 0$  and  $\beta' \Sigma_{22} \beta_{(1)} = 0$  to the optimizing function  $w$  and carrying out the optimization. We will then realize that these additional conditions are redundant and that the original optimizing equations are recovered, as was observed in the case of Principal Components. Similarly, we could incorporate the conditions  $\alpha' \Sigma_{11} \alpha_{(r)} = 0$ ,  $r = 1, \dots, j - 1$  when constructing  $\alpha_{(j)}$  and similar conditions when constructing  $\beta_{(j)}$ . However, these uncorrelatedness conditions will become redundant in the optimization procedure. Note that  $\lambda_{(1)} = \rho_{(1)}^2$  is the square of the first canonical correlation. Thus, the first canonical correlation is denoted by  $\rho_{(1)}$ . Similarly  $\lambda_{(r)} = \rho_{(r)}^2$  is the square of the  $r$ -th canonical correlation,  $r = 1, \dots, p$  when  $p \leq q$  and  $\Sigma_{12}$  is of full rank  $p$ . That is,  $\rho_{(1)}, \dots, \rho_{(p)}$ , the  $p$  nonzero roots of (10.1.1) when  $p \leq q$  and  $\Sigma_{12}$  is of full rank  $p$ , are canonical correlations,  $\rho_{(r)}$  being called the  $r$ -th canonical correlation which is the  $r$ -th largest root of the determinantal equation (10.1.1). If  $p \leq q$  and  $\Sigma_{12}$  is not of full rank  $p$ , then there will be fewer nonzero canonical correlations.

**Example 10.2.1.** Let  $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$  be a  $5 \times 1$  real vector random variable where  $X$  is  $3 \times 1$  and  $Y$  is  $2 \times 1$ . Let the covariance matrix of  $Z$  be  $\Sigma$  where

$$\Sigma = \begin{bmatrix} 3 & 1 & 0 & 1 & 0 \\ 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 3 & 1 & 1 \\ 1 & 0 & 1 & 2 & 1 \\ 0 & 1 & 1 & 1 & 2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \Sigma_{11} = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix},$$

$$\Sigma_{22} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \Sigma_{21} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad \Sigma_{12} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

Construct the pairs of canonical variables.

**Solution 10.2.1.** We need the following quantities:

$$\begin{aligned} \Sigma_{22}^{-1} &= \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad \Sigma_{11}^{-1} = \frac{1}{15} \begin{bmatrix} 6 & -3 & 0 \\ -3 & 9 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \\ \Sigma_{22}^{-1} \Sigma_{21} &= \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix}, \\ \Sigma_{11}^{-1} \Sigma_{12} &= \frac{1}{15} \begin{bmatrix} 6 & -3 & 0 \\ -3 & 9 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} = \frac{1}{15} \begin{bmatrix} 6 & -3 \\ -3 & 9 \\ 5 & 5 \end{bmatrix}; \end{aligned}$$

$$A = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21},$$

$$\begin{aligned} B &= \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} = \frac{1}{45} \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ -3 & 9 \\ 5 & 5 \end{bmatrix} \\ &= \frac{1}{45} \begin{bmatrix} 20 & -10 \\ -7 & 26 \end{bmatrix}. \end{aligned}$$

Let us compute the eigenvalues of  $B$  since it is  $2 \times 2$  whereas  $A$  is  $3 \times 3$ . The characteristic equation of  $45B$  is  $(20 - \lambda)(26 - \lambda) - 70 = 0 \Rightarrow \lambda^2 - 46\lambda + 450 = 0$ . The roots are  $\lambda_1 = 23 + \sqrt{79}$ ,  $\lambda_2 = 23 - \sqrt{79}$ . Hence, the eigenvalues of  $B$  are  $\rho_j = \frac{\lambda_j}{45}$ , that



is,  $\rho_1 = \frac{23+\sqrt{79}}{45}$ ,  $\rho_2 = \frac{23-\sqrt{79}}{45}$ . We have denoted the second set of real scalar random variables by  $Y$ ,  $Y' = [y_1, y_2]$ . An eigenvector corresponding to  $\rho_1$  is available from  $(B - \rho_1 I)Y = O$ . Since the right-hand side is null, we may omit the denominator. The first equation is then  $(-3 - \sqrt{79})y_1 - 10y_2 = 0$ . Taking  $y_1 = 1$ ,  $y_2 = -\frac{1}{10}(3 + \sqrt{79})$ . It is easily verified that these values will also satisfy the second equation in  $(B - \rho_1 I)Y = O$ . An eigenvector, denoted by  $\beta_1$ , is the following:

$$\beta_1 = \begin{bmatrix} 1 \\ -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix}.$$

We normalize  $\beta_1$  through  $\beta_1' \Sigma_{22} \beta_1 = 1$ . To this end, consider

$$\begin{aligned} \beta_1' \Sigma_{22} \beta_1 &= [1, -\frac{1}{10}(3 + \sqrt{79})] \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix} \\ &= \frac{1}{25}(79 - 2\sqrt{79}). \end{aligned}$$

Hence a normalized  $\beta_1$ , denoted by  $\beta_{(1)}$ , and the corresponding canonical variable  $v_1$  are the following:

$$\beta_{(1)} = \frac{5}{\sqrt{79 - 2\sqrt{79}}} \begin{bmatrix} 1 \\ -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix}, \quad v_1 = \frac{5}{\sqrt{79 - 2\sqrt{79}}} [y_1 - \frac{1}{10}(3 + \sqrt{79})y_2].$$

The second eigenvalue of  $B$  is  $\rho_2 = \frac{1}{45}(23 - \sqrt{79})$ . An eigenvector corresponding to  $\rho_2$  is available from the equation  $(B - \rho_2 I)Y = O$ . The second equation gives  $-7y_1 + (3 + \sqrt{79})y_2 = 0$ . Taking  $y_2 = 1$ ,  $y_1 = \frac{1}{7}(3 + \sqrt{79})$ . Hence, an eigenvector corresponding to  $\rho_2$ , denoted by  $\beta_2$ , is the following:

$$\beta_2 = \begin{bmatrix} \frac{1}{7}(3 + \sqrt{79}) \\ 1 \end{bmatrix}.$$

We normalize this vector through the constraint  $\beta_2' \Sigma_{22} \beta_2 = 1$ . Consider

$$\begin{aligned} \beta_2' \Sigma_{22} \beta_2 &= \left[ \frac{1}{7}(3 + \sqrt{79}), 1 \right] \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{7}(3 + \sqrt{79}) \\ 1 \end{bmatrix} \\ &= \frac{1}{49}(316 + 26\sqrt{79}). \end{aligned}$$

Hence, the normalized eigenvector, denoted by  $\beta_{(2)}$ , and the corresponding canonical variable  $v_2$  are

$$\beta_{(2)} = \frac{7}{\sqrt{316 + 26\sqrt{79}}} \begin{bmatrix} \frac{1}{7}(3 + \sqrt{79}) \\ 1 \end{bmatrix}, \quad v_2 = \frac{7}{\sqrt{316 + 26\sqrt{79}}} \left[ \frac{1}{7}(3 + \sqrt{79})y_1 + y_2 \right].$$

We will obtain the eigenvectors resulting from  $(A - \rho_1 I)X = O$  from the eigenvector  $\beta_1$  instead of solving the equation relating to  $A$ , as the presence of the term  $3 + \sqrt{79}$  can make the computations tedious. From equation (ii) of Sect. 10.1, we have

$$\begin{aligned}\alpha_1 &= \frac{1}{\rho_1} \Sigma_{11}^{-1} \Sigma_{12} \beta_1 \\ &= \frac{\sqrt{45}}{\sqrt{23 + \sqrt{79}}} \left( \frac{1}{15} \right) \begin{bmatrix} 6 & -3 \\ -3 & 9 \\ 5 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix} \\ &= \frac{\sqrt{45}}{15(\sqrt{23 + \sqrt{79}})} \begin{bmatrix} 6 + \frac{3}{10}(3 + \sqrt{79}) \\ -3 - \frac{9}{10}(3 + \sqrt{79}) \\ 5 - \frac{5}{10}(3 + \sqrt{79}) \end{bmatrix}.\end{aligned}$$

Let us normalize this vector by requiring that  $\alpha_1' \Sigma_{11} \alpha_1 = 1$  or  $\alpha_1' \Sigma_{11} \alpha_1 = \frac{1}{\rho_1^2} \beta_1' \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \beta_1 = \gamma_1^2$ , say:

$$\begin{aligned}\gamma_1^2 &= \frac{45}{15(23 + \sqrt{79})} \begin{bmatrix} 1, -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \\ &\quad \times \begin{bmatrix} 6 & -3 \\ -3 & 9 \\ 5 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix} \\ &= \frac{45}{15(23 + \sqrt{79})} \begin{bmatrix} 1, -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix} \begin{bmatrix} 11 & 2 \\ 2 & 14 \end{bmatrix} \begin{bmatrix} 1 \\ -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix} \\ &= \frac{45}{15(23 + \sqrt{79})(25)} [553 + 11\sqrt{79}].\end{aligned}$$

Hence, the normalized  $\alpha_1$ , denoted by  $\alpha_{(1)}$ , is the following:

$$\alpha_{(1)} = \frac{\alpha_1}{|\gamma_1|} = \frac{5}{\sqrt{(15)}\sqrt{(553 + 11\sqrt{79})}} \begin{bmatrix} 6 + \frac{3}{10}(3 + \sqrt{79}) \\ -3 - \frac{9}{10}(3 + \sqrt{79}) \\ 5 - \frac{5}{10}(3 + \sqrt{79}) \end{bmatrix},$$

so that the corresponding canonical variable is

$$\begin{aligned}u_1 &= \frac{5}{\sqrt{(15)}\sqrt{(553 + 11\sqrt{79})}} \left\{ \left[ 6 + \frac{3}{10}(3 + \sqrt{79}) \right] x_1 - \left[ 3 + \frac{9}{10}(3 + \sqrt{79}) \right] x_2 \right. \\ &\quad \left. + \left[ 5 - \frac{1}{2}(3 + \sqrt{79}) \right] x_3 \right\}.\end{aligned}$$

Now, from the formula  $\alpha_2 = \frac{1}{\rho_2} \Sigma_{11}^{-1} \Sigma_{12} \beta_2$ , we have

$$\begin{aligned} \alpha_2 &= \frac{\sqrt{45}}{\sqrt{(23 - \sqrt{79})}} \left( \frac{1}{15} \right) \begin{bmatrix} 6 & -3 \\ -3 & 9 \\ 5 & 5 \end{bmatrix} \begin{bmatrix} \frac{1}{7}(3 + \sqrt{79}) \\ 1 \end{bmatrix} \\ &= \frac{\sqrt{45}}{15(\sqrt{23 - \sqrt{79}})} \begin{bmatrix} \frac{6}{7}(3 + \sqrt{79}) - 3 \\ -\frac{3}{7}(3 + \sqrt{79}) + 9 \\ \frac{5}{7}(3 + \sqrt{79}) + 5 \end{bmatrix}. \end{aligned}$$

Let us normalize this vector via the constraint  $\alpha_2' \Sigma_{11} \alpha_2 = 1$  or

$$\alpha_2' \Sigma_{11} \alpha_2 = \frac{1}{\rho_2^2} \beta_2' \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \beta_2 = \gamma_2^2,$$

say. Thus,

$$\begin{aligned} \gamma_2^2 &= \frac{45}{15(23 - \sqrt{79})} \begin{bmatrix} \frac{1}{7}(3 + \sqrt{79}), 1 \end{bmatrix} \begin{bmatrix} 11 & 2 \\ 2 & 14 \end{bmatrix} \begin{bmatrix} \frac{1}{7}(3 + \sqrt{79}) \\ 1 \end{bmatrix} \\ &= \frac{45}{15(23 - \sqrt{79})} \left[ \frac{1}{49} (1738 + 94\sqrt{79}) \right], \end{aligned}$$

and the normalized vector  $\alpha_2$ , denoted by  $\alpha_{(2)}$ , is

$$\alpha_{(2)} = \frac{\alpha_2}{|\gamma_2|} = \frac{7}{\sqrt{15}\sqrt{(1738 + 94\sqrt{79})}} \begin{bmatrix} \frac{6}{7}(3 + \sqrt{79}) - 3 \\ -\frac{3}{7}(3 + \sqrt{79}) + 9 \\ \frac{5}{7}(3 + \sqrt{79}) + 5 \end{bmatrix},$$

so that the second canonical variable is

$$\begin{aligned} u_2 &= \frac{7}{\sqrt{15}\sqrt{(1738 + 94\sqrt{79})}} \left\{ \left[ \frac{6}{7}(3 + \sqrt{79}) - 3 \right] x_1 - \left[ \frac{3}{7}(3 + \sqrt{79}) + 9 \right] x_2 \right. \\ &\quad \left. + \left[ \frac{5}{7}(3 + \sqrt{79}) + 5 \right] x_3 \right\}. \end{aligned}$$

Hence, the canonical pairs are  $(u_1, v_1)$ ,  $(u_2, v_2)$  where  $u_j$  is the best predictor of  $v_j$  and vice versa for  $j = 1, 2$ . The pair of canonical variables  $(u_2, v_2)$  has the second largest canonical correlation. It is easy to verify that  $\text{Cov}(u_1, u_2) = 0$  and  $\text{Cov}(v_1, v_2) = 0$ .

### 10.3. Estimation of the Canonical Correlations and Canonical Variables

Consider a simple random sample of size  $n$  from a population designated by the  $(p + q) \times 1$  real vector  $\begin{pmatrix} X \\ Y \end{pmatrix}$ . Let the (corrected) sample sum of products matrix be denoted by

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad S_{11} \text{ is } p \times p, \quad S_{22} \text{ is } q \times q,$$

where  $S_{11}$  is the sample sum of products matrix corresponding to the sample from the subvector  $X$ , whose  $(i, j)$ th element is of the form  $\sum_{k=1}^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$  with the matrix  $(x_{ik})$  denoting a sample of size  $n$  from  $X$ ,  $S_{22}$  is the sample sum of products matrix corresponding to the subvector  $Y$  and  $\frac{1}{n}S_{12}$  is the sample covariance between  $X$  and  $Y$ . Thus, denoting the estimates by hats, the estimates of  $\Sigma_{11}$ ,  $\Sigma_{22}$  and  $\Sigma_{12}$  are  $\hat{\Sigma}_{11} = \frac{1}{n}S_{11}$ ,  $\hat{\Sigma}_{22} = \frac{1}{n}S_{22}$  and  $\hat{\Sigma}_{12} = \frac{1}{n}S_{12}$ , respectively. These will also be the maximum likelihood estimates if we assume normality, that is, if

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{p+q}(\nu, \Sigma), \quad \Sigma > O, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad (10.3.1)$$

where  $\Sigma_{11} = \text{Cov}(X) > O$ ,  $\Sigma_{22} = \text{Cov}(Y) > O$  and  $\Sigma_{12} = \text{Cov}(X, Y)$ . For the estimates of these submatrices, equation (10.1.1) will take the following form:

$$\begin{vmatrix} -t\hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & -t\hat{\Sigma}_{22} \end{vmatrix} = 0 \Rightarrow \begin{vmatrix} -tS_{11} & S_{12} \\ S_{21} & -tS_{22} \end{vmatrix} = 0 \quad (10.3.2)$$

where  $t$  is the sample canonical correlation; the reader may also refer to Mathai and Haubold (2017b). Letting  $\hat{\rho} = t$  be the estimated canonical correlation, whenever  $p \leq q$ ,  $t^2$  is an eigenvalue of the sample canonical correlation matrix given by

$$\hat{\Sigma}_{11}^{-\frac{1}{2}} \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-\frac{1}{2}} = S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}} = R_{11}^{-\frac{1}{2}} R_{12} R_{22}^{-1} R_{21} R_{11}^{-\frac{1}{2}}. \quad (10.3.3)$$

Note that we have chosen the symmetric format for the sample canonical correlation matrix. Observe that the sample size  $n$  is omitted from the middle expression in (10.3.3) as it gets canceled. As well, the middle expression is expressed in terms of sample correlation matrices in the last expression appearing in (10.3.3). The conversion from a sample covariance matrix to a sample correlation matrix has previously been explained. Letting  $S = (s_{ij})$  denote the sample sum of products matrix, we can write

$$S = S_1 R S_1, \quad S_1 = \text{diag}(\sqrt{s_{11}}, \dots, \sqrt{s_{p+q, p+q}}), \quad R = (r_{ij}) = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix},$$

where  $r_{ij}$  is the  $(i, j)$ th sample correlation coefficient, and for example,  $R_{11}$  is the  $p \times p$  submatrix within the  $(p + q) \times (p + q)$  matrix  $R$ . We will examine the distribution of  $t^2$  when the population covariance submatrix  $\Sigma_{12} = O$ , as well as when  $\Sigma_{12} \neq O$ , in the case of a  $(p + q)$ -variate real Gaussian population as given in (10.3.1), after considering an example to illustrate the computations of the canonical correlation  $\rho$  resulting from (10.1.1) and presenting an iterative procedure.

**Example 10.3.1.** Let  $X$  and  $Y$  be two real bivariate vector random variables and  $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$ . Consider the following simple random sample of size 5 from  $Z$ :

$$Z_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \end{bmatrix}, Z_2 = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \end{bmatrix}, Z_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \end{bmatrix}, Z_4 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, Z_5 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}.$$

Construct the sample pairs of canonical variables.

**Solution 10.3.1.** Let us use the standard notation. The sample matrix is  $\mathbf{Z} = [Z_1, \dots, Z_5]$ , the sample average  $\bar{Z} = \frac{1}{5}[Z_1 + \dots + Z_5]$ , the matrix of sample averages is  $\bar{\mathbf{Z}} = [\bar{Z}, \dots, \bar{Z}]$ , the deviation matrix is  $\mathbf{Z}_d = [Z_1 - \bar{Z}, \dots, Z_5 - \bar{Z}]$  and the sample sum of products matrix is  $S = \mathbf{Z}_d \mathbf{Z}'_d$ . These quantities are the following:

$$\mathbf{Z} = \begin{bmatrix} 1 & 2 & 2 & 0 & 0 \\ 2 & 0 & 1 & 1 & 1 \\ 1 & -1 & 0 & 1 & -1 \\ -1 & 1 & -1 & 0 & 1 \end{bmatrix}, \bar{\mathbf{Z}} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix},$$

$$\mathbf{Z}_d = \begin{bmatrix} 0 & 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 1 & -1 \\ -1 & 1 & -1 & 0 & 1 \end{bmatrix}, S = \begin{bmatrix} 4 & -1 & -1 & -1 \\ -1 & 2 & 2 & -2 \\ -1 & 2 & 4 & -3 \\ -1 & -2 & -3 & 4 \end{bmatrix},$$

where, as per our notation,

$$S_{11} = \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix}, S_{12} = \begin{bmatrix} -1 & -1 \\ 2 & -2 \end{bmatrix},$$

$$S_{21} = \begin{bmatrix} -1 & 2 \\ -1 & -2 \end{bmatrix}, S_{22} = \begin{bmatrix} 4 & -3 \\ -3 & 4 \end{bmatrix}.$$

We need to compute the following items:

$$\begin{aligned} S_{11}^{-1} &= \frac{1}{7} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}, \quad S_{22}^{-1} = \frac{1}{7} \begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix}, \\ S_{11}^{-1} S_{12} &= \frac{1}{7} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 2 & -2 \end{bmatrix} = \frac{1}{7} \begin{bmatrix} 0 & -4 \\ 7 & -9 \end{bmatrix}, \\ S_{22}^{-1} S_{21} &= \frac{1}{7} \begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ -1 & -2 \end{bmatrix} = \frac{1}{7} \begin{bmatrix} -7 & 2 \\ -7 & -2 \end{bmatrix}. \end{aligned}$$

The matrices  $A$  and  $B$  are then the following (using the same notation as for the population values for convenience):

$$\begin{aligned} A &= S_{11}^{-1} S_{12} S_{22}^{-1} S_{21} = \frac{1}{7^2} \begin{bmatrix} 0 & -4 \\ 7 & -9 \end{bmatrix} \begin{bmatrix} -7 & 2 \\ -7 & -2 \end{bmatrix} = \frac{2}{7^2} \begin{bmatrix} 14 & 4 \\ 7 & 16 \end{bmatrix}, \\ B &= S_{22}^{-1} S_{21} S_{11}^{-1} S_{12} = \frac{1}{7^2} \begin{bmatrix} -7 & 2 \\ -7 & -2 \end{bmatrix} \begin{bmatrix} 0 & -4 \\ 7 & -9 \end{bmatrix} = \frac{2}{7^2} \begin{bmatrix} 7 & 5 \\ -7 & 23 \end{bmatrix}. \end{aligned}$$

If the population covariance matrix of  $Z$  is  $\Sigma$ , an estimate of  $\Sigma$  is  $\frac{S}{n}$  where  $S$  is the sample sum of products matrix and  $n$  is the sample size, which is also the maximum likelihood estimate of  $\Sigma$  if  $Z$  is Gaussian distributed. Instead of using  $\frac{S}{n}$ , we will work with  $S$  since the normalized eigenvectors of  $S$  and  $\frac{S}{n}$  are identical, although the eigenvalues of  $\frac{S}{n}$  are  $\frac{1}{n}$  times the eigenvalues of  $S$ .

The eigenvalues of  $A$  are  $\frac{2}{7^2}$  times the solution of  $(14 - \lambda)(16 - \lambda) - 28 = 0 \Rightarrow \lambda_1 = 15 + \sqrt{29}$ ,  $\lambda_2 = 15 - \sqrt{29}$ , so that the eigenvalues of  $A$ , denoted by  $\lambda_{11}$  and  $\lambda_{12}$ , are  $\lambda_{11} = (\frac{2}{7^2})(15 + \sqrt{29})$ ,  $\lambda_{12} = \frac{2}{7^2}(15 - \sqrt{29})$ . The eigenvalues of  $B$  are  $\frac{2}{7^2}$  times the solutions of  $(7 - \nu)(23 - \nu) + 35 = 0 \Rightarrow \nu_1 = 15 + \sqrt{29}$ ,  $\nu_2 = 15 - \sqrt{29}$ , so that the eigenvalues of  $B$ , denoted by  $\nu_{21}$  and  $\nu_{22}$ , are  $\nu_{21} = \frac{2}{7^2}(15 + \sqrt{29})$ ,  $\nu_{22} = \frac{2}{7^2}(15 - \sqrt{29})$ , which, as expected, are the same as those of  $A$ . Corresponding to  $\lambda_{11}$ , an eigenvector from  $A$  is available from the equation

$$\begin{bmatrix} 14 - (15 + \sqrt{29}) & 4 \\ 7 & 16 - (15 + \sqrt{29}) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

deleting  $\frac{2}{7^2}$  from both sides. Thus, one solution is

$$\alpha_1 = \begin{bmatrix} 4/[1 + \sqrt{29}] \\ 1 \end{bmatrix}.$$

Let us normalize this vector through the constraint  $\alpha_1' S_{11} \alpha_1 = 1$ . Since

$$\alpha_1' S_{11} \alpha_1 = \left[ \frac{4}{1 + \sqrt{29}}, 1 \right] \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} \frac{4}{1 + \sqrt{29}} \\ 1 \end{bmatrix} = \frac{2}{7^2} (116 - 11\sqrt{29}),$$

the normalized eigenvector, denoted by  $\alpha_{(1)}$ , and the corresponding sample canonical variable, denoted by  $u_1$ , are

$$\alpha_{(1)} = \frac{7}{\sqrt{2}\sqrt{116 - 11\sqrt{29}}} \begin{bmatrix} \frac{4}{1 + \sqrt{29}} \\ 1 \end{bmatrix} \text{ and } u_1 = \frac{7}{\sqrt{2}\sqrt{116 - 11\sqrt{29}}} \left[ \frac{4}{1 + \sqrt{29}} x_1 + x_2 \right].$$

The eigenvalues of  $B$  are also the same as  $v_1 = 15 + \sqrt{29}$ ,  $v_2 = 15 - \sqrt{29}$ . Let us compute an eigenvector corresponding to the eigenvalue  $v_1$  obtained from  $B$ . This eigenvector can be determined from the equation

$$\begin{bmatrix} -8 - \sqrt{29} & 5 \\ -7 & 8 - \sqrt{29} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which gives one solution as

$$\beta_1 = \begin{bmatrix} 5/(8 + \sqrt{29}) \\ 1 \end{bmatrix}.$$

Let us normalize under the constraint  $\beta_1' S_{22} \beta_1 = 1$ . Since

$$\beta_1' S_{22} \beta_1 = \left[ \frac{5}{8 + \sqrt{29}}, 1 \right] \begin{bmatrix} 4 & -3 \\ -3 & 4 \end{bmatrix} \begin{bmatrix} \frac{5}{8 + \sqrt{29}} \\ 1 \end{bmatrix} = \frac{2}{7^2} (116 - 11\sqrt{29}),$$

the normalized eigenvector, denoted by  $\beta_{(1)}$ , and the corresponding canonical variable denoted by  $v_1$ , are the following:

$$\beta_{(1)} = \frac{7}{\sqrt{2}\sqrt{116 - 11\sqrt{29}}} \begin{bmatrix} \frac{5}{8 + \sqrt{29}} \\ 1 \end{bmatrix} \text{ and } v_1 = \frac{7}{\sqrt{2}\sqrt{116 - 11\sqrt{29}}} \left[ \frac{5}{8 + \sqrt{29}} y_1 + y_2 \right].$$

Therefore, one pair of canonical variables is  $(u_1, v_1)$  where  $u_1$  is the best predictor of  $v_1$  and vice versa. Now, consider  $\lambda_2 = 15 - \sqrt{29}$  and  $v_2 = 15 - \sqrt{29}$ . Proceed as in the above case to obtain the second pair of canonical variables  $(u_2, v_2)$ .

### 10.3.1. An iterative procedure

Without any loss of generality, let  $p \leq q$ . We will illustrate the procedure for the population values for convenience. Consider the matrix  $A$  as previously defined, that is,  $A = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ , and  $\rho$ , the canonical correlation which is a solution of (10.1.1) with  $\rho^2 = \lambda$  where  $\lambda$  is an eigenvalue of  $A$ . When  $p$  is small, we may directly solve the determinantal equation (10.1.1) and evaluate the roots which are the canonical correlations. We are now illustrating the computations for the population values. When  $p$  is large, direct evaluation could prove tedious without resorting to computational software packages. In this case, the following iterative procedure may be employed. Let  $\lambda$  be an eigenvalue of  $A$  and  $\alpha$  the corresponding eigenvector. We want to evaluate  $\lambda = \rho^2$  and  $\alpha$ , but we cannot solve (10.1.1) directly when  $p$  is large. In that case, take an initial trial vector  $\gamma_0$  and normalize it via the constraint  $\gamma_0' \Sigma_{11} \gamma_0 = 1$  so that  $\alpha_0' \Sigma_{11} \alpha_0 = 1$ ,  $\alpha_0$  being the normalized  $\gamma_0$ . Then,  $\alpha_0 = \frac{1}{\sqrt{\gamma_0' \Sigma_{11} \gamma_0}} \gamma_0$ . Now, consider the equation

$$A \alpha_0 = \gamma_1.$$

If  $\alpha_0$  happens to be the eigenvector  $\alpha_{(1)}$  corresponding to the largest eigenvalue  $\lambda_{(1)}$  of  $A$  then  $A \alpha_0 = \lambda_{(1)} \alpha_{(1)}$ ;  $A \alpha_0 = \gamma_1 \Rightarrow \gamma_1' \Sigma_{11} \gamma_1 = \lambda_{(1)}^2 \alpha_{(1)}' \Sigma_{11} \alpha_{(1)} = \lambda_{(1)}^2$  since  $\alpha_{(1)}' \Sigma_{11} \alpha_{(1)} = 1$ . Then  $\rho_{(1)}^2 = \lambda_{(1)} = \sqrt{\gamma_1' \Sigma_{11} \gamma_1}$ . This gives the motivation for the iterative procedure. Consider the equation

$$A \alpha_i = \gamma_{i+1}, \quad \alpha_i = \frac{1}{\sqrt{\gamma_i' \Sigma_{11} \gamma_i}} \gamma_i, \quad i = 0, 1, \dots \quad (i)$$

Continue the iteration process. At each stage compute  $\delta_i = \alpha_i' \Sigma_{11} \alpha_i$  while ensuring that  $\delta_i$  is increasing. Halt the iteration when  $\gamma_j = \gamma_{j-1}$  approximately, that is, when  $\alpha_j = \alpha_{j-1}$  approximately, which indicates that  $\gamma_j$  converges to some vector  $\gamma$ . At this stage, the normalized  $\gamma$  is  $\alpha_{(1)}$ , the eigenvector corresponding to the largest eigenvalue  $\lambda_{(1)}$  of  $A$ . Then, the largest eigenvalue  $\lambda_{(1)}$  of  $A$  is given by  $\lambda_{(1)} = \sqrt{\gamma' \Sigma_{11} \gamma}$ . Thus, as a result of the iteration process specified by equation (i),

$$\lim_{j \rightarrow \infty} \alpha_j = \alpha_{(1)} \text{ and } + \sqrt{\lim_{j \rightarrow \infty} \gamma_j' \Sigma_{11} \gamma_j} = \lambda_{(1)}. \quad (ii)$$

These initial iterations produce the largest eigenvalue  $\lambda_{(1)} = \rho_{(1)}^2$  and the corresponding eigenvector  $\alpha_{(1)}$ . From (10.1.1), we have

$$\Sigma_{22}^{-1} \Sigma_{21} \alpha = \rho \beta \Rightarrow \frac{1}{\rho} \Sigma_{22}^{-1} \Sigma_{21} \alpha = \beta. \quad (iii)$$



Substitute the computed  $\rho_{(1)}$  and  $\alpha_{(1)}$  in (iii) to obtain  $\beta_{(1)}$ , the eigenvector corresponding to the largest eigenvalue  $\lambda_{(1)}$  of  $B = \Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}$ . This completes the first stage of the iteration process. Now, consider  $A_2 = A - \lambda_{(1)} \alpha_{(1)} \alpha'_{(1)}$ . Observe that  $\alpha_{(1)} \alpha'_{(1)}$  is a  $p \times p$  matrix. In general, we can express a symmetric matrix in terms of its eigenvalues and normalized eigenvectors as follows:

$$A = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} = \lambda_{(1)} \alpha_{(1)} \alpha'_{(1)} + \lambda_{(2)} \alpha_{(2)} \alpha'_{(2)} + \cdots + \lambda_{(p)} \alpha_{(p)} \alpha'_{(p)}, \quad (iv)$$

as explained in Chapter 1 or Mathai and Haubold (2017a). Carry out the second stage of the iteration process on  $A_2$  as indicated in (i). This will produce the second largest eigenvalue  $\lambda_{(2)}$  of  $A$  and the corresponding eigenvector  $\alpha_{(2)}$ . Then, compute the corresponding  $\beta_{(2)}$  via the procedure given in (iii). This will complete the second stage. For the next stage, consider

$$A_3 = A_2 - \lambda_{(2)} \alpha_{(2)} \alpha'_{(2)} = A - \lambda_{(1)} \alpha_{(1)} \alpha'_{(1)} - \lambda_{(2)} \alpha_{(2)} \alpha'_{(2)}$$

and perform the iterative steps (i) to (iv). This will produce  $\lambda_{(3)}$ ,  $\alpha_{(3)}$  and  $\beta_{(3)}$ . Keep on iterating until all the  $p$  eigenvalues  $\lambda_{(1)}, \dots, \lambda_{(p)}$  of  $A$  as well as  $\alpha_{(j)}$  and  $\beta_{(j)}$ , the corresponding eigenvectors of  $A$  and  $B$  are obtained for  $j = 1, \dots, p$ .

In the case of sample eigenvalues and eigenvectors, start with the sample matrices

$$\begin{aligned} \hat{A} &= \hat{\Sigma}_{11}^{-\frac{1}{2}} \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-\frac{1}{2}} = R_{11}^{-\frac{1}{2}} R_{12} R_{22}^{-1} R_{21} R_{11}^{-\frac{1}{2}} \\ \hat{B} &= \hat{\Sigma}_{22}^{-\frac{1}{2}} \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-\frac{1}{2}} = R_{22}^{-\frac{1}{2}} R_{21} R_{11}^{-1} R_{12} R_{22}^{-\frac{1}{2}}. \end{aligned}$$

Carry out the iteration steps (i) to (iv) on  $\hat{A}$  to obtain the sample eigenvalues, denoted by  $\hat{\rho}_{(j)} = t_{(j)}$ ,  $j = 1, \dots, p$  for  $p \leq q$ , where  $t_{(j)}$  is the  $j$ -th sample canonical correlation, and the corresponding eigenvectors of  $\hat{A}$  denoted by  $a_{(j)}$  as well as those of  $\hat{B}$  denoted by  $b_{(j)}$ .

**Example 10.3.2.** Consider the real vectors  $X' = (x_1, x_2)$ ,  $Y' = (y_1, y_2, y_3)$ , and let  $Z' = (X', Y')$  where  $x_1, x_2, y_1, y_2, y_3$  are real scalar random variables. Let the covariance matrix of  $Z$  be  $\Sigma > O$  where

$$\begin{aligned} \text{Cov}(Z) &= \text{Cov} \begin{bmatrix} X \\ Y \end{bmatrix} = \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \Sigma_{12} = \Sigma'_{21}, \\ \text{Cov}(X) &= \Sigma_{11}, \quad \text{Cov}(Y) = \Sigma_{22}, \quad \text{Cov}(X, Y) = \Sigma_{12}, \end{aligned}$$

with

$$\Sigma_{11} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \Sigma_{12} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}, \quad \Sigma_{22} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Consider the problem of predicting  $X$  from  $Y$  and vice versa. Obtain the best predictors by constructing pairs of canonical variables.

**Solution 10.3.2.** Let us first compute the inverses  $\Sigma_{11}^{-1}$ ,  $\Sigma_{22}^{-1}$  and  $\Sigma_{11}^{-1}\Sigma_{12}$ ,  $\Sigma_{22}^{-1}\Sigma_{21}$ . We are taking the non-symmetric form of  $A$  as the symmetric form requires more calculations. Either way, the eigenvalues are identical. On directly applying the formula  $C^{-1} = \frac{1}{|C|} \times [\text{the matrix of cofactors}]'$ , we have

$$\begin{aligned}\Sigma_{11}^{-1} &= \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad \Sigma_{22}^{-1} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \\ \Sigma_{11}^{-1}\Sigma_{12} &= \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 3 & 1 \\ 1 & -3 & 1 \end{bmatrix}, \\ \Sigma_{22}^{-1}\Sigma_{21} &= \begin{bmatrix} 2 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 0 & -2 \\ 1 & 1 \end{bmatrix}.\end{aligned}$$

Thus, the non-symmetric forms of  $A$  and  $B$  are

$$\begin{aligned}A &= \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \frac{1}{3} \begin{bmatrix} 1 & 3 & 1 \\ 1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & -2 \\ 1 & 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & -2 \\ 2 & 10 \end{bmatrix}, \\ B &= \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \frac{1}{3} \begin{bmatrix} 1 & 3 \\ 0 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 1 \\ 1 & -3 & 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 4 & -6 & 4 \\ -2 & 6 & -2 \\ 2 & 0 & 2 \end{bmatrix}.\end{aligned}$$

Let us compute the eigenvalues of  $3A$ . Consider  $\begin{vmatrix} 2-\lambda & -2 \\ 2 & 10-\lambda \end{vmatrix} = 0 \Rightarrow (2-\lambda)(10-\lambda) + 4 = 0$ , which gives

$$\lambda = \frac{12 \pm \sqrt{(12)^2 - 4(24)}}{2} = 6 + 2\sqrt{3}, 6 - 2\sqrt{3},$$

the eigenvalues of  $A$  being  $\lambda_{(1)} = 2 + \frac{2}{3}\sqrt{3}$ ,  $\lambda_{(2)} = 2 - \frac{2}{3}\sqrt{3}$ . These are the squares of the canonical correlation coefficient  $\rho$  resulting from (10.1.1). Let us determine the eigenvectors corresponding to  $\lambda_{(1)}$  and  $\lambda_{(2)}$ . Our notations for the linear functions of  $X$  and  $Y$  are  $u = \alpha'X$  and  $v = \beta'Y$ ; in this case,  $\alpha' = (\alpha_1, \alpha_2)$  and  $\beta' = (\beta_1, \beta_2, \beta_3)$ . Then, the eigenvector  $\alpha$ , corresponding to  $\lambda_{(1)}$  is obtained from the equation

$$\begin{bmatrix} \frac{2}{3} - (2 + \frac{2}{3}\sqrt{3}) & -\frac{2}{3} \\ \frac{2}{3} & \frac{10}{3} - (2 + \frac{2}{3}\sqrt{3}) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \quad (i)$$

$$-(2 + \sqrt{3})\alpha_1 - \alpha_2 = 0 \Rightarrow \alpha_1 = 1, \alpha_2 = -(2 + \sqrt{3}).$$

Observe that since (i) is a singular system of linear equations, we need only consider one equation and we can preassign a value for  $\alpha_1$  or  $\alpha_2$ . Taking  $\alpha_1 = 1$ , let us normalize the resulting vector via the constraint  $\alpha' \Sigma_{11} \alpha = 1$ . Since

$$\begin{aligned} \alpha' \Sigma_{11} \alpha &= [1 \quad -(2 + \sqrt{3})] \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -(2 + \sqrt{3}) \end{bmatrix} \\ &= 12 + 6\sqrt{3} \equiv \gamma_1, \end{aligned}$$

the normalized  $\alpha$ , denoted by  $\alpha_{(1)}$ , is

$$\alpha_{(1)} = \frac{1}{\sqrt{\gamma_1}} \begin{bmatrix} 1 \\ -(2 + \sqrt{3}) \end{bmatrix} \Rightarrow u_1 = \frac{1}{\sqrt{\gamma_1}} [x_1 - (2 + \sqrt{3})x_2]. \quad (ii)$$

Now, the eigenvector corresponding to the second eigenvalue  $\lambda_{(2)}$  is such that

$$\begin{bmatrix} \frac{2}{3} - (2 - \frac{2}{3}\sqrt{3}) & -\frac{2}{3} \\ \frac{2}{3} & \frac{10}{3} - (2 - \frac{2}{3}\sqrt{3}) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow$$

$$(-2 + \sqrt{3})\alpha_1 - \alpha_2 = 0 \Rightarrow \alpha_1 = 1, \alpha_2 = -2 + \sqrt{3}.$$

Since

$$\begin{aligned} \alpha' \Sigma_{11} \alpha &= [1 \quad -2 + \sqrt{3}] \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -2 + \sqrt{3} \end{bmatrix} \\ &= 12 - 6\sqrt{3} \equiv \gamma_2, \end{aligned}$$

the normalized  $\alpha$  such that  $\alpha' \Sigma_{11} \alpha = 1$  is

$$\alpha_{(2)} = \frac{1}{\sqrt{\gamma_2}} \begin{bmatrix} 1 \\ -2 + \sqrt{3} \end{bmatrix} \Rightarrow u_2 = \frac{1}{\sqrt{\gamma_2}} [x_1 + (-2 + \sqrt{3})x_2]. \quad (iii)$$

Observe that computing the eigenvalues of  $B$  from the equation  $|B - \lambda_{(j)}I| = 0$  will be difficult. However, we know that they are  $\lambda_{(1)}$  and  $\lambda_{(2)}$  as given above, the third one being equal to zero. So, let us first verify that  $3\lambda_{(1)}$  is an eigenvalue of  $3B$ . Consider

$$\begin{aligned}
|3B - (6 + 2\sqrt{3})I| &= \begin{vmatrix} 4 - (6 + 2\sqrt{3}) & -6 & 4 \\ -2 & 6 - (6 + 2\sqrt{3}) & -2 \\ 2 & 0 & 2 - (6 + 2\sqrt{3}) \end{vmatrix} \\
&= 2 \times 2 \times 2 \begin{vmatrix} 1 & \sqrt{3} & 1 \\ -(1 + \sqrt{3}) & -3 & 2 \\ 1 & 0 & -(2 + \sqrt{3}) \end{vmatrix} \\
&= 8 \begin{vmatrix} 1 & \sqrt{3} & 1 \\ 0 & \sqrt{3} & 3 + \sqrt{3} \\ 0 & -\sqrt{3} & -(3 + \sqrt{3}) \end{vmatrix} = 0.
\end{aligned}$$

The operations performed are the following: Taking out 2 from each row; interchanging the second and the first rows; adding  $(1 + \sqrt{3})$  times the first row to the second row and adding minus one times the first row to the third row. Similarly, it can be verified that  $\lambda_{(2)}$  is also an eigenvalue of  $B$ . Moreover, since the third row of  $B$  is equal to the sum of its first two rows,  $B$  is singular, which means that the remaining eigenvalue must be zero. In Example 10.2.1, we made use of the formula resulting from (ii) of Sect. 10.1 for determining the second set of canonical variables. In this case, they will be directly computed from  $B$  to illustrate a different approach. Let us now determine the eigenvectors with respect to  $B$ , corresponding to  $\lambda_{(1)}$  and  $\lambda_{(2)}$ :

$$B = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} = \frac{1}{3} \begin{bmatrix} 4 & -6 & 4 \\ -2 & 6 & -2 \\ 2 & 0 & 2 \end{bmatrix};$$

$$(B - \lambda_{(1)}I)\beta = O$$

$$\begin{aligned}
&\Rightarrow \begin{bmatrix} \frac{4}{3} - (\frac{6}{3} + \frac{2}{3}\sqrt{3}) & -\frac{6}{3} & \frac{4}{3} \\ -\frac{2}{3} & \frac{6}{3} - (\frac{6}{3} + \frac{2}{3}\sqrt{3}) & -\frac{2}{3} \\ \frac{2}{3} & 0 & \frac{2}{3} - (\frac{6}{3} + \frac{2}{3}\sqrt{3}) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\
&\Rightarrow \begin{bmatrix} -(1 + \sqrt{3}) & -3 & 2 \\ -1 & -\sqrt{3} & -1 \\ 1 & 0 & -(2 + \sqrt{3}) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.
\end{aligned}$$

This yields the equations

$$-(1 + \sqrt{3})\beta_1 - 3\beta_2 + 2\beta_3 = 0 \quad (iv)$$

$$-\beta_1 - \sqrt{3}\beta_2 - \beta_3 = 0 \quad (v)$$

$$\beta_1 - (2 + \sqrt{3})\beta_3 = 0, \quad (vi)$$

whose solution in terms of an arbitrary  $\beta_3$  is  $\beta_2 = -(1 + \sqrt{3})\beta_3$  and  $\beta_1 = (2 + \sqrt{3})\beta_3$ . Taking  $\beta_3 = 1$ , we have the solution,  $\beta_3 = 1$ ,  $\beta_2 = -(1 + \sqrt{3})$ ,  $\beta_1 = (2 + \sqrt{3})$ . Let us normalize the resulting vector via the constraint  $\beta' \Sigma_{22} \beta = 1$ :

$$\begin{aligned} \beta' \Sigma_{22} \beta &= \begin{bmatrix} 2 + \sqrt{3} & -(1 + \sqrt{3}) & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 + \sqrt{3} \\ -(1 + \sqrt{3}) \\ 1 \end{bmatrix} \\ &= 6 + 2\sqrt{3} = \delta_1. \end{aligned}$$

Thus, the normalized  $\beta$ , denoted by  $\beta_{(1)}$ , is

$$\begin{aligned} \beta_{(1)} &= \frac{1}{\sqrt{\delta_1}} \begin{bmatrix} 2 + \sqrt{3} \\ -(1 + \sqrt{3}) \\ 1 \end{bmatrix} \Rightarrow \\ v_1 &= \frac{1}{\sqrt{\delta_1}} [(2 + \sqrt{3})y_1 - (1 + \sqrt{3})y_2 + y_3]. \quad (vii) \end{aligned}$$

Observe that we could also have utilized (iii) of Sect. 10.3.1 to evaluate  $\beta_{(1)}$  and  $\beta_{(2)}$  from  $\alpha_{(1)}$  and  $\alpha_{(2)}$ . Consider the second eigenvalue  $\lambda_{(2)} = \frac{6}{3} - \frac{2}{3}\sqrt{3}$  and the equation  $(B - \lambda_{(2)}I)\beta = O$ , that is,

$$\begin{aligned} \begin{bmatrix} \frac{4}{3} - (\frac{6}{3} - \frac{2}{3}\sqrt{3}) & -\frac{6}{3} & \frac{4}{3} \\ -\frac{2}{3} & \frac{6}{3} - (\frac{6}{3} - \frac{2}{3}\sqrt{3}) & -\frac{2}{3} \\ \frac{2}{3} & 0 & \frac{2}{3} - (\frac{6}{3} - \frac{2}{3}\sqrt{3}) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \\ \begin{bmatrix} -1 + \sqrt{3} & -3 & 2 \\ -1 & \sqrt{3} & -1 \\ 1 & 0 & -2 + \sqrt{3} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \end{aligned}$$

which leads to the equations

$$(-1 + \sqrt{3})\beta_1 - 3\beta_2 + 2\beta_3 = 0 \quad (\text{viii})$$

$$-\beta_1 + \sqrt{3}\beta_2 - \beta_3 = 0 \quad (\text{ix})$$

$$\beta_1 + (-2 + \sqrt{3})\beta_3 = 0. \quad (\text{x})$$

Thus, when  $\beta_3 = 1$ ,  $\beta_1 = 2 - \sqrt{3}$  and  $\beta_2 = \sqrt{3} - 1$ . Subject to the constraint  $\beta' \Sigma_{22} \beta = 1$ , we have

$$\begin{aligned} \beta' \Sigma_{22} \beta &= [(2 - \sqrt{3}) \quad (\sqrt{3} - 1) \quad 1] \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 - \sqrt{3} \\ \sqrt{3} - 1 \\ 1 \end{bmatrix} \\ &= 6 - 2\sqrt{3} = \delta_2. \end{aligned}$$

Hence, the normalized  $\beta$  is

$$\begin{aligned} \beta_{(2)} &= \frac{1}{\sqrt{\delta_2}} \begin{bmatrix} 2 - \sqrt{3} \\ \sqrt{3} - 1 \\ 1 \end{bmatrix} \Rightarrow \\ v_2 &= \frac{1}{\sqrt{\delta_2}} [(2 - \sqrt{3})y_1 + (\sqrt{3} - 1)y_2 + y_3]. \quad (\text{xi}) \end{aligned}$$

The reader may also verify that this solution for  $\beta_{(2)}$  is identical to that coming from (iii) of Sect. 10.3.1. Thus, the canonical pairs are the following: From (ii) and (vii), we have the first canonical pair  $(u_1, v_1)$ , the second pair  $(u_2, v_2)$  resulting from (iii) and (xi). This means  $u_1$  is the best predictor of  $v_1$  and vice versa, and that  $u_2$  is the second best predictor of  $v_2$  and vice versa.

Let us ensure that no computational errors have been committed. Consider

$$\begin{aligned} \alpha'_{(1)} \Sigma_{12} \beta_{(1)} &= \frac{1}{\sqrt{\gamma_1 \delta_1}} [1, -(2 + \sqrt{3})] \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 + \sqrt{3} \\ -(1 + \sqrt{3}) \\ 1 \end{bmatrix} \\ &= \frac{1}{\sqrt{\gamma_1 \delta_1}} 4(3 + 2\sqrt{3}), \end{aligned}$$

with

$$\gamma_1 \delta_1 = 6(2 + \sqrt{3})2(3 + \sqrt{3}) = 12(9 + 5\sqrt{3}),$$

so that

$$\begin{aligned} \frac{[\alpha'_{(1)} \Sigma_{12} \beta_{(1)}]^2}{\gamma_1 \delta_1} &= \frac{16(3 + 2\sqrt{3})^2}{12(9 + 5\sqrt{3})} \\ &= \frac{16(21 + 12\sqrt{3})}{12(9 + 5\sqrt{3})} = \frac{4(7 + 4\sqrt{3})}{9 + 3\sqrt{3}} = \frac{4(7 + 4\sqrt{3})(9 - 5\sqrt{3})}{6} \\ &= \frac{2}{3}(3 + \sqrt{3}) = 2 + \frac{2}{\sqrt{3}} = 2 + \frac{2}{3}\sqrt{3} = \lambda_{(1)} : \text{the largest eigenvalue of } A, \end{aligned}$$

which corroborates the results obtained for  $\alpha_{(1)}$ ,  $\beta_{(1)}$  and  $\lambda_{(1)}$ . Similarly, it can be verified that  $\alpha_{(2)}$ ,  $\beta_{(2)}$  and  $\lambda_{(2)}$  have been correctly computed.

#### 10.4. The Sampling Distribution of the Canonical Correlation Matrix

Consider a simple random sample of size  $n$  from  $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$ . Let the  $(p + q) \times (p + q)$  sample sum of products matrix be denoted by  $S$  and let  $Z$  have a real  $(p + q)$ -variate standard Gaussian density. Then  $S$  has a real  $(p + q)$ -variate Wishart distribution with the identity matrix as its parameter matrix and  $m = n - 1$  degrees of freedom,  $n$  being the sample size. Letting the density of  $S$  be denoted by  $f(S)$ ,

$$f(S) = \frac{|S|^{\frac{m}{2} - \frac{p+q-1}{2}}}{2^{\frac{m(p+q)}{2}} \Gamma_{p+q}(\frac{m}{2})} e^{-\frac{1}{2}\text{tr}(S)}, \quad S > O, \quad m \geq p + q. \quad (10.4.1)$$

Let us partition  $S$  as follows:

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad S_{11} \text{ is } p \times p, \quad S_{22} \text{ is } q \times q,$$

and let  $dS = dS_{11} \wedge dS_{22} \wedge dS_{12}$ . Note that  $\text{tr}(S) = \text{tr}(S_{11}) + \text{tr}(S_{22})$  and

$$\begin{aligned} |S| &= |S_{22}| |S_{11} - S_{12} S_{22}^{-1} S_{21}| \\ &= |S_{22}| |S_{11}| |I - S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}}|. \end{aligned}$$

Letting  $U = S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-\frac{1}{2}}$ ,  $dU = |S_{11}|^{-\frac{q}{2}} |S_{22}|^{-\frac{p}{2}} dS_{12}$  for fixed  $S_{11}$  and  $S_{22}$ , so that the joint density of  $S_{11}$ ,  $S_{22}$  and  $S_{12}$  is given by

$$\begin{aligned} f_1(S) dS_{11} \wedge dS_{22} \wedge dS_{12} &= \frac{|S_{11}|^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(S_{11})}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} dS_{11} \\ &\times \frac{|S_{22}|^{\frac{m}{2} - \frac{q+1}{2}} e^{-\frac{1}{2}\text{tr}(S_{22})}}{2^{\frac{mq}{2}} \Gamma_q(\frac{m}{2})} dS_{22} \\ &\times \frac{\Gamma_p(\frac{m}{2}) \Gamma_q(\frac{m}{2})}{\Gamma_{p+q}(\frac{m}{2})} |I - UU'|^{\frac{m}{2} - \frac{p+q-1}{2}} dU. \end{aligned} \quad (10.4.2)$$

It is seen from (10.4.2) that  $S_{11}$ ,  $S_{22}$  and  $U$  are mutually independently distributed, and so are  $S_{11}$ ,  $S_{22}$  and  $W = UU'$ . Further,  $S_{11} \sim W_p(m, I)$  and  $S_{22} \sim W_q(m, I)$ . Note that  $W = UU' = S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}}$  is the sample canonical correlation matrix. It follows from Theorem 4.2.3 of Chapter 4, that for  $p \leq q$  and  $U$  of full rank  $p$ ,

$$dU = \frac{\pi^{\frac{pq}{2}}}{\Gamma_p(\frac{q}{2})} |W|^{\frac{q}{2} - \frac{p+1}{2}} dW. \quad (10.4.3)$$

After integrating out  $S_{11}$  and  $S_{22}$  from (10.4.2) and substituting for  $dS_{12}$ , we obtain the following representation of the density of  $W$ :

$$f_2(W) = \frac{\pi^{\frac{pq}{2}}}{\Gamma_p(\frac{q}{2})} \frac{\Gamma_p(\frac{m}{2}) \Gamma_q(\frac{m}{2})}{\Gamma_{p+q}(\frac{m}{2})} |W|^{\frac{q}{2} - \frac{p+1}{2}} |I - W|^{\frac{m-q}{2} - \frac{p+1}{2}},$$

where

$$\frac{\Gamma_q(\frac{m}{2})}{\Gamma_{p+q}(\frac{m}{2})} = \frac{\pi^{\frac{q(q-1)}{4}}}{\pi^{\frac{(p+q)(p+q-1)}{4}}} \frac{\Gamma(\frac{m}{2}) \cdots \Gamma(\frac{m}{2} - \frac{q-1}{2})}{\Gamma(\frac{m}{2}) \cdots \Gamma(\frac{m}{2} - \frac{p+q-1}{2})} = \frac{1}{\pi^{\frac{pq}{2}} \Gamma_p(\frac{m-q}{2})}.$$

Hence, the density of  $W$  is

$$f_2(W) = \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{q}{2}) \Gamma_p(\frac{m-q}{2})} |W|^{\frac{q}{2} - \frac{p+1}{2}} |I - W|^{\frac{m-q}{2} - \frac{p+1}{2}}. \quad (10.4.4)$$

Thus, the following result:

**Theorem 10.4.1.** *Let  $Z$ ,  $S$ ,  $S_{11}$ ,  $S_{22}$ ,  $S_{12}$ ,  $U$  and  $W$  be as defined above. Then, for  $p \leq q$  and  $U$  of full rank  $p$ , the  $p \times p$  canonical correlation matrix  $W = UU'$  has the real matrix-variate type-I beta density with the parameters  $(\frac{q}{2}, \frac{m-q}{2})$  that is specified in (10.4.4) with  $m \geq p + q$ ,  $m = n - 1$ .*



When  $q \leq p$  and  $S_{21}$  is of full rank  $q$ , the canonical correlation matrix  $W_1 = U'U = S_{22}^{-\frac{1}{2}} S_{21} S_{11}^{-1} S_{12} S_{22}^{-\frac{1}{2}}$  will have the density given in (10.4.4) with  $p$  and  $q$  interchanged. Suppose that  $p \leq q$  and we would like to consider the density of  $W_1 = U'U$ . In this case  $U'U$  is real positive semi-definite as the rank of  $U$  is  $p \leq q$ . However, on expanding the following determinant in two different ways:

$$\begin{vmatrix} I_p & U \\ U' & I_q \end{vmatrix} = |I_p - UU'| = |I_q - U'U|,$$

it follows from (10.4.2) that the  $q \times q$  matrix  $U'U$  has a distribution that is equivalent to that of the  $p \times p$  matrix  $UU'$ , as given in (10.4.4). The distribution of the sample canonical correlation matrix has been derived in Mathai (1981) for a Gaussian population under the assumption that  $\Sigma_{12} \neq O$ .

**10.4.1. The joint density of the eigenvalues and eigenvectors**

Without any loss of generality, let  $p \leq q$  and  $U$  be of full rank  $p$ . Let  $W$  denote the sample canonical correlation matrix whose density is as given in (10.4.4) for the case when the population canonical matrix is a null matrix. Let the eigenvalues of  $W$  be distinct and such that  $1 > v_1 > v_2 > \dots > v_p > 0$ . Observe that  $v_j = r_{(j)}^2$  where  $r_{(j)}$ ,  $j = 1, \dots, p$  are the sample canonical correlations. For a unique  $p \times p$  orthonormal matrix  $Q$ ,  $QQ' = I$ ,  $Q'Q = I$ , we have  $Q'WQ = \text{diag}(v_1, \dots, v_p) \equiv D$ . Consider the transformation from  $W$  to  $D$  and the normalized eigenvectors of  $W$ , which constitute the columns of  $Q$ . Then, as is explained in Theorem 8.2.1 or Theorem 4.4 of Mathai (1997),

$$dW = \left[ \prod_{i < j} (v_i - v_j) \right] dD \wedge h(Q) \tag{10.4.5}$$

where  $h(Q) = \wedge[(dQ)Q']$  is the differential element associated with  $Q$ , and we have the following result:

**Theorem 10.4.2.** *The joint density of the distinct eigenvalues  $1 > v_1 > v_2 > \dots > v_p > 0$ ,  $p \leq q$ , of  $W = UU'$  whose density is specified in (10.4.4),  $U$  being assumed to be of full rank  $p$ , and the normalized eigenvectors corresponding to  $v_1, \dots, v_p$ , denoted by  $f_3(D, Q)$ , is the following:*

$$\begin{aligned} f_3(D, Q)dD \wedge h(Q) &= \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{q}{2})\Gamma_p(\frac{m-q}{2})} \left[ \prod_{j=1}^p v_j^{\frac{q-p+1}{2}} \right] \left[ \prod_{j=1}^p (1 - v_j)^{\frac{m-q}{2} - \frac{p+1}{2}} \right] \\ &\quad \times \left[ \prod_{i < j} (v_i - v_j) \right] dD \wedge h(Q) \end{aligned} \tag{10.4.6}$$

where  $h(Q)$  is as defined in (10.4.5). To obtain the joint density of the squares of the sample canonical correlations  $r_{(j)}^2$  and the corresponding canonical vectors, it suffices to replace  $v_j$  by  $r_{(j)}^2$ ,  $1 > r_{(1)}^2 > \cdots > r_{(p)}^2 > 0$ ,  $-1 < r_{(j)} < 1$ ,  $j = 1, \dots, p$ .

The joint density of the eigenvalues can be determined by integrating out  $h(Q)$  from (10.4.6) in this real case. It follows from Theorem 4.2.2 that

$$\int_{O_p} h(Q) = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})}, \quad (10.4.7)$$

this result being also stated in Mathai (1997). For the complex case, the expression on the right-hand side of (10.4.7) is  $\pi^{p(p-1)}/\tilde{\Gamma}_p(p)$ . Hence, the joint density of the eigenvalues or, equivalently, the density of  $D$  and the density of  $Q$  are the following:

**Theorem 10.4.3.** *When  $p \leq q$  and  $U$  is of full rank  $p$ , the joint density of the distinct eigenvalues  $1 > v_1 > \cdots > v_p > 0$  of the canonical correlation matrix  $W$  in (10.4.4), which is available from (10.4.6) and denoted by  $f_4(D)$ , is*

$$\begin{aligned} f_4(D) &= \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{q}{2})\Gamma_p(\frac{m-q}{2})} \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \left[ \prod_{j=1}^p v_j^{\frac{q}{2} - \frac{p+1}{2}} \right] \\ &\times \left[ \prod_{j=1}^p (1 - v_j)^{\frac{m-q}{2} - \frac{p+1}{2}} \right] \left[ \prod_{i < j} (v_i - v_j) \right], \end{aligned} \quad (10.4.8)$$

and the joint density of the normalized eigenvectors associated with  $W$ , denoted by  $f_5(Q)$ , is given by

$$f_5(Q) = \frac{\Gamma_p(\frac{p}{2})}{\pi^{\frac{p^2}{2}}} h(Q) \quad (10.4.9)$$

where  $h(Q)$  is as defined in (10.4.5).

To obtain the joint density of the squares of the sample canonical correlations  $r_{(j)}^2$ , one should replace  $v_j$  by  $r_{(j)}^2$ ,  $j = 1, 2, \dots, p$ , in (10.4.8).

**Example 10.4.1.** Verify that (10.4.8) is a density for  $p = 2$ ,  $m - q = p + 1$ , with  $q$  being a free parameter.

**Solution 10.4.1.** For  $m - q = p + 1$ ,  $p = 2$ , the right-hand side of (10.4.8) becomes

$$\begin{aligned} & \frac{\Gamma_p\left(\frac{q+p+1}{2}\right)}{\Gamma_p\left(\frac{q}{2}\right)\Gamma_p\left(\frac{p+1}{2}\right)} \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p\left(\frac{p}{2}\right)} (v_1 v_2)^{\frac{q-(p+1)}{2}} (v_1 - v_2) \\ &= \frac{\Gamma_2\left(\frac{q+3}{2}\right)}{\Gamma_2\left(\frac{q}{2}\right)\Gamma_2\left(\frac{p+1}{2}\right)} \frac{\pi^2}{\Gamma_2(1)} (v_1 v_2)^{\frac{m-3}{2}} (v_1 - v_2). \end{aligned} \quad (i)$$

The constant part simplifies as follows:

$$\frac{\Gamma_2\left(\frac{q+3}{2}\right)}{\Gamma_2\left(\frac{q}{2}\right)\Gamma_2\left(\frac{p+1}{2}\right)} \frac{\pi^2}{\Gamma_2(1)} = \frac{\Gamma\left(\frac{q+3}{2}\right)\Gamma\left(\frac{q+2}{2}\right)}{\sqrt{\pi}\left[\Gamma\left(\frac{q}{2}\right)\Gamma\left(\frac{q-1}{2}\right)\right]\left[\Gamma\left(\frac{3}{2}\right)\Gamma(1)\right]} \frac{\pi^2}{\sqrt{\pi}\Gamma(1)\Gamma\left(\frac{1}{2}\right)}; \quad (ii)$$

now, noting that

$$\Gamma\left(\frac{q+3}{2}\right) = \left(\frac{q+1}{2}\right)\left(\frac{q-1}{2}\right)\Gamma\left(\frac{q-1}{2}\right) \text{ and } \Gamma\left(\frac{q+2}{2}\right) = \frac{q}{2}\Gamma\left(\frac{q}{2}\right),$$

and substituting these values in (ii), the constant part becomes

$$\frac{\left(\frac{q+1}{2}\right)\left(\frac{q-1}{2}\right)\left(\frac{q}{2}\right)}{\sqrt{\pi}\left(\frac{1}{2}\right)\sqrt{\pi}} \frac{\pi^2}{\sqrt{\pi}\sqrt{\pi}} = \frac{(q-1)q(q+1)}{4}. \quad (iii)$$

Let us show that the total integral equals 1. The integral part is the following:

$$\begin{aligned} & \int_{v_1=0}^1 \int_{v_2=0}^{v_1} (v_1 v_2)^{\frac{q-3}{2}} (v_1 - v_2) dv_1 \wedge dv_2 \\ &= \int_0^1 v_1^{\frac{q-1}{2}} \left[ \int_{v_2=0}^{v_1} v_2^{\frac{q-3}{2}} dv_2 \right] dv_1 - \int_0^1 v_1^{\frac{q-3}{2}} \left[ \int_{v_2=0}^{v_1} v_2^{\frac{q-1}{2}} dv_2 \right] dv_1 \\ &= \int_0^1 \frac{v_1^{q-1}}{\left(\frac{q-1}{2}\right)} dv_1 - \int_0^1 \frac{v_1^{q-1}}{\left(\frac{q+1}{2}\right)} dv_1 = \frac{1}{q\left(\frac{q-1}{2}\right)} - \frac{1}{q\left(\frac{q+1}{2}\right)} \\ &= \frac{4}{(q-1)q(q+1)}. \end{aligned} \quad (iv)$$

The product of (iii) and (iv) being equal to 1, this verifies that (10.4.8) is a density for  $m - q = p + 1$ ,  $p = 2$ . This completes the computations.

### 10.4.2. Testing whether the population canonical correlations equal zero

In its symmetric form, whenever  $\Sigma_{12} = O$ , the population canonical correlation matrix is a null matrix, that is,  $\Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} = O$ . Thus, when  $\Sigma_{12} = O$ , the canonical correlations are equal to zero and vice versa. As was explained in Sect. 6.8.2, we have a one-to-one function of  $u_4$ , the likelihood ratio criterion for testing this hypothesis in the case of a Gaussian distributed population. It was established that

$$u_4 = \frac{|S|}{|S_{11}| |S_{22}|} = |I - S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}}| = |I - UU'| = |I - W| = \prod_{j=1}^p (1 - r_{(j)}^2) \quad (10.4.10)$$

where  $r_{(j)}$ ,  $j = 1, \dots, p$ , are the sample canonical correlations. It can also be seen from (10.4.2) that, when  $U$  is of full rank  $p$ ,  $U$  has a rectangular matrix-variate type-1 beta distribution and  $W = UU'$  has a real matrix-variate type-1 beta distribution. Since it has been determined in Sect. 6.8.2, that under  $H_o$ , the  $h$ -th moment of  $u_4$  for an arbitrary  $h$  is given by

$$E[u_4^h | H_o] = c \frac{\prod_{j=p+1}^{p+q} \Gamma(\frac{m}{2} - \frac{j-1}{2} + h)}{\prod_{j=1}^q \Gamma(\frac{m}{2} - \frac{j-1}{2} + h)}, \quad m = n - 1, \quad (10.4.11)$$

where  $n$  is the sample size and  $c$  is such that  $E[u_4^0 | H_o] = 1$ , the density of  $u_4$  is expressible in terms of a G-function. It was also shown in the same section that  $-n \ln u_4$  is asymptotically distributed as a real chisquare random variable having  $\frac{(p+q)(p+q-1)}{2} - \frac{p(p-1)}{2} - \frac{q(q-1)}{2} = pq$  degrees of freedom, which corresponds to the number of parameters restricted by the hypothesis  $\Sigma_{12} = O$  since there are  $pq$  free parameters in  $\Sigma_{12}$ . Thus, the following result:

**Theorem 10.4.4.** *Consider the hypothesis  $H_o: \rho_{(1)} = \dots = \rho_{(p)} = 0$ , that is, the population canonical correlations  $\rho_{(j)}$ ,  $j = 1, \dots, p$ , are all equal to zero, which is equivalent to the hypothesis  $H_o: \Sigma_{12} = O$ . Let  $u_4$  denote the  $(2/n)$ -th root of the likelihood ratio criterion for testing this hypothesis. Then, as the sample size  $n \rightarrow \infty$ , under  $H_o$ ,*

$$-n \ln u_4 = -2 \ln(\text{the likelihood ratio criterion}) \rightarrow \chi_{pq}^2, \quad (10.4.12)$$

$\chi_v^2$  denoting a real chisquare random variable having  $v$  degrees of freedom.

An illustrative numerical example has already been presented in Chap. 6.

**Note 10.1.** We have initially assumed that  $\Sigma > O$ ,  $\Sigma_{11} > O$  and  $\Sigma_{22} > O$ . However,  $\Sigma_{12} = \Sigma'_{21}$  may or may not be of full rank or some of its elements could be equal to zero. Note that  $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$  is either positive definite or positive semi-definite. Whenever  $p \leq q$  and  $\Sigma_{12}$  is of rank  $p$ ,  $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} > O$  and, in this instance, all the  $p$  canonical correlations are positive. If  $\Sigma_{12}$  is not of full rank, then some of the eigenvalues of  $W$  as previously defined, as well as the corresponding canonical correlations will be equal to zero and, in the event that  $q \leq p$ , similar statements would hold with respect to  $\Sigma_{21}$ ,  $W'$  and the resulting canonical correlations. This aspect will not be further investigated from an inferential standpoint.

**Note 10.2.** Consider the regression of  $X$  on  $Y$ , that is,  $E[X|Y]$ , when  $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$  has the following real  $(p + q)$ -variate normal distribution:

$$Z \sim N_{p+q}(\mu, \Sigma), \Sigma > O, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

$$\Sigma_{11} = \text{Cov}(X) > O \text{ is } p \times p, \Sigma_{22} = \text{Cov}(Y) > O \text{ is } q \times q.$$

Then, from equation (3.3.5), we have

$$E[X|Y] = \mu_{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(Y - \mu_{(2)})$$

where  $\mu' = (\mu'_{(1)}, \mu'_{(2)})$  and

$$\text{Cov}(X|Y) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Regression analysis is performed on the conditional space where  $Y$  is either composed of non-random real scalar variables or given values of real scalar random variables, whereas canonical correlation analysis is carried out in the entire space of  $Z$ . Clearly, these techniques involve distinct approaches. When  $Y$  is given values of random variables, then  $\Sigma_{12}$  and  $\Sigma_{22}$  can make sense. In this instance, the hypothesis  $H_o : \Sigma_{12} = O$ , in which case the regression coefficient matrix is a null matrix or, equivalently, the hypothesis that  $Y$  does not contribute to predicting  $X$ , implies that the canonical correlation matrix  $\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}$  is as well a null matrix. Accordingly, in this case, the ‘no regression’ hypothesis  $\Sigma_{12} = O$  (no contribution of  $Y$  in predicting  $X$ ) is equivalent to the hypothesis that the canonical correlations are equal to zero and vice versa.

### 10.5. The General Sampling Distribution of the Canonical Correlation Matrix

Let the  $(p + q) \times 1$  real vector random variable  $Z = \begin{bmatrix} X \\ Y \end{bmatrix} \sim N_{p+q}(\mu, \Sigma), \Sigma > O$ . Consider a simple random sample of size  $n$  from this Gaussian population and let the

sample sum of products matrix  $S$  be partitioned as in the preceding section. Let the sample canonical correlation matrix be denoted by  $R$  and the corresponding population canonical correlation, by  $P$ , that is,

$$R = S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}} \text{ and } P = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}.$$

We now examine the distribution of  $R$ , assuming that  $P \neq O$ . Letting the determinant of  $I - R$  be denoted by  $u$ , we have

$$u = |I - R| = |S_{11} - S_{12} S_{22}^{-1} S_{21}| |S_{11}|^{-1} = \frac{|S|}{|S_{11}| |S_{22}|}.$$

Thus, the  $h$ -th moment of  $u$  is

$$E[u^h] = E\left[\frac{|S|}{|S_{11}| |S_{22}|}\right]^h = E[|S|^h |S_{11}|^{-h} |S_{22}|^{-h}].$$

Since  $S$ ,  $S_{11}$  and  $S_{22}$  are functions of  $S$ , we can integrate out over the density of  $S$ , namely the Wishart density with  $m = n - 1$  degrees of freedom and parameter matrix  $\Sigma > O$ . Then for  $m \geq p + q$ ,

$$E[u^h] = \frac{1}{|2\Sigma|^{\frac{m}{2}} \Gamma_{p+q}(\frac{m}{2})} \int_{S>O} \left[\frac{|S|}{|S_{11}| |S_{22}|}\right]^h |S|^{\frac{m}{2} - \frac{p+q+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)} dS. \quad (10.5.1)$$

Let us substitute  $S$  to  $\frac{1}{2}S$  so that 2 will vanish from the factors containing 2, and let us replace  $|S_{11}|^{-h}$  and  $|S_{22}|^{-h}$  by equivalent integrals:

$$\begin{aligned} |S_{11}|^{-h} &= \frac{1}{\Gamma_p(h)} \int_{Y_1>O} |Y_1|^{h - \frac{p+1}{2}} e^{-\text{tr}(Y_2 S_{11})} dY_1, \quad \Re(h) > \frac{p-1}{2}; \\ |S_{22}|^{-h} &= \frac{1}{\Gamma_q(h)} \int_{Y_2>O} |Y_2|^{h - \frac{q+1}{2}} e^{-\text{tr}(Y_2 S_{22})} dY_2, \quad \Re(h) > \frac{q-1}{2}. \end{aligned}$$

Then,

$$\begin{aligned} E[u^h] &= \frac{1}{|\Sigma|^{\frac{m}{2}} \Gamma_{p+q}(\frac{m}{2})} \frac{1}{\Gamma_p(h) \Gamma_q(h)} \int_{Y_1>O} \int_{Y_2>O} |Y_1|^{h - \frac{p+1}{2}} |Y_2|^{h - \frac{q+1}{2}} \int_{S>O} |S|^{\frac{m}{2} + h - \frac{p+q+1}{2}} \\ &\quad \times e^{-\text{tr}(\Sigma^{-1}S + YS)} dS \wedge dY_1 \wedge dY_2 \quad (10.5.2) \end{aligned}$$

where

$$\text{tr}(YS) = \text{tr}\left\{ \begin{pmatrix} Y_1 & O \\ O & Y_2 \end{pmatrix} \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \right\} = \text{tr}(Y_1 S_{11}) + \text{tr}(Y_2 S_{22}), \quad Y = \begin{pmatrix} Y_1 & O \\ O & Y_2 \end{pmatrix}.$$

Integrating over  $S$  in (10.5.2) gives

$$E[u^h] = \frac{\Gamma_{p+q}(\frac{m}{2} + h)}{\Gamma_{p+q}(\frac{m}{2})} \frac{1}{|\Sigma|^{\frac{m}{2}} \Gamma_p(h) \Gamma_q(h)} \int_{Y_1 > 0} \int_{Y_2 > 0} |Y_1|^{h - \frac{p+1}{2}} |Y_2|^{h - \frac{q+1}{2}} \times |\Sigma^{-1} + Y|^{-(\frac{m}{2} + h)} dY_1 \wedge dY_2.$$

Let

$$\Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \Rightarrow [\Sigma^{-1} + Y] = \begin{bmatrix} \Sigma^{11} + Y_1 & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} + Y_2 \end{bmatrix}.$$

Then, the determinant can be expanded as follows:

$$\begin{aligned} |\Sigma^{-1} + Y| &= |\Sigma^{22} + Y_2| |\Sigma^{11} + Y_1 - \Sigma^{12}(\Sigma^{22} + Y_2)^{-1} \Sigma^{21}| \\ &= |\Sigma^{22} + Y_2| |Y_1 + B|, \quad B = \Sigma^{11} - \Sigma^{12}(\Sigma^{22} + Y_2)^{-1} \Sigma^{21}, \end{aligned}$$

so that

$$|\Sigma^{-1} + Y|^{-(\frac{m}{2} + h)} = |\Sigma^{22} + Y_2|^{-(\frac{m}{2} + h)} |I + B^{-1} Y_1|^{-(\frac{m}{2} + h)} |B|^{-(\frac{m}{2} + h)}.$$

Collecting the factors containing  $Y_1$  and integrating out, we have

$$\frac{1}{\Gamma_p(h)} \int_{Y_1 > 0} |Y_1|^{h - \frac{p+1}{2}} |I + B^{-1} Y_1|^{-(\frac{m}{2} + h)} dY_1 = \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{m}{2} + h)} |B|^h, \quad \Re(h) > \frac{p-1}{2},$$

and  $|B|^{-(\frac{m}{2} + h)} |B|^h = |B|^{-\frac{m}{2}}$ . Noting that

$$\begin{vmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} + Y_2 \end{vmatrix} = \begin{cases} |\Sigma^{11}| |\Sigma^{22} + Y_2 - \Sigma^{21}(\Sigma^{11})^{-1} \Sigma^{12}| \\ |\Sigma^{22} + Y_2| |\Sigma^{11} - \Sigma^{12}(\Sigma^{22} + Y_2)^{-1} \Sigma^{21}| \end{cases} = |\Sigma^{22} + Y_2| |B|,$$

$|B|$  can be expressed in the following form:

$$|B| = \frac{|\Sigma^{11}| |Y_2 + C|}{|\Sigma^{22} + Y_2|}, \quad C = \Sigma^{22} - \Sigma^{21}(\Sigma^{11})^{-1} \Sigma^{12} = \Sigma_{22}^{-1}, \quad (i)$$

so that,

$$\begin{aligned} |B|^{-\frac{m}{2}} &= |\Sigma^{11}|^{-\frac{m}{2}} |Y_2 + \Sigma_{22}^{-1}|^{-\frac{m}{2}} |Y_2 + \Sigma^{22}|^{\frac{m}{2}} \Rightarrow \\ |B|^{-\frac{m}{2}} |Y + \Sigma^{22}|^{-(\frac{m}{2} + h)} &= |\Sigma^{11}|^{-\frac{m}{2}} |Y_2 + \Sigma_{22}^{-1}|^{-\frac{m}{2}} |Y_2 + \Sigma^{22}|^{-h}. \end{aligned} \quad (ii)$$

Collecting all the factors containing  $Y_2$  and integrating out, we have the following:

$$\begin{aligned} & \frac{1}{\Gamma_q(h)} \int_{Y_2 > O} |Y_2|^{h-\frac{q+1}{2}} |Y_2 + \Sigma_{22}^{-1}|^{-\frac{m}{2}} |Y_2 + \Sigma^{22}|^{-h} dY_2 \\ &= \frac{|\Sigma_{22}|^{\frac{m}{2}+h}}{\Gamma_q(h)} \int_{Y_2 > O} |Y_2|^{h-\frac{q+1}{2}} |I + \Sigma_{22}^{\frac{1}{2}} Y_2 \Sigma_{22}^{\frac{1}{2}}|^{-\frac{m}{2}} \\ & \quad \times |\Sigma_{22}^{\frac{1}{2}} Y_2 \Sigma_{22}^{\frac{1}{2}} + \Sigma_{22}^{\frac{1}{2}} \Sigma^{22} \Sigma_{22}^{\frac{1}{2}}|^{-h} dY_2 \\ &= \frac{|\Sigma_{22}|^{\frac{m}{2}}}{\Gamma_q(h)} \int_{W > O} |W|^{h-\frac{q+1}{2}} |I + W|^{-\frac{m}{2}} |W + \Sigma_{22}^{\frac{1}{2}} \Sigma^{22} \Sigma_{22}^{\frac{1}{2}}|^{-h} dW, \quad W = \Sigma_{22}^{\frac{1}{2}} Y_2 \Sigma_{22}^{\frac{1}{2}}, \end{aligned} \tag{iii}$$

as  $W = \Sigma_{22}^{\frac{1}{2}} Y_2 \Sigma_{22}^{\frac{1}{2}} \Rightarrow dY_2 = |\Sigma_{22}|^{-\frac{q+1}{2}} dW$ . Now, letting  $W = U^{-1} - I$ , so that  $dW = |U|^{-(q+1)} dU$  with  $O < U < I$ , the expression in (iii), denoted by  $\delta$ , becomes

$$\delta = \frac{|\Sigma_{22}|^{\frac{m}{2}}}{\Gamma_q(h)} \int_{O < U < I} |U|^{\frac{m}{2}-\frac{q+1}{2}} |I - U|^{h-\frac{q+1}{2}} |I - AU|^{-h} dU$$

where  $A = I - \Sigma_{22}^{\frac{1}{2}} \Sigma^{22} \Sigma_{22}^{\frac{1}{2}}$ . Note that since  $|\Sigma|^{\frac{m}{2}} = |\Sigma_{22}|^{\frac{m}{2}} |\Sigma_{11} - \Sigma_{12}(\Sigma_{22})^{-1} \Sigma_{21}|^{\frac{m}{2}} = |\Sigma_{22}|^{\frac{m}{2}} |\Sigma^{11}|^{-\frac{m}{2}}$ ,  $|\Sigma|^{\frac{m}{2}}$  in the denominator of the constant part gets canceled out, the remaining constant expression being

$$\frac{\Gamma_{p+q}(\frac{m}{2} + h)}{\Gamma_{p+q}(\frac{m}{2})} \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{m}{2} + h)}. \tag{iv}$$

The integral part of  $\delta$  can be evaluated by making use of Euler’s representation of a Gauss’ hypergeometric function of matrix argument, which as given in formula (5.2.15) of Mathai (1997), is

$$\frac{\Gamma_q(a)\Gamma_q(c-a)}{\Gamma_q(c)} {}_2F_1(a, b; c; X) = \int_{O < Z < I} |Z|^{a-\frac{q+1}{2}} |I - Z|^{c-a-\frac{q+1}{2}} |I - XZ|^{-b} dZ$$

where  $O < Z < I$  and  $O < X < I$  are  $q \times q$  real matrices. Thus,  $\delta$  can be expressed as the follows:

$$\delta = \frac{\Gamma_q(\frac{m}{2})\Gamma_q(h)}{\Gamma_q(\frac{m}{2} + h)\Gamma_q(h)} {}_2F_1\left(\frac{m}{2}, h; \frac{m}{2} + h; I - \Sigma_{22}^{\frac{1}{2}} \Sigma^{22} \Sigma_{22}^{\frac{1}{2}}\right),$$

so that

$$E[u^h] = \frac{\Gamma_{p+q}(\frac{m}{2} + h)\Gamma_p(\frac{m}{2})}{\Gamma_{p+q}(\frac{m}{2})\Gamma_p(\frac{m}{2} + h)} \frac{\Gamma_q(\frac{m}{2})}{\Gamma_q(\frac{m}{2} + h)} {}_2F_1\left(\frac{m}{2}, h; \frac{m}{2} + h; I - \Sigma_{22}^{\frac{1}{2}} \Sigma^{22} \Sigma_{22}^{\frac{1}{2}}\right). \tag{10.5.3}$$



For  $p \leq q$ , it follows from the definition of the matrix-variate gamma function that  $\Gamma_{p+q}(\alpha) = \pi^{-p q/2} \Gamma_q(\alpha) \Gamma_p(\alpha - q/2)$ . Thus, the constant part in (10.5.3) simplifies to

$$\frac{\Gamma_p(\frac{m}{2} - \frac{q}{2} + h)}{\Gamma_p(\frac{m}{2} - \frac{q}{2})} \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{m}{2} + h)}.$$

Then,

$$E[u^h] = \frac{\Gamma_p(\frac{m-q}{2} + h)}{\Gamma_p(\frac{m-q}{2})} \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{m}{2} + h)} {}_2F_1\left(\frac{m}{2}, h; \frac{m}{2} + h; I - \Sigma_{22}^{\frac{1}{2}} \Sigma^{22} \Sigma_{22}^{\frac{1}{2}}\right) \quad (10.5.4)$$

for  $m \geq p + q$ ,  $\Re(h) > -\frac{m}{2} + \frac{p-1}{2} + \frac{q}{2}$ ,  $p \leq q$ . Had  $Y_2$  been integrated out first instead of  $Y_1$ , we would have ended up with a hypergeometric function having  $I - \Sigma_{11}^{\frac{1}{2}} \Sigma^{11} \Sigma_{11}^{\frac{1}{2}}$  as its argument, that is,

$$E[u^h] = \frac{\Gamma_p(\frac{m-q}{2} + h)}{\Gamma_p(\frac{m-q}{2})} \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{m}{2} + h)} {}_2F_1\left(\frac{m}{2}, h; \frac{m}{2} + h; I - \Sigma_{11}^{\frac{1}{2}} \Sigma^{11} \Sigma_{11}^{\frac{1}{2}}\right). \quad (10.5.5)$$

### 10.5.1. The sampling distribution of the multiple correlation coefficient

When  $p = 1$  and  $q > 1$ ,  $r_{1(1\dots q)}^2$  is equal to the square of the sample multiple correlation coefficient  $r_{1(1\dots q)}$ . In this case, the argument in (10.5.5) is a real scalar quantity that is equal to  $1 - \sigma_{11} \sigma^{11}$ , the real matrix-variate  $\Gamma_p(\cdot)$  functions are simply  $\Gamma(\cdot)$  functions and  $u = 1 - r_{1(1\dots q)}^2$ . Letting  $y = r_{1(1\dots q)}^2$ ,  $E[1 - y]^h$  is available from (10.5.5) for  $p = 1$  and the argument of the  ${}_2F_1$  hypergeometric function is then

$$1 - \sigma_{11} \sigma^{11} = 1 - \sigma_{11} (\sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21})^{-1} = -\frac{\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}{\sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}. \quad (10.5.6)$$

By taking the inverse Mellin transform of (10.5.5) for  $h = s - 1$  and  $p = 1$ , we can express the density  $f(y)$  of the square of the sample multiple correlation as follows:

$$f(y) = \frac{(1 - \rho^2)^{\frac{m}{2}} \Gamma(\frac{m}{2})}{\Gamma(\frac{m-q}{2}) \Gamma(\frac{q}{2})} y^{\frac{q}{2}-1} (1 - y)^{\frac{m-q}{2}-1} {}_2F_1\left(\frac{m}{2}, \frac{m}{2}; \frac{q}{2}; \rho^2 y\right) \quad (10.5.7)$$

where  $\rho^2$  is the population multiple correlation squared, that is,  $\rho^2 = [\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}] / \sigma_{11}$ . We can verify the result by computing the  $h$ -th moment of  $1 - y$

in (10.5.7). The  $h$ -th moment can be determined as follows by expanding the  ${}_2F_1$  function and then integrating:

$$E[1 - y]^h = \frac{(1 - \rho^2)^{\frac{m}{2}} \Gamma(\frac{m}{2})}{\Gamma(\frac{m-q}{2}) \Gamma(\frac{q}{2})} \sum_{k=0}^{\infty} \frac{(\frac{m}{2})_k (\frac{m}{2})_k (\rho^2)^k}{(\frac{q}{2})_k k!} \\ \times \int_0^1 y^{\frac{q}{2}+k-1} (1 - y)^{\frac{m-q}{2}+h-1} dy,$$

the integral part being

$$\frac{\Gamma(\frac{q}{2} + k) \Gamma(\frac{m-q}{2} + h)}{\Gamma(\frac{m}{2} + h + k)} = \frac{\Gamma(\frac{q}{2}) \Gamma(\frac{m-q}{2} + h)}{\Gamma(\frac{m}{2} + h)} \frac{(\frac{q}{2})_k}{(\frac{m}{2} + h)_k},$$

so that

$$E[1 - y]^h = (1 - \rho^2)^{\frac{m}{2}} \frac{\Gamma(\frac{m-q}{2} + h)}{\Gamma(\frac{m-q}{2})} \frac{\Gamma(\frac{m}{2})}{\Gamma(\frac{m}{2} + h)} {}_2F_1\left(\frac{m}{2}, \frac{m}{2}; \frac{m}{2} + h; \rho^2\right). \quad (10.5.8)$$

On applying the relationship,

$${}_2F_1(a, b; c; z) = (1 - z)^{-b} {}_2F_1\left(c - a, b; c; \frac{z}{z - 1}\right), \quad (10.5.9)$$

we have

$${}_2F_1\left(\frac{m}{2}, \frac{m}{2}; \frac{m}{2} + h; \rho^2\right) = (1 - \rho^2)^{-\frac{m}{2}} {}_2F_1\left(h, \frac{m}{2}; \frac{m}{2} + h; \frac{\rho^2}{\rho^2 - 1}\right),$$

with

$$\frac{\rho^2}{\rho^2 - 1} = -\frac{\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}{\sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}},$$

which agrees with (10.5.6). Observe that  $(1 - \rho^2)^{\frac{m}{2}}$  gets canceled out so that (10.5.8) agrees with (10.5.5) for  $p = 1$ .

We can also obtain a representation of the density of the sample canonical correlation matrix whose M-transform is as given in (10.5.5) for  $p \leq q$ . This can be achieved by duplicating the steps utilized for the particular case considered in this section, which yields the following density:

$$f(R) = \frac{|I - P|^{\frac{m}{2}} \Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{m-q}{2}) \Gamma_p(\frac{q}{2})} |R|^{\frac{q-p+1}{2}} |I - R|^{\frac{m-q-p+1}{2}} \times {}_2F_1\left(\frac{m}{2}, \frac{m}{2}; \frac{q}{2}; P^{\frac{1}{2}} R P^{\frac{1}{2}}\right) \tag{10.5.10}$$

where  $P = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}$  is the population canonical correlation matrix. Note that a function giving rise to a certain M-transform need not be unique. However, by making use of the Laplace transform and its inverse in the real matrix-variate case, Mathai (1981) has shown that the function specified in (10.5.10) is actually the unique density of  $R$ .

**Exercises 10**

**10.1.** In Example 10.3.2, verify that

$$\frac{[\alpha'_{(2)} \Sigma_{12} \beta_{(2)}]^2}{\gamma_2 \delta_2} = \lambda_2$$

where  $\lambda_2$  is the second largest eigenvalue of the canonical correlation matrix  $A$ .

**10.2.** In Example 10.3.2, use equation (10.1.1) or equation (ii) preceding it with  $\rho_1 = \rho_2 = \rho$  and evaluate  $\beta_{(1)}$  and  $\beta_{(2)}$  from  $\alpha_{(1)}$  and  $\alpha_{(2)}$ . Obtain  $\beta$  first, normalize it subject to the constraint  $\beta' \Sigma_{22} \beta = 1$  and then obtain  $\beta_{(1)}$  and  $\beta_{(2)}$ . Then verify the results

$$\frac{[\alpha'_{(1)} \Sigma_{12} \beta_{(1)}]^2}{\gamma_1 \delta_1} = \lambda_1 \text{ and } \frac{[\alpha'_{(2)} \Sigma_{12} \beta_{(2)}]^2}{\gamma_2 \delta_2} = \lambda_2$$

where  $\lambda_1$  and  $\lambda_2$  are the largest and second largest eigenvalues of the canonical correlation matrix  $A$ .

**10.3.** Let

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \text{Cov}(X) = \Sigma_{11} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix},$$

$$\text{Cov}(Y) = \Sigma_{22} = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}, \text{Cov}(X, Y) = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 1 & 0 \end{bmatrix},$$

where  $x_1, x_2, x_3, y_1, y_2$  are real scalar random variables. Evaluate the following where the notations of this chapter are utilized: (1): The canonical correlations  $\rho_{(1)}$  and  $\rho_{(2)}$ ; (2): The first pair of canonical variables  $(u_1, v_1)$  by direct evaluation as done in Example 10.3.2; (3): Verify that

$$\frac{[\beta'_{(1)} \Sigma_{21} \alpha_{(1)}]^2}{\gamma_1 \delta_1} = \lambda_1 : \text{the largest eigenvalue of } B$$

where  $B = \Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}$ ; (4): Evaluate the second pair of canonical variables  $(u_2, v_2)$  by using equation (10.1.1) for constructing  $\alpha_{(1)}$  and  $\alpha_{(2)}$  after obtaining  $\beta_{(1)}$  and  $\beta_{(2)}$ ; (5): Verify that

$$\frac{[\beta'_{(2)} \Sigma_{21} \alpha_{(2)}]^2}{\gamma_2 \delta_2} = \lambda_2 : \text{the second largest eigenvalue of } B.$$

**10.4.** Repeat Problem 10.3 with  $X, Y$  and their associated covariance matrices defined as follows:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \text{Cov}(X) = \Sigma_{11} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 3 \end{bmatrix},$$

$$\text{Cov}(Y) = \Sigma_{22} = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \text{Cov}(X, Y) = \Sigma_{12} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix}$$

where  $x_1, x_2, x_3, y_1, y_2, y_3$  are real scalar random variables. As well, compute the three pairs of canonical variables.

**10.5.** Show that the M-transform in (10.5.5) is available from the density specified in (10.5.10).

## References

- A.M. Mathai (1981): Distribution of the canonical correlation matrix, *Annals of the Institute of Statistical Mathematics*, **33**(A), 35–43.
- 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.

**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.

