Multivariate skewness and kurtosis measures with an application in ICA

Tõnu Kollo

Institute of Mathematical Statistics, University of Tartu, J. Liivi 2, 50409 Tartu, Estonia

Received 24 May 2006
Available online 10 March 2008

Abstract

In this paper skewness and kurtosis characteristics of a multivariate $p$-dimensional distribution are introduced. The skewness measure is defined as a $p$-vector while the kurtosis is characterized by a $p \times p$-matrix. The introduced notions are extensions of the corresponding measures of Mardia [K.V. Mardia, Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (1970) 519–530] and Móri, Rohatgi & Székely [T.F. Móri, V.K. Rohatgi, G.J. Székely, On multivariate skewness and kurtosis, Theory Probab. Appl. 38 (1993) 547–551]. Basic properties of the characteristics are examined and compared with both the above-mentioned results in the literature. Expressions for the measures of skewness and kurtosis are derived for the multivariate Laplace distribution. The kurtosis matrix is used in Independent Component Analysis (ICA) where the solution of an eigenvalue problem of the kurtosis matrix determines the transformation matrix of interest [A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, New York, 2001].

AMS 2000 subject classifications: 62E20; 62H10; 65C60

Keywords: Independent component analysis; Multivariate cumulants; Multivariate kurtosis; Multivariate moments; Multivariate skewness

1. Introduction, motivation

Since Mardia [1] introduced measures of multivariate skewness $\beta_{1,p}$ and kurtosis $\beta_{2,p}$ they have become somewhat standard characteristics of a multivariate distribution. Let $X_1$ and $X_2$ be
independent identically distributed copies of a random $p$-vector $X$ with expectation $E X = \mu$ and dispersion matrix $DX = \Sigma$.

The population measures of $p$-variate skewness and kurtosis are respectively (Mardia [1,3])

$$\beta_{1,p}(X) = E[(X_1 - \mu)'\Sigma^{-1}(X_2 - \mu)]^3; \quad (1.1)$$

$$\beta_{2,p}(X) = E[(X - \mu)'\Sigma^{-1}(X - \mu)]^2. \quad (1.2)$$

Shortcomings of Mardia’s characteristic (1.1) are carefully examined in Gutjahr et al. [4]. They point out that Mardia’s skewness measure $\beta_{1,p}$ equals zero not only in the case of multivariate normality, but also within the much wider class of elliptically symmetric distributions. Therefore using it in normality-tests must be done with precaution. A need for a different skewness measure is motivated by the following circumstances. In recent years several new distribution families have been introduced for modelling skewed data. The multivariate skew normal distribution was introduced in [5], different skew elliptical distributions are presented in the collective monograph [6]. The asymmetric multivariate Laplace distribution is examined with applications in [7] while the multivariate skew $t$-distribution is under consideration in [8]. Typically these families are characterized by three parameters: one vector as the shift parameter, another vector-parameter controlling location and skewness and a matrix scale parameter. Unfortunately parameter estimation creates problems for all these families. The maximum likelihood method can give wrong answers (the skew symmetric normal distribution) or cannot be applied as an explicit density expression is missing (the multivariate asymmetric Laplace distribution). The moment method can be applied but at least the first three moments are needed. It would be good to have a characteristic of skewness as a $p$-vector to determine one vector-parameter. Univariate Mardia characteristics (1.1) and (1.2) do not help much in this context: these characteristics can have the same numerical values for distributions with different shapes. This is demonstrated in Fig. 1 where density estimates of two bivariate Laplace distributions with the same Mardia characteristics are sketched. For both distributions $\beta_{1,p} = 5.63$ and $\beta_{2,p} = 20$.

Several suggestions and generalizations for modifying multivariate skewness and kurtosis measures have been made. For recent treatments we refer to Klar [9], who gives a thorough overview of the problems, introduces a modified skewness measure and examines asymptotic distributions of different multivariate skewness and kurtosis characteristics. Malkovich and Afifi [10] suggest examining univariate skewness and kurtosis of linear combinations of the
coordinates of the initial vector \( X \). A different approach is suggested in Móri, Rohatgi and Székely [2] who define multivariate skewness as a \( p \)-vector:

\[
s(X) = E(\|Y\|^2 Y)
\]

and multivariate kurtosis as a \( p \times p \)-matrix

\[
K(X) = E(YY'YY') - (p + 2)I_p,
\]

where

\[
Y = \Sigma^{-1/2}(X - \mu)
\]

and \( \Sigma^{1/2} \) is any symmetric square root of \( \Sigma \). The kurtosis matrix \( K(X) \) has been used in Independent Component Analysis [11]. We shall examine some properties of these characteristics later on.

One disadvantage of Mardia’s kurtosis \( \beta_{2,p} \) was pointed out by Koziol [12]. He noticed that not all fourth order mixed central moments are taken into account in (1.2) and suggested a new characteristic

\[
\beta^*_{2,p}(X) = \left\{ \sum_{i,j,k,l} E[(Y_i Y_j Y_k Y_l)]^2 \right\}^{1/2}.
\]

We are going to introduce multivariate skewness and kurtosis characteristics analogous to (1.3) and (1.4). A kurtosis characteristic as a \( p \times p \)-matrix can be applied for at least two purposes. The first is the aforementioned application in ICA. The second concerns parameter estimation. There exist several \( p \)-dimensional distribution families with two \( p \times p \) parameter matrices (see [6], Chapter 2, for example). One of them can be estimated using the sample covariance matrix, for another one we could use a \( p \times p \)-matrix constructed from the fourth order mixed moments. In our definitions all mixed third and fourth order moments will be taken into account. At the same time, compressing mixed moments is necessary as the huge matrices of the third and fourth order mixed moments are not perspicuous and do not help in solving the aforementioned estimation problems without transforming the matrices. The introduced notions will be compared with the existing ones. In Section 2 we shall give necessary notation and notions. In Section 3 we shall deal with skewness measures, and Section 4 is devoted to the kurtosis. Finally in Section 5 we apply the new kurtosis measure to determine the ICA transformation matrix.

2. Notation and notions

We need some basic notions from matrix algebra and probability theory. The Kronecker product \( A \otimes B \) of \( A : m \times n \) and \( B : p \times q \) is defined as a partitioned matrix

\[
A \otimes B = [a_{ij}B], \quad i = 1, \ldots, m; \quad j = 1, \ldots, n,
\]

\( \text{vec} \ A \) denotes an \( mn \)-vector, obtained from an \( m \times n \)-matrix \( A \) by stacking its columns in the natural order. \( K_{p,q} \) stands for the \( pq \times pq \) commutation matrix consisting of \( q \times p \)-blocks, where the \( ij \)th element in the \( jk \)th block equals 1 and all the other elements are zeros, \( i = 1, \ldots, m; \quad j = 1, \ldots, n \). For the properties of the vec-operator, the commutation matrix, the Kronecker product and related matrix algebra the interested reader is referred to Harville [13], Schott [14], Magnus and Neudecker [15] or Kollo and von Rosen [16].
In Sections 3 and 4 we define skewness and kurtosis measures with the help of the star product of matrices. The star product was introduced in [17], where some basic properties were proved.

**Definition 1.** Let \( A \) be an \( m \times n \)-matrix and \( B \) an \( mr \times ns \) partitioned matrix consisting of \( r \times s \)-blocks \( B_{ij}, i = 1, \ldots, m; j = 1, \ldots, n \). The star product \( A \star B \) of \( A \) and \( B \) is an \( r \times s \)-matrix

\[
A \star B = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} B_{ij}.
\]

In the following bold capital letters \( X, Y, Z \) stand for random vectors and lower-case bold letters \( x, y, z \) for their realizations. Lower-case letters \( t, u, v, \ldots \) are used for arbitrary constant vectors. Constant matrices are denoted by capital letters \( A, B, C, \ldots \) If \( X \) is a continuous random \( p \)-vector then the density function of \( X \) is denoted by \( f_X(x) \).

Let \( X \) be a continuous random \( p \)-vector. The characteristic function \( \varphi_X(t) \) of \( X \) is defined as an expectation:

\[
\varphi_X(t) = E \exp(it^T X), \quad t \in \mathbb{R}^p
\]

and the cumulant function \( \psi_X(t) \):

\[
\psi_X(t) = \ln \varphi_X(t).
\]

Moments and cumulants of \( X \) can be obtained by differentiation of \( \varphi_X(t) \) and \( \psi_X(t) \) respectively (see [18] or [16], for example). The third and the fourth moments of \( X \) can be presented in the form

\[
m_3(X) = E(X \otimes X' \otimes X), \quad m_4(X) = E((XX') \otimes (XX')).
\]

The corresponding central moments are

\[
\overline{m}_3(X) = E[(X - \mu) \otimes (X - \mu)'], \quad \overline{m}_4(X) = E[((X - \mu)(X - \mu))' \otimes ((X - \mu)(X - \mu))'].
\]

The third and the fourth cumulants can be expressed through the corresponding central moments (2.3) and (2.4):

\[
c_3(X) = \overline{m}_3(X), \quad c_4(X) = \overline{m}_4(X) - (I_p \otimes \Sigma + K_{p,p}) \Sigma \Sigma' - \text{vec} \Sigma \text{vec}' \Sigma.
\]

It is worth noticing that if \( Z \sim N_p(\mu, \Sigma) \), then

\[
c_4(X) = \overline{m}_4(X) - \overline{m}_4(Z).
\]

**3. Skewness**

In Mardia [3] the skewness measure (1.1) is also presented in an alternative way:

\[
\beta_{1,p}(X) = \sum_{i,j,k} (E[Y_i Y_j Y_k])^2.
\]
where \( Y = (Y_1, \ldots, Y_p)' \) is defined in (1.5). Kollo and Srivastava [19] have shown that \( \beta_{1,p} \) can be represented via the third order multivariate moments or cumulants given by (2.1) and (2.5) respectively:

\[
\beta_{1,p}(X) = \text{tr}[m_3'(Y)m_3(Y)] = \text{tr}[c_3'(Y)c_3(Y)].
\]

This means that all third order central moments of \( X \) have been used for finding \( \beta_{1,p} \). Let us write out the measure \( s(X) \) in (1.3) in terms of the coordinates of \( Y \):

\[
s(X) = E\|Y\|^2 = E[(Y'Y)Y] = \sum_{i=1}^{p} E[Y_i^2 Y] = \sum_{i=1}^{p} E\left[ \sum_{j=1}^{p} Y_i^2 Y_j \right].
\]

One can conclude that not all third order mixed moments appear in the expression of \( s(X) \). Now we shall define a skewness characteristic as a \( p \)-vector which includes all mixed moments of the third order.

**Definition 2.** Let \( X \) be a random \( p \)-vector and \( Y \) be defined by (1.5). Then a \( p \)-vector \( b(X) \) is called a skewness vector of \( X \), if

\[
b(X) = \mathbf{1}_{p \times p} \star m_3(Y).
\]

In terms of the coordinates of \( Y \) we have the following representation:

\[
b(X) = E \left[ \sum_{i,j} (Y_i Y_j)Y \right].
\]

**Definition 2** is in good agreement with Mardia’s skewness measure \( \beta_{1,p} \):

\[
\|b(X)\|^2 = \sum_{i,j,k} [E(Y_i Y_j Y_k)]^2 = \beta_{1,p}(X).
\]

Notice that a similar relation does not hold for \( s(X) \). A sample estimate \( \widehat{b}(X) \) of the skewness vector (3.1) has the following form:

\[
\widehat{b}(X) = \frac{1}{n} \mathbf{1}_{p \times p} \star \sum_{i=1}^{n} (y_i y'_i \otimes y_i),
\]

where

\[
y_i = S^{-1/2}(x_i - \overline{x}),
\]

where, in turn, \( \overline{x} \) and \( S \) are the sample mean and the sample covariance matrix of the initial sample \( (x_1, \ldots, x_n) \).

**4. Kurtosis**

In [19] it is shown that Mardia’s kurtosis measure \( \beta_{2,p}(X) \) can be represented as the trace of the fourth moment of the vector \( Y \) given by (2.2) or through the fourth cumulant (formulae (2.6))
\[ \beta_{2,p}(X) = \text{tr}[E(YY' \otimes YY')]. \]

Rewritten as a sum of expectations this expression becomes

\[ \beta_{2,p}(X) = \sum_{i,j=1}^{p} E(Y_i^2 Y_j^2). \]

This means that not all mixed fourth order moments are taken into account in the definition of Mardia’s kurtosis (1.2). This was pointed out by Koziol [12] who suggested the use of a modified characteristic \( \beta_{2,p}^*(X) \) given in (1.6) instead of Mardia’s measure \( \beta_{2,p} \). Paper [2] introduced a \( p \times p \)-matrix \( K(X) \) as a characteristic of kurtosis of a \( p \)-vector \( X \):

\[ K(X) = E(YY'YY') - (p + 2)I_p. \]

As was shown by Móri et al. [2], the trace of \( K(X) \) can be expressed with the aid of Mardia’s \( \beta_{2,p} \):

\[ \beta_{2,p}(X) = \text{tr}[K(X) + (p + 2)I_p]. \]

In order to have all mixed fourth moments included into the notion of kurtosis we define it in the following way.

**Definition 3.** Let \( X \) be a random \( p \)-vector and \( Y = \Sigma^{-1/2}(X - \mu) \), where \( E(X) = \mu; D(X) = \Sigma \). Then the kurtosis matrix of \( X \) is a \( p \times p \)-matrix

\[ B(X) = 1_p \otimes m_4(Y), \]

where \( 1_p \otimes m_4 \) is a \( p \times p \)-matrix consisting of ones.

A sample estimate \( \hat{B}(X) \) of the kurtosis matrix then is:

\[ \hat{B}(X) = \frac{1}{n} 1_p \otimes \sum_{i=1}^{n} (y_i y_i' \otimes y_i y_i'). \]

where, again,

\[ y_i = S^{-1/2}(x_i - \bar{x}), \]

and \( \bar{x} \) and \( S \) are the sample mean and the sample covariance matrix of the initial sample \( (x_1, \ldots, x_n) \).

It is straightforward to check that the [2] kurtosis matrix can be represented as a star product in the following way:

\[ K(X) = 1_p \otimes m_4(Y) - (p + 2)I_p. \]

From (2.6) and (2.7) it follows that the matrix \( K(X) \) has a representation through cumulants:

\[ K(X) = 1_p \otimes c_4(Y). \]  \hspace{1cm} (4.1)

Applying the definition of the star product we get a representation of the matrix \( B(X) \) as a sum:

\[ B(X) = \sum_{i,j=1}^{p} E(Y_i^2 Y_j^2). \]
Now the trace function can be easily calculated:

$$\text{tr}(B) = \sum_{i,j,k=1}^{p} E(Y_i Y_j Y_k^2).$$

Clearly this characteristic does not include all fourth order moments. It turns out that the norm of the matrix $B(X)$ equals Koziol’s measure of kurtosis (1.6):

$$\|B(X)\| = \sqrt{\text{tr}[B'(X)B(X)]} = \sqrt{\text{vec}'(B(X))\text{vec}(B(X))}$$

$$= \left[ \sum_{i,j,k,l=1}^{p} (E(Y_i Y_j Y_k Y_l))^2 \right]^{1/2}.$$  

**Example 1.** Let $Z \sim N_p(\mu, \Sigma)$. It follows from (2.7) that

$$m_4(Z) = (I_p^2 + K_{p,p})(\Sigma \otimes \Sigma) + \Sigma \text{vec}'\Sigma.$$  

Denote $Z_0 \sim N_p(0, I_p)$. The vector $Z_0$ is an analog of the vector $Y$ for $X$. Then

$$m_4(Z_0) = m_4(\Sigma_0) = (I_p^2 + K_{p,p}) + \text{vec}I_p\text{vec}'I_p.$$  

Let us find the kurtosis matrix $B(Z)$. By **Definition 3**

$$B(Z) = 1_{p \times p} * m_4(Z_0).$$

Applying the star product to the three terms one by one we get the following sum:

$$B(Z) = pI_p + 21_p1_p'.$$

For the cumulants we introduce a $p \times p$-matrix $C(X)$ analogous to the matrix $B(X)$:

$$C(X) = 1_{p \times p} * c_4(Y).$$

The analog of $C(X)$ in the approach of Móri et al. [2] is the matrix $K(X)$ in (4.1). As matrices of the fourth order moments and cumulants are simple functions of each other (see (2.6) and (2.7)), we can rewrite the kurtosis matrix $B(X)$ in terms of $C(X)$ using the expression of $B(Z)$ from **Example 1**:

$$B(X) = C(X) + pI_p + 21_p1_p'.$$  

Let us examine in an example how the different notions vary in the case of Laplace distributions.

**Example 2.** Consider the multivariate Laplace distribution $X \sim ML_p(\theta, \Sigma)$ in the following parametrization [19]. A $p$-vector $X$ has multivariate Laplace distribution with parameters $\theta$ and $\Sigma$ if the characteristic function of $X$ is of the form:

$$\varphi_X(t) = \frac{1}{1 + it'\theta - \frac{1}{2}(t'\theta)^2 + \frac{1}{2}t'\Sigma t}.$$  

In this parametrization $E(X) = \theta$; $D(X) = \Sigma$. 
Kollo and Srivastava [19] have shown that
\[ \beta_{1,p}(X) = a(a^2 - 6a + 3(p + 2)); \tag{4.4} \]
\[ \beta_{2,p}(X) = 2(p + a)(p + 2) - 3a^2, \tag{4.5} \]
where \( a = \theta' \Sigma^{-1} \theta \). The measures of skewness \( s(X) \) and kurtosis \( K(X) \) (formulae (1.3) and (1.4)) of Móri et al. [2] can be found after some algebra:
\[ s(X) = (p + 2 - a)c; \tag{4.6} \]
\[ K(X) = (2p + 4 + a)I_p + (p + 4 - 3a)cc', \tag{4.7} \]
where \( a \) is the same as in (4.4) and (4.5) and \( c = \Sigma^{-1/2} \theta \). The new characteristics of skewness (3.1) and kurtosis (4.3) are of the form
\[ b(X) = (p - tr(D))c; \tag{4.8} \]
\[ B(X) = (2p + tr(D))I_p + 4I_{p \times p} + 2(D + D') + (p - 3tr(D))cc', \tag{4.9} \]
where \( a \) and \( c \) are defined as above and \( D = 1_{p \times p}cc' \). To see the differences between the different characteristics let us consider the bivariate distribution \( X \sim \text{ML}_2(\theta, \Sigma) \). We first take a symmetric distribution with parameters
\[ \theta = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 1 \end{pmatrix}. \]
The different skewness and kurtosis measures have the following values:
Mardia ((4.4) and (4.5)):
\[ \beta_{12}(X) = 0, \quad \beta_{22}(X) = 16; \]
Móri, Rohatgi and Székely ((4.6) and (4.7)):
\[ s(X) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad K(X) = \begin{pmatrix} 8 & 0 \\ 0 & 8 \end{pmatrix}; \]
new characteristics ((4.8) and (4.9)):
\[ b(X) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad B(X) = \begin{pmatrix} 8 & 4 \\ 4 & 8 \end{pmatrix}. \]
Next consider an asymmetric distribution with parameters
\[ \theta = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 1.5 & 1.5 \\ 1.5 & 3 \end{pmatrix}. \]
Then the same characteristics are the following:
Mardia ((4.4) and (4.5)):
\[ \beta_{12}(X) = 5.63, \quad \beta_{22}(X) = 20; \]
Móri, Rohatgi & Székely ((4.6) and (4.7)):
\[ s(X) = \begin{pmatrix} -2.43 \\ -1.22 \end{pmatrix} \quad \text{and} \quad K(X) = \begin{pmatrix} 10.8 & 1.07 \\ 1.07 & 9.2 \end{pmatrix}; \]
new characteristics ((4.8) and (4.9)):

\[ b(X) = \begin{pmatrix} -2.78 \\ 2.48 \end{pmatrix} \quad \text{and} \quad B(X) = \begin{pmatrix} 11.55 & 5.97 \\ 5.97 & 10.59 \end{pmatrix}. \]

5. Application to ICA

Independent Component Analysis (ICA) is a method for finding initial components from multivariate statistical data [11]. In the basic form the ICA model can be presented in the following way. Denote an unknown \( p \)-dimensional multivariate signal by \( S \), an unknown mixing \( p \times p \)-matrix by \( A \) and a known centred \( p \)-vector by \( X \). It is assumed that the coordinates of the initial vector \( S \) are independent and non-normally distributed. The aim is to find the unknown \( S \) and \( A \). To simplify the problem we assume that our data is whitened before analysis, i.e. instead of the \( p \)-vector \( X \) we have the \( p \)-vector \( Z \) with uncorrelated coordinates \( Z_i, DZ_i = 1 \):

\[ Z = VX. \]

Through our initial vector \( S \) we have the representation

\[ Z = VAS = W'S, \]

where \( W \) is orthogonal. To solve the problem we have to find the orthogonal transformation matrix \( W \). Then the initial signal will be detected up to the order of coordinates, as

\[ S = WZ. \]

In [11] several methods are suggested for finding the orthogonal matrix \( W \). One of them is based on the tensor of the fourth order cumulants (Hyvärinen, et al. [11], Chapter 11). In matrix representation the method can be reduced to the following eigenvalue problem:

\[ K(Z)W = WA; \]
\[ W'W = I_p, \]

where \( K(Z) \) is the [2] kurtosis matrix in (1.4). As explained by Hyvärinen, et al. [11] this simple method does not always give a satisfactory solution. One important situation when the method does not work is the case of identically distributed random variables \( S_i \). Then \( (p - 1) \) eigenvalues in \( A \) are equal and it is not possible to identify the original signal \( S \) by the orthogonal matrix \( W \) of the eigenvectors of \( K(Z) \). We guess that a complication may arise because not all mixed fourth order cumulants have been taken into account in the matrix \( K(Z) \). Our aim is to work out an analog of the method based on the eigenvalue problem described above. Instead of the matrix \( K(Z) \) we start from the matrix \( C(Z) \) which includes all fourth order mixed cumulants. Fortunately in this set-up we are able to find the desired transformation in the case of identically distributed signals although we end up with a different eigenvalue problem.

Summarizing, we have the following starting point:

\[ Z = W'S, \]

where \( W \) is orthogonal. The coordinates of \( S \) are independent and \( E(Z) = 0, D(Z) = I_p \). The fourth order moment of \( Z \) equals:

\[ m_4(Z) = E(Z'Z' \otimes ZZ') = E[(W'SS'W) \otimes (W'SS'W)] \]
\[ E[(W \otimes W)'(SS' \otimes SS')(W \otimes W)] = (W \otimes W)'m_4(S)(W \otimes W). \]

A similar equality holds for the fourth order cumulant matrix, defined in (2.6):

\[ c_4(Z) = (W \otimes W)'c_4(S)(W \otimes W). \]

Because of independence of the coordinates of \( S \) we have \( c_4(S) \) in a very simple form: the \( p \times p \)-matrix \( c_4(S) \) is block-diagonal with the \( i \)th diagonal block of the form \( c_4(S_i)(e_i)_d \), \( i = 1, \ldots, p \) (the notation \( (e_i)_d \) is explained before the next lemma).

Applying equality (4.2) we get the matrix

\[ C(Z) = \sum_{i,j=1}^{p} (W \otimes w_i)'c_4(S)(w_j \otimes W). \]

The expression thus obtained can be simplified with some matrix algebra. For that we need some additional notation.

Let \( A_d \) stand for the diagonalized matrix \( A \) and let \( a_d \) denote a diagonal matrix with the components of \( a \) on the main diagonal. The notation above is used in the next lemma.

**Lemma.** Let \( M, N, U \) and \( V \) be \( p \times p \)-matrices and \( a, b \) be \( p \)-vectors. Then

\[ (M \otimes a')K_d(U \otimes V)K_d(b \otimes N) = Ma_d(U \otimes V)b_dN, \]

where \( \circ \) denotes the elementwise (Hadamard) product of matrices and \( K_d = (K_{p,p})_d \). (5.1)

The statement follows from the fact that the corresponding elements of the matrices on the left- and right-hand sides are equal, as can be checked in a straightforward manner.

We are going to apply the lemma to the expression of \( C(X) \). Note first that the diagonal matrix \( c_4(S) \) can be expressed as a product:

\[ c_4(S) = K_d\left( \begin{pmatrix} c_4(S_1) & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & c_4(S_p) \end{pmatrix} \otimes I_p \right) K_d. \]

Denote

\[ (c_4(S))_d = \begin{pmatrix} c_4(S_1) & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & c_4(S_p) \end{pmatrix}. \]

Then using (5.1) and (5.2) we get

\[ C(Z) = \sum_{i,j=1}^{p} W'(w_i)_d(c_4(S))_d(w_j)_d W = \sum_{i,j=1}^{p} W'(w_i)_d(w_j)_d(c_4(S))_d W = W'DW, \]
where
\[
D = (c_4(S))_d \left( \left( \sum_{i=1}^{p} w_i \right) \odot \left( \sum_{i=1}^{p} w_i \right) \right)_{d}.
\]

Let
\[
V = W^{1/2}.
\]
Then \( C(Z) = VV' \) with \( V'V = D \).

Therefore the vectors \( v_i \) and \( v_j \) are orthogonal if \( i \neq j \) and \( v_i'v_i = d_i \). This means that \( v_i \) is an eigenvector of \( C(Z) \) with length \( \sqrt{d_i} \) corresponding to the eigenvalue \( d_i \). The matrix \( V \) can be found as a solution of an eigenvalue problem. Our transformation matrix of interest \( W \) can be expressed as the product
\[
W = D^{-1/2}V'.
\]

Acknowledgments

The author is grateful to the Ph.D. student Helle Kilgi who made the calculations for the examples and to the Referees and the Associate Editor for valuable remarks which considerably improved the presentation of the paper. The author is grateful to the Estonian Science Foundation for support through the grants ETF5686 and ETF7435.

References