Orthogonalization of Vectors with Minimal Adjustment

PAUL H. GARTHWAITE, FRANK CRITCHLEY, KARIM ANAYA-IZQUIERDO and EMMANUEL MUBWANDARIKWA

Department of Mathematics and Statistics, The Open University, Milton Keynes, MK7 6AA, U.K.

p.h.garthwaite@open.ac.uk f.critchley@open.ac.uk

k.anaya@open.ac.uk e.mubwandarikwa@open.ac.uk

Summary

Two transformations are proposed that give orthogonal components with a one-to-one correspondence between the original vectors and the components. The aim is that each component should be close to the vector with which it is paired, orthogonality imposing a constraint. Applications of the transformations are diverse and form new statistical methods. These include a unified approach to the identification and diagnosis of collinearities, a method of setting prior weights for Bayesian model averaging, and regression on orthogonal components. A means of calculating an upper bound for a multivariate Chebyshev inequality is also obtained. One of the transformations has the property that duplicating a vector, perhaps several times, has no effect on the orthogonal components that correspond to non-duplicated vectors. This transformation is determined using a new algorithm that also provides the decomposition of a positive definite matrix in terms of a diagonal matrix and a correlation matrix. The algorithm is shown to converge to a global optimum.

Some key words: Collinearity; Cos-max; Cos-square; Dilution; Matrix decomposition; Multivariate Chebychev inequality; Orthogonal components; Prior weight; Transformation; Variance inflation factor.

1. Introduction

This paper addresses the following range of practically important problems. Despite their apparent disparity, a single line of research links them.

Problem 1. Suppose a set of non-orthogonal vectors are to be adjusted by minimal amounts so as to transform them to a set of orthonormal components. That is, from a set of vectors $x_1, \ldots, x_m$ we want to construct components $u_1, \ldots, u_m$ such that x_i is "close" to u_i for each i, and $U'U$ is an identity matrix where $U = (u_1, \ldots, u_m)$. Under such a transformation the x_i vectors retain their identities; x_i is associated with u_i. This contrasts with principal components or factor analysis, where a component typically relates to a number of x-variables and where an x-variable may relate to more than one component. The first problem is to find an intuitively reasonable definition of closeness or minimal adjustment between $x_1, \ldots, x_m$ and $u_1, \ldots, u_m$, such that a usable, useful transformation results.

Problem 2. Variance inflation factors, VIFs, are commonly used to determine if collinearities are present among a set of variables. If collinearities are detected, then eigenvectors are generally examined in order to identify which variables cause them. This is not a unified approach, as VIFs are not closely related to eigenvectors. Hence, for example, a VIF is not linked to a particular eigenvector and there is no direct relationship between the quantity used to identify a collinearity and the quantity used to determine its cause. The second problem is to find quantities that link to VIFs in close one-to-one relationships, and that are better than eigenvectors at identifying which variables underly any large VIF.

Problem 3. In Bayesian model averaging, a prior weight must be given to each of the models under consideration. The most common choice is a uniform prior in which each model is given the same prior weight. However, it can be argued that models that are very similar to each other should be given smaller weights. Quoting Clyde (1999), 'It is not clear that the independent uniform prior on the model space

is sensible. ... Suppose we start off with one [explanatory] variable, X_1, and consider two models ($\{1\}$, $\{1, X_1\}$) where $\{1\}$ represents the model with just an intercept. Assign both models equal prior probabilities 0.5. Now consider adding a second variable X_2 that is highly correlated (or even perfectly correlated) with X_1, with possible models ($\{1\}$, $\{1, X_1\}$, $\{1, X_2\}$, $\{1, X_1, X_2\}$) and uniform prior probabilities (0.25, 0.25, 0.25, 0.25). The total prior probability mass of the last three models is 0.75, while, if X_2 is really a proxy for X_1, the mass should be closer to 0.5, as these three models are approximately equivalent (or exactly with perfect collinearity), and should have the same weight as in the original model space with just X_1.'

The need to reduce the prior weights given to similar models was first advocated by George (1999), who referred to it as the dilution of prior weights. George also argued that if a model is duplicated, then the prior weight formerly given to the model that has been duplicated should be divided between that model and its duplicates, while the prior probability given to other models should be unchanged. The third problem we address is to find a weighting scheme that has these properties.

Problem 4. Suppose $(Y_1, \ldots, Y_m)'$ is a random vector with mean $0'_m$ and nonsingular covariance matrix $\Sigma = (\sigma_{ij})$. Olkin & Pratt (1958) consider the task of generalizing the standard univariate Chebychev inequality so as to obtain an upper bound for $\mathrm{pr}(|Y_i| \geq l_i\sigma_{ii}, \text{ for some } i)$, where $l_1, \ldots, l_m$ are specified positive constants. To calculate the bound that they obtain, a positive-definite symmetric matrix V must be decomposed as $V = SDS$, where D is a diagonal matrix and S is a correlation matrix. Olkin & Pratt prove that this decomposition is unique, but note that it cannot be performed by standard matrix operations. The problem of interest here is to find an efficient method of performing the decomposition.

Problem 5. Bolla et al. (1998) considered the problem of maximizing $\sum_{i=1}^{m} u_i' G u_i$ under the constraint that $U'U$ is an identity matrix, where $G_1, \ldots, G_m$ are symmetric, positive-definite matrices, and $U = (u_1, \ldots, u_m)$. They gave an algorithm to search

for U and showed that the algorithm converges. However, while they believe the algorithm converges to a global maximum, it has only been shown that the algorithm converges to a local maximum. Here we consider a related problem where each G_i equals $x_i x_i'$ for some vector x_i. Thus the G_i are non-negative definite matrices, rather than positive-definite. The algorithm of Bolla et al. may nevertheless be used to search for U. We address the problem of showing that the algorithm converges to a unique global maximum for this case.

Throughout we let $(x_1, \ldots, x_m)$ be a set of $n \times 1$ vectors and assume that $X'X$ is non-singular, where $X = (x_1, \ldots, x_m)$. For Problem 1 above, we consider two measures of closeness between X and $U = u_1, \ldots, u_m$:

$$\psi = \sum_{i=1}^{m} x_i' u_i \tag{1}$$

and

$$\phi = \sum_{i=1}^{m} (x_i' u_i)^2. \tag{2}$$

We devise transformations from X to U such that either ψ or ϕ is maximized subject to the following conditions:

Condition 1. $U = XA$ for some $m \times m$ matrix A.

Condition 2. $U'U = I_m$, the $m \times m$ identity matrix.

Condition 3. $x_i' u_i > 0$ for $i = 1, \ldots, m$.

Condition 1 implies that $u_1, \ldots, u_m$ are obtained from a linear transformation of $(x_1, \ldots, x_m)$. Condition 2 implies that $u_1, \ldots, u_m$ are a set of standardized orthogonal components. Condition 3 removes an obvious sign indeterminacy in maximizing ϕ.

When $x_1, \ldots, x_m$ are standardized vectors, so that $x_i' x_i = 1$ for each i, then x_i and u_i each denote directional vectors in n-space and $x_i' u_i$ is the cosine of the angle between them. The magnitude of the cosine gets larger as x_i and u_i become more co-linear. Hence in maximizing ψ or ϕ we are seeking a set of orthonormal components $u_1, \ldots, u_m$ such that, for each i, u_i is close to x_i. If each x-vector is standardized, then

$x_i'u_i$ is also the correlation between x_i and u_i, so the sum of these correlations or their squares is also maximized. We will refer to the transformation from X to U as the cos-max transformation when ψ is maximized, and as the cos-square transformation when ϕ is maximized. In applications, the matrices that yield the transformations are useful as well as the orthogonal components that are obtained.

Turning to Problem 2, variance inflation factors can be obtained from the matrix A that satisfies $U = XA$ in Condition 1. With both the cos-max and cos-square transformations, if a_i' is the ith row of A $(i = 1, \ldots, m)$, then the m VIFs for X are $a_1'a_1, \ldots, a_m'a_m$. Moreover, if the ith VIF is large, then the variables that cause that collinearity correspond to those components of a_i that are large in magnitude. Thus, the use of the a_i unifies the tasks of determining collinearities and identifying their causes.

The focus of this paper is the cos-square transformation because it has an unusual invariance property. Suppose the set of x-vectors is increased by adding duplicates of one of them. For definiteness, let $x_{m+1}, \ldots, x_{m+k}$ be duplicates of x_m and suppose that $x_m, x_{m+1}, \ldots, x_{m+k}$ all differ very slightly, so that $(X^*)'X^*$ is positive-definite, where $X^* = (x_1, \ldots, x_{m+k})$. Then the first $m-1$ orthogonal components, $u_1, \ldots, u_{m-1}$, are the same when the cos-square transformation is applied to the set of vectors $x_1, \ldots, x_m$ as when it is applied to the augmented set of vectors $x_1, \ldots, x_{m+k}$. That is, under this transformation, duplicating a vector does not change the orthogonal components associated with the other vectors. We call this property the duplicate invariance property.

With the cos-square transformation, the u_i components cannot be obtained from a simple formula and they are calculated using an iterative algorithm. The algorithm finds a diagonal matrix C such that $Q\Lambda Q' = CX'XC$ is the spectral decomposition of $CX'XC$, and the diagonal elements of C^2 and $Q\Lambda^{1/2}Q'$ are equal. For Bayesian model averaging, Garthwaite & Mubwandakiwa (2010) consider allocating weights to models on the basis of a correlation matrix, where highly correlated models receive

smaller weight. If $X'X$ is equated to the correlation matrix for m models, then one weighting scheme gives the ith model the prior weight,

$$p_i = c_i^2 / \sum_{j=1}^{m} c_j^2, \tag{3}$$

where $c_1, \ldots, c_m$ are the diagonal elements C. This weighting scheme meets the requirements of Problem 3. In particular, a consequence of the duplicate invariance property is that duplicating a model does not change the weights given to other models; the duplicated model simply shares its weight with its duplicates.

The algorithm used for the cos-square transformation also solves Problem 4, as it yields the decomposition of a positive-definite matrix V as $V = SDS$, where D is a diagonal matrix and S is a correlation matrix. To obtain this decomposition, the algorithm is applied with $X'X$ set equal to V. When the algorithm converges, the decomposition is obtained by putting $D = C^2$ and $S = C^{-1}Q\Lambda^{1/2}Q'C^{-1}$. Olkin & Pratt (1958) prove V has a unique decomposition as $V = SDS$. This leads to the result that the algorithm converges to a unique global maximum, rather than a local maximum, which solves Problem 5.

2. The transformations

Theorem 1 underpins both the cos-max and cos-square transformations. Proofs of theorems are given in an appendix.

Theorem 1. *Suppose C is a diagonal $m \times m$ matrix of positive constants and let c_i denote its (i, i) element ($i = 1, \ldots, m$). Let $Q\Lambda Q' = CX'XC$ be the spectral decomposition of $CX'XC$, where Λ is the diagonal matrix of eigenvalues and Q is the matrix whose columns are the corresponding normalized eigenvectors. If $U = XA$ satisfies conditions 1–3, then $tr(CX'U) = \sum_{i=1}^{m} c_i x_i' u_i$ is maximized when $A = CQ\Lambda^{-1/2}Q'$.*

To transform X to U using the cos-max transformation, $\sum_{i=1}^{m} x_i' u_i$ must be maximized under Conditions 1–3. Let $Q_1\Lambda_1 Q_1'$ be the spectral decomposition of $X'X$.

Applying Theorem 1 with $C = I_m$ gives

$$U = XQ_1\Lambda_1^{-1/2}Q_1'. \tag{4}$$

Now $Q_1\Lambda_1^{-1/2}Q_1'$ is unique as it is the symmetric square-root of $(X'X)^{-1}$. Hence equation (4) determines U uniquely.

The following iterative algorithm exploits Theorem 1 to find U for the cos-square transformation, when $\sum_{i=1}^m (x_i'u_i)^2$ must be maximized under conditions 1–3. Step 6 of the algorithm gives U. The algorithm repeatedly maximizes $\sum_{i=1}^m c_i x_i'u_i$ until convergence, at each iteration setting each c_i equal to the most recent estimate of $x_i'u_i$.

Algorithm 1. *Algorithm for the cos-square transformation.*

Step 1. Set C_1 equal to the $m \times m$ identity matrix and put $i = 1$.

Step 2. At the ith iteration, perform a spectral decomposition of $C_iX'XC_i$, giving $C_iX'XC_i = Q_i\Lambda_iQ_i'$, where Q_i is an orthogonal matrix and Λ_i is a diagonal matrix.

Step 3. Set C_{i+1} equal to a diagonal matrix, with diagonal equal to the diagonal of $C_i^{-1}Q_i\Lambda_i^{1/2}Q_i'$.

Step 4. Repeat Steps 2 and 3 until convergence, when $C_{i+1} \approx C_i$.

Step 5. Set $(u_1, \ldots, u_m)$ equal to $XC_iQ_i\Lambda_i^{-1/2}Q_i'$.

The transformation matrix, A, is given by $A = Q_1\Lambda_1^{-1/2}Q_1'$ for the cos-max transformation, where $Q_1\Lambda_1Q_1' = X'X$, and by $A = CQ\Lambda^{-1/2}Q'$ for the cos-square transformations, where $Q\Lambda Q' = CX'XC$. Both transformations are determined by $X'X$, rather than X. It will often be appropriate to centre X so that each of its rows has a mean of zero and then base transformations on the covariance matrix $X'X$ or, if standardization of $x_1, \ldots, x_m$ is required, the corresponding correlation matrix.

The remainder of this section relates solely to the cos-square transformation, focusing on the algorithm that gives the transformation.

Remark 1. Suppose the diagonal elements of C_i are positive. Then the diagonal elements of $C_i^{-1}Q_i\Lambda_i^{1/2}Q_i'$ are also positive and, from Step 3 of the algorithm, so are those of C_{i+1}. Since C_1 has positive diagonal elements, so do all C_i, which is a condition in Theorem 1.

Remark 2. The rationale behind the algorithm is that in its ith iteration it finds the matrix U_i that maximizes $\mathrm{tr}(C_iX'U_i)$ and meets Conditions 1–3. From Theorem 1, $U_i = XA_i$ where $A_i = C_iQ_i\Lambda_i^{-1/2}Q_i'$. Thus $X'U_i = C_i^{-1}C_iX'XC_iQ_i\Lambda_i^{-1/2}Q_i' = C_i^{-1}Q_i\Lambda_iQ_i'Q_i\Lambda_i^{-1/2}Q_i' = C_i^{-1}Q_i\Lambda_i^{1/2}Q_i'$. Hence, Step 3 sets the diagonal elements of C_{i+1} equal to the diagonal elements of $X'U_i$.

Remark 3. At the ith iteration of the algorithm, U^* is chosen to maximize $\mathrm{tr}(C_iX'U^*)$ under the constraint that U^* must satisfy Conditions 1–3. The maximum value of $\mathrm{tr}(C_iX'U^*)$ is $\mathrm{tr}(C_iC_{i+1})$. The option of putting $U^* = U_{i-1}$ was available, when $\mathrm{tr}(C_iX'U^*)$ would equal $\mathrm{tr}(C_iC_{i-1})$. Hence $\mathrm{tr}(C_iC_{i+1}) \geq \mathrm{tr}(C_iC_{i-1}) = \mathrm{tr}(C_{i-1}C_i)$. Consequently, $\mathrm{tr}(C_iC_{i+1})$ is monotonic non-decreasing as i increases. Also, $\mathrm{tr}(C_iC_{i+1})$ is bounded above as $\sum_{i=1}^m (x_i'u_i)^2$ is bounded above. Hence the algorithm converges.

Remark 4. From the Cauchy-Schwarz inequality, $\mathrm{tr}(C_iC_{i+1}) \leq \{\mathrm{tr}(C_iC_i)\,\mathrm{tr}(C_{i+1}C_{i+1})\}^{1/2}$, with equality only if $C_i = C_{i+1}$. Also, from the ith iteration of the algorithm, $\mathrm{tr}(C_iX'U^*)$ has a maximum value of $\mathrm{tr}(C_iC_{i+1})$, so $\mathrm{tr}(C_iC_{i+1}) \geq \mathrm{tr}(C_iC_i)$. Hence $\mathrm{tr}(C_iC_{i+1}) \leq \mathrm{tr}(C_{i+1}C_{i+1})$ and so, at convergence, $C_i = C_{i+1}$, as stated in Step 4.

Remark 5. We refer to C, Q and Λ as the cosine matrix, eigenvector matrix and eigenvalue matrix of the cos-square transformation, where these are the values of C_i, Q_i and Λ_i at convergence. We also refer to $\langle C, Q, \Lambda\rangle$ as the matrix triple of the transformation.

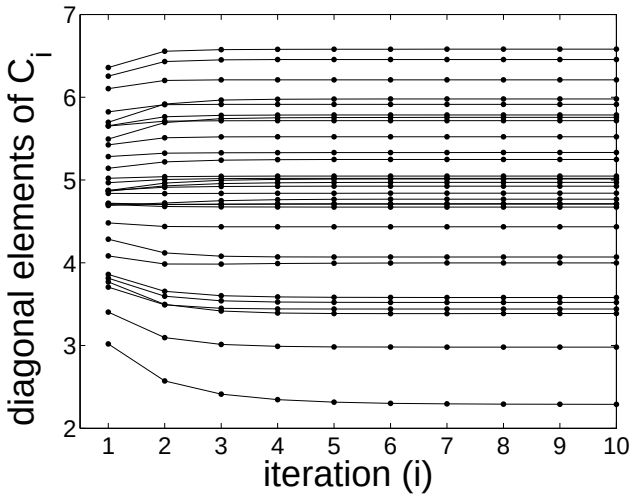

Fig. 1. Diagonal elements of C_i in the first 10 iterations of the algorithm for a 30×30 matrix.

The algorithm converges quickly, near-convergence typically being obtained in 4–10 iterations, although convergence can be slowed by extreme collinearities in $x_1, \ldots, x_m$. The algorithm may be viewed as a fixed point iteration method that takes C as its fixed point. While the aim is to estimate a matrix U with m^2 unknown elements, the diagonal matrix C has only m unknown elements, which benefits the convergence rate.

To illustrate the speed of convergence, a 30×30 matrix X was constructed, each of whose elements were randomly generated from a standard normal distribution. This form of matrix was also used by Roberts & Rosenthal (2009), who note that $X'X$ is almost singular, making it a challenging example. The algorithm was applied to X and the values taken by the diagonal elements of C_i were recorded at each iteration until convergence. Fig. 1 plots the path of each of these elements for the first ten iterations. It can be seen that every element had virtually converged within six iterations. Their values after six iterations were compared with their values after 50 iterations, and none had changed by more than 0.6%. Thus the algorithm converged quickly with this moderately large, ill-conditioned matrix. Our experience indicates that the speed of convergence is almost unaffected by the dimension of X.

It is obviously important that the algorithm should converge to a global maximum, rather than a local maximum. The following lemma and theorem establish that this is always the case. Their proofs are given in the appendix.

Lemma 1. *Let $\langle C, Q, \Lambda \rangle$ be the matrix triple when the cos-square transformation is applied to X. Put*

$$D = C^2 \tag{5}$$

and

$$S = C^{-1} Q \Lambda^{1/2} Q' C^{-1}. \tag{6}$$

Then D is a diagonal matrix, S is a correlation matrix and $X'X = SDS$.

Theorem 2. *At convergence of the algorithm for the cos-square transformation, the matrix triple $\langle C, Q, \Lambda \rangle$ is unique, apart from the ordering of eigenvectors and eigenvalues in Q and Λ. Also, the transformation matrix is unique and $\sum_{j=1}^{m} (x_j' u_j)^2$ attains its global maximum subject to conditions 1–3.*

Any positive definite matrix may be equated to $X'X$. It follows from Lemma 1 that the algorithm is a means of decomposing a positive definite matrix as SDS, where D is a diagonal matrix and S is a correlation matrix. The simplicity of this decomposition is attractive and an application where it proves useful is given later. Equations (5) and (6) give D and S from $\langle C, Q, \Lambda \rangle$.

Bolla et al. (1998) gave an algorithm for choosing $U = (u_1, \ldots, u_m)$ so as to maximize $\sum_{j=1}^{m} u_j' G_j u_j$ under the constraint that $U'U = I_m$, where each G_j is a given matrix. Their algorithm iteratively constructs a convergent sequence of matrices $U^{(1)}, U^{(2)}, \ldots$. Although Bolla et al. assume that each G_j is positive-definite, suppose the G_j are allowed to be positive semi-definite and that $G_j = x_j x_j'$ for $j = 1, \ldots, m$. Then it may be shown that their algorithm and our algorithm are equivalent, with $U^{(i+1)} = X C_{i+1} (C_{i+1} X' X C_{i+1})^{-1/2}$. The question of whether the algorithm of Bolla et al. converges to a global maximum for any positive-definite $G_1, \ldots, G_m$ has been considered by Bolla et al. (1998), Bolla (2001) and Rapcsak (2001, 2002), but the

result has not been proved. The work of Olkin & Pratt (1958) is key to proving convergence to a global maximum for the case considered in Theorem 2, where each G_j is of rank 1. By extending their work and adapting the proof of Theorem 2, it may be possible to prove convergence to a global maximum for the case where the G_j are positive-definite matrices.

3. The duplicate invariance property

Let $\{x_1, \ldots, x_m\}$ and $\{x_1, \ldots, x_{m+k}\}$ be two sets of vectors, so that the latter set is formed from the former by adding $x_{m+1}, \ldots, x_{m+k}$. In general, if some transformation is applied separately to each set, then the transformed values of $x_1, \ldots, x_m$ may not be the same in the two cases. Indeed, if the vectors are transformed to sets of orthogonal vectors, then their transformed values will almost certainly differ in the two cases, unless $x_{m+1}, \ldots, x_{m+k}$ are orthogonal to $x_1, \ldots, x_m$. This is true of the cos-max and cos-square transformations, and principal components, for example.

Suppose, however, that each of the vectors $x_{m+1}, \ldots, x_{m+k}$ is virtually a duplicate of x_m. Then, with some transformations, perhaps the transformed values of $x_1, \ldots, x_{m-1}$ will not depend on whether or not x_m is duplicated. If that is true when any single vector is duplicated, then we say that the transformation has the duplicate invariance property. The following defines the property more precisely.

Definition 1. *Let $u_1, \ldots, u_m$ and $u_1^*, \ldots, u_{m+k}^*$ denote the first m orthonormal components when a transformation is applied to $\langle x_1, \ldots, x_m \rangle$ and $\langle x_1, \ldots, x_{m+k} \rangle$, respectively. Assume $x_m \neq x_i$ for $i = 1, \ldots, m-1$. Then the transformation has the duplicate invariance property if $u_i^* \to u_i$ for $i = 1, \ldots, m-1$, as $x_{m+j} \to x_m$ for $j = 1, \ldots, k$.*

Almost any example shows that the cos-max transformation does not have the duplicate invariance property. However, in an appendix we prove the following theorem, which shows that the cos-square transformation does have the property. The theorem also gives related invariance properties. Applications in Section 4 illustrate

benefits of the properties.

THEOREM 3. *Suppose that k duplicates $x_{m+1}, \ldots, x_{m+k}$ are added to the set of vectors $x_1, \ldots, x_m$. Let $X^* = (x_1, \ldots, x_{m+k})$. Assume the duplicates are virtually identical to x_m while $(X^*)'X^*$ is positive-definite. Let $u_1, \ldots, u_m$ and $u_1^*, \ldots, u_{m+k}^*$ denote the components resulting from the cos-square transformation when it is applied to $\langle x_1, \ldots, x_m \rangle$ and $\langle x_1, \ldots, x_{m+k} \rangle$, respectively. Also, let A and A^* be the corresponding transformation matrices. Then, as $x_{m+j} \to x_m$ for $j = 1, \ldots, k$,*

(i) $u_i^* \to u_i$ *for* $i = 1, \ldots, m-1$,

(ii) $\sum_{i=1}^{m+k} (x_i' u_i^*)^2 \to \sum_{i=1}^{m} (x_i' u_i)^2$, *and*

(iii) if

$$A = \begin{pmatrix} A_{11} & a_{12} \\ a_{21}' & a_{22} \end{pmatrix} \quad \text{and} \quad A^* = \begin{pmatrix} A_{11}^* & A_{12}^* \\ A_{21}^* & A_{22}^* \end{pmatrix},$$

where A_{11} and A_{11}^* are $(m-1) \times (m-1)$ matrices, then $A_{11}^* \to A_{11}$.

Remark 6. In principle, $U'U$ should equal the identity matrix. However, in the limit $x_m = x_{m+1} = \ldots = x_{m+k}$, so then $U'U = A'X'XA$ is singular and the components $u_{m+1}, \ldots, u_{m+k}$ become ill-defined. Hence, Theorem 3 describes limiting behaviour when behaviour at the limit itself is ill-defined.

Remark 7. Suppose $\langle C, Q, \Lambda \rangle$ and $\langle C^*, Q^*, \Lambda^* \rangle$ are the matrix triples when the cos-square transformation is applied to $(x_1, \ldots, x_m)$ and $(x_1, \ldots, x_{m+k})$, respectively. From (ii) of Theorem 3, $\mathrm{tr}(C^*C^*) \to \mathrm{tr}(C^2)$ as $x_{m+j} \to x_m$ for $j = 1, \ldots, k$. This result is used in Section 4.4.

4. METHODOLOGICAL APPLICATIONS

The applications in this section mostly give new statistical methods. The work is less developmental in the last application, on a multivariate Chebychev inequality,

but a fifty-year-old problem is solved. The diversity of the applications suggests that the transformations and their associated matrices should prove widely useful.

4.1. *Detection and identification of collinearities*

A well-recommended approach for detecting collinearities is to calculate a variance inflation factor for each column of X (Farrar & Glauber, 1967). Suppose $x_1, \ldots, x_m$ are observations of variables $X_1, \ldots, X_m$ and let R_i^2 denote the multiple correlation coefficient when X_i is regressed on the other X-variables. The variance inflation factor for X_i, VIF_i say, is defined to be

$$VIF_i = (1 - R_i^2)^{-1}. \tag{7}$$

VIF_i will be large if X_i is involved in a collinearity. Hence, the VIF may be calculated for each variable and used as a diagnostic to identify which variables are involved in collinearities. However, the VIFs do not indicate the number of collinearities or directly identify which variables are associated with each of them (Wetherill, 1986, p. 87). The most common approach for identifying which variables form a collinearity is to determine the eigenvectors and eigenvalues of $X'X$. Near-zero eigenvalues suggest a collinearity and, in principle, the corresponding eigenvectors identify the X-variables involved; components of the eigenvector that are large in magnitude should correspond to those X-variables that are most influencing the collinearity.

As in Section 2, let $U = XA$, where A is the transformation matrix for either the cos-max or cos-square transformation of X and $U = (u_1, \ldots, u_m)$. Then A may be used both as a diagnostic for determining the number of collinearities and as a means of identifying the variables that contribute to each collinearity. The transformations aim to find orthogonal vectors $u_1, \ldots, u_m$, where u_i links strongly to x_i but not to the other x-vectors. Consequently, most of the off-diagonal elements of A will be close to 0. However, if there is a collinearity between k of the X-variables, then some of the x-vectors must be rotated substantially to transform them to k orthogonal

u-vectors, causing some off-diagonal elements of A to be quite far from 0. Thus a collinearity is indicated by large off-diagonal elements in A.

Suppose now that the X-variables had been standardized so that $X'X$ is a correlation matrix. Then the transformation matrix also gives the values of the VIFs. It is well known that VIF_i equals the (i,i) diagonal element of $(X'X)^{-1}$ when $X'X$ is a correlation matrix (e.g. Farrar & Glauber, 1967). For the cos-max transformation, $A = Q_1\Lambda_1^{-1/2}Q_1'$ with $Q_1\Lambda_1Q_1' = X'X$, so $AA' = Q_1\Lambda_1^{-1}Q_1' = (X'X)^{-1}$. For the cos-square transformation, $A = CQ\Lambda^{-1/2}Q'$ with $Q\Lambda Q' = CX'XC$, so $AA' = C(Q\Lambda^{-1}Q')C = C(C^{-1}(X'X)^{-1}C^{-1})C = (X'X)^{-1}$. In either case, if a_i' is the ith row of A, then

$$VIF_i = a_i'a_i. \tag{8}$$

That is, the rows of A determine the VIFs, with large values of $a_i'a_i$ indicating a collinearity. Moreover, most components of a_i will typically be small; the values that have greater magnitude cause VIF_i to be large and correspond to the variables that form the collinearity.

Examining A provides more information about collinearities than examining the eigenvectors that correspond to small eigenvalues. This is partly because each variable is associated with a separate row of A. Hence, if a collinearity involves three variables, for example, then three rows of A provide information about it. With eigenvectors, in contrast, a collinearity may relate to just one eigenvector. In addition, an eigenvector may combine two collinearities. For instance, if there is a collinearity between X_1 and X_2, and a second collinearity between X_3 and X_4, then an eigenvector may suggest that a collinearity involves X_1, X_2, X_3 and X_4. These points are illustrated in the following example.

The example concerns data on 180 pitprops cut from Corsican pine (Jeffers, 1967). The data have been widely analyzed, most commonly in the context of principal components analysis, where components are formed from 13 physical variables. Table 1 gives the sample correlation matrix for these variables, $X_1, \ldots, X_{13}$. There are quite

strong correlations between X_1 and X_2, between X_3 and X_4, and between X_6 and X_7, as well as a number of moderate correlations. When $X'X$ is set equal to this correlation matrix, the transformation matrix is given for the cos-max transformation in Table 2 and for the cos-square transformation in Table 3. Values above 0.5 are given in bold-face type. The last columns in the two tables give the VIF for each variable; eg. 13.71 is the VIF for X_2 and equals $a_2'a_2$. These columns are identical, of course, but other values in the two tables are also very similar to each other, so that the two transformations yield the same qualitative inferences. A VIF value above 10 is often treated as indicative of a collinearity (Neter, 1983, p. 392) and these values are also given in bold-face in the tables. On this basis, both transformations indicate that there is a collinearity between X_1 and X_2, and another between X_3 and X_4. From the components of a_3, a_4 and a_5, there is a suggestion in the tables that the latter collinearity also involves X_5; from practical considerations that is plausible, as X_4 and X_5 are the specific gravity of a pitprop before and after being oven-dried, while X_3 is the initial moisture content of the prop. X_7 also has a VIF above 10. It seems to be moderately collinear with X_6 and, to a lesser extent, X_{10}.

[Tables 1, 2 and 3 about here]

In comparison, eigenvalues and eigenvectors provide more limited information about collinearities. The smallest three eigenvalues for the pitprop data are 0.05, 0.04 and 0.04, and the next smallest is 0.19. Hence there appear to be three collinearities. The eigenvectors corresponding to the three smallest eigenvalues are:
$(-0.00, -0.05, 0.12, -0.02, 0.01, -0.54, 0.76, 0.03, -0.05, -0.32, -0.05, 0.05, 0.04)$,
$(0.39, \quad -0.41, 0.53, -0.59, 0.20, 0.08, -0.04, -0.05, 0.05, 0.06, 0.00, 0.00, 0.01)$ and
$(-0.57, \quad 0.58, \quad 00.41, -0.38, 0.12, 0.06, 0.00, 0.02, -0.06, 0.00, -0.01, \quad 0.00, -0.01)$.
The first of these eigenvectors suggests a collinearity between X_6 and X_7 and, perhaps, X_{10}. This concurs with one of the VIF findings from the cos-max and cos-square transformations. The other two eigenvectors both suggest a collinearity that involves X_1, X_2, X_3 and X_4. However, in contrast to the VIF analyses from

Tables 2 and 3, the eigenvectors do not suggest that there is one collinearity between X_1 and X_2 and another between X_3 and X_4.

4.2. *Bayesian prior weights and other weighting schemes*

The application that motivated the present work arises in Bayesian model averaging. Suppose we have models $M_1, \ldots, M_m$ and that p_i is the prior probability, or prior weight, that M_i is the true model. Suppose also that we have data $\mathcal{D}$ and $f_i(\mathcal{D} \mid M_i)$ is the likelihood of the data if M_i is true. Then in forming a Bayesian weighted model average, the posterior weight given to M_i is

$$w_i = \frac{p_i \, f_i(\mathcal{D} \mid M_i)}{\sum_{j=1}^{m} p_j \, f_j(\mathcal{D} \mid M_j)}, \tag{9}$$

for $i = 1, \ldots, m$. As this formula shows, the influence of the prior probabilities $p_1, \ldots, p_m$ do not dissipate as data are gathered, but instead have a multiplicative effect on the w_i. Hence, these prior probabilities should be chosen with care.

In practice, the most common choice for the p_i is to set each of them equal to $1/m$. However, this does not seem the best choice if some models are very similar to each other while other models are quite distinct. In regression, for example, if a small subset of the explanatory variables are believed important, then many models might be constructed by combining that subset of variables with two or three of the other explanatory variables. Then giving the same prior probability to each model would strongly favour the subset of variables believed important. Instead, the probabilities given to models that are similar should be reduced because of that similarity. Otherwise too little probability may be placed "...on good, but unique models, as a consequence of massing excess probability on large sets of bad, but similar models" (Chipman et al., 2001, p. 79).

Garthwaite & Mubwandarikwa (2010) suggest using the cos-square transformation to choose prior probabilities. Let $M_1, \ldots, M_m$ form the set of models to which prior probabilities must be assigned. They suppose an $m \times m$ correlation matrix R is determined that reflects the similarity between models. For example, R might be

formed from the correlations between predictions given by the models at a range of design points. The matrix triple $\langle C, Q, \Lambda \rangle$ is determined for the cos-square transformation with $X'X$ set equal to R. If $c_1, \ldots, c_m$ are the diagonal of C, then M_i is given the prior probability $p_i = c_i^2 / \sum_{j=1}^{m} c_j^2$. This is called the cos-square weighting scheme and it assigns smaller prior probabilities to models that are correlated more highly with other models. As an illustration, suppose five models $M_1, \ldots, M_5$ have the following correlation matrix:

M_1	1	0.9	0.8	0.7	0.6
M_2	0.9	1	0.7	0.6	0.5
M_3	0.8	0.7	1	0.5	0.4
M_4	0.7	0.6	0.5	1	0.3
M_5	0.6	0.5	0.4	0.3	1
	M_1	M_2	M_3	M_4	M_5

(10)

Then the prior probabilities given to these models are 0.10, 0.19, 0.22, 0.24 and 0.25, respectively. Thus M_1 receives the lowest prior probability, M_5 the highest, and the full order is M_1, M_2, M_3, M_4, M_5. Ranking the models by the strength of their correlations gives precisely the reverse ordering, with M_5 having the lowest correlations and M_1 the highest.

A related issue concerns how prior probabilities should be revised if the set of models is augmented by adding models that are virtually identical to a model already in the set. George (2010), Clyde (1999), and Chipman et al. (2001) argue that the prior probabilities given to the models that have not been duplicated should be unchanged. The remaining probability, which was formerly allocated to just the duplicated model, should be divided between that model and its duplicates. Methods of allocating weights that meet this requirement are said to have the strong dilution property (Garthwaite & Mubwandarikwa, 2010).

The desirability of this property is illustrated in the following example, given by

George (1999). Suppose a regression problem initially involves two uncorrelated predictors X_1 and X_2. These yield three regression models $\{X_1\}$, $\{X_2\}$ and $\{X_1, X_2\}$. A new predictor variable is introduced, X_3, and this variable is very highly correlated with X_2, but not with X_1. It is argued that the probability allocated to $\{X_1\}$ should be unchanged when X_3 is introduced as a potential predictor, whereas the probability allocated to $\{X_2\}$ and $\{X_1, X_2\}$ should be 'diluted' across all the new models containing X_3. George (1999, p. 176) writes, 'X_3 has not really added any new models to the mix. Instead, models containing X_3 are merely equivalent substitutes for the corresponding models containing X_2. Introducing X_3 has essentially resulted in relabelling a set of equivalent models. The probability of such a set should not increase as a result of this relabelling, and it is dilution that prevents this from happening.'

The following useful result about dilution is almost immediate from Theorem 3. It is proved in the appendix.

COROLLARY 3.1. *The cos-square weighting scheme has the strong dilution property.*

To illustrate the result, suppose a model M_6 is added to the models whose correlation matrix is given in (10). If M_6 were virtually identical to M_5, then the prior probabilities given to $M_1, \ldots, M_4$ would be unchanged, while the probability of 0.25 previously given to M_5 would be divided between M_5 and M_6.

The cos-square transformation is also useful for forming weighted averages in contexts other than choosing Bayesian prior weights. For example, suppose a variety of different measurements is made on each of a set of objects and a univariate similarity measure is to be constructed from both these measurements and, perhaps, functions of these measures that are thought to be useful, such as ratios, products and linear combinations. There may well be high correlations between some quantities and it seems sensible to take these correlations into account in forming a similarity mea-

sure. The cos-square transformation would yield suitable weights based just on the correlation structure, though the weights would generally need to be adjusted to reflect other factors.

The cos-max transformation has not been considered in this section because it does not appear to lead to any weighting scheme that has the strong dilution property.

4.3. *Regression and design*

In multiple regression, there is a preference for explanatory variables that are orthogonal. Methods such as principal component regression, latent root regression and partial least squares are advocated not only for their dimension reduction properties, but also because they give orthogonal components. An obvious advantage of orthogonal explanatory variables is that the relationship between each explanatory variable and the dependent variable is not affected by which other explanatory variables are included in the model. For example, forward stepwise regression and backwards stepwise regression will lead to the same model if predictors are orthogonal.

Strategies have been proposed for using regression on orthogonal components to select variables. For example, Jolliffe (1972, 1973) gives methods of selecting variables following principal component regression. However, methods that give orthogonal components do not, in general, lend themselves naturally to the task of variable selection. The cos-max and cos-square transformations are exceptions, since these retain a one-to-one correspondence between the original variables and the orthogonal components.

The pitprop data reported by Jeffers (1967) contained a dependent variable, the maximum compression stress of a prop (Y), as well as the thirteen X-variables whose correlation matrix is given in Table 1. Correlations between Y and the X-variables are given in Jeffers (1967), Mardia et al. (1979, p. 178) and elsewhere. If Y is regressed against the orthogonal components $u_1, \ldots, u_{13}$ obtained from the cos-max transformation, the percentages of the variance of Y accounted for by each of these

components are 7.89 (7.00), 0.43 (0.26), 40.38 (41.15), 8.05 (7.92), 2.00 (2.34), 2.96 (3.12), 2.34 (1.82), 6.30 (6.51), 1.25 (1.35), 0.48 (0.41), 0.09 (0.05), 0.57 (0.67) and 0.34 (0.49). Corresponding figures for the cos-square transformation are given in brackets and it can be seen that they are very similar. Jeffers (1967) and Mardia et al. (1979, p.246) both performed a principal components regression for the pitprops data. Jeffers selected five principal components which together accounted for 64.0% of the variation in Y while Mardia et al. selected eight principal components which together accounted for 72.8% of this variation. By comparison, with both the cos-max and cos-square transformations the best five predictors from $u_1, \ldots, u_{13}$ account for 66% of the variation in Y, which is slightly better than best five principal components, while the best eight predictors account for 71%, which is slightly poorer than the best eight principal components. Hence, there is disparity as to whether the constructed orthogonal components or an equal number of principal components make better predictors but, in any case, differences seem small. Regarding variable selection, X_1, X_3, X_4, X_6, X_8 are the variables that correspond to the best five components with both the cos-max and cos-square transformations. These variables form a good set of predictor variables; Mardia et al (1979, p. 178) give $(X_1, X_3, X_6, X_8, X_{11})$ as the optimal set of five predictors. However, differences are greater when the number of predictor variables is increased from five to eight: three of the variables chosen by the cos-max or cos-square transformations are not in the optimum set of eight predictor variables. Hence, regression on orthogonal components can clearly give a different perspective in variable selection problems.

Experimental design is another area where orthogonalization is often sought. Our transformations should prove useful if a design is to be adjusted to make it orthogonal and small adjustment is preferred. As an example, suppose n units are meant to represent a population of m-dimensional objects. Then n objects might be sampled from the population and taken as representative of it. However, if x_i denotes the vector of sample values for the ith dimension, then $x_1, \ldots, x_m$ are unlikely to be

a set of orthogonal vectors. If orthogonality is required, then the cos-max or cos-square transformation could be applied to $x_1, \ldots, x_m$. After rescaling, the resulting orthogonal components should be reasonably similar to $x_1, \ldots, x_m$.

4.4. *Multivariate Chebychev inequality*

Suppose $(Y_1, \ldots, Y_m)'$ is a random vector with mean 0_m and nonsingular covariance matrix $\Sigma = (\sigma_{ij})$. Olkin & Pratt (1958) considered the problem of finding an upper bound for $P(|Y_i| \geq l_i\sigma_{ii}, \text{ for some } i)$, where $l_1, \ldots, l_m$ are specified positive constants. Define Ω to be the positive definite matrix whose (i,j)th element is $\sigma_{ij}/(\sigma_{ii}\sigma_{jj}l_il_j)$. Let T be the unique correlation matrix such that $T\Omega^{-1}T$ is diagonal. Olkin & Pratt showed that

$$P(|Y_i| \geq l_i\sigma_{ii}, \text{ for some } i) \leq \text{tr}(T^{-1}\Omega T^{-1}), \tag{11}$$

which they referred to as a multivariate Chebychev inequality. In the univariate case, where $m = 1$, this reduces to the usual Chebychev inequality, $P(|Y_1| \geq l_1\sigma_{11}) \leq 1/l_1^2$.

As noted by Olkin & Pratt (1958), T cannot be obtained from Ω by standard matrix operations. Moreover, Olkin & Pratt did not suggest a method for calculating T and we did not find one in a search of the literature. Hence an efficient method for determining T has long been needed. The algorithm for the cos-square transformation provides one. If the transformation is applied with $X'X$ set equal to Ω and $\langle C, Q, \Lambda \rangle$ is the matrix triple of the transformation, then $\Omega = SDS$, where S and D are defined by equations (5) and (6). As $S\Omega^{-1}S$ is diagonal and S is a correlation matrix, we have that $T = S$. Also, $T^{-1}\Omega T^{-1} = D = C^2$, so the upper bound given by (11) is simply $\text{tr}(C^2)$.

Suppose $Y_{m+1}, \ldots, Y_{m+k}$ are virtual duplicates of one or more of $Y_1, \ldots, Y_m$. Then the true upper bounds of $P(|Y_i| \geq l_i\sigma_{ii}$, for some i; $i = 1, \ldots, m)$ and $P(|Y_i| \geq l_i\sigma_{ii}$, for some i; $i = 1, \ldots, m + k)$ should be virtually equal. It is not obvious that the bound of Olkin & Pratt has this property when the bound is expressed as $\text{tr}(T^{-1}\Omega T^{-1})$, since T becomes singular when the set $(Y_1, \ldots, Y_{m+k})$ contains du-

plicates. However, this property follows from Remark 7 on expressing the bound as $\mathrm{tr}(C^2)$.

5. Concluding comments

We have not directly addressed the questions of whether, or when, one of the two transformations that we have constructed is to be preferred to the other. There are applications in which the duplicate invariance property or the strong dilution property are necessary or highly desirable, such as when choosing prior weights in Bayesian model averaging, and then the cos-square transformation is appropriate. Also, only the cos-square transformation yields the useful decomposition of V as $V = SDS$, with D diagonal and S a correlation matrix. These results suggest that the cos-square transformation has a richer set of qualities than the cos-max transformation. However, there are applications in which the two transformations yield identical results, such as in calculating variance inflation factors, or they will typically give similar results, as when identifying the cause of large variance inflation factors. Then the cos-max transformation might be preferred because it is simpler and can be computed without recourse to an algorithm.

The method of assigning prior weights for Bayesian model averaging has been examined further with encouraging results (Garthwaite & Mubwandarikwa, 2010). Other methods proposed here that use the transformations also merit further work. These methods concern varied areas of statistics, so the number of methods using the transformations seems likely to grow.

Acknowledgement

Much of this work was conducted while PHG was a visitor at the University of New South Wales. EM was supported by a PhD studentship from the Open University.

Appendix

Proof of Theorem 1. Put

$$B = \Lambda^{1/2} Q' C^{-1} A Q, \tag{12}$$

where $\Lambda^{1/2}$ denotes a diagonal matrix with non-negative diagonal elements and $\Lambda^{1/2}\Lambda^{1/2} = \Lambda$. Then $B'B = Q'A'C^{-1}Q\Lambda Q'C^{-1}AQ = Q'A'X'XAQ = Q'U'UQ$. Now Q is an orthogonal matrix. Hence, $B'B = I_m$ if and only if $U'U = I_m$. Thus condition 2 is equivalent to $B'B = I_m$. We seek to maximize $\sum_{i=1}^{m} c_i x_i' u_i = \mathrm{tr}(CX'U) = \mathrm{tr}(CX'XA) = \mathrm{tr}(CX'XCC^{-1}A) = \mathrm{tr}(Q\Lambda Q'C^{-1}A) = \mathrm{tr}(\Lambda Q'C^{-1}AQ) = \mathrm{tr}(\Lambda^{1/2}B)$. Thus, $\mathrm{tr}(\Lambda^{1/2}B)$ is to be maximized under the constraint $B'B = I_m$. This has solution $B = I_m$ so, from equation (12), $A = CQ\Lambda^{-1/2}Q'$. □

Proof of Lemma 1. As C is diagonal and $D = C^2$, we have that D is diagonal. Now, in Steps 3 and 4 of our algorithm, the diagonal of C_{i+1} is set equal to the diagonal of $C_i^{-1}Q_i\Lambda_i^{1/2}Q_i'$. Hence, the diagonals of C and $C^{-1}Q\Lambda^{1/2}Q'$ are equal, where C, Q and Λ are the values of C_i, Q_i and Λ_i at convergence. Consequently $S = C^{-1}Q\Lambda^{1/2}Q'C^{-1}$ is a correlation matrix as it is positive-definite and its diagonal elements all equal 1. Also $SDS = (C^{-1}Q\Lambda^{1/2}Q'C^{-1})C^2(C^{-1}Q\Lambda^{1/2}Q'C^{-1}) = C^{-1}Q\Lambda Q'C^{-1} = X'X$. □

Proof of Theorem 2. Let V be any specified positive definite matrix and consider the decomposition $V = SDS$. Impressive work by Olkin & Pratt (1958, Theorem 3.6) showed that D and S are determined uniquely by the the constraints that (i) D is a diagonal matrix, and (ii) S is a correlation matrix. From Lemma 1, $X'X = SDS$ where $D = C^2$ and $\langle C, Q, \Lambda\rangle$ is the matrix triple for the cos-square transformation of X. As all the elements of C are positive, it follows that C is unique. Then, since $Q\Lambda Q'$ is the spectral decomposition of $CX'XC$, the matrix triple $\langle C, Q, \Lambda\rangle$ is also unique, apart from the ordering of eigenvectors and eigenvalues in Q and Λ. In

addition, $Q\Lambda^{1/2}Q'$ is unique as it is the symmetric square-root of $CX'XC$. It follows that the transformation matrix, A, is also unique, as $A = C(Q\Lambda^{1/2}Q')^{-1}$. Hence there is only one maximum to which the algorithm converges so that maximum is necessarily a global maximum. □

Proof of Theorem 3. The vector x_m will be duplicated. Partitioning $X'X$ to separate x_m from $x_1, \ldots, x_{m-1}$, put

$$X'X = \begin{pmatrix} R_{11} & r \\ r' & r_m \end{pmatrix},$$

where R_{11} is a $(m-1) \times (m-1)$ matrix, r is a vector and r_m is a scalar. Duplicates $x_{m+1}, \ldots, x_{m+k}$ are added to $x_1, \ldots, x_m$ to form the matrix $X^* = (x_1, \ldots . x_{m+k})$. Each duplicate differs from x_m by only a small amount and we put $x_{m+i} = x_m + \alpha\xi_i$ for $i = 1, \ldots, k$. We are interested in the transformation of $x_1, \ldots, x_{m+k}$ as $\alpha \to 0$. Write $(X^*)'X^*$ as

$$(X^*)'X^* = \begin{pmatrix} R_{11} & R_{12} + O(\alpha) \\ R'_{12} + O(\alpha) & R_{22} + O(\alpha) \end{pmatrix}, \tag{13}$$

where each column of R_{12} equals r and each element of R_{22} is r_m. We assume $(X^*)'X^*$ is a positive-definite matrix for any $\alpha > 0$.

Let $\langle C^*, Q^*, \Lambda^* \rangle$ denote the matrix triple at convergence when the algorithm for the transformation is applied to $x_1, \ldots, x_{m+k}$. We assume that the diagonal of Λ^* contains the eigenvalues in descending order of size. C^*, Q^* and Λ^* are $(m+k) \times (m+k)$ matrices that depend on α. Put

$$C^* = \begin{pmatrix} C_{(1)}^{(m-1)\times(m-1)} & 0_{(m-1)\times(k+1)} \\ 0_{(k+1)\times(m-1)} & C_{(2)}^{(k+1)\times(k+1)} \end{pmatrix}$$

and

$$Q^* = \begin{pmatrix} Q_{11}^{(m-1)\times m} & Q_{12}^{(m-1)\times k} \\ Q_{21}^{(k+1)\times m} & Q_{22}^{(k+1)\times k} \end{pmatrix}.$$

Since, $x_{m+i} = x_m + O(\alpha)$ for $i = 1, \ldots, k$, we can put

$$\Lambda^* = \begin{pmatrix} \Lambda_{(1)}^{m\times m} & 0_{m\times k} \\ 0_{k\times m} & \Lambda_{(2)}^{k\times k} \end{pmatrix}.$$

where each diagonal element $\Lambda_{(2)}$ is of order $O(\alpha)$. Λ^* and C^* are diagonal matrices. Let $c_0, \ldots, c_k$ denote the diagonal elements of $C_{(2)}$. The theorem is proved via Propositions 1–6.

PROPOSITION 1. *Choose $q = (q_1, \ldots, q_m)'$ so that the first row of Q_{21} is $c_0 q$. Then, as $\alpha \to 0$, the $(i+1)$th row of $Q_{21} \to c_i q'$ for $i = 0, \ldots, k$.*

Proof. For $j = 1, \ldots, m$, let s_j and t_j be the jth columns of Q_{11} and Q_{21}, respectively. We have that $C^*(X^*)'X^*C^*(s', t')' = (\lambda_j s_j', \lambda_j t_j')'$, where λ_j is the jth eigenvalue of Λ^*. Hence, as $\alpha \to 0$,

$$C_{(2)}R_{12}'C_{(1)}s_j + C_{(2)}R_{22}'C_{(2)}t_j \to \lambda_j t_j. \tag{14}$$

Now the rows of R_{12}' are identical to each other, as are those of R_{22}'. Hence, elements in the vector $R_{12}'C_{(1)}s_j$ are identical and so are those in $R_{22}'C_{(2)}t_j$. Consequently, if $t_j = (t_{j0}, \ldots, t_{jk})'$, then $t_{ji}/t_{j0} \to c_i/c_0$ for $i = 0, \ldots, k$. As c_i/c_0 does not depend on which column of Q_{21} was chosen as t_j, Proposition 1 follows. □

Define

$$C = \begin{pmatrix} C_{(1)} & 0_{m-1} \\ 0_{m-1}' & c_m \end{pmatrix}, \text{ and } Q = \begin{pmatrix} Q_{11} \\ c_m q' \end{pmatrix}, \tag{15}$$

where $c_m = (\sum_{i=0}^k c_i^2)^{1/2}$.

PROPOSITION 2. *As $\alpha \to 0$, the spectral decomposition of $CX'XC \to Q\,\Lambda_{(1)}Q'$, where $X = (x_1, \ldots, x_m)$.*

Proof. Let $\alpha \to 0$ and let $c = (c_0, \ldots, c_k)'$. From Proposition 1, if t_j is the jth column of Q_{21}, then $t_j \approx q_j c$ and the $(i+1)$th row of equation (14) gives $c_i r C_{(1)}s_j + c_i r_m c' t_j \approx \lambda_j c_i q_j$. Now $c't_j \approx c_m^2 q_j$, so

$$rC_{(1)}s_j + r_m c_m^2 q_j \to \lambda_j q_j. \tag{16}$$

From $C^*(X^*)'X^*C^*(s',t')' = (\lambda_j s_j', \lambda_j t_j')'$, we also have that $C_{(1)}R_{11}C_{(1)}s_j + C_{(1)}R_{12}C_{(2)}t_j \to \lambda_j s_j$. As $R_{12}C_{(2)}t_j = r(c't_j) \approx r(c_m^2 q_j)$,

$$C_{(1)}R_{11}C_{(1)}s_j + C_{(1)}rc_m^2 q_j \to \lambda_j t_j. \tag{17}$$

From equations (16) and (17), as $\alpha \to 0$,

$$\begin{pmatrix} C_{(1)} & 0_{m-1} \\ 0'_{m-1} & c_m \end{pmatrix} \begin{pmatrix} R_{11} & r \\ r' & r_m \end{pmatrix} \begin{pmatrix} C_{(1)} & 0_{m-1} \\ 0'_{m-1} & c_m \end{pmatrix} \begin{pmatrix} s_j \\ c_m q_j \end{pmatrix} \to \begin{pmatrix} \lambda_j s_j \\ \lambda_j c_m q_j \end{pmatrix}.$$

Using equation (13), as $\alpha \to 0$, λ_j is an eigenvalue of $CX'XC$ and $(s_j', c_m q_j)'$ is its corresponding eigenvector. Now, $(s_j', c_m q_j)(s_j', c_m q_j)' = s_j's_j + c_m^2 q_j^2 = s_j's_j + (q_j c')(cq_j) \approx s_j's_j + t_j't_j$. As $(s_j', t_j')'$ is the jth column of the orthogonal matrix Q^*, it follows that $(s_j', c_m q_j)(s_j', c_m q_j)' \approx 1$. Hence the jth column of Q [i.e. $(s_j', c_m q_j)$] is a standardized eigenvector as $\alpha \to 0$. □

PROPOSITION 3. *As $\alpha \to 0$, diagonal$(C) \to$ diagonal$(C^{-1}Q\Lambda_{(1)}^{1/2}Q')$.*

Proof. As $\langle C^*, Q^*, \Lambda^* \rangle$ is the matrix triple for the transformation of $x_1, \ldots, x_{m+k}$, diagonal(C^*C^*) = diagonal$\{Q^*(\Lambda^*)^{1/2}(Q^*)'\}$. Hence as $\alpha \to 0$,

$$\text{diagonal} \begin{pmatrix} C_{(1)}C_{(1)} & 0_{(m-1)\times(k+1)} \\ 0_{(k+1)\times(m-1)} & C_{(2)}C_{(2)} \end{pmatrix} \to \text{diagonal} \begin{pmatrix} Q_{11}\Lambda_{(1)}^{1/2}Q_{11}' & Q_{11}\Lambda_{(1)}^{1/2}Q_{21}' \\ Q_{21}\Lambda_{(1)}^{1/2}Q_{11}' & Q_{21}\Lambda_{(1)}^{1/2}Q_{21}' \end{pmatrix},$$

so

$$\text{diagonal}(C_{(1)}C_{(1)}) \to \text{diagonal}(Q_{11}\Lambda_{(1)}^{1/2}Q_{11}'). \tag{18}$$

Also, $\text{tr}(C_{(2)}C_{(2)}) \to \text{tr}(Q_{21}\Lambda_{(1)}^{1/2}Q_{21}')$. Now $\text{tr}(C_{(2)}C_{(2)}) = \sum_{i=0}^{k} c_i^2 = c_m^2$ and $\text{tr}(Q_{21}\Lambda_{(1)}^{1/2}Q_{21}') \approx \sum_{i=0}^{k} c_i q'\Lambda_{(1)}^{1/2} q c_i = (\sum_{i=0}^{k} c_i^2) q'\Lambda_{(1)}^{1/2} q = c_m^2 q'\Lambda_{(1)}^{1/2} q$. Thus, as $\alpha \to 0$,

$$c_m^2 \to c_m^2 q'\Lambda_{(1)}^{1/2} q. \tag{19}$$

From equations (18) and (19), as $\alpha \to 0$,

$$\text{diagonal} \begin{pmatrix} C_{(1)}C_{(1)} & 0_{m-1} \\ 0'_{m-1} & c_m^2 \end{pmatrix} \to \text{diagonal} \begin{pmatrix} Q_{11}\Lambda_{(1)}^{1/2}Q_{11}' & c_m Q_{11}\Lambda_{(1)}^{1/2} q \\ c_m q'\Lambda_{(1)}^{1/2}Q_{11}' & c_m^2 q'\Lambda_{(1)}^{1/2} q \end{pmatrix}.$$

□

There is a unique set of matrices that satisfy the conditions to be the matrix triple for the transformation of $x_1, \ldots, x_m$. Propositions 2 and 3 show that $\langle C, Q, \Lambda_{(1)} \rangle$ is that matrix triple as $\alpha \to 0$.

PROPOSITION 4. *Let $\theta = q'\Lambda_{(1)}^{1/2}q$. Then the $(m-1) \times (m-1)$ matrix $Q_{11}(\Lambda_{(1)}^{1/2} - \Lambda_{(1)}^{1/2}qq'\Lambda_{(1)}^{1/2}/\theta)Q_{11}'$ is of full rank.*

Proof. Let $\Gamma = \Lambda_{(1)}^{1/2} - \Lambda_{(1)}^{1/2}qq'\Lambda_{(1)}^{1/2}/\theta$. As $\Lambda_{(1)}^{1/2}$ is of rank m and $\Lambda_{(1)}^{1/2}qq'\Lambda_{(1)}^{1/2}$ is of rank 1, the rank of Γ is at least $m-1$. Hence, as Γ is an $m \times m$ matrix, if g_1 and g_2 are two non-zero vectors such that $g_1'\Gamma = g_2'\Gamma = 0_m'$, then g_1 must be proportional to g_2. Now $q'\Gamma = q'\Lambda_{(1)}^{1/2} - q'\Lambda_{(1)}^{1/2} = 0_m'$, so for any vector g, if $g'Q_{11}\Gamma = 0_m'$, then $g'Q_{11} = 0_m'$ or $g'Q_{11} = bq$ for some non-zero constant b. Since $(Q_{11}', q)'$ is an orthogonal matrix, $(g', b)(Q_{11}', q)' \neq 0_m'$. Hence, if $g'Q_{11}\Gamma = 0_m'$, then $g'Q_{11} = 0_m'$. But if $g'Q_{11} = 0_m'$, then $(g', 0)(Q_{11}', q)' = 0_m'$, implying $g = 0_{m-1}$. Hence, if $g'Q_{11}\Gamma = 0_m'$, then $g = 0_{m-1}$ and $Q_{11}\Gamma$ is of rank $m-1$. Since $Q_{11}\Gamma q = 0_{m-1}$, a similar argument shows that $Q_{11}\Gamma Q_{11}'$ is of rank $m-1$. □

PROPOSITION 5. *If $u_1, \ldots, u_m$ and $u_1^*, \ldots, u_{m+k}^*$ denote the components resulting from the cos-square transformation when it is applied to $\langle x_1, \ldots, x_m \rangle$ and $\langle x_1, \ldots, x_{m+k} \rangle$, respectively, then $u_i \to u_i^*$ for $i = 1, \ldots, m-1$, as $x_{m+j} \to x_m$ for $j = 1, \ldots, k$.*

Proof. Put $X_{(1)} = (x_1, \ldots, x_{m-1})$, $X_{(2)} = (x_m, \ldots, x_{m+k})$, $U_{(1)}^* = (u_1^*, \ldots, u_{m-1}^*)$, $U_{(2)}^* = (u_m^*, \ldots, u_{m+k}^*)$ and $U_{(1)} = (u_1, \ldots, u_{m-1})$. From Theorem 1, $(U_{(1)}^*, U_{(2)}^*) = (X_{(1)}, X_{(2)})A^*$, where $A^* = CQ^*(\Lambda^*)^{-1/2}(Q^*)'$. Hence

$$(X_{(1)}, X_{(2)}) = (U_{(1)}^*, U_{(2)}^*)Q^*(\Lambda^*)^{-1/2}(Q^*)'(C^*)^{-1}. \tag{20}$$

Let $\alpha \to 0$. Then

$$Q^*(\Lambda^*)^{1/2}(Q^*)'(C^*)^{-1} \approx \begin{pmatrix} Q_{11}\Lambda_{(1)}^{1/2}Q_{11}'C_{(1)}^{-1} & Q_{11}\Lambda_{(1)}^{1/2}Q_{21}'C_{(2)}^{-1} \\ Q_{21}\Lambda_{(1)}^{1/2}Q_{11}'C_{(1)}^{-1} & Q_{21}\Lambda_{(1)}^{1/2}Q_{21}'C_{(2)}^{-1} \end{pmatrix}, \tag{21}$$

and each column of $Q'_{21}C^{-1}_{(2)} \to q$. Also, each column of $X_{(2)} \to x_m$. Hence, from (20) and (21),

$$x_m \approx U^*_{(1)}Q_{11}\Lambda^{1/2}_{(1)}q + U^*_{(2)}Q_{21}\Lambda^{1/2}_{(1)}q \tag{22}$$

and

$$X_{(1)} \approx U^*_{(1)}Q_{11}\Lambda^{1/2}_{(1)}Q'_{11}C^{-1}_{(1)} + U^*_{(2)}Q_{21}\Lambda^{1/2}_{(1)}Q'_{11}C^{-1}_{(1)}. \tag{23}$$

From Proposition 1, $U^*_{(2)}Q_{21} \approx U^*_{(2)}(c_0, \ldots, c_k)'q'$. Hence, from (22), $U^*_{(2)}(c_0, \ldots, c_k)' \approx (x_m - U^*_{(1)}Q_{11}\Lambda^{1/2}_{(1)}q)/\theta$ where, as in Proposition 4, $\theta = q'\Lambda^{1/2}_{(1)}q$. Hence, from (23),

$$\begin{aligned} X_{(1)} &\approx U^*_{(1)}Q_{11}\Lambda^{1/2}_{(1)}Q'_{11}C^{-1}_{(1)} + (x_m - U^*_{(1)}Q_{11}\Lambda^{1/2}_{(1)}q)q'\Lambda^{1/2}_{(1)}Q'_{11}C^{-1}_{(1)}/\theta \\ &= U^*_{(1)}Q_{11}(\Lambda^{1/2}_{(1)} - \Lambda^{1/2}_{(1)}qq'\Lambda^{1/2}_{(1)}/\theta)Q'_{11}C^{-1}_{(1)} + x_m q'\Lambda^{1/2}_{(1)}Q'_{11}C^{-1}_{(1)}/\theta. \end{aligned} \tag{24}$$

Similar computations from the equation $(U_{(1)}, u_m) = (X_{(1)}, x_m)A$, where $A = CQ\Lambda^{-1/2}_{(1)}Q'$, give

$$x_m \approx U_{(1)}Q_{11}\Lambda^{1/2}_{(1)}q + c_m u_m \theta \tag{25}$$

and

$$X_{(1)} \approx U_{(1)}Q_{11}\Lambda^{1/2}_{(1)}Q'_{11}C^{-1}_{(1)} + c_m u_m q'\Lambda^{1/2}_{(1)}Q'_{11}C^{-1}_{(1)}. \tag{26}$$

Using (25) to substitute for $c_m u_m$ in (26),

$$X_{(1)} \approx U_{(1)}Q_{11}(\Lambda^{1/2}_{(1)} - \Lambda^{1/2}_{(1)}qq'\Lambda^{1/2}_{(1)}/\theta)Q'_{11}C^{-1}_{(1)} + x_m q'\Lambda^{1/2}_{(1)}Q_{11}C^{-1}_{(1)}/\theta. \tag{27}$$

From (24) and (27),

$$0_{n\times(m-1)} \approx (U^*_{(1)} - U_{(1)})Q_{11}(\Lambda^{1/2}_{(1)} - \Lambda^{1/2}_{(1)}qq'\Lambda^{1/2}_{(1)}/\theta)Q'_{11}C^{-1}_{(1)}.$$

From Proposition 4, $Q_{11}(\Lambda^{1/2}_{(1)} - \Lambda^{1/2}_{(1)}qq'\Lambda^{1/2}_{(1)}/\theta)Q'_{11}C^{-1}_{(1)}$ is of full rank. Hence $U^*_{(1)} - U_{(1)} \approx 0_{n\times(m-1)}$. □

PROPOSITION 6. *As $\alpha \to 0$, $Q_{12}\Lambda^{-1/2}_{(2)}Q'_{12} \to 0_{(m-1)\times(m-1)}$.*

Proof. Put $Z = (x_m, \xi_1, \ldots, \xi_k)$ and

$$H_\alpha = \begin{pmatrix} 1 & 1'_k \\ 0 & \alpha I_k \end{pmatrix}.$$

Then H_α is non-singular and $X_{(2)} = ZH_\alpha$. As $X^* = (X_{(1)}, X_{(2)})$,

$$[(X^*)'X^*]^{-1} = \begin{pmatrix} X'_{(1)}X_{(1)} & X'_{(1)}ZH_\alpha \\ H'_\alpha Z'X_{(1)} & H'_\alpha Z'ZH_\alpha \end{pmatrix}^{-1}.$$

Let Γ_{11} denote the top-left $m \times m$ sub-matrix of $[(X^*)'X^*]^{-1}$. Then $\Gamma_{11} = (X'_{(1)}X_{(1)} - X'_{(1)}ZH_\alpha(H'_\alpha Z'ZH_\alpha)^{-1}H'_\alpha Z'X_{(1)})^{-1} = (X'_{(1)}X_{(1)} - X'_{(1)}Z(Z'Z)^{-1}Z'X_{(1)})^{-1}$. Hence Γ_{11} does not depend on α. Also, $[(X^*)'X^*]^{-1} = C^*Q^*\Lambda^*(Q^*)'C^*$, so $\Gamma_{11} = C_{(1)}[Q_{11}\Lambda^{-1}_{(1)}Q'_{11} + Q_{12}\Lambda^{-1}_{(2)}Q'_{12}]C_{(1)}$. As Γ_{11} does not depend on α and both $C_{(1)}$ and $Q_{11}\Lambda^{-1}_{(1)}Q'_{11}$ are finite, $Q_{12}\Lambda^{-1}_{(2)}Q'_{12}$ is finite for all α. If q_{ij} denotes the (i, j) element of Q^*, the ith diagonal element of $Q_{12}\Lambda^{-1}_{(2)}Q'_{12}$ is

$$\sum_{j=1}^{k} q^2_{i(m+j)}/\lambda_{m+j} \tag{28}$$

for $i = 1, \ldots, (m-1)$. Each term in the summation in (28) is positive, so each term is finite as their sum is finite. Since $\lambda_{m+j} \to 0$ as $\alpha \to 0$,

$$q^2_{i(m+j)}/\lambda^{1/2}_{m+j} \to 0 \tag{29}$$

as $\alpha \to 0$ for $i = 1, \ldots, (m-1)$; $j = 1, \ldots, k$. The (h, i) element of $Q_{12}\Lambda^{-1/2}_{(2)}Q'_{12}$ is

$$\sum_{j=1}^{k} q_{h(m+j)}q_{i(m+j)}/\lambda^{1/2}_{m+j} = \sum_{j=1}^{k}[(q^2_{h(m+j)}/\lambda^{1/2}_{m+j})(q^2_{i(m+j)}/\lambda^{1/2}_{m+j})]^{1/2}.$$

Hence, from (29), $Q_{12}\Lambda^{-1/2}_{(2)}Q'_{12} \to 0_{(m-1)\times(m-1)}$ as $\alpha \to 0$. □

Proposition 5 is part (i) of Theorem 3. By definition, $c_m = (\sum_{i=0}^{k} c_i^2)^{1/2}$ so $\sum_{i=m}^{m+k}(x'_i u_i^*)^2 \to (x'_m u_m)^2$. Consequently, $\sum_{i=1}^{m+k}(x'_i u_i^*)^2 \to \sum_{i=1}^{m}(x'_i u_i)^2$, giving (ii). As $A = CQ\Lambda^{-1/2}_{(1)}Q'$. it follows from equation (14) that $A_{11} = C_{(1)}Q_{11}\Lambda^{-1/2}_{(1)}Q'_{11}$. Also, $A^* = C^*Q^*(\Lambda^*)^{-1/2}(Q^*)'$ so $A^*_{11} = C_{(1)}(Q_{11}\Lambda^{-1/2}_{(1)}Q'_{11} + Q_{12}\Lambda^{-1/2}_{(2)}Q'_{12})$. Thus, from Proposition 6, $A^*_{11} \to A_{11}$ as $\alpha \to 0$. This gives part (iii), completing the proof of Theorem 3. □

Proof of Corollary 3.1. In the notation of Theorem 3, let p_i and p_i^* denote the prior probabilities that the cos-square weighting scheme gives to M_i when the set

of vectors is $x_1, \ldots, x_m$ and $x_{m+1}, \ldots, x_{m+k}$, respectively. Then, $p_i = c_i^2 / \sum_{j=1}^m c_j^2 = (x_i' u_i)^2 / \sum_{j=1}^m (x_j' u_j)^2$ for $j = 1, \ldots m$. Similarly, $p_i^* = (x_i' u_i^*)^2 / \sum_{j=1}^{m+k} (x_j' u_j)^2$ for $j = 1, \ldots m+k$. From Theorem 3, $p_i^* \to p_i$ $(i = 1, \ldots, m-1)$ as $x_{m+j} \to x_m$ for $j = 1, \ldots, k$. □

References

Bolla, M. (2001). Parallel factoring of strata. In *Proceedings of the 23rd International Conference on Information Technology Interfaces*, Ed. D. Kalpic and V. H. Dobric, pp. 250–266. Pula, Croatia.

Bolla, M., Michaletzky, G., Tusnady, G. & Ziermann, M. (1998). Extrema of sums of heterogeneous quadratic forms. *Linear Algebra Appl.* **269**, 331–365.

Chipman, H. A., George, E.I. & McCulloch, R. E. (2001). The practical implementation of Bayesian model selection (with discussion). In *Model Selection*, Ed. P. Lahiri, pp. 65–134. Beachwood, OH: IMS.

Clyde, M. (1999). Bayesian model averaging and model search strategies. In *Bayesian Statistics 6*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp. 157–185. Oxford: Oxford University Press.

Farrar, D. B. & Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *Rev. Econ. Statist.* **73**, 191–193.

George, E. I. (1999). Discussion of "Bayesian model averaging and model search strategies," by M. Clyde. In *Bayesian Statistics 6*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp. 175–178. Oxford: Oxford University Press.

George, E. I. (2010). Dilution priors: Compensating for model space redundancy. In *Borrowing Strength: Theory Powering Applications – A Festschrift for Lawrence D. Brown*, Ed. J. O. Berger, T. T. Cai and I. M. Johnstone, pp. 158–165. Ohio: IMS Collections.

Garthwaite, P. H. & Mubwandarikwa, E. (2010). Selection of prior weights for weighted model averaging. *Aust. N. Z. J. Stat.* In press. (Early view: on line at http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1467-842X/earlyview)

Jeffers, J. N. R. (1967). Two case studies in the application of principal component analysis. *Appl. Statist.* **16**, 225–236.

Jolliffe, I. T. (1972). Discarding variables in principal component analysis I: artificial data. *Appl. Statist.* **21**, 160–173.

Jolliffe, I. T. (1972). Discarding variables in principal component analysis II: real data. *Appl. Statist.* **22**, 21–31.

Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979). *Multivariate Analysis.* London: Academic Press.

Neter, J., Wasserman, W. & Kutner, M. H. (1983). *Applied Linear Regression Models.* Illinois: Irwin.

Olkin, I. & Pratt, J. W. (1958). A multivariate Tchebycheff inequality. *Ann. Math. Statist.* **29**, 226–234.

Rapcsak, T. (2001). On minimization of sums of heterogeneous quadratic forms on Stiefel manifolds. In *From Local to Global Optimization*, Ed. A. Migdalas, P. Pardalos and P. Varbrand, pp. 277–290. Dordrechy: Kluwer Academic Publishers.

Rapcsak, T. (2002). On minimization on Stiefel manifolds. *European J. Oper. Res.* **143**, 365–376.

Roberts, G. O. & Rosenthal, J. S. (2009). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, **16**, 351–367.

Wetherill, G. B. (1986). *Regression Analysis with Applications.* London: Chapman and Hall.

Table 1. *Correlation matrix for the physical properties of pitprops*

X_1												
0.954	X_2											
0.364	0.297	X_3										
0.342	0.284	0.882	X_4									
-0.129	-0.118	-0.148	0.220	X_5								
0.313	0.291	0.153	0.381	0.364	X_6							
0.496	0.503	-0.029	0.174	0.296	0.813	X_7						
0.424	0.419	-0.05	-0.059	0.004	0.090	0.372	X_8					
0.592	0.648	0.125	0.137	-0.039	0.211	0.465	0.482	X_9				
0.545	0.569	-0.081	-0.014	0.037	0.274	0.675	0.557	0.526	X_{10}			
0.084	0.076	0.162	0.097	-0.091	-0.036	-0.113	0.061	0.085	-0.319	X_{11}		
-0.019	-0.036	0.220	0.169	-0.145	0.024	-0.232	-0.357	-0.127	-0.368	0.029	X_{12}	
0.134	0.144	0.126	0.015	-0.208	-0.329	-0.424	-0.202	-0.076	-0.291	0.007	0.184	X_{13}

Table 2. *Cos-max transformation matrix and VIFs for the data on pitprops*

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	VIF
$\mathbf{a}_1$	**3.04**	**-1.91**	-0.25	-0.10	0.12	-0.06	-0.22	-0.20	0.01	-0.20	-0.08	-0.07	-0.18	**13.14**
$\mathbf{a}_2$	**-1.91**	**3.11**	0.00	-0.05	0.05	0.03	-0.22	0.00	-0.38	-0.29	-0.10	-0.06	-0.26	**13.71**
$\mathbf{a}_3$	-0.25	0.00	**2.72**	**-1.95**	**0.59**	0.02	0.26	-0.03	-0.02	0.08	-0.05	-0.03	0.02	**11.66**
$\mathbf{a}_4$	-0.10	-0.05	**-1.95**	**2.84**	**-0.66**	-0.33	0.02	0.12	-0.03	-0.01	0.01	-0.04	0.01	**12.42**
$\mathbf{a}_5$	0.12	0.05	**0.59**	**-0.66**	**1.30**	-0.11	-0.13	-0.04	0.02	0.07	0.04	0.09	0.02	2.53
$\mathbf{a}_6$	-0.06	0.03	0.02	-0.33	-0.11	**2.10**	**-1.48**	0.04	0.10	0.39	0.04	-0.15	0.07	6.93
$\mathbf{a}_7$	-0.22	-0.22	0.26	0.02	-0.13	**-1.48**	**2.98**	0.00	-0.21	**-0.79**	-0.02	0.17	0.30	**12.03**
$\mathbf{a}_8$	-0.20	0.00	-0.03	0.12	-0.04	0.04	0.00	**1.29**	-0.17	-0.25	-0.10	0.15	0.09	1.85
$\mathbf{a}_9$	0.01	-0.38	-0.02	-0.03	0.02	0.10	-0.21	-0.17	**1.36**	-0.12	-0.08	-0.02	0.02	2.10
$\mathbf{a}_{10}$	-0.20	-0.29	0.08	-0.01	0.07	0.39	**-0.79**	-0.25	-0.12	**1.97**	0.41	0.19	0.21	5.12
$\mathbf{a}_{11}$	-0.08	-0.10	-0.05	0.01	0.04	0.04	-0.02	-0.10	-0.08	0.41	**1.14**	0.06	0.10	1.51
$\mathbf{a}_{12}$	-0.07	-0.06	-0.03	-0.04	0.09	-0.15	0.17	0.15	-0.02	0.19	0.06	**1.14**	-0.00	1.43
$\mathbf{a}_{13}$	-0.18	-0.26	0.02	0.01	0.02	0.07	0.30	0.09	0.02	0.21	0.10	-0.00	**1.23**	1.77

Table 3. *Cos-square transformation matrix and VIFs for the data on pitprops*

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	VIF
$\mathbf{a}_1$	**3.04**	**-1.92**	-0.23	-0.09	0.10	-0.07	-0.22	-0.17	-0.01	-0.21	-0.07	-0.06	-0.15	**13.14**
$\mathbf{a}_2$	**-1.88**	**3.14**	0.00	-0.05	0.04	0.01	-0.22	-0.00	-0.33	-0.28	-0.08	-0.05	-0.21	**13.71**
$\mathbf{a}_3$	-0.26	0.01	**2.69**	**-2.00**	**0.51**	0.02	0.28	-0.02	-0.01	0.08	-0.05	-0.03	0.01	**11.66**
$\mathbf{a}_4$	-0.10	-0.06	**-1.92**	**2.88**	**-0.57**	-0.31	0.03	0.10	-0.03	-0.02	0.00	-0.04	0.01	**12.42**
$\mathbf{a}_5$	0.13	0.06	**0.60**	**-0.70**	**1.27**	-0.13	-0.13	-0.03	0.02	0.07	0.04	0.09	0.03	2.53
$\mathbf{a}_6$	-0.09	0.02	0.02	-0.34	-0.12	**1.99**	**-1.64**	0.04	0.08	0.36	0.03	-0.13	0.09	6.93
$\mathbf{a}_7$	-0.21	-0.21	0.23	0.03	-0.09	**-1.26**	**3.12**	0.01	-0.18	**-0.71**	-0.02	0.13	0.23	**12.03**
$\mathbf{a}_8$	-0.22	-0.00	-0.03	0.12	-0.03	0.04	0.01	**1.28**	-0.18	-0.27	-0.09	0.15	0.09	1.85
$\mathbf{a}_9$	-0.01	-0.42	-0.02	-0.04	0.02	0.08	-0.24	-0.18	**1.34**	-0.15	-0.07	-0.02	0.01	2.10
$\mathbf{a}_{10}$	-0.23	-0.32	0.08	-0.02	0.06	0.33	**-0.85**	-0.24	-0.13	**1.96**	0.36	0.17	0.19	5.12
$\mathbf{a}_{11}$	-0.09	-0.12	-0.06	0.00	0.04	0.04	-0.03	-0.09	-0.08	0.44	**1.12**	0.05	0.09	1.51
$\mathbf{a}_{12}$	-0.08	-0.07	-0.04	-0.05	0.09	-0.15	0.19	0.16	-0.02	0.21	0.05	**1.13**	-0.01	1.43
$\mathbf{a}_{13}$	-0.20	-0.29	0.01	0.01	0.03	0.10	0.33	0.10	0.01	0.23	0.09	-0.01	**1.21**	1.77