Pak. J. Statist.
2009, Vol. 25(4), 473-503

ON RELATIONSHIPS BETWEEN UNCENTRED AND COLUMN-CENTRED PRINCIPAL COMPONENT ANALYSIS

Jorge Cadima[1] and **Ian Jolliffe**[2]

[1] Departamento de Matematica, Instituto Superior de Agronomia, Universidade Tecnica de Lisboa, Portugal
Email: jcadima@isa.utl.pt
[2] Exeter Climate Systems, School of Engineering, Computing and Mathematics, University of Exeter, UK
Email: i.t.jolliffe@ex.ac.uk

ABSTRACT

Principal component analysis (PCA) can be seen as a singular value decomposition (SVD) of a column-centred data matrix. In a number of applications, no pre-processing of the data is carried out, and it is the uncentred data matrix that is subjected to an SVD, in what is often called an uncentred PCA. This paper explores the relationships between the results from both the standard, column-centred, PCA, and its uncentred counterpart. In particular, it obtains both exact results and bounds relating the eigenvalues and eigenvectors of the covariation matrices, as well as the principal components, in both types of analysis. These relationships highlight how the eigenvalues of both the covariance matrix and the matrix of non-central second moments contain much information that is highly informative for a comparative assessment of PCA and its uncentred variant. The relations and the examples also suggest that the results of both types of PCA have more in common than might be supposed.

KEYWORDS

Principal component analysis, singular value decomposition, uncentred principal components.

2000 Mathematics Subject Classification: 62H25, 15A18

1 Introduction

Principal component analysis (PCA) can be seen as a singular value decomposition (SVD) of a column-centred data matrix. In a variety of applications, no pre-processing of the data is carried out, and the uncentred data matrix is subjected to an SVD. The SVD of a matrix that has not been previously column-centred is often referred to as an uncentred PCA. We shall use this terminology, although from a historical point of view it is debatable whether the adjective *principal* should be used in this context. Whereas the standard, column-centred, PCs are uncorrelated linear combinations of the column-centred variables which successively maximize variance, uncentred PCs are linear combinations of the uncentred variables which successively maximize non-central second moments, subject to having their crossed non-central second moments equal to zero. Despite warnings [3] there has been a tendency in some fields to equate the name singular value decomposition only with the uncentred version of PCA.

Uncentred PCAs have been used in many areas of application, such as climatology [19], astronomy [4], the study of microarrays [1], neuroimaging data [6], ecology [18], chemistry [10] and geology [15], among others.

In most cases, the use of either variant of PCA seeks to describe and explore the data, rather than to model it. Hence there is a need for a greater understanding of the nature of uncentred PCA and of its relations with the standard, column-centred, PCA. In particular, quantitative relations between both types of PCAs are essential to clarify these relations, and they do not seem to be currently available.

This paper compares the results from standard, column-centred, principal component analysis, and from its uncentred counterpart. In particular, it obtains both exact relationships and bounds involving the eigenvalues and eigenvectors of the matrices of covariances and of non-central second moments, as well as the principal components in both types of analysis. The eigenvalues of both covariation matrices tell us a great deal about the comparative behaviour of both kinds of PCAs.

Section 2 introduces the basic concepts and the notation that will be used. Section 3 presents the main results regarding the relations between column-centred and uncentred PCAs, the proofs of which are left to an Appendix. In Section 4 these results are applied in the discussion of two examples. Finally, Section 5 presents a few conclusions.

2 Preliminaries and notation

Standard principal component analysis replaces a set of p variables with p new variables, called principal components (PCs), which are uncorrelated linear combinations of the original (centred) variables. The first PC is the linear combination of maximum variance (with a unit-norm vector of coefficients); subsequent PCs are the linear combinations of maximum variance, but with the additional requirement of zero correlation with all previously defined PCs. As is well known [11], the vectors of coefficients that define the PCs (the PC loadings) are the eigenvectors of the covariance matrix of the p variables (or correlation matrix, if we work with standardized variables). Each eigenvalue of the covariance matrix is the variance of the PC defined by the corresponding eigenvector. Eigenvectors and PCs only define directions: they can be arbitrarily multiplied by -1. Some authors confusingly call the vectors of PC loadings "principal components" (e.g., [6]), but we reserve this name for the linear combinations of the observed variables that are defined by those loadings, in keeping with Hotelling [8].

Alternatively, PCA can be given via a singular value decomposition [5, 11] of the column-centred $n \times p$ data matrix, whose n rows are associated with the observation units ("individuals") and whose p columns are associated with the observed variables. Column-centring means that the raw data matrix is pre-processed, so that the mean of each column is subtracted from all observations on that matrix column, which therefore becomes a zero-mean vector of observations.

Throughout, an $m \times m$ identity matrix will be denoted by $\mathbf{I}_m$. An m-dimensional vector of ones is represented by $\mathbf{1}_m$. The subspace spanned by the columns of a given $n \times k$ matrix $\mathbf{A}$ - i.e., the rangespace of $\mathbf{A}$ - will be represented by $\mathcal{R}(\mathbf{A})$ and its orthogonal complement – i.e., the subspace of all elements of $\mathbb{R}^n$ that are orthogonal to *every* element of $\mathcal{R}(\mathbf{A})$ – by $\mathcal{R}(\mathbf{A})^{\perp}$. $\mathbf{P_A}$ indicates the matrix of orthogonal projections onto the subspace $\mathcal{R}(\mathbf{A})$, so that $\mathbf{P_A} = \mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t$, assuming $\mathbf{A}$ is of rank k. If $\mathbf{P_A}$ is an $n \times n$ matrix of orthogonal projections on $\mathcal{R}(\mathbf{A})$, then $\mathbf{I}_n - \mathbf{P_A}$ is the matrix of orthogonal projections on $\mathcal{R}(\mathbf{A})^{\perp}$.

An uncentred $n \times p$ data matrix will be represented by $\mathbf{X}_{uc}$ and its column-centred counterpart by $\mathbf{X}_{cc}$. Unless otherwise stated, it is assumed that $n > p$ and that any multi-collinearities in the data were previously removed, so that the rank of $\mathbf{X}_{cc}$ is p. It is assumed throughout that $\mathbf{X}_{uc} \neq \mathbf{X}_{cc}$, i.e., that the data columns do not all originally have zero means. In algebraic terms, column-centring results from pre-multiplying a raw data matrix $\mathbf{X}_{uc}$ by the $n \times n$ matrix $\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n}$, where $\mathbf{P}_{\mathbf{1}_n} = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^t$, i.e., $\mathbf{X}_{cc} = (\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})\mathbf{X}_{uc}$.

Let a singular value decomposition (SVD) of a rank r matrix $\frac{1}{\sqrt{n}}\mathbf{X}_{cc}$ be given by

$$\tfrac{1}{\sqrt{n}}\mathbf{X}_{cc} \;=\; \mathbf{U}\Delta\mathbf{W}^t \;=\; \sum_{i=1}^{r}\sigma_i\mathbf{u}_i\mathbf{w}_i^t, \tag{2.1}$$

where Δ is the diagonal matrix of singular values $\{\sigma_i\}_{i=1}^r$, $\mathbf{U}$ is an $n\times r$ matrix whose columns are the left singular vectors $\{\mathbf{u}_i\}_{i=1}^r$, and $\mathbf{W}$ a $p\times r$ matrix whose columns are the right singular vectors $\{\mathbf{w}_i\}_{i=1}^r$. The columns of both $\mathbf{U}$ and $\mathbf{W}$ form orthonormal sets, so that $\mathbf{U}^t\mathbf{U} = \mathbf{W}^t\mathbf{W} = \mathbf{I}_r$. If the columns of the column-centred matrix $\mathbf{X}_{cc}$ are linearly independent (i.e., if $r = p$), then $\mathbf{W}$ is an orthogonal matrix ($\mathbf{W}^t\mathbf{W} = \mathbf{W}\mathbf{W}^t = \mathbf{I}_p$). But even when there are no multicollinearities, the product $\mathbf{U}\mathbf{U}^t$ is an $n\times n$ matrix that is *not* an identity matrix: it can have no more than $r = p$ non-zero eigenvalues, since its rank must equal the rank of $\mathbf{X}_{cc}$. The multiplicative constant $\frac{1}{\sqrt{n}}$ in equation (2.1) is useful because the matrix product $(\frac{1}{\sqrt{n}}\mathbf{X}_{cc})^t(\frac{1}{\sqrt{n}}\mathbf{X}_{cc})$ produces the covariance matrix $\mathbf{S}$ of the original data. Therefore

$$\mathbf{S} \;=\; (\mathbf{W}\Delta\mathbf{U}^t)(\mathbf{U}\Delta\mathbf{W}^t) \;=\; \mathbf{W}\Delta^2\mathbf{W}^t\;, \tag{2.2}$$

is a spectral decomposition [7, p.171] for $\mathbf{S}$. The eigenvectors and the (non-zero) eigenvalues of the covariance matrix $\mathbf{S}$ are, respectively, the right singular vectors and the squared singular values of $\frac{1}{\sqrt{n}}\mathbf{X}_{cc}$. The j-th principal component (i.e., vector of PC scores) is given by $\boldsymbol{\eta}_j = \mathbf{X}_{cc}\mathbf{w}_j \in \mathbb{R}^n$, and again resorting to the SVD of $\frac{1}{\sqrt{n}}\mathbf{X}_{cc}$, we have $\boldsymbol{\eta}_j = \sqrt{n}\sum_{i=1}^{p}\sigma_i\mathbf{u}_i\mathbf{w}_i^t\cdot\mathbf{w}_j = \sqrt{n}\sigma_j\mathbf{u}_j$. Thus, the left singular vectors of $\mathbf{X}_{cc}$ are the principal components re-scaled to have unit length with the ℓ_2 norm.

Similar comments apply to the counterpart of the covariance matrix $\mathbf{S}$, if one begins with an uncentred data matrix, which is the matrix of non-central second moments,

$$\mathbf{T} \;=\; \left(\tfrac{1}{\sqrt{n}}\mathbf{X}_{uc}\right)^t\left(\tfrac{1}{\sqrt{n}}\mathbf{X}_{uc}\right)\;. \tag{2.3}$$

The matrices $\mathbf{S}$ and $\mathbf{T}$ will be collectively called *covariation* matrices. The uncentred correlation coefficients are sometimes referred to as *congruence coefficients* [9]. Superscripts $[c]$ and $[u]$ are used to distinguish the column-centred and the uncentred case, in the eigen- or singular value decompositions.

3 Main results

A first result relates the singular values of the column-centred and the uncentred data matrices, and therefore the eigenvalues of the covariance (correlation) matrix $\mathbf{S}$ and its uncentred counterpart $\mathbf{T}$, as well as the proportion of variation explained by the PCs in both cases. This result is a direct application of the following theorem which Takane and Shibayama [17] attribute to Yanai and Takeuchi [21, in Japanese], and which is related to the eigenvalue interlacing theorems for symmetric matrices whose difference is of reduced rank [7, p.184].

Theorem 1 (Yanai & Takeuchi Separation Theorem)**.** *Let* $\mathbf{Y}$ *be an* $n \times p$ *matrix. Let* $\mathbf{P}_n$ *and* $\mathbf{P}_p$ *be, respectively,* $n \times n$ *and* $p \times p$ *orthogonal projection matrices of rank* r *and* s*. Then,*

$$\sigma_{j+t}(\mathbf{Y}) \;\leq\; \sigma_j(\mathbf{P}_n \mathbf{Y} \mathbf{P}_p) \;\leq\; \sigma_j(\mathbf{Y}) \,, \tag{3.1}$$

where $\sigma_j(\cdot)$ *indicates the* j*-th largest singular value of the matrix and* $t = n + p - (r + s)$*.*

Corollary 3.1. Let $\mathbf{S}$ be the covariance matrix for a given data set, and $\mathbf{T}$ its corresponding matrix of non-central second moments. Let $\lambda_j(\cdot)$ be the j-th largest eigenvalue of either matrix, and $\pi_j(\cdot) = \frac{\lambda_j(\cdot)}{\sum_{i=1}^{p} \lambda_i(\cdot)}$ the proportion of total variation accounted for by the corresponding PC. Then $\lambda_{j+1}(\mathbf{T}) \leq \lambda_j(\mathbf{S}) \leq \lambda_j(\mathbf{T})$. Equivalently, $\pi_{j+1}(\mathbf{T}) \cdot \alpha \leq \pi_j(\mathbf{S}) \leq \pi_j(\mathbf{T}) \cdot \alpha$. where $\alpha = \frac{tr(\mathbf{T})}{tr(\mathbf{S})} = \frac{1}{1 - \|\mathbf{c}_m\|^2 / tr(\mathbf{T})}$, with $tr(\cdot)$ the matrix trace, $\mathbf{c}_m = \frac{1}{n}\mathbf{X}_{uc}^t \mathbf{1}_n$ and $\|\mathbf{c}_m\|^2 = \mathbf{c}_m^t \mathbf{c}_m$. The p-dimensional vector $\mathbf{c}_m$ is the vector of column means of the uncentred data matrix $\mathbf{X}_{uc}$ and $\|\mathbf{c}_m\|^2$ is the sum of squares of those column means.

Hence, the eigenvalues of $\mathbf{S}$ are interlaced with those of $\mathbf{T}$: an eigenvalue of $\mathbf{S}$ cannot be larger than the eigenvalue of $\mathbf{T}$ of equal rank - nor smaller than $\mathbf{T}$'s next ranking eigenvalue. However, since $\alpha \geq 1$, the proportion of variance accounted for by the j-th column-centred PC can be either smaller or larger than the proportion of total variation (sum of non-central second moments) of the uncentred PC of equal rank. In our context, it is possible to go further than the above separation theorem and obtain exact relations between the eigenvalues of matrices $\mathbf{S}$ and $\mathbf{T}$, thus providing a fuller understanding of the effects of column-centring.

The following Theorem addresses the eigendecomposition of the difference in the $p \times p$ covariation matrices, $\mathbf{T} - \mathbf{S}$, which has important consequences for the comparative study of column-centred and uncentred PCAs. This is a symmetric matrix, with a full set of orthonormal real eigenvectors associated with its p real eigenvalues.

Theorem 2. *Let* $\mathbf{X}_{uc}$ *be an* $n \times p$ *data matrix, whose vector of column means is* $\mathbf{c}_m = \frac{1}{n}\mathbf{X}_{uc}^t \mathbf{1}_n$ *and whose matrix of non-central second moments is* $\mathbf{T} = \frac{1}{n}\mathbf{X}_{uc}^t \mathbf{X}_{uc}$*. Let* $\mathbf{S}$ *be the covariance matrix of the data, and* $\mathbf{X}_{cc}$ *be the column-centred data matrix so that* $\mathbf{S} = \frac{1}{n}\mathbf{X}_{cc}^t \mathbf{X}_{cc}$*. Then,*

1. ***Eigendecomposition of*** $\mathbf{T}-\mathbf{S}$*: The matrix* $\mathbf{T}-\mathbf{S}$ *can be written as*

$$\mathbf{T}-\mathbf{S} \quad = \quad \mathbf{c}_m \cdot \mathbf{c}_m^t \ , \tag{3.2}$$

It is a matrix of rank one, whose only nonzero eigenvalue is $\|\mathbf{c}_m\|^2$*, associated with the unit-norm eigenvector* $\mathbf{c}_m/\|\mathbf{c}_m\|$*. Any vector* $\mathbf{y} \in \mathcal{R}(\mathbf{c}_m)^{\perp}$ *is an eigenvector of* $\mathbf{T}-\mathbf{S}$*, with eigenvalue zero.*

2. *For any vector* $\mathbf{x} \in \mathbb{R}^p$*, the difference in the Rayleigh-Ritz ratios of* $\mathbf{T}$ *and* $\mathbf{S}$ *is:*

$$\frac{\mathbf{x}^t\mathbf{T}\mathbf{x}}{\mathbf{x}^t\mathbf{x}} - \frac{\mathbf{x}^t\mathbf{S}\mathbf{x}}{\mathbf{x}^t\mathbf{x}} \ = \ \|\mathbf{c}_m\|^2 \cdot \cos^2(\mathbf{c}_m, \mathbf{x}) \ . \tag{3.3}$$

3. *If one of the eigenvectors of either* $\mathbf{T}$ *or* $\mathbf{S}$ *is exactly equal to the (unit-norm) vector of means of the columns of* $\mathbf{X}_{uc}$*, the vector* $\mathbf{c}_m/\|\mathbf{c}_m\|$*, then:*

   - $\mathbf{S}$ *and* $\mathbf{T}$ *have a common set of eigenvectors (including the vector* $\mathbf{c}_m/\|\mathbf{c}_m\|$*);*
   - *The* $p-1$ *eigenvalues of* $\mathbf{S}$ *and* $\mathbf{T}$ *that are not associated with the common eigenvector* $\mathbf{c}_m/\|\mathbf{c}_m\|$ *are shared;*
   - *If the common eigenvector* $\frac{\mathbf{c}_m}{\|\mathbf{c}_m\|}$ *has eigenvalue* $\lambda^{[u]}$ *in* $\mathbf{T}$ *and* $\lambda^{[c]}$ *in* $\mathbf{S}$*, then* $\lambda^{[c]} = \lambda^{[u]} - \|\mathbf{c}_m\|^2$*.*

   Note that the rank of this final eigenvalue of $\mathbf{S}$ *need not be the same as in* $\mathbf{T}$*.*

A few comments on the statements in this Theorem:

- It is a direct consequence of the first point of Theorem 2 that

$$tr(\mathbf{T}) - tr(\mathbf{S}) \ = \ \|\mathbf{c}_m\|^2 \ . \tag{3.4}$$

- The matrix $(\mathbf{T}-\mathbf{S})/\|\mathbf{c}_m\|^2$ is a matrix of orthogonal projections onto the one-dimensional subspace $\mathcal{R}(\mathbf{c}_m)$, but $\mathbf{T}-\mathbf{S}$ is not: it does not preserve the vectors in that subspace, but rather re-scales them.

- It is a direct consequence of (3.3) that for any vector $\mathbf{x}$ in the $(p-1)$-dimensional subspace $\mathcal{R}(\mathbf{c}_m)^{\perp}$, then $\mathbf{Tx} = \mathbf{Sx}$. If any such vector is an eigenvector of one of the matrices, it must also be an eigenvector (with a common eigenvalue) of the other matrix.

- It is known that the eigenvalues of a symmetric matrix successively maximize its Rayleigh-Ritz ratio, among all vectors $\mathbf{x}$ that are orthogonal to previously determined solutions [7, p.176]. Equation (3.3) tells us that maximizing $\mathbf{S}$'s Rayleigh-Ritz ratio is a trade-off between maximizing the Rayleigh-Ritz ratio of $\mathbf{T}$ and staying as orthogonal as possible to vector $\mathbf{c}_m$, with $\|\mathbf{c}_m\|^2$ playing the role of a penalizing constant for excessive closeness to $\mathbf{c}_m$.

- The above remarks help explain why many eigenvectors in column-centred and uncentred PCAs are often similar, namely when they have approximately equal-sized eigenvalues (as distinct from the proportions of variation which they account for). This is illustrated by the examples in Section 4.

- It is unlikely that any of the matrices will have an eigenvector that is exactly colinear with $\mathbf{c}_m$, but it is very often the case that $\mathbf{T}$'s first eigenvector is close to that direction, in $\mathbb{R}^p$, which unites the origin and the centre of gravity of the n-point scatter that is defined by the rows of $\mathbf{X}_{uc}$. The final point in the Theorem can therefore be useful as a benchmark against which to compare what happens when $\mathbf{T}$'s first eigenvector diverges from the direction of $\mathbf{c}_m$.

We now consider the implications of these results for the relations between both types of PCAs.

Proposition 3.1. *Let $\mathbf{X}_{uc}$, $\mathbf{X}_{cc}$, $\mathbf{c}_m$, $\mathbf{S}$ and $\mathbf{T}$ be defined as in Theorem 2. Let $\mathbf{S}$ and $\mathbf{T}$ have eigenvalues $\left\{\lambda_i^{[c]}\right\}_{i=1}^p$ and $\left\{\lambda_i^{[u]}\right\}_{i=1}^p$, and eigenvectors $\left\{\mathbf{w}_i^{[c]}\right\}_{i=1}^p$ and $\left\{\mathbf{w}_j^{[u]}\right\}_{j=1}^p$, respectively. Let $\boldsymbol{\eta}_i^{[c]}$ denote the i-th column-centred PC and $\boldsymbol{\eta}_j^{[u]}$ its j-th uncentred counterpart. Then,*

1. *The mean score, $\bar{s}_j$, on the j-th uncentred principal component is $\bar{s}_j = \|\mathbf{c}_m\| \cdot \cos(\mathbf{w}_j^{[u]}, \mathbf{c}_m)$. Equivalently,*
$$\cos(\mathbf{w}_j^{[u]}, \mathbf{c}_m) \;=\; \frac{\bar{s}_j}{\|\mathbf{c}_m\|} \;, \tag{3.5}$$

2. *The variance of the scores on the j-th uncentred principal component is*
$$\mathrm{var}(\boldsymbol{\eta}_j^{[u]}) \;=\; \lambda_j^{[u]} - \|\mathbf{c}_m\|^2 \cdot \cos^2(\mathbf{c}_m, \mathbf{w}_j^{[u]}) \;. \tag{3.6}$$

   The covariance of the scores on the i-th and j-th uncentred Principal Components ($i \neq j$) is
$$\mathrm{cov}(\boldsymbol{\eta}_i^{[u]}, \boldsymbol{\eta}_j^{[u]}) \;=\; -\|\mathbf{c}_m\|^2 \cdot \cos(\mathbf{c}_m, \mathbf{w}_i^{[u]}) \cos(\mathbf{c}_m, \mathbf{w}_j^{[u]}) \;. \tag{3.7}$$

The covariance matrix of the uncentred PCs is $\mathbf{W}^{[u]^t}\mathbf{S}\mathbf{W}^{[u]}$*, where* $\mathbf{W}^{[u]}$ *is the matrix whose columns are the eigenvectors of* $\mathbf{T}$*. This matrix has the same eigenvalues as* $\mathbf{S}$*, the data's covariance matrix. The coordinates of the* j*-th eigenvector of* $\mathbf{W}^{[u]^t}\mathbf{S}\mathbf{W}^{[u]}$ *are the cosines of the angles which the* j*-th eigenvector of* $\mathbf{S}$ *forms with each of* $\mathbf{T}$*'s eigenvectors.*

3. *The cross-covariance between the* i*-th column-centred PC and the* j*-th uncentred PC is*

$$\mathrm{cov}(\boldsymbol{\eta}_i^{[c]},\boldsymbol{\eta}_j^{[u]}) \;=\; \lambda_i^{[c]}\cdot\cos(\mathbf{w}_i^{[c]},\mathbf{w}_j^{[u]})\;. \tag{3.8}$$

4. *The correlation between the* i*-th column-centred PC and the* j*-th uncentred PC is*

$$r_{\boldsymbol{\eta}_i^{[c]},\boldsymbol{\eta}_j^{[u]}} = \sqrt{\frac{\mathrm{var}(\boldsymbol{\eta}_i^{[c]})}{\mathrm{var}(\boldsymbol{\eta}_j^{[u]})}}\cdot\cos(\mathbf{w}_i^{[c]},\mathbf{w}_j^{[u]}) = \sqrt{\frac{\lambda_i^{[c]}}{\lambda_j^{[u]}-\|\mathbf{c}_m\|^2\cdot\cos^2(\mathbf{c}_m,\mathbf{w}_j^{[u]})}}\cdot\cos(\mathbf{w}_i^{[c]},\mathbf{w}_j^{[u]})\;. \tag{3.9}$$

5. *The cosine of the angle in* $\mathbb{R}^n$ *between the* i*-th column-centred PC and the* j*-th uncentred PC is*

$$\cos(\boldsymbol{\eta}_i^{[c]},\boldsymbol{\eta}_j^{[u]}) \;=\; \sqrt{\frac{\lambda_i^{[c]}}{\lambda_j^{[u]}}}\cdot\cos(\mathbf{w}_i^{[c]},\mathbf{w}_j^{[u]})\;. \tag{3.10}$$

6. *The standard Euclidean distance in* $\mathbb{R}^n$ *between the* i*-th column-centred PC and the* j*-th uncentred PC is the smallest of the two distances*

$$\|\boldsymbol{\eta}_i^{[c]}\pm\boldsymbol{\eta}_j^{[u]}\| \;=\; \sqrt{n\lambda_i^{[c]}\cdot\|\mathbf{w}_i^{[c]}\pm\mathbf{w}_j^{[u]}\|^2+n\,(\lambda_j^{[u]}-\lambda_i^{[c]})}\;. \tag{3.11}$$

7. *The standard Euclidean distance in* $\mathbb{R}^n$ *between the* i*-th unit-norm column-centred PC,* $\mathbf{u}_i^{[c]}=\frac{\boldsymbol{\eta}_i^{[c]}}{\|\boldsymbol{\eta}_i^{[c]}\|}$*, and the* j*-th unit-norm uncentred PC,* $\mathbf{u}_j^{[u]}=\frac{\boldsymbol{\eta}_j^{[u]}}{\|\boldsymbol{\eta}_j^{[u]}\|}$*, is the smallest of*

$$\|\mathbf{u}_i^{[c]}\pm\mathbf{u}_j^{[u]}\| \;=\; \sqrt{2\left[1\pm\sqrt{\frac{\lambda_i^{[c]}}{\lambda_j^{[u]}}}\cdot\cos(\mathbf{w}_i^{[c]},\mathbf{w}_j^{[u]})\right]}\;. \tag{3.12}$$

8. *We have*

$$\left(\lambda_j^{[u]}-\lambda_i^{[c]}\right)\cdot\cos\left(\mathbf{w}_i^{[c]},\mathbf{w}_j^{[u]}\right) \;=\; \|\mathbf{c}_m\|^2\cdot\cos(\mathbf{w}_i^{[c]},\mathbf{c}_m)\cdot\cos(\mathbf{w}_j^{[u]},\mathbf{c}_m) \tag{3.13}$$

which, when $\cos(\mathbf{w}_i^{[c]},\mathbf{w}_j^{[u]})\neq 0$*, is equivalent to*

$$\lambda_j^{[u]} \;=\; \lambda_i^{[c]} \;+\; \frac{\|\mathbf{c}_m\|^2\cdot\cos(\mathbf{w}_i^{[c]},\mathbf{c}_m)\cdot\cos(\mathbf{w}_j^{[u]},\mathbf{c}_m)}{\cos(\mathbf{w}_i^{[c]},\mathbf{w}_j^{[u]})}\;. \tag{3.14}$$

9. *Let A_k (a_k) denote the largest (smallest) sum of variances of any k uncentred PCs. Then, the sums of the k largest and the k smallest eigenvalues of* **S** *satisfy:*

$$\sum_{i=1}^{k} \lambda_i^{[c]} \geq A_k \qquad \text{and} \qquad \sum_{i=p-k+1}^{p} \lambda_i^{[c]} \leq a_k \tag{3.15}$$

In particular, we have the following lower bound for the largest eigenvalue of **S***:*

$$\lambda_1^{[c]} \;\geq\; \max_j \{\lambda_j^{[u]} - \|\mathbf{c}_m\|^2 \cdot \cos^2(\mathbf{c}_m, \mathbf{w}_j^{[u]})\} \;=\; \max_j \{\text{var}(\boldsymbol{\eta}_j^{[u]})\} \,. \tag{3.16}$$

10. *Let $\xi_i = \lambda_i^{[c]} + \|\mathbf{c}_m\|^2 \cdot \cos^2(\mathbf{c}_m, \mathbf{w}_i^{[c]})$ ($i = 1, ..., p$). Let B_k (b_k) denote the largest (smallest) sum of any k different ξ_i's. Then, the sums of the k largest and the k smallest eigenvalues of* **T** *satisfy:*

$$\sum_{j=1}^{k} \lambda_j^{[u]} \geq B_k \qquad \text{and} \qquad \sum_{j=p-k+1}^{p} \lambda_j^{[u]} \leq b_k \tag{3.17}$$

In particular, we have the following lower bound for the largest eigenvalue of **T***:*

$$\lambda_1^{[u]} \;\geq\; \max_i \{\lambda_i^{[c]} + \|\mathbf{c}_m\|^2 \cdot \cos^2(\mathbf{c}_m, \mathbf{w}_i^{[c]})\} \,. \tag{3.18}$$

11. *For any $k = 1, ..., p$, we have the following bounds for sums involving the k largest eigenvalues of* **T** *and* **S***:*

$$\sum_{i=1}^{k} \cos^2\left(\mathbf{w}_i^{[c]}, \mathbf{c}_m\right) \;\leq\; \sum_{i=1}^{k} \frac{\lambda_i^{[u]} - \lambda_i^{[c]}}{\|\mathbf{c}_m\|^2} \;\leq\; \sum_{i=1}^{k} \cos^2\left(\mathbf{w}_j^{[u]}, \mathbf{c}_m\right) \tag{3.19}$$

In particular, the first eigenvectors for each centring, and their associated eigenvalues, satisfy the following bounds:

$$\cos^2\left(\mathbf{w}_1^{[c]}, \mathbf{c}_m\right) \;\leq\; \frac{\lambda_1^{[u]} - \lambda_1^{[c]}}{\|\mathbf{c}_m\|^2} \;\leq\; \cos^2\left(\mathbf{w}_1^{[u]}, \mathbf{c}_m\right) \tag{3.20}$$

For any $k = 1, ..., p$, we have the following bounds for sums involving the k smallest eigenvalues of **T** *and* **S***:*

$$\sum_{j=p-k+1}^{p} \cos^2\left(\mathbf{w}_j^{[u]}, \mathbf{c}_m\right) \;\leq\; \sum_{j=p-k+1}^{p} \frac{\lambda_i^{[u]} - \lambda_i^{[c]}}{\|\mathbf{c}_m\|^2} \;\leq\; \sum_{i=p-k+1}^{p} \cos^2\left(\mathbf{w}_i^{[c]}, \mathbf{c}_m\right) \tag{3.21}$$

12. *The largest eigenvalue of* **T** *and the largest and smallest eigenvalues of* **S** *satisfy the following bounds:*

$$\|\mathbf{c}_m\|^2 + \lambda_p^{[c]} \quad \leq \quad \lambda_1^{[u]} \quad \leq \quad \|\mathbf{c}_m\|^2 + \lambda_1^{[c]} \tag{3.22}$$

13. *The cosine of the angle in* $\mathbb{R}^p$ *between the i-th eigenvector of* **S** *and the j-th eigenvector of* **T** *satisfies the following bound:*

$$\left|\cos(\mathbf{w}_i^{[c]}, \mathbf{w}_j^{[u]})\right| \leq \sqrt{\frac{\lambda_j^{[u]} - \|\mathbf{c}_m\|^2 \cdot \cos^2(\mathbf{c}_m, \mathbf{w}_j^{[u]})}{\lambda_i^{[c]}}} = \sqrt{\frac{\mathrm{var}(\boldsymbol{\eta}_j^{[u]})}{\mathrm{var}(\boldsymbol{\eta}_i^{[c]})}} . \tag{3.23}$$

14. *The cosine of the angle between the first eigenvector of each of* **T** *and* **S** *satisfies:*

$$\cos^2(\mathbf{w}_1^{[c]}, \mathbf{w}_1^{[u]}) \geq \frac{\cos^2(\mathbf{w}_1^{[c]}, \mathbf{c}_m)}{\cos^2(\mathbf{w}_1^{[u]}, \mathbf{c}_m)} . \tag{3.24}$$

3.1 Implications of Proposition 3.1

- Equation (3.5) states that the cosine between an eigenvector of **T** and the vector $\mathbf{c}_m$ of column means of $\mathbf{X}_{uc}$ is directly proportional to the mean score on the uncentred PC which that eigenvector defines. In particular, the mean score of an uncentred PC is zero if and only if the vector of PC loadings is orthogonal to $\mathbf{c}_m$. As was seen in the discussion following Theorem 2, eigenvectors of **T** that are orthogonal to $\mathbf{c}_m$ are also eigenvectors of **S** with a common eigenvalue and from equation (3.10), define a common PC. Therefore, an uncentred PC can only have a zero mean score if it coincides with some column-centred PC.

- Expressions (3.6) and (3.7) highlight that uncentred PCs are different from their column-centred counterparts in important respects. They are not uncorrelated among themselves, in general, and their associated eigenvalues do not indicate PC variances, but only the non-central second moments of those PCs, as can be seen from equation (3.6).

- Equation (3.9) implies the condition for strongly correlated PCs of both types:

$$r^2_{\boldsymbol{\eta}_i^{[c]}, \boldsymbol{\eta}_j^{[u]}} \geq \varepsilon \quad \Longleftrightarrow \quad \cos^2(\mathbf{w}_i^{[c]}, \mathbf{w}_j^{[u]}) \geq \varepsilon \cdot \frac{\mathrm{var}(\boldsymbol{\eta}_j^{[u]})}{\mathrm{var}(\boldsymbol{\eta}_i^{[c]})} .$$

This condition cannot be met if $\lambda_i^{[c]} < \varepsilon \cdot \text{var}(\boldsymbol{\eta}_j^{[u]})$. The PCs cannot be perfectly correlated unless $\cos^2(\mathbf{w}_i^{[c]}, \mathbf{w}_j^{[u]}) = \frac{\text{var}(\boldsymbol{\eta}_j^{[u]})}{\text{var}(\boldsymbol{\eta}_i^{[c]})} = \frac{\left(\lambda_j^{[u]} - \|\mathbf{c}_m\|^2 \cos^2(\mathbf{c}_m, \mathbf{w}_j^{[u]})\right)}{\lambda_i^{[c]}}$. Equation (3.10) tells us that the cosine of the angle between the n-dimensional vectors of PC scores is related in a more direct way to their associated eigenvalues.

- The Euclidean distance (3.11) between a column-centred PC and an uncentred PC provides a measure for the similarity of the scores on both components, but also a measure of the similarity of the "functional forms" of both principal components, which is useful whenever the nature of the data suggests that the patterns of these components are meaningful. Since PCs can be arbitrarily multiplied by -1, we must consider the norms of both $\boldsymbol{\eta}_i^{[c]} - \boldsymbol{\eta}_j^{[u]}$ and $\boldsymbol{\eta}_i^{[c]} + \boldsymbol{\eta}_j^{[u]}$. The analogous expression (3.12) is useful for comparing such "waveforms" regardless of size, since the PCs have been standardized to have unit norm.

- Equality (3.13) has many interesting implications. It indicates that if $\mathbf{S}$ and $\mathbf{T}$ share a common eigenvalue, $\lambda_i^{[c]} = \lambda_j^{[u]}$, then either the associated eigenvector of $\mathbf{S}$, $\mathbf{w}_i^{[c]}$, is orthogonal to $\mathbf{c}_m$, in which case (as we saw in the discussion following Theorem 2) it is also an eigenvector of $\mathbf{T}$ with the same eigenvalue; or $\mathbf{T}$'s associated eigenvector $\mathbf{w}_j^{[u]}$ is orthogonal to $\mathbf{c}_m$, in which case it is also an eigenvector of $\mathbf{S}$ with the same eigenvalue. If none of the matrices have repeated eigenvalues, then $\lambda_i^{[c]} = \lambda_j^{[u]}$ implies $\mathbf{w}_i^{[c]} = \mathbf{w}_j^{[u]}$. Recall, from Theorem 1, that it is only possible for $\lambda_i^{[c]} = \lambda_j^{[u]}$ if $i = j$ or $i = j+1$, unless there are repeated eigenvalues in either matrix. If any of the matrices has repeated eigenvalues, there is an element of arbitrariness in choosing their associated eigenvectors, but it is always possible to choose those eigenvectors in such a way that $\mathbf{S}$ and $\mathbf{T}$ share an eigenvector. This line of reasoning remains approximately valid when the two eigenvalues in question are only approximately equal, although the discussion in this case is not entirely straightforward, as will be seen in Sections 4 and 5.

- By a similar argument, if $\mathbf{w}_i^{[c]}$ is orthogonal to $\mathbf{w}_j^{[u]}$, then (3.13) implies that either $\mathbf{w}_i^{[c]} \perp \mathbf{c}_m$, or $\mathbf{w}_j^{[u]} \perp \mathbf{c}_m$ (or both). If $\mathbf{w}_i^{[c]} \perp \mathbf{c}_m$, then $\mathbf{T}\mathbf{w}_i^{[c]} = \mathbf{S}\mathbf{w}_i^{[c]} = \lambda_i^{[c]}\mathbf{w}_i^{[c]}$, with a similar result in the case of $\mathbf{w}_j^{[u]}$. Hence, an eigenvector of $\mathbf{S}$ can only be orthogonal to an eigenvector of $\mathbf{T}$ if at least one of them is a common eigenvector for both matrices, with the same eigenvalue.

- But common eigenvectors of $\mathbf{T}$ and $\mathbf{S}$ (apart from a multiplication by -1) may be associated with different eigenvalues. If $\mathbf{w}_i^{[c]} = \pm\mathbf{w}_j^{[u]}$, the column-centred PC $\boldsymbol{\eta}_i^{[c]}$ is the result of column-centring the uncentred PC $\boldsymbol{\eta}_j^{[u]}$. From equations (3.13) and (3.6), $\cos(\mathbf{w}_i^{[c]}, \mathbf{w}_j^{[u]}) = \pm 1$ implies

$$\lambda_j^{[u]} - \lambda_i^{[c]} \;=\; \|\mathbf{c}_m\|^2 \cos^2(\mathbf{c}_m, \mathbf{w}_j^{[u]}) \quad \Longleftrightarrow \quad \mathrm{var}(\boldsymbol{\eta}_j^{[u]}) \;=\; \mathrm{var}(\boldsymbol{\eta}_i^{[c]}) \;,$$

and $\lambda_i^{[c]} \neq \lambda_j^{[u]}$, unless $\mathbf{w}_i^{[c]}, \mathbf{w}_j^{[u]} \perp \mathbf{c}_m$. The PCs $\boldsymbol{\eta}_j^{[u]}$ and $\boldsymbol{\eta}_i^{[c]}$ are not colinear, unless $\mathbf{w}_i^{[c]}, \mathbf{w}_j^{[u]} \perp \mathbf{c}_m$, since equation (3.10) gives $\cos^2(\boldsymbol{\eta}_j^{[u]}, \boldsymbol{\eta}_i^{[c]}) = \lambda_i^{[c]}/\lambda_j^{[u]}$, which can only equal 1 if $\mathbf{w}_j^{[u]} \perp \mathbf{c}_m$.

- In bounds (3.15), there is no guarantee that the variances of the uncentred PCs are ranked in the same order as their corresponding eigenvalues. The nature of the inequalities implies, of course, that similar bounds are valid if A_k and a_k are replaced by the sum of variances of *any k* uncentred PCs, but the bounds given in (3.15) are the optimal ones. A similar comment applies to the bounds (3.17). Relaxations of this kind produce the bounds (3.19) which will therefore, in general, be less tight (though possibly more useful) than the previous set of bounds. When $k = p$, all three members in the inequalities (3.19) and (3.21) will have value 1.

- The inequalities (3.20) imply that the vector $\mathbf{c}_m$ always forms a smaller angle with the direction of **T**'s first eigenvector, than with the direction of the first eigenvector of **S**. The former angle is small if $\lambda_1^{[u]} \gg \lambda_1^{[c]}$ (as often happens).

- The upper bound (3.22) is less tight than the upper bound obtained from the right-hand inequality in (3.20), but it does not depend on the eigenvector. The bounds (3.22) are also useful to highlight that, if $\|\mathbf{c}_m\|^2 \gg tr(\mathbf{S})$, then most of the excess in the trace of **T**, vis-a-vis the trace of **S**, will be associated with **T**'s largest eigenvalue. Since none of the eigenvalues of **S** can exceed equal-rank eigenvalues of **T**, the remaining eigenvalues of both matrices must, in this case, be similar.

- The bound (3.24) implies that the first eigenvector of **S** must form a smaller angle with **T**'s first eigenvector than with the vector $\mathbf{c}_m$ of column means.

4 Examples

Consider the n-point scatter in $\mathbb{R}^p$ which is defined by the n rows of an uncentred data matrix $\mathbf{X}_{uc}$. Equation (3.4) indicates that the total variation of the data set about the origin, as measured by $tr(\mathbf{T})$, can be decomposed into the sum of the squared distance from the origin of the cloud's centre of gravity ($\|\mathbf{c}_m\|^2$), plus the within-cloud variability of the points around that center of gravity, as measured by $tr(\mathbf{S})$. A fairly trivial situation will not be considered in great detail: if the cloud's center of gravity is close to the origin when compared to the within-cloud variability, i.e., if $\|\mathbf{c}_m\|^2 \approx 0$ ($tr(\mathbf{T}) \approx tr(\mathbf{S})$), the two variation matrices will be similar ($\mathbf{T} \approx \mathbf{S}$) and both eigendecompositions and PCAs will be similar. The two examples that are now considered illustrate situations where (i) the cloud's center of gravity is very far from the origin, when compared with the variability around the centre of gravity, i.e., $\|\mathbf{c}_m\|^2 \gg tr(\mathbf{S})$ and $\|\mathbf{c}_m\|^2 \approx tr(\mathbf{T})$ (first example); or (ii) the inner-cloud variability, $tr(\mathbf{S})$, is comparable in size to the distance of the cloud's centre of gravity from the origin: $tr(\mathbf{S}) \approx \|\mathbf{c}_m\|^2$ (second example).

4.1 The crayfish data

Somers [16] discusses a 13-variable dataset of morphometric measurements on $n = 63$ crayfish from Lake Opeongo, in Canada. We shall consider a correlation matrix PCA of the data. For our purposes, this means that the column-centred data matrix $\mathbf{X}_{cc}$ is the matrix of standardized data, whereas the columns of the uncentred data matrix, $\mathbf{X}_{uc}$ are the observed values, divided by their standard deviation, but not mean-centred. Hence, matrix $\mathbf{S}$ is the correlation matrix of the original variables and matrix $\mathbf{T}$ is the matrix of non-central second moments of the scaled data. The trace of the 13×13 matrix $\mathbf{T}$ is $tr(\mathbf{T}) = 2857.302$. The vector of column means of the uncentred data matrix has norm squared $\|\mathbf{c}_m\|^2 = 2844.302$, that is, over 99.5% of the trace of $\mathbf{T}$. Such extremely high values of this ratio are not uncommon for morphometric data sets, reflecting the fact that individual variability is small, compared to the overall (scaled) size of the individuals. From equation (3.3), the non-central variability associated with the direction $\mathbf{x} = \mathbf{c}_m$ is given by $\frac{\mathbf{c}_m^t \mathbf{T} \mathbf{c}_m}{\mathbf{c}_m^t \mathbf{c}_m} = \frac{\mathbf{c}_m^t \mathbf{S} \mathbf{c}_m}{\mathbf{c}_m^t \mathbf{c}_m} + \|\mathbf{c}_m\|^2 \geq \|\mathbf{c}_m\|^2$. Hence, the proportion of total non-central variability associated with that direction must exceed $\|\mathbf{c}_m\|^2 / tr(\mathbf{T}) = 0.993$. The main direction of variability about the origin in the 13-dimensional scatter of $n = 63$ points must therefore be very close to the direction that unites the scatter's centre of gravity with the origin, reflecting the dominance of overall size. In fact, the cosine of the angle between $\mathbf{T}$'s first eigenvector and $\mathbf{c}_m$ is almost unity: $|\cos(\mathbf{w}_1^{[u]}, \mathbf{c}_m)| = 0.9999998$. Necessarily, the other $p-1$ eigenvectors of $\mathbf{T}$ are nearly orthogonal to $\mathbf{c}_m$ and therefore, as was seen in the discussion of Theorem 2, we should expect fairly similar sets of eigenvalues and eigenvectors to those of the data's correlation matrix, $\mathbf{S}$.

Table 1: The eigenvalues of the correlation matrix (**S**) and of the matrix of non-central second moments of the scaled crayfish data (**T**), and the absolute values of the cosines of the angles between each eigenvector and the vector $\mathbf{c}_m$. The second column gives the variances of each uncentred PC. Column seven gives the values ξ_i, used in the bounds of point 10) of Proposition 3.1.

	Uncentred		Eigenvalues		Column-centred	
Rank	$\text{var}(\boldsymbol{\eta}_j^{[u]})$	$\lvert\cos(\mathbf{w}_j^{[u]},\mathbf{c}_m)\rvert$	**T**	**S**	$\lvert\cos(\mathbf{w}_i^{[c]},\mathbf{c}_m)\rvert$	ξ_i
1	7.60997	1.00000	2851.91089	8.00809	0.96851	2676.01891
2	2.38198	0.00045	2.38255	2.10800	0.20844	125.68760
3	0.73956	0.00014	0.73961	0.71934	0.05167	8.31378
4	0.53522	0.00001	0.53522	0.53519	0.00185	0.54489
5	0.38785	0.00014	0.38790	0.37425	0.03321	3.51137
6	0.30207	0.00019	0.30218	0.28851	0.01145	0.66134
7	0.28607	0.00007	0.28608	0.26208	0.05121	7.72230
8	0.24008	0.00008	0.24010	0.22377	0.05652	9.31028
9	0.18121	0.00009	0.18124	0.16867	0.03946	4.59844
10	0.15549	0.00006	0.15551	0.14149	0.06425	11.88184
11	0.10074	0.00006	0.10075	0.09098	0.05553	8.86101
12	0.04757	0.00001	0.04757	0.04744	0.00623	0.15777
13	0.03218	0.00000	0.03218	0.03218	0.00017	0.03226

Table 1 confirms the overall similarity in the eigenvalues of **T** and **S**, except for the largest eigenvalue of each matrix, which are associated with eigenvectors that are close to the direction of $\mathbf{c}_m$. For these eigenvectors, as indicated in point 3) of Theorem 2, the similarity is between $\lambda_1^{[c]} = 8.0081$ and $\lambda_1^{[u]} - \|\mathbf{c}_m\|^2 = 7.6100$. This difference is large enough to be associated with an angle of over 14 degrees between the directions of $\mathbf{c}_m$ and of the first eigenvector of **S**, a fact which is partially concealed by the non-linear cosine transformation which reads out as a more impressive value of 0.9685. This latter value is an approximate lower bound for the absolute value of the cosine of the angle formed by the first eigenvector of each matrix (inequality 3.24). Whereas all of **T**'s eigenvectors from rank 2 onwards are almost orthogonal to $\mathbf{c}_m$, the cosine involving $\mathbf{c}_m$ and the second eigenvector of **S** looks fairly large: $|\cos(\mathbf{c}_m,\mathbf{w}_2^{[c]})| = 0.2084$. It corresponds to a an angle of almost 78

Table 2: The cosines of the angles between the eigenvectors of the correlation matrix **S** and those of the non-central covariation matrix **T** for the scaled crayfish data.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
T1	0.97	0.21	0.05	0.00	0.03	0.01	-0.05	0.06	-0.04	0.06	-0.06	0.01	0.00
T2	0.22	-0.97	-0.04	0.00	-0.02	-0.01	0.03	-0.03	0.02	-0.04	0.03	0.00	0.00
T3	0.05	0.06	-0.99	0.00	-0.04	-0.01	0.04	-0.04	0.03	-0.04	0.03	0.00	0.00
T4	0.00	0.00	0.00	-1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
T5	-0.05	-0.05	-0.06	0.00	0.96	0.05	-0.16	0.14	-0.07	0.10	-0.07	0.01	0.00
T6	-0.07	-0.06	-0.07	0.00	-0.25	0.45	-0.69	0.39	-0.16	0.22	-0.14	0.01	0.00
T7	-0.02	-0.02	-0.02	0.00	-0.07	-0.89	-0.40	0.17	-0.06	0.08	-0.05	0.00	0.00
T8	-0.03	-0.03	-0.03	0.00	-0.06	-0.06	0.54	0.81	-0.13	0.15	-0.09	0.01	0.00
T9	0.03	0.03	0.02	0.00	0.04	0.03	-0.16	0.34	0.81	-0.42	0.16	-0.01	0.00
T10	-0.02	-0.02	-0.02	0.00	-0.03	-0.02	0.09	-0.15	0.53	0.81	-0.15	0.01	0.00
T11	0.02	0.02	0.01	0.00	0.02	0.01	-0.05	0.08	-0.10	0.26	0.95	-0.02	0.00
T12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.01	-0.01	0.03	1.00	0.00
T13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-1.00

degress, that is, an angle of over 12 degrees between $\mathbf{w}_2^{[c]}$ and the subspace $\mathcal{R}(\mathbf{c}_m)^{\perp}$. These perturbations at the top of the scale smooth out for subsequent eigenvectors. None of the vectors $\mathbf{w}_j^{[u]}$,with $j \geq 3$ forms an angle larger than 4 degrees with the subspace $\mathcal{R}(\mathbf{c}_m)^{\perp}$.

What has just been said might suggest that the eigenvectors of both matrix **T** and matrix **S** are similar. But Table 2 shows that the situation is not entirely straightforward.

While there is a fairly strong association between equal-rank eigenvectors of both matrices, at both ends of the scale, things are much fuzzier in the middle-ranking range. For example, the sixth eigenvector of **S** forms a smaller angle with **T**'s seventh eigenvector than with **T**'s eigenvector of equal rank, and the seventh eigenvector of **S** does not form small angles with any of **T**'s eigenvectors. Figure 1 illustrates how this is connected with the pattern of eigenvalues of both matrices. When the eigenvalues of **S** are almost equal to their same-rank counterparts in **T**, the dots in each box of Figure 1 are located near the upper right-hand corner of the box, as occurs with eigenvalues 3 to 5 and 11 to 13.

If eigenvalue i of **S** is almost equal to eigenvalue $i+1$ of **T**, the dots appear in the bottom left-hand corner, as occurs with eigenvalue $i = 6$ of **S**. When the dots are near the center of the box (e.g., $i = 7$), $\lambda_i(\mathbf{S})$ is about half-way between $\lambda_i(\mathbf{T})$ and $\lambda_{i+1}(\mathbf{T})$. This Figure

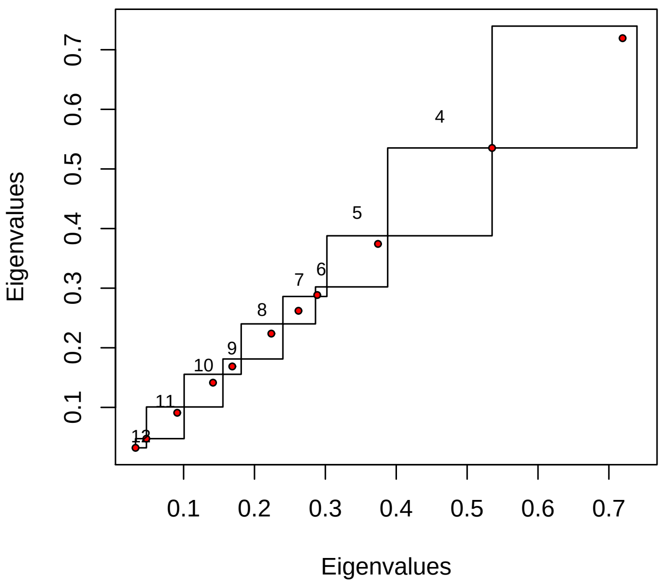

Figure 1: Each box has its upper right corner $(\lambda_j^{[u]}, \lambda_j^{[u]})$, where $\lambda_j^{[u]}$ is **T**'s j-th eigenvalue ($j \geq 3$), for the crayfish data. The lower left-hand corners are the points $(\lambda_{j+1}^{[u]}, \lambda_{j+1}^{[u]})$. These are the bounds given in Corollary 3.1 for the eigenvalues of the crayfish correlation matrix, **S**. The dots inside each box are the points $(\lambda_i^{[c]}, \lambda_i^{[c]})$, where $\lambda_i^{[c]}$ denotes the i-th eigenvalue of **S**. For reasons of scale, the two largest eigenvalues of **T** and **S** have been left out.

should be compared with the cosines of the angles between the corresponding eigenvectors, in Table 2, which follow similar patterns.

This situation, where the eigenvalues of **S** and **T** offer an indication on the relations between their associated eigenvectors, has been observed in other data sets and seems to be fairly general. But the fact that the eigenvalues of both matrices are interlaced imposes a structural constraint on how far these differences can go and implies that a matrix of cosines between eigenvectors, such as Table 2, essentially consists of a diagonal band of larger magnitude values, with values nearer to zero as the row- and column-numbers of the matrix diverge.

The variances of the uncentred PCs (column 2 of Table 1) are very close to **T**'s eigenvalues from rank 2 onwards, since the eigenvectors (vectors of PC loadings) which define them are almost orthogonal to $\mathbf{c}_m$ (see equation 3.6). This is not surprising. Unlike the eigenvalues of **S**, the eigenvalues of **T** indicate non-central second moments, rather than variance, of the uncentred PCs. But the concepts of central and non-central second moment coincide for zero-mean data, which is approximately the case in the latter uncentred PCs (see equation 3.5). Equation (3.9) ensures that, in this case, the correlations between the uncentred and column-centred PCs defined by the eigenvectors of **S** and **T** are essentially the same as the cosines given in Table 2 (slightly smaller), a situation which need not occur in general. For similar reasons, the correlation between the first PC of each kind is slightly larger (0.99) than the cosine between their vectors of loadings (0.97). In this example many uncentred PCs account for approximately the same non-central variation as their nearly-equivalent column-centred PCs account for variance, but the situation has changed dramatically in terms of the proportion of total variability which is accounted for. Whereas the second column-centred PC accounts for over 16.2% of total variance, the uncentred counterpart which accounts for almost the same variability represents less than $1/1000$ of total (non-central) variability. These differences in proportion of total variability often conceal the noticeable similarities that exist between the eigendecompositions of **T** and of **S**.

Given the very small trace of matrix **S**, when compared with $\|\mathbf{c}_m\|^2$, the bounds (3.22) are fairly tight, ensuring that $\lambda_1^{[u]} \in [2844.334, 2852.310]$.

4.2 An mRNA data set

The second example deals with a data set from the field of DNA microarrays that has been discussed by Alter and Golub [1].The data matrix consists of measurements of relative abundance levels of gene transcripts in yeast. It has $n = 6776$ rows (observations) corresponding to different genes and $p = 30$ columns (variables) corresponding to different gel slices. Alter and Golub [1] find the singular value decomposition of this data matrix. They note that one of the matrices in the decomposition contains the eigenvectors of what they call the 'correlation matrix' for the data, but which is, in fact, the 30×30 matrix of non-central second moments of the unstandardized data. They call these eigenvectors, which arise from an uncentred PCA of the data, 'eigengenes'. The results in [1] are now compared with those from a standard column-centred PCA.

In this example, the trace of the 30×30 matrix of non-central second moments is

$tr(\mathbf{T}) = 111.0344$, while the centre of gravity of the n-point cloud in $\mathbb{R}^{30}$ is at a squared distance $\|\mathbf{c}_m\|^2 = 36.6781$ from the origin. Thus, the total variability of the points around their centre of gravity is about twice as big as the squared norm of $\mathbf{c}_m$: $tr(\mathbf{S}) = 111.0344 - 36.6781 = 74.3563$. However, **T**'s first eigenvector still points in a direction that is not too different from that of $\mathbf{c}_m$, since all the extra variability in **T** (*vis a vis* that of **S**) is due to that direction (equations 3.3 and 3.4), and $\mathbf{c}_m$ accounts for a proportion of at least $\|\mathbf{c}_m\|^2/tr(\mathbf{T}) \approx 1/3$ of total non-central variability. Table 3 confirms that **T**'s first eigenvector is indeed close to the direction of vector $\mathbf{c}_m$ (it forms an angle of just under 7 degrees). The proportion of total variability associated with it is 0.3827.

As noted above, Alter and Golub (2006) are concerned with the eigenvectors of matrix **T**, which they call "eigengenes". Figure 2 depicts the first eigengene of the data and the unit-norm vector of column means, confirming a good deal of agreement between both vectors.

From inequality (3.20), the first eigenvalue of **S** is bounded below by the variance of the first uncentred PC: 6.3393. But inequality (3.16) ensures that $\lambda_1^{[c]}$ is bounded below by the variance of any uncentred PC, and the variance of the second uncentred PC provides a much better lower bound: 12.7800. Likewise, the right-hand side of inequality (3.19) ensures that $\lambda_1^{[c]} + \lambda_2^{[c]}$ is bounded below by $\text{var}(\boldsymbol{\eta}_1^{[u]}) + \text{var}(\boldsymbol{\eta}_2^{[u]}) = 19.1197$. But the better bounds (3.15) guarantee that the sum of the two largest eigenvalues of **S** is bounded below by the sum of variances of any two uncentred PCs, and so by the largest of them all: $\text{var}(\boldsymbol{\eta}_2^{[u]}) + \text{var}(\boldsymbol{\eta}_3^{[u]}) = 22.0122$. Note how the variance of the second uncentred PC is larger than that of the first uncentred PC. By definition, this could not happen with column-centred PCs.

The eigenvalues of both matrices are broadly speaking similar, as in the previous example, and this is related to the fact that both sets of eigenvalues are interlaced (Corollary 3.1). In particular, for ranks 12, 17 and 23 through 30, there are very small differences in the eigenvalues which, from the discussion in Section 3.1, suggests that there may be similar eigenvectors. This is usually the case with eigenvectors associated with near-zero eigenvalues, since they tend to be very close to $\mathcal{R}(\mathbf{c}_m)^{\perp}$ – see bounds (3.21). Eigenvalues 1, 3, 4, 5, 7 and 8 of **S** are closer to the eigenvalues of **T** of rank larger by one, than to their equal-rank counterparts, suggesting closer agreement of those pairs of eigenvectors.

The cosine of the angle between any given eigenvectors of **T** and of **S** can be computed from the information provided in Table 3, using (3.13). The most striking difference in relation to the previous example is that **T**'s first eigenvector is not similar to any eigenvector of **S**: the cosine of the angle it forms with any eigenvector of **S** is never larger, in absolute

Table 3: Results for the mRNA data set (table structure as in Table 1).

	Uncentred		Eigenvalues		Column-centred	
Rank	$\text{var}(\boldsymbol{\eta}_j^{[u]})$	$\lvert\cos(\mathbf{w}_j^{[u]}, \mathbf{c}_m)\rvert$	**T**	**S**	$\lvert\cos(\mathbf{w}_i^{[c]}, \mathbf{c}_m)\rvert$	ξ_i
1	6.3393	0.9928	42.4898	13.1083	0.2440	15.2914
2	12.7800	0.0302	12.8138	11.5702	0.4279	18.2859
3	9.2317	0.0955	9.5665	8.2055	0.1175	8.7122
4	8.0117	0.0333	8.0526	6.8557	0.1988	8.3052
5	6.5449	0.0338	6.5869	4.9294	0.2087	6.5267
6	4.8863	0.0050	4.8873	4.7148	0.3972	10.5002
7	3.9925	0.0318	4.0298	3.5107	0.0975	3.8596
8	3.4962	0.0037	3.4967	3.3836	0.2467	5.6162
9	3.1892	0.0159	3.1985	2.7665	0.3766	7.9687
10	2.5958	0.0069	2.5976	2.5091	0.2839	5.4647
11	2.0286	0.0140	2.0358	1.9040	0.2248	3.7576
12	1.7564	0.0020	1.7565	1.7538	0.0374	1.8050
13	1.5260	0.0148	1.5340	1.4197	0.1604	2.3632
14	1.3506	0.0066	1.3522	1.2921	0.1970	2.7154
15	1.0432	0.0087	1.0460	1.0037	0.1184	1.5182
16	0.9333	0.0056	0.9345	0.9091	0.1111	1.3621
17	0.8169	0.0028	0.8172	0.8136	0.0347	0.8578
18	0.7386	0.0100	0.7422	0.6804	0.1077	1.1057
19	0.6690	0.0021	0.6692	0.6520	0.1540	1.5216
20	0.5326	0.0042	0.5332	0.5195	0.0840	0.7784
21	0.4598	0.0045	0.4605	0.4469	0.0689	0.6212
22	0.4346	0.0025	0.4348	0.4227	0.1099	0.8656
23	0.3188	0.0005	0.3188	0.3186	0.0122	0.3241
24	0.2226	0.0016	0.2227	0.2211	0.0264	0.2467
25	0.1669	0.0040	0.1675	0.1567	0.0729	0.3516
26	0.1469	0.0002	0.1469	0.1468	0.0087	0.1496
27	0.0638	0.0005	0.0638	0.0636	0.0085	0.0663
28	0.0562	0.0008	0.0563	0.0558	0.0152	0.0643
29	0.0170	0.0002	0.0170	0.0170	0.0027	0.0172
30	0.0055	0.0008	0.0056	0.0051	0.0136	0.0120

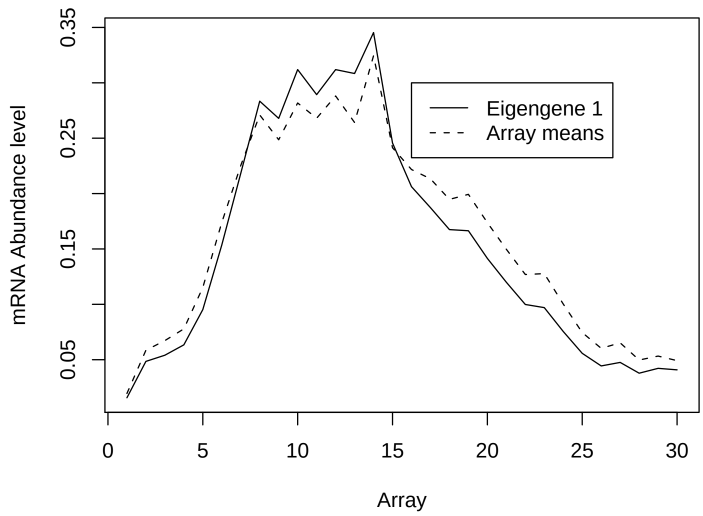

Figure 2: The first eigenvector (eigengene) of matrix **T** (solid line), and the unit-norm vector of column (array) means of the uncentred data matrix (dashed line), plotted against the ranks of their coordinates. As in Alter and Golub (2006), the coefficients of the eigenvector have been given positive signs and the points in both vectors are joined by line segments, to highlight their "form".

value, than 0.50. This is in agreement with the information in the sixth column of Table 3. The cosines (in absolute value) of the angles formed by **T**'s eigenvectors of rank 2 to 9 with the eigenvectors of **S** of ranks 1 to 8, are, respectively: 0.92, 0.75, 0.94, 0.92, 0.91, 0.68, 0.95, 0.77. The comparative graphs of the first six of these pairs are given in Figure 3. Even for the pairs with largest difference there is substantial visual similarity. The patterns described in Alter and Golub (2006) have been essentially preserved.

The correlations between the PCs of both kinds which these eigenvectors define are larger: 0.93, 0.84, 0.95, 0.94, 0.91, 0.74, 0.95, 0.80. This reflects the fact that the standard deviations are larger in these column-centred components (see equation 3.9 and Table 3). From rank 10 onwards, equal-rank eigenvectors in both matrices form angles with cosines: 0.81, 0.87, 0.99, 0.76, 0.79, 0.89, 0.91, 0.98, 0.64, 0.68, 0.95, 0.83, 0.84, 1.00, 1.00, 0.99,

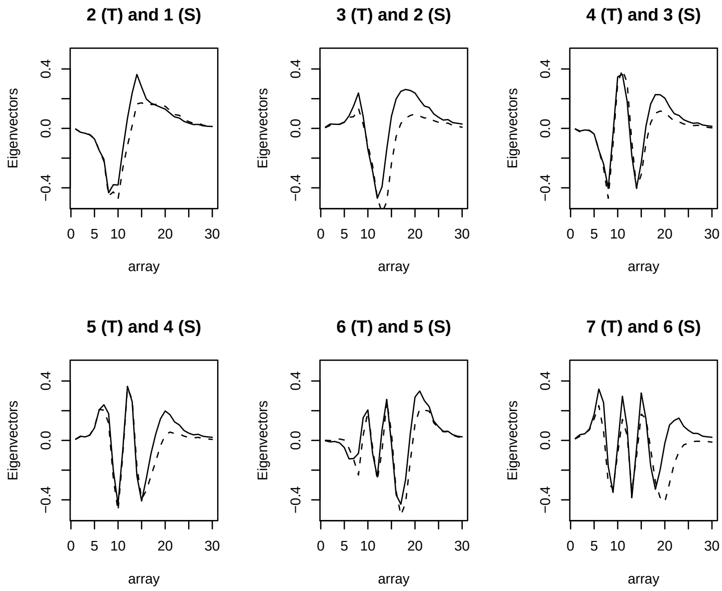

Figure 3: Eigenvectors (eigengenes) 2 to 7 of matrix **T** (solid line) and eigenvectors 1 to 6 of matrix **S** (dashed lines), are plotted against the ranks of their coordinates. Eigenvectors may have been multiplied by -1 to highlight their similarity.

1.00, 1.00, 1.00, 1.00, 1.00. The smaller values for pairs 13/14 and 18/19 are, once again, associated with situations in which the difference in a pair of eigenvalues from both matrices is smaller for different ranking eigenvalues than for their equal-rank counterparts. The ninth eigenvector of **S** never forms an angle with cosine in excess of 0.56 with any eigenvector of **T**, in much the same way as its eigenvalue does not pair up with any eigenvalue of **T**.

5 Discussion

Carrying out a singular value decomposition on an uncentred data matrix can be viewed as a variant of principal component analysis that produces "uncentred principal components" - linear combinations of the uncentred variables, whose loadings are given by the eigenvectors of the matrix of the data's non-central second moments.

From a strictly mathematical point of view, there is nothing wrong with carrying out an uncentred PCA, that is, a singular value decomposition of a given matrix. Whether or not it is statistically appropriate to do so in any given context, and what conclusions can be drawn from such an analysis, are a different matter. In an uncentred PCA it is variability about the origin, rather than about the centre of gravity of the n-point scatter in $\mathbb{R}^p$, that will be of concern. Unless the origin is an important point of reference for the p variables, an uncentred PCA may be an artificial procedure that merely highlights the size of the variable means relative to the origin [11]. On the other hand, centring columns may destroy meaningful patterns in the rows of the data matrix, which would therefore also be lost in the eigenvectors of a covariance matrix. This is a relevant concern in the second of the examples considered in Section 4.

In both variants of PCA, the underlying variables are pre-processed in different ways, the criterion for choosing a linear combination of the variables is different (maximizing central or non-central second moments) and the additional requirements are also different (uncorrelatedness, or zero crossed non-central second moments). Despite all this, the relationships between a standard, column-centred, PCA and its uncentred counterpart are strong, and have so far not been the object of a quantitative discussion.

Unsurprisingly, the vector $\mathbf{c}_m$ of column means of $\mathbf{X}_{uc}$, which is a measure of the overall size of the data, plays an important role in the comparative study of both types of PCA. The proportion of $\mathbf{T}$'s total variation that is associated with the direction of $\mathbf{c}_m$ always exceeds $\|\mathbf{c}_m\|^2/tr(\mathbf{T})$. Three types of situation can arise:

- If the norm squared of $\mathbf{c}_m$ is very small, when compared with the trace of matrix $\mathbf{T}$, then the eigendecompositions of $\mathbf{T}$ and of the data's covariance matrix $\mathbf{S}$ are similar and the two PCAs will also be similar.

- If the norm squared of vector $\mathbf{c}_m$ is very large when compared with $tr(\mathbf{T})$, as in the first of the examples considered in Section 4, then the characterization of the eigenstructure of $\mathbf{T}$ and $\mathbf{S}$ made in point 3) of Theorem 2 remains, broadly speaking, valid. However, some perturbations can arise even when $\mathbf{T}$'s first eigenvector is extremely close to the vector of variable means. Even if both matrices essentially share all their eigenvectors, the rank of the eigenvector close to $\mathbf{c}_m$ can remain the same (as in the first example), but can also drop.

- In between these two extremes, the situation becomes less clear and there is more scope for differences between both eigendecompositions and PCAs. However, as the second example illustrates, it is often still the case that $\mathbf{T}$'s first eigenvector is not too divergent from $\mathbf{c}_m$ and that many eigenvalues and eigenvectors in both matrices are similar.

In general, the inspection of both matrices' eigenvalues contains much valuable information regarding both PCAs. Whenever both matrices have very similar eigenvalues, the corresponding eigenvectors and PCs are also similar. It will often be the case that the eigenvalues of equal rank in both **S** and **T** are the most similar ones, also as a result of the fact that both sets of eigenvalues are interlaced (Corollary 3.1). But when an eigenvalue of **S**, say, is half-way between two eigenvalues of **T**, or even closer to the next-ranking eigenvalue of **T** than to its equal-rank counterpart, that situation tends be reflected in the corresponding eigenvectors. The ocurrence of a situation of this kind is associated with a perturbation which can affect one or more subsequent eigenvectors.

An important issue remains: when and why do such perturbations arise? The correlations between the uncentred principal components which are given by combining expressions (3.6) and (3.7) provide a clue to what lies behind these perturbations. Although uncentred PCs with loadings that are approximately orthogonal to $\mathbf{c}_m$ are nearly uncorrelated among themselves, the uncentred PC defined by a vector that is close to $\mathbf{c}_m$ may have a non-negligible correlation with the remaining uncentred PCs. The implication is that the eigenvectors of matrix **S** cannot remain exactly the same since, in that case, the PCs which they define would not be uncorrelated (recall that, for equal vectors of loadings, the correlation remains the same, whether we consider linear combinations of uncentred, or of column-centred, variables).

The many relations stated in Section 3 can provide a better understanding of the fairly substantial links between the standard, column-centred, and the uncentred principal components, as well as ensuring a better understanding of the nature and implications of the latter variant of principal component analysis. Despite the differences between the analyses, our experience with a number of examples shows that the constraints imposed by the results of Section 3 often lead to a remarkable degree of similarity between many of the eigenvectors and components obtained from the two techniques, although their relative importance may sometimes differ dramatically.

Many of the results reported in this paper can be adapted with more or less effort to related situations. For instance, the discussion in this paper has been entirely for the conventional case where $n > p$. In a non-trivial number of applications, the number of variables p exceeds the number of observations n. A classic example is in atmospheric science where a meteorological variable is measured at p geographical locations and n time points. If the spatial domain is large and the time series short, it can happen that $n \ll p$ [20, Section

11.6]. With larger and larger data sets being collected in areas ranging from microarrays to the spectral energy of blazars [4], $n < p$ is increasingly common. Another related situation concerns the use of hybrid varieties of PCA, in which constants other than the column means are subtracted from each column ([12], [13], [14]). Yet another possible variant revolves around the use SVDs of row-centred data matrices, or matrices that are both row- and column-centred ([2], [3]). Row-centring may make sense when the rows of the matrix can be viewed as variables, or when they are individuals for which variables (columns) on a common scale have been measured and it is desired to filter out some measure of overall size of individuals prior to the analysis. It is most frequent together with column-centring. Details of the relevant relations in such situations will be reported elsewhere.

A Proofs

A.1 Proof of Corollary 3.1

The first pair of inequalities results from a direct application of the Yanai & Takeuchi Separation Theorem, with $\mathbf{Y} = \frac{1}{\sqrt{n}}\mathbf{X}_{uc}$, $\mathbf{P}_n = \mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n}$ and $\mathbf{P}_p = \mathbf{I}_p$. Thus, $r = n-1$ and $s = p$, and so $t = 1$. The second pair of inequalities is a direct result of dividing through by the trace of $\mathbf{S}$. The second expression for α results from the fact that, by definition,

$$\mathbf{S} = \tfrac{1}{n}\mathbf{X}_{uc}^t(\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})\mathbf{X}_{uc} = \mathbf{T} - \mathbf{c}_m \cdot \mathbf{c}_m^t \ , \tag{A.1}$$

and so $tr(\mathbf{S}) = tr(\mathbf{T}) - tr(\mathbf{c}_m \cdot \mathbf{c}_m^t) \ = \ tr(\mathbf{T}) - \|\mathbf{c}_m\|^2$ by the circularity property of the trace. ∇

A.2 Proof of Theorem 2

1. The results follow directly from relation (A.1).

2. The difference in the Rayleigh-Ritz ratios of $\mathbf{T}$ and $\mathbf{S}$ is a direct consequence of equation (3.2).

3. Assume that $\{\mathbf{w}_j^{[u]}\}_{j=1}^p$ are the eigenvectors of $\mathbf{T}$, with eigenvalues $\{\lambda_j^{[u]}\}_{j=1}^p$, and that one of $\mathbf{T}$'s unit-norm eigenvectors, say the first, is $\mathbf{w}_1^{[u]} = \mathbf{c}_m/\|\mathbf{c}_m\|$. From equation

(3.2) and the spectral decomposition of $\mathbf{T}$ we have

$$\begin{aligned}\mathbf{S} \;=\; \mathbf{T}-\mathbf{c}_m\cdot\mathbf{c}_m^t &= \sum_{j=1}^{p}\lambda_j^{[u]}\,\mathbf{w}_j^{[u]}\mathbf{w}_j^{[u]^t} - \|\mathbf{c}_m\|^2\,\mathbf{w}_1^{[u]}\mathbf{w}_1^{[u]^t} \\ &= \left(\lambda_1^{[u]} - \|\mathbf{c}_m\|^2\right)\mathbf{w}_1^{[u]}\mathbf{w}_1^{[u]^t} + \sum_{j\neq 1}\lambda_j^{[u]}\,\mathbf{w}_j^{[u]}\mathbf{w}_j^{[u]^t}\;,\end{aligned}$$

which is a spectral decomposition of $\mathbf{S}$.

∇

A.3 Proof of Proposition 3.1

1. By definition, the mean score is $\frac{1}{n}\boldsymbol{\eta}_j^{[u]^t}\mathbf{1}_n = \frac{1}{n}\mathbf{w}_j^{[u]^t}\mathbf{X}_{uc}^t\mathbf{1}_n = \mathbf{w}_j^{[u]^t}\mathbf{c}_m = \|\mathbf{c}_m\|\cdot\cos(\mathbf{c}_m,\mathbf{w}_j^{[u]})$.

2. The covariance of any pair of linear combinations $\mathbf{X}_{uc}\mathbf{a}$ and $\mathbf{X}_{uc}\mathbf{b}$ of the columns of $\mathbf{X}_{uc}$ is given by $\mathbf{a}^t\mathbf{S}\mathbf{b}$. The i-th uncentred PC results from taking $\mathbf{a}=\mathbf{w}_i^{[u]}$. Now, from equation (A.1), the covariance between the i-th and j-th uncentred PCs is $\mathbf{w}_i^{[u]^t}\mathbf{S}\mathbf{w}_j^{[u]} = \mathbf{w}_i^{[u]^t}(\mathbf{T}-\mathbf{c}_m\mathbf{c}_m^t)\mathbf{w}_j^{[u]} = \lambda_j^{[u]}\mathbf{w}_i^{[u]^t}\mathbf{w}_j^{[u]} - (\mathbf{w}_i^{[u]^t}\mathbf{c}_m)(\mathbf{w}_j^{[u]^t}\mathbf{c}_m)$. Considering, first $i=j$ and then $i\neq j$, gives expressions (3.6) and (3.7). In matrix terms, the covariance matrix for the uncentred PCs is just $\mathbf{W}^{[u]^t}\mathbf{S}\mathbf{W}^{[u]}$. Let the spectral decomposition of matrix $\mathbf{S}$ be $\mathbf{S}=\mathbf{W}^{[c]}\Lambda^{[c]}\mathbf{W}^{[c]^t}$, where $\Lambda^{[c]}$ is the diagonal matrix of eigenvalues $\lambda_i^{[c]}$ of $\mathbf{S}$ and $\mathbf{W}^{[c]}$ has the eigenvectors $\mathbf{w}_i^{[c]}$ in its columns. Then the covariance matrix of the uncentred PCs is

   $$\mathbf{W}^{[u]^t}\mathbf{S}\mathbf{W}^{[u]} = (\mathbf{W}^{[u]^t}\mathbf{W}^{[c]})\,\Lambda^{[c]}\,(\mathbf{W}^{[u]^t}\mathbf{W}^{[c]})^t\;. \qquad \text{(A.2)}$$

   Since both $\mathbf{W}^{[u]}$ and $\mathbf{W}^{[c]}$ are orthogonal matrices and the product of orthogonal matrices is an orthogonal matrix, equation (A.2) gives the spectral decomposition of the covariance matrix of the uncentred PCs. The eigenvalues of this covariance matrix are therefore the same as those of $\mathbf{S}$, and its eigenvectors are the columns of the matrix $\mathbf{W}^{[u]^t}\mathbf{W}^{[c]}$, which has as its (i,j)-th element the cosine of the angle between the i-th eigenvector of $\mathbf{T}$ and the j-th eigenvector of $\mathbf{S}$.

3. If $\mathbf{y}_1$ and $\mathbf{y}_2$ are any two linear combinations of the columns of either the uncentred, or the column-centred data matrices, $\mathbf{y}_1=\mathbf{X}_{\star c}\mathbf{a}$, $\mathbf{y}_2=\mathbf{X}_{\star c}\mathbf{b}$, (where $\mathbf{X}_{\star c}$ denotes either $\mathbf{X}_{uc}$ or $\mathbf{X}_{cc}$), the covariance between them is again given by:

   $$\operatorname{cov}(\mathbf{X}_{\star c}\mathbf{a},\mathbf{X}_{\star c}\mathbf{b}) = \mathbf{a}^t\mathbf{S}\,\mathbf{b}\;. \qquad \text{(A.3)}$$

The i-th column-centred PC results from taking $\mathbf{a} = \mathbf{w}_i^{[c]}$ with $\mathbf{X}_{\star c} = \mathbf{X}_{cc}$, and the j-th uncentred PC from taking $\mathbf{b} = \mathbf{w}_j^{[u]}$ with $\mathbf{X}_{\star c} = \mathbf{X}_{uc}$. Therefore, $\mathrm{cov}(\boldsymbol{\eta}_i^{[c]}, \boldsymbol{\eta}_j^{[u]}) = \mathbf{w}_i^{[c]^t} \mathbf{S} \mathbf{w}_j^{[u]} = \lambda_i^{[c]} \mathbf{w}_i^{[c]^t} \mathbf{w}_j^{[u]}$, which is the result we wished to prove.

4. A direct result from expressions (3.8) and (3.6), as well as from the fact that the variance of the i-th standard PC is $\lambda_i^{[c]}$. Note that, in general, if $\mathbf{y}_1$ and $\mathbf{y}_2$ are any two linear combinations of the columns of either the uncentred, or the column-centred data matrices, $\mathbf{y}_1 = \mathbf{X}_{\star c}\mathbf{a}$, $\mathbf{y}_2 = \mathbf{X}_{\star c}\mathbf{b}$, (where $\mathbf{X}_{\star c}$ denotes either $\mathbf{X}_{uc}$ or $\mathbf{X}_{cc}$), the correlation between them is:

$$r_{(\mathbf{X}_{\star c}\mathbf{a}, \mathbf{X}_{\star c}\mathbf{b})} = \frac{\mathbf{a}^t \mathbf{S} \mathbf{b}}{\sqrt{\mathbf{a}^t \mathbf{S} \mathbf{a} \cdot \mathbf{b}^t \mathbf{S} \mathbf{b}}} \,. \tag{A.4}$$

5. If $\mathbf{y}_1$ is any linear combination of the columns of the uncentred data matrix, $\mathbf{y}_1 = \mathbf{X}_{uc}\mathbf{a}$, and $\mathbf{y}_2$ is any linear combination of the columns of the column-centred data matrix, $\mathbf{y}_2 = \mathbf{X}_{cc}\mathbf{b}$, then the cosine of the angle between them is:

$$\cos(\mathbf{X}_{uc}\mathbf{a}, \mathbf{X}_{cc}\mathbf{b}) = \frac{<\mathbf{X}_{uc}\mathbf{a}, \mathbf{X}_{cc}\mathbf{b}>}{\|\mathbf{X}_{uc}\mathbf{a}\| \cdot \|\mathbf{X}_{cc}\mathbf{b}\|} = \frac{\mathbf{a}^t \mathbf{S} \mathbf{b}}{\sqrt{\mathbf{a}^t \mathbf{T} \mathbf{a} \cdot \mathbf{b}^t \mathbf{S} \mathbf{b}}} \,. \tag{A.5}$$

The j-th uncentred PC results from taking $\mathbf{a} = \mathbf{w}_j^{[u]}$, and the i-th column-centred PC from taking $\mathbf{b} = \mathbf{w}_i^{[c]}$. Therefore,

$$\begin{aligned} \cos(\boldsymbol{\eta}_j^{[u]}, \boldsymbol{\eta}_i^{[c]}) &= \frac{\mathbf{w}_j^{[u]^t} \mathbf{S} \mathbf{w}_i^{[c]}}{\sqrt{\mathbf{w}_j^{[u]^t} \mathbf{T} \mathbf{w}_j^{[u]} \cdot \mathbf{w}_i^{[c]^t} \mathbf{S} \mathbf{w}_i^{[c]}}} = \frac{\lambda_i^{[c]} \mathbf{w}_j^{[u]^t} \mathbf{w}_i^{[c]}}{\sqrt{\lambda_j^{[u]} \cdot \lambda_i^{[c]}}} \\ &= \sqrt{\frac{\lambda_i^{[c]}}{\lambda_j^{[u]}}} \cdot \cos(\mathbf{w}_j^{[u]}, \mathbf{w}_i^{[c]}) \,. \end{aligned}$$

6. The ℓ_2 norm of the difference between $\boldsymbol{\eta}_i^{[c]} = \mathbf{X}_{cc}\mathbf{w}_i^{[c]}$ and $\pm \boldsymbol{\eta}_j^{[u]} = \pm \mathbf{X}_{uc}\mathbf{w}_j^{[u]}$ is

$$\begin{aligned} \|\boldsymbol{\eta}_i^{[c]} \mp \boldsymbol{\eta}_j^{[u]}\| &= \sqrt{\mathbf{w}_i^{[c]^t} \mathbf{X}_{cc}^t \mathbf{X}_{cc} \mathbf{w}_i^{[c]} \mp 2\, \mathbf{w}_i^{[c]^t} \mathbf{X}_{cc}^t \mathbf{X}_{uc} \mathbf{w}_j^{[u]} + \mathbf{w}_j^{[u]^t} \mathbf{X}_{uc}^t \mathbf{X}_{uc} \mathbf{w}_j^{[u]}} \\ &= \sqrt{n\, \mathbf{w}_i^{[c]^t} \mathbf{S} \mathbf{w}_i^{[c]} \mp 2n\, \mathbf{w}_i^{[c]^t} \mathbf{S} \mathbf{w}_j^{[u]} + n\, \mathbf{w}_j^{[u]^t} \mathbf{T} \mathbf{w}_j^{[u]}} \\ &= \sqrt{n \lambda_i^{[c]} \mp 2n \lambda_i^{[c]} \cos(\mathbf{w}_i^{[c]}, \mathbf{w}_j^{[u]}) + n \lambda_j^{[u]}} \\ &= \sqrt{n \lambda_i^{[c]} \left[2(1 \mp \cos(\mathbf{w}_i^{[c]}, \mathbf{w}_j^{[u]}))\right] + n\, (\lambda_j^{[u]} - \lambda_i^{[c]})} \\ &= \sqrt{n \lambda_i^{[c]} \, \|\mathbf{w}_i^{[c]} \mp \mathbf{w}_j^{[u]}\|^2 + n\, (\lambda_j^{[u]} - \lambda_i^{[c]})} \,, \end{aligned}$$

as we wished to prove.

7. As with any unit-norm vector, the distance between the re-sized principal components $\mathbf{u}_i^{[c]} = \boldsymbol{\eta}_i^{[c]}/\|\boldsymbol{\eta}_i^{[c]}\|$ and $\pm\mathbf{u}_j^{[u]} = \boldsymbol{\eta}_j^{[u]}/\|\boldsymbol{\eta}_j^{[u]}\|$ is given by $\|\mathbf{u}_i^{[c]} \mp \mathbf{u}_j^{[u]}\| = \sqrt{2\left[1 \mp \cos(\mathbf{u}_i^{[c]}, \mathbf{u}_j^{[u]})\right]}$. But the angle between the unit-norm components is the same as that between their original counterparts, and using expression (3.10) we obtain the desired result.

8. From the definition of eigenvalues/vectors,

$$\mathbf{T}\mathbf{w}_j^{[u]} = \lambda_j^{[u]}\mathbf{w}_j^{[u]} \iff \mathbf{w}_j^{[u]} = \lambda_j^{[u]}\mathbf{T}^{-1}\mathbf{w}_j^{[u]} \tag{A.6}$$

$$\mathbf{S}\mathbf{w}_i^{[c]} = \lambda_i^{[c]}\mathbf{w}_i^{[c]} \iff \mathbf{w}_i^{[c]} = \frac{1}{\lambda_i^{[c]}}\mathbf{S}\mathbf{w}_i^{[c]} \tag{A.7}$$

Therefore, the cosine of the angle between $\mathbf{w}_j^{[u]}$ and $\mathbf{w}_i^{[c]}$ can be written as

$$\cos\left(\mathbf{w}_j^{[u]}, \mathbf{w}_i^{[c]}\right) = \mathbf{w}_j^{[u]t}\mathbf{w}_i^{[c]} = \frac{\lambda_j^{[u]}}{\lambda_i^{[c]}} \cdot \mathbf{w}_j^{[u]t}\mathbf{T}^{-1}\mathbf{S}\mathbf{w}_i^{[c]} . \tag{A.8}$$

Now, for any vector $\mathbf{x}$ that is orthogonal to $\mathbf{c}_m$, from equation (3.2) we have $\mathbf{T}^{-1}\mathbf{S}\mathbf{x} = \mathbf{T}^{-1}(\mathbf{T} - \mathbf{c}_m\mathbf{c}_m^t)\mathbf{x} = \mathbf{I}\mathbf{x} = \mathbf{x}$ (in other words, $\mathbf{x} \in \mathcal{R}(\mathbf{c}_m)^\perp$ implies that $\mathbf{x}$ is an eigenvector of $\mathbf{T}^{-1}\mathbf{S}$, with eigenvalue 1). Thus, if $\mathbf{w}_i^{[c]}$ is orthogonal to $\mathbf{c}_m$, then equation (A.8) becomes:

$$\cos\left(\mathbf{w}_j^{[u]}, \mathbf{w}_i^{[c]}\right) = \frac{\lambda_j^{[u]}}{\lambda_i^{[c]}} \cdot \cos(\mathbf{w}_j^{[u]}, \mathbf{w}_i^{[c]}) ,$$

and we must therefore have, either $\lambda_j^{[u]} = \lambda_i^{[c]}$, or $\mathbf{w}_j^{[u]} \perp \mathbf{w}_i^{[c]}$, or both. In any case, the result is proved. In the more general case, when $\mathbf{w}_i^{[c]}$ is not orthogonal to $\mathbf{c}_m$, we take the decomposition $\mathbf{w}_i^{[c]} = \mathbf{P}_{\mathbf{c}_m}\mathbf{w}_i^{[c]} + (\mathbf{I}_p - \mathbf{P}_{\mathbf{c}_m})\mathbf{w}_i^{[c]}$, where $\mathbf{P}_{\mathbf{c}_m} = \mathbf{c}_m(\mathbf{c}_m^t\mathbf{c}_m)^{-1}\mathbf{c}_m^t$ is the matrix of orthogonal projections onto the subspace spanned by the vector $\mathbf{c}_m$. The term $(\mathbf{I}_p - \mathbf{P}_{\mathbf{c}_m})\mathbf{w}_i^{[c]}$ is now a vector in the subspace $\mathcal{R}(\mathbf{c}_m)^\perp$, and therefore an eigenvector of $\mathbf{T}^{-1}\mathbf{S}$ with eigenvalue 1. Thus, (A.8) becomes

$$\begin{aligned}
\cos\left(\mathbf{w}_j^{[u]}, \mathbf{w}_i^{[c]}\right) &= \frac{\lambda_j^{[u]}}{\lambda_i^{[c]}}\mathbf{w}_j^{[u]t}\mathbf{T}^{-1}\mathbf{S}\mathbf{P}_{\mathbf{c}_m}\mathbf{w}_i^{[c]} + \frac{\lambda_j^{[u]}}{\lambda_i^{[c]}}\mathbf{w}_j^{[u]t}(\mathbf{I}_n - \mathbf{P}_{\mathbf{c}_m})\mathbf{w}_i^{[c]} \\
\Leftrightarrow \left(1 - \frac{\lambda_j^{[u]}}{\lambda_i^{[c]}}\right)\cos\left(\mathbf{w}_j^{[u]}, \mathbf{w}_i^{[c]}\right) &= \frac{\lambda_j^{[u]}}{\lambda_i^{[c]}}\mathbf{w}_j^{[u]t}\left(\mathbf{T}^{-1}\mathbf{S} - \mathbf{I}_p\right)\mathbf{P}_{\mathbf{c}_m}\mathbf{w}_i^{[c]} . \qquad \text{(A.9)}
\end{aligned}$$

Now, from equation (3.2), we have $\mathbf{T}^{-1}\mathbf{S} - \mathbf{I}_p = -\mathbf{T}^{-1}\mathbf{c}_m\mathbf{c}_m^t$. In addition, $\mathbf{P}_{\mathbf{c}_m}\mathbf{w}_i^{[c]} = \frac{\mathbf{c}_m^t\mathbf{w}_i^{[c]}}{\mathbf{c}_m^t\mathbf{c}_m}\mathbf{c}_m$. Substituting, and keeping in mind the relation (A.6), we have

$$\left(1 - \frac{\lambda_j^{[u]}}{\lambda_i^{[c]}}\right)\cos\left(\mathbf{w}_j^{[u]}, \mathbf{w}_i^{[c]}\right) = -\frac{\lambda_j^{[u]}}{\lambda_i^{[c]}}\mathbf{c}_m^t\mathbf{w}_i^{[c]} \cdot \frac{1}{\lambda_j^{[u]}}\mathbf{w}_j^{[u]t}\mathbf{c}_m ,$$

which is the result we wished to prove.

9. Following the definition in [7, p.192], given two vectors $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^p$, we say that vector $\boldsymbol{\beta}$ *majorizes* vector $\boldsymbol{\alpha}$ if the sum of $\boldsymbol{\beta}$'s k smallest coordinates is always greater than, or equal to, the sum of $\boldsymbol{\alpha}$'s k smallest coordinates, for all $k = 1, ..., p$ and with equality when $k = p$. Assuming that the coordinates in $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are given in decreasing order, this means that $\sum_{j=p-k+1}^{p} \beta_j \geq \sum_{i=p-k+1}^{p} \alpha_i$, for all $k = 1, ..., p$, with equality when $k = p$. For any symmetric matrix $\mathbf{A}$, its vector of diagonal elements majorizes its vector of eigenvalues [7, p.193]. Since the covariance matrix of the uncentred PCs, $\mathbf{W}^{[u]^t}\mathbf{S}\mathbf{W}^{[u]}$, has the same eigenvalues as the data's covariance matrix $\mathbf{S}$ (see point 2 of this Theorem), the vector of variances of the uncentred PCs majorizes the vector of eigenvalues of $\mathbf{S}$. This is the right-hand side statement in (3.15). The lower bound on the sum of the largest k eigenvalues of $\mathbf{S}$ is also a direct result of the majorization of the vectors. Given the equality of the sum of all vector coordinates in the definition of vector majorization, the defining condition is equivalent to saying that the sum of the k largest coordinates of vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ satisfy the bounds $\sum_{j=1}^{k} \beta_j \leq \sum_{i=1}^{k} \alpha_i$, for all $k = 1, ..., p$, with equality when $k = p$.

10. The proof of this result is analogous to the proof of the previous point if we replace matrix $\mathbf{S}$ with matrix $\mathbf{T}$ and matrix $\mathbf{W}^{[u]^t}\mathbf{S}\mathbf{W}^{[u]}$ with matrix $\mathbf{W}^{[c]^t}\mathbf{T}\mathbf{W}^{[c]}$. Note that the latter is a matrix whose diagonal elements are $\mathbf{w}_i^{[c]^t}\mathbf{T}\mathbf{w}_i^{[c]} = \mathbf{w}_i^{[c]^t}(\mathbf{S} + \mathbf{c}_m\mathbf{c}_m^t)\mathbf{w}_i^{[c]} = \lambda_i^{[c]} + \|\mathbf{c}_m\|^2 \cdot \cos^2(\mathbf{c}_m, \mathbf{w}_i^{[c]})$. On the other hand, the eigenvalues of matrix $\mathbf{W}^{[c]^t}\mathbf{T}\mathbf{W}^{[c]}$ are the same as those of matrix $\mathbf{T}$, since a spectral decomposition of the matrix $\mathbf{W}^{[c]^t}\mathbf{T}\mathbf{W}^{[c]}$ is given by $(\mathbf{W}^{[c]^t}\mathbf{W}^{[u]})\,\Lambda^{[u]}\,(\mathbf{W}^{[c]^t}\mathbf{W}^{[u]})^t$.

11. The first bounds given in (3.15) imply that the sum of the k largest eigenvalues of $\mathbf{S}$ is larger than, or equal to, the sum of the variances of *any* k uncentred PCs. This means that we can use the sum of variances of the first k uncentred PCs (although this bound may not be the best one available), and write bounds involving equal-ranking eigenvalues from both matrices:

$$\sum_{i=1}^{k} \lambda_i^{[c]} \geq \sum_{i=1}^{k} \left[\lambda_i^{[u]} - \|\mathbf{c}_m\|^2 \cdot \cos^2(\mathbf{c}_m, \mathbf{w}_i^{[u]})\right] \; = \; \sum_{i=1}^{k} \mathrm{var}(\boldsymbol{\eta}_i^{[u]}) \; .$$

The upper bound in (3.19) now follows directly. Likewise, the sum of the smallest k eigenvalues of $\mathbf{S}$ is bounded above by the sum of variances of the last k uncentred PCs (the right-most bound in (3.15) implies that it is bounded above by the sum of variances of *any* k uncentred PCs). Thus, we obtain the lower bound in (3.21). The lower bound in (3.19) and the upper bound in (3.21) appear in the same way, working from the bounds in (3.17).

12. The right-hand side bound in (3.22) results directly from relaxing the right-hand side bound in (3.20). The left-hand side bound in (3.22) results from (3.3), since:

$$\lambda_1^{[u]} \geq \frac{\mathbf{x}^t\mathbf{T}\mathbf{x}}{\mathbf{x}^t\mathbf{x}} = \frac{\mathbf{x}^t\mathbf{S}\mathbf{x}}{\mathbf{x}^t\mathbf{x}} + \cos^2(\mathbf{x},\mathbf{c}_m)\cdot\|\mathbf{c}_m\|^2 \geq \lambda_p^{[c]} + \cos^2(\mathbf{x},\mathbf{c}_m)\cdot\|\mathbf{c}_m\|^2 \ .$$

Since these bounds are valid for any vector $\mathbf{x} \in \mathbb{R}^p$, they are valid in particular when $\mathbf{x} = \mathbf{c}_m$.

13. This is a direct result of bounding the lefthand side of equation (3.9).

14. From inequality (3.20) and the more general relation (3.13), we obtain the bounds:

$$\begin{array}{ccccc} \cos^2(\mathbf{w}_1^{[c]},\mathbf{c}_m) & \leq & \frac{\cos(\mathbf{w}_1^{[c]},\mathbf{c}_m)\cdot\cos(\mathbf{w}_1^{[u]},\mathbf{c}_m)}{\cos(\mathbf{w}_1^{[c]},\mathbf{w}_1^{[u]})} & \leq & \cos^2(\mathbf{w}_1^{[u]},\mathbf{c}_m) \\ \frac{1}{\cos^2(\mathbf{w}_1^{[c]},\mathbf{c}_m)} & \geq & \frac{\cos(\mathbf{w}_1^{[c]},\mathbf{w}_1^{[u]})}{\cos(\mathbf{w}_1^{[c]},\mathbf{c}_m)\cdot\cos(\mathbf{w}_1^{[u]},\mathbf{c}_m)} & \geq & \frac{1}{\cos^2(\mathbf{w}_1^{[u]},\mathbf{c}_m)} \end{array}$$

Squaring and multiplying through by the denominator in the middle term produces the inequalities:

$$\frac{\cos^2(\mathbf{w}_1^{[u]},\mathbf{c}_m)}{\cos^2(\mathbf{w}_1^{[c]},\mathbf{c}_m)} \ \geq \ \cos^2(\mathbf{w}_1^{[c]},\mathbf{w}_1^{[u]}) \ \geq \ \frac{\cos^2(\mathbf{w}_1^{[c]},\mathbf{c}_m)}{\cos^2(\mathbf{w}_1^{[u]},\mathbf{c}_m)} \ .$$

Only the right-hand side inequality is useful, since the left-hand side bound exceeds unity.

REFERENCES

[1] O. Alter and G.H. Golub. Singular value decomposition of genome-scale mRNA lenghts distribution reveals asymmetry in RNA gel electrophoresis band broadening. *Proceedings of the National Academy of Science*, 103(32):11828–11833, 2006.

[2] S.T. Buckland and A.J.B. Anderson. Multivariate analysis of atlas data. In B.J.T. Morgan and P.M. North, editors, *Statistics in Ornithology*, pages 93–112. Springer-Verlag, Berlin, 1985.

[3] R. Cangelosi and A. Goriely. Component retention in principal component analysis with application to cDNA microarray data. *Biology Direct*, 2:doi:10.1186/1745–6150–2–2, 2007.

[4] S. Clements. Principal component analysis applied to the spectral energy distributions of blazars. In H.R. Miller, J.R.Webb, and J.C. Noble, editors, *Blazar Continuum Variability*, volume 110 of *Astronomical Society of the Pacific Conference Series*, pages 455–461, 1996.

[5] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218, 1936.

[6] F. Gwadry, C. Berenstein, J. van Horn, and A. Braun. Implementation and application of principal component analysis on functional neuroimaging data. Technical Research Report TR 2001-47, Institute for Systems Research, 2001.

[7] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[8] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441 and 498–520, 1933.

[9] R. Huth and L. Pokorná. Simultaneous analysis of climatic trends in multiple variables: an example of application of multivariate statistical methods. *International Journal of Climatology*, 25:469–484, 2005.

[10] J.E. Jackson. *A User's Guide to Principal Components*. Wiley, New York, 1991.

[11] I.T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 2002.

[12] R.S. Mann, M.E.; Bradley and M.K. Hughes. Global-scale temperature patterns and climate forcing over the past six centuries. *Nature*, 392:779–787, 1998.

[13] S. McIntyre and R. McKitrick. Hockey sticks, principal components, and spurious significance. *Geophysical Research Letters*, 32:doi:10.1029/2004GL021750, 2005.

[14] S. McIntyre and R. McKitrick. The M&M critique of the MBH98 northern hemisphere climate index: update and implications. *Energy & Environment*, 16(1):69–100, 2005.

[15] R.A. Reyment and K.G. Jöreskog. *Applied Factor Analysis in the Natural Sciences*. Cambridge University Press, 1993.

[16] K.M. Somers. Multivariate allometry and removal of size with principal components analysis. *Systematic Zoology*, 35:359–368, 1986.

[17] Y. Takane and T. Shibayama. Principal component analysis with external information on both subjects and variables. *Psychometrika*, 56(1):97–120, 1991.

[18] C. ter Braak. Principal components biplots and alpha and beta diversity. *Ecology*, 64:454–462, 1983.

[19] H. van den Dool. *Empirical Methods in Short-Term Climate Prediction*. Oxford University Press, 2007.

[20] D. S. Wilks. *Statistical Methods in the Atmospheric Sciences*. Academic Press, second edition, 2006.

[21] H. Yanai and K. Takeuchi. *Projection matrices, generalized inverse and singular value decomposition*. University of Tokyo Press, 1983. (in Japanese).