Stat
The ISI's Journal for the Rapid
Dissemination of Statistics Research
(wileyonlinelibrary.com) DOI: 10.100X/sta.0000

Correlated Components

Trevor F. Cox[a]*, David S. Arnold[b]

Received 00 Month 2015; Accepted 00 Month 2015

Principal components analysis is a much used and practical technique for analysing multivariate data, finding a particular set of linear compounds of the variables under consideration, such that covariances between all pairs are zero. An alternative view is that when the variables are considered as axes in a Cartesian coordinate system, then principal components analysis is the particular orthogonal rotation of the axes that makes all the pairwise covariances equal to zero. It is this view that is taken here, but instead of finding the rotation that makes all covariances equal to zero, an orthogonal rotation is found that maximizes the sum of the covariances. The rotation is not unique, except for the two or three component case and so another criterion can be used alongside so that it too can also be optimized. The motivation is that two highly correlated components will tend to measure the same latent variable but with interesting differences due to the orthogonality between them. Theory is given for identifying the correlated components as well as algorithms for finding them. Two illustrative examples are provided, one involving gene expression data and the other consumer questionnaire data.

Keywords: correlated components; factor analysis; gene expression data; parallel coordinates plot; principal components

1. Introduction

Principal components analysis is a well established and a corner-stone technique in multivariate data analysis and is described in most books on multivariate analysis, see for example Cox (2005) or Everitt & Hothorn (2011). Jolliffe (2002) gives a detailed authoritative account of the subject with a brief history and describes how the origins go back to Pearson (1901) and Hotelling (1933) and possibly further. Principal components analysis can be described as a dimension-reducing technique that transforms a set of p original variables $\mathbf{x} = (x_1, \ldots, x_p)^T$ to a new set of uncorrelated variables $\mathbf{y} = (y_1, \ldots, y_p)^T$. First, the vector $\mathbf{u}_1$ giving the linear combination, $y_1 = \mathbf{u}_1^T\mathbf{x}$ ($\mathbf{u}_1^T\mathbf{u}_1 = 1$) is found so that y_1 has maximum possible variance. This is the first principal component. Then the next linear combination, $y_2 = \mathbf{u}_2^T\mathbf{x}$ ($\mathbf{u}_2^T\mathbf{u}_2 = 1$) is found that has maximum variance, but subject to being uncorrelated with y_1, giving the second principal component. The process is repeated until all p principal components are found, all uncorrelated with each

[a]**Departments of Molecular and Clinical Cancer Medicine and Biostatistics, University of Liverpool, Liverpool, L69 3GL, UK.**
[b]**Unilever R&D Port Sunlight, Bebington, CH63 3JW, UK.**
*__Email: coxt@liv.ac.uk__

Prepared using staauth.cls [Version: 2012/05/12 v1.00]

other and labelled according to decreasing order of magnitude of variance. The dimension of the original space is said to be reduced from p to q when the first q ($q < p$) principal components are selected for further analysis.

Let $\mathbf{\Sigma}$ be the covariance matrix of $\mathbf{x}$. It is easily shown that $\{\mathbf{u}_i\}$ are given by the eigenvectors of $\mathbf{\Sigma}$ with $\text{var}(y_i) = \lambda_i$, the ith eigenvalue of $\mathbf{\Sigma}$ and where the eigenvalues are labelled in decreasing order of magnitude. Let $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_p)$ be the orthogonormal matrix of eigenvectors of $\mathbf{\Sigma}$ and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_p)$ the diagonal matrix of eigenvalues of $\mathbf{\Sigma}$. Then

$$\mathbf{y} = \mathbf{U}^T\mathbf{x}, \quad \mathbf{\Sigma} = \mathbf{U}^T\mathbf{\Lambda}\mathbf{U}, \quad \mathbf{\Lambda} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^T, \quad \mathbf{U}^T\mathbf{U} = \mathbf{I}.$$

In practice principal components analysis is carried out on the sample covariance matrix, $\mathbf{S}$, or the sample correlation matrix $\mathbf{R}$, the choice being made according to the problem at hand.

Another way of viewing principal components analysis is that, at the population level, it finds the orthogonal rotation, $\mathbf{U}$, of the Cartesian axes defined by $\mathbf{x}$ to a new set of axes, $\mathbf{y}$, so that the correlations between pairs of rotated variables are all zero. At the sample level, the observations existing in the $\mathbf{x}$-space are rotated to be observations in the $\mathbf{y}$-space that have pairwise sample correlations all equal to zero. It is this geometric view that we will be taking in this paper where various orthogonal rotations will be sought that optimize specific criteria.

The elements of the $\mathbf{u}_i$ are often called loadings and it can be difficult to interpret these in practice. How important a particular loading is to a principal component is often measured by its absolute value compared to the other loadings, but as Cadima & Jolliffe (2001) point out, interpretation is not so simple and other factors should be considered such as the correlations of the principal components with the original variables. However, these issues will not be discussed here.

Principal components factor analysis starts with the first q principal components, scales these so that all have unit variance, and then orthogonally rotates them to be more interpretable than the original principal components. More interpretable may mean that loadings, if possible, should be very small or very large but not an in-between value. To achieve this, the varimax procedure (Kaiser, 1958) finds the rotation such that the sum of the variances of the squared loadings is maximized, but this is just one of a set of orthogonal rotations that can be undertaken to meet specific optimal criteria. Non-orthogonal rotations have been proposed in the literature, but these are not considered here, see for example, Basilevsky (1994) and Browne (2001) for further details. Rotation of principal components is widely used in climatology (Richman, 1986; Mestas-Nunez, 2000) and is discussed by Jolliffe (2002).

This paper seeks a different set of components from principal components. Components are sought that have maximum covariance in some sense to be defined. The rationale behind this idea is as follows using the example of measuring intelligence often used to motivate the ideas of factor analysis. If a sample of people are given several tests to complete, say, mathematics tests, logic tests and comprehension tests, then a factor analysis based on a rotation of test scores might identify a factor that could be labelled intelligence. A second factor, uncorrelated to the intelligence factor, would measure another aspect of the test scores, for instance, it might contrast mathematics scores with comprehension scores. However, if a rotation is sought that maximizes the covariance of factors, then it could be argued that the factors, although orthogonal, are attempting to measure the same thing. So there would be two or more ways of measuring intelligence with interesting differences between the factors induced by their orthogonality.

Section 2 shows how the orthogonal rotations are found in order to maximize the total sum of pairwise covariances and Section 3 for the total sum of squared pairwise covariances. Note that covariances are used in preference to correlations since two components can have high correlation, but each have very small variance making them unimportant in the analysis. Notwithstanding the fact that covariances are used in finding components, the components are called correlated components. Seeking correlated components is not to be confused with canonical correlation analysis where there are two groups of variables and linear compounds are found, one for each group, that have maximum correlation

Härdle & Simar (2007). Section 4 illustrates the new techniques of finding correlated components on a classic set of gene expression data collected by Golub (1999) and on some consumer questionnaire data.

Rotations are found that maximize the sum of all pairwise covariances between the new variables, named Criterion 1, and unlike principal components analysis, it is not a straight forward task to find the rotations. The covariance matrix $\mathbf{\Sigma}$ of the original variables $\mathbf{x}$ could either be a population covariance matrix or a sample covariance matrix. If the variables are scaled to have unit variance then $\mathbf{\Sigma}$ would be a correlation matrix. Note, orthogonal rotations preserve eigenvalues, so if $\mathbf{R}$ is the rotation matrix, and $\mathbf{V}$ is the covariance matrix of the rotated components, then

$$|\mathbf{V} - \lambda\mathbf{I}| \equiv |\mathbf{R}^T\mathbf{\Sigma}\mathbf{R} - \lambda\mathbf{I}| \equiv |\mathbf{R}^T(\mathbf{\Sigma} - \lambda\mathbf{I})\mathbf{R}| \equiv |(\mathbf{\Sigma} - \lambda I)\mathbf{R}\mathbf{R}^T| \equiv |\mathbf{\Sigma} - \lambda\mathbf{I}|.$$

An orthogonal rotation by a matrix $\mathbf{R}_1$ followed by another orthogonal rotation, $\mathbf{R}_2$ results in an overall orthogonal rotation, or put another way, any orthogonal rotation can be split into two separate orthogonal rotations. Thus the rotation that is being sought here can be composed of a first orthogonal rotation to principal components using $\mathbf{U}$, followed by a rotation, $\mathbf{A}$, from the principal components to the correlated components. This implies that the only case that needs to be considered is that of rotating to correlated components from the principal components. This has the consequence that as the covariance matrix of the principal components is $\mathbf{\Lambda}$ and the covariance matrix of the correlated components is $\mathbf{V} = \mathbf{A}^T\mathbf{\Lambda}\mathbf{A}$, then the elements of $\mathbf{V}$ can only be functions of the eigenvalues of $\mathbf{\Sigma}$.

The method is not to look for the rotation matrix $\mathbf{A}$ directly, but to find $\mathbf{V}$ first and then $\mathbf{A}$ is found from the spectral decomposition of $\mathbf{V}$. Hence the overall rotation matrix from the original variables is $\mathbf{UA}$. Note, here the term orthogonal rotation will also include improper rotations which can include reflection as well as rotation.

2. Rotation to correlated components (Criterion 1)

2.1. Rotations to maximize the sum of covariances

The criterion to be maximized is M, where

$$2M = \sum_{\substack{i=1 \\ i\neq j}}^{p}\sum_{j=1}^{p} \mathrm{cov}(z_i, z_j) = \mathbf{1}^T\mathbf{V}\mathbf{1} - \mathrm{trace}(\mathbf{V}) \tag{1}$$

and $\{z_i\}$ are the correlated components after the rotation from principal components, $\mathbf{V}$ their covariance matrix and $\mathbf{1}$ is a vector of 1's.

Now $\mathbf{V} = \mathbf{A}^T\mathbf{\Lambda}\mathbf{A}$ and hence

$$2M = \mathbf{1}^T\mathbf{A}^T\mathbf{\Lambda}\mathbf{A}\mathbf{1} - \mathrm{trace}(\mathbf{A}^T\mathbf{\Lambda}\mathbf{A}) = \mathbf{a}^T\mathbf{\Lambda}\mathbf{a} - \mathrm{trace}(\mathbf{\Lambda})$$

where $\mathbf{a} = \mathbf{A}\mathbf{1}$. Note $\mathbf{a}^T\mathbf{a} = \mathbf{1}^T\mathbf{A}^T\mathbf{A}\mathbf{1} = \mathbf{1}^T\mathbf{1} = p$.

Introducing a Lagrange multiplier, μ, for the constraint $\mathbf{a}^T\mathbf{a} = p$ and then differentiating $2M$ with respect to $\mathbf{a}$ and equating to $\mathbf{0}$, gives $(\mathbf{\Lambda} - \mu\mathbf{I})\mathbf{a} = \mathbf{0}$ and hence μ is an eigenvalue of $\mathbf{\Lambda}$ and $\mathbf{a}$ the corresponding eigenvector. Since $2M = \mathbf{a}^T\mathbf{\Lambda}\mathbf{a} - \mathrm{trace}(\mathbf{\Lambda})$ and since the ith eigenvector of $\mathbf{\Lambda}$ has elements all zero apart from the ith, $\mathbf{a}$ must be the eigenvector corresponding to the largest eigenvalue, i.e. $\mathbf{a} = (\sqrt{p}, 0, \ldots, 0)^T$. Thus

$$M = \big(p\lambda_1 - \sum_{i=0}^{p}\lambda_i\big)/2 \tag{2}$$

and this maximum value is achieved by any rotation matrix, $\mathbf{A}$, such that $\mathbf{A1} = \mathbf{a}$, i.e. any orthonormal matrix with first row sum equal to $\sqrt{p}$ and the remaining row sums equal to zero.

Now write $\mathbf{A}$ in terms of its column vectors, $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_{p-1}, \mathbf{a}_p)$, where the last column vector can be written as $\mathbf{a}_p = (\sqrt{p}, 0, \ldots, 0)^T - \sum_{j=1}^{p-1} \mathbf{a}_j$. Now

$$\mathbf{a}_i^T \mathbf{a}_p = 0 = \sqrt{p} a_{i1} - \sum_{j=1}^{p-1} \mathbf{a}_i^T \mathbf{a}_j = \sqrt{p} a_{i1} - 1 \qquad (i = 1, \ldots, p-1),$$

implying that $a_{i1} = 1/\sqrt{p}$. Thus all the elements of the first column of $\mathbf{A}$ are equal to $1/\sqrt{p}$.

The row sums of $\mathbf{V}$ are given by

$$\mathbf{V1} = \mathbf{A}^T \mathbf{\Lambda A 1} = \mathbf{A}^T \mathbf{\Lambda a} = \mathbf{A}^T (\lambda_1 \sqrt{p}, 0, \ldots, 0)^T = (\lambda_1, \ldots, \lambda_1)^T,$$

the last step using the fact that the elements of the first column vector of $\mathbf{A}$ are all equal to $1/\sqrt{p}$. Thus all the row sums of $\mathbf{V}$ are equal to the largest eigenvalue.

In practice, it may be more desirable to find a q ($q < p$) correlated components, where $M = \sum \sum_{i\,j=1, i \neq j}^{q} \mathrm{cov}(z_i, z_j)/2$ is maximised. The solution for this case is shown in the Appendix and makes extensive use of Lagrange multipliers.

2.2. Finding the optimal rotations

In order for any matrix, $\mathbf{A}$, to be orthogonal, within its p^2 elements, there are $p + p(p-1)/2$ constraints: p length constraints and $p(p-1)/2$ orthogonality constraints. This leaves $p(p-1)/2$ degrees of freedom. There are $p + p(p-1)/2$ variances and covariances in an arbitrary covariance matrix $\mathbf{V}$, but here the p eigenvalues of $\mathbf{V}$ need to be equal to those for $\mathbf{\Sigma}$, and so this leaves $p(p-1)/2$ degrees of freedom for $\mathbf{V}$, the same as for $\mathbf{A}$. For the optimal $\mathbf{V}$, the row sums are fixed and so this gives an extra p constraints, leaving $p(p-3)/2$ degrees of freedom, implying that for $p > 3$ there will not be a unique solution for $\mathbf{V}$ and equivalently, $\mathbf{A}$. Thus, in practice $p(p-3)/2$ elements of $\mathbf{V}$ can be chosen arbitrarily and then $\mathbf{A}$ found. However, the choice of the values for the $p(p-3)/2$ elements must be such that a solution actually exits. For instance $p(p-3)/2$ elements could be put equal to zero and a solution sought but if one does not exist, the same elements could be given a very small value and then a solution possibly found. However, in the following section it is shown how solutions can be easily generated.

Three methods are given for finding optimal solutions for maximizing 1. The first method involves choosing values for $p(p-3)/2$ elements of $\mathbf{V}$ and finding the remaining elements of $\mathbf{V}$ by solving the equations

$$\sum_{j=1}^{p} V_{ij} - \lambda_1 = 0 \quad (i = 1, \ldots, p), \qquad \mathrm{trace}(\mathbf{V}^k) = \mathrm{trace}(\mathbf{\Lambda}^k) \quad (k = 1, \ldots, p),$$

using a Newton-Raphson approach. However, the method is sensitive to starting values and it is difficult to know whether there is no solution or a bad starting point has been used.

The second method again needs $p(p-3)/2$ elements of $\mathbf{V}$ to be set and then uses a perturbation algorithm that at the current iteration compares the eigenvalues, $\{\mu_i\}$, of the current $\mathbf{V}$ with $\{\lambda_i\}$ by calculating the loss function

$$\sum_{i=1}^{p} (\lambda_i - \mu_i)^2.$$

Now all the variances, $\{V_{ii}\}$, must lie in the range (λ_1, λ_p) and can be ordered $V_{11} > \ldots > V_{pp}$. The algorithm proceeds by taking each V_{ii} in turn and perturbing it by a small increment and also decrement. Then the row sum equations are solved to yield updated covariances $\{V_{ij}\}$. The eigenvalues of the updated $\mathbf{V}$ are calculated and then the particular perturbation that gave the best improvement to the loss is chosen. The process repeats until the loss becomes zero. If the loss never achieves the value zero, then this indicates that a solution does not exist for the $p(p-3)/2$ values initially chosen for $\mathbf{V}$.

The third method always produces a solution. Firstly an arbitrary $p \times p$ symmetric matrix, $\mathbf{G}$, is chosen where the row sums are made to be equal. The matrix must be such that it can not be partitioned into a 2×2 block matrix structure, where the off-diagonal block contains all zeros. An orthonomal basis is found for $\mathbf{G}$ using the spectral decomposition $\mathbf{G} = \mathbf{H}^T\mathbf{\Theta}\mathbf{H}$. When $\mathbf{H}$ is used as an orthogonal rotation matrix for $\mathbf{\Lambda}$ then $\mathbf{V} = \mathbf{H}^T\mathbf{\Lambda}\mathbf{H}$ has row sums equal and is a solution for Criterion 1. This can be shown using similar arguments to those at the end of Section 2.1 for the row sums of $\mathbf{V}$.

3. Rotation to correlated components (Criterion 2)

3.1. Rotations to maximize the sum of squared covariances

Maximizing M for Criterion 1 does allow negative covariances, but these will be swamped by the positive covariances. If it is the absolute value of the covariances that are important, then the sum of the squared pairwise covariances can be maximized: Criterion 2. Let M be given by

$$M = \frac{1}{2}\sum_{\substack{i=1 \\ i\neq j}}^{p}\sum_{j=1}^{p} \mathrm{cov}(z_i, z_j)^2.$$

It is easily seen that

$$2M = \mathrm{trace}(\mathbf{V}^2) - \sum_{i=1}^{p} V_{ii}^2.$$

Now $\mathbf{V} = \mathbf{A}^T\mathbf{\Lambda}\mathbf{A}$ and so

$$2M = \mathrm{trace}\{(\mathbf{A}^T\mathbf{\Lambda}\mathbf{A})(\mathbf{A}^T\mathbf{\Lambda}\mathbf{A})\} - \sum_{i=1}^{p} V_{ii}^2 = \mathrm{trace}(\mathbf{\Lambda}^2) - \sum_{i=1}^{p} V_{ii}^2.$$

The first term is fixed and it is easily shown that the second in minimized when all the V_{ii}'s are equal and so $V_{ii} = \mathrm{trace}(\mathbf{\Lambda})/p$. Thus

$$2M = \mathrm{trace}(\mathbf{\Lambda}^2) - \{\mathrm{trace}(\mathbf{\Lambda})\}^2/p = \sum_{i=1}^{p} \lambda_i^2 - \Big(\sum_{i=1}^{p} \lambda_i\Big)^2/p. \tag{3}$$

The Appendix shows how the solution for the case where q ($q < p$) correlated components are required.

3.2. Finding the optimal rotations

An algorithm was developed to find optimal rotations using the fact that all variances of $\mathbf{V}$ have to be equal. An Euler angle approach is used to rotate pairs of axes to make their variances equal. Starting with the covariance matrix,

$\mathbf{V} = \mathbf{\Lambda}$, with the variances placed in descending order along the diagonal, an Euler matrix, $\mathbf{E}(\theta)$, is defined as a $p \times p$ identity matrix, but with $(1,1)$th element $\cos(\theta)$, (p,p)th element $\cos(\theta)$, $(1,p)$th element $\sin(\theta)$ and $(p,1)$th element $-\sin(\theta)$. If θ is chosen as $\arctan\{(V_{11} - V_{pp})/2V_{1p}\}/2$, assuming $V_{1p} \neq 0$, then rotation by the Euler rotation matrix will give rise to a new covariance matrix, $\mathbf{V}'$ with $V'_{11} = V'_{pp}$. The rows and columns are interchanged so that the variances are again in descending order along the diagonal. This process is repeated until all variances are equal and thus arriving at $\mathbf{V}$. Different solutions can be found by first applying a random rotation matrix $\mathbf{W}$ to obtain a new starting covariance matrix $\mathbf{W}^T\mathbf{\Lambda W}$ from which to start the algorithm. A random rotation matrix can be easily constructed by generating a $p \times p$ matrix, $\mathbf{B}$, of random numbers, then using this to form the symmetric matrix, $\mathbf{C} = \mathbf{B} + \mathbf{B}^T$, and finding the spectral decomposition, $\mathbf{C} = \mathbf{W}^T\mathbf{\Psi W}$, giving $\mathbf{W}$ as the random rotation matrix.

4. Examples

4.1. Correlated components for gene expression data

Golub (1999) collected gene expression data for acute leukemia that has become a classic dataset in bioinformatics. It is used extensively in the on-line book by Krijnen (2009) to illustrate the use of the statistical computing and graphics package R in bioinformatics. There are gene expression levels on 3051 genes for thirty eight leukemia patients, twenty seven with acute lymphoblastic leukemia (ALL) and eleven with acute myeloid leukemia (AML). Here we use a subset of thirty four genes selected by Krijnen (2009) that are involved in S-phase cell cycle progression. These genes were extracted from the database and are listed in Table 1.

A principal components analysis was carried out on the thirty four genes using the covariance matrix. Table 2 shows the first six principal components which account for 80% of the total variation.

TABLE 1 ABOUT HERE

TABLE 2 ABOUT HERE

Genes 8, 16, 23 and 26 (CD9, CD22, CD19 antigens and CD24 signal transducer) weight highly positive on PC1; for PC2, genes 4, 14, 22 and 25 (CD3Z, CD3G, CD7 precursor, CD2) are contrasted with genes 8, 13, 18 and 23 (CD9, Catalase (EC1.11.1.6), ME491 and CD24) and similarly for the other PCs. Figure 1a shows the PC scores plot (PC2 versus PC1) for the thirty eight patients. The two leukemia groups are well separated.

The subspace spanned by the first six PCs was considered to appropriately summarize the variation in the gene expression data. An orthogonal rotation was found using Criterion 1 to find just two highly correlated components. The unique solution has variance matrix given by

$$\mathbf{V} = \begin{bmatrix} 3.63 & 2.99 \\ 2.99 & 3.63 \end{bmatrix},$$

giving the correlation between the two CCs as 0.82. The two CCs are given in Table 2 as CC1a and CC2a. They explain 50% of the total variance within the subspace. The first CC is dominated by genes 1, 12, 15 and 23 (PRKCD, TFRC, CD37 and CD24) and the second by genes 8, 9, 15, 16, 18 and 26 (CD9, CD72, CD37, CD22, ME491 and CD19). Figure 1b shows a plot of the CC2a scores against the CC1a scores for the thirty eight patients. Again the two leukemia groups separate.

The same six-dimensional subspace was rotated where this time a rotation matrix was found using Criterion 2 that also maximized one particular squared covariance between two of the CCs. The algorithm used ten thousand random

starts. The covariance matrix obtained was

$$\mathbf{V} = \begin{bmatrix} 2.44 & 1.79 & -1.30 & -0.52 & -0.86 & -1.16 \\ 1.79 & 2.44 & -1.36 & -0.41 & -0.78 & -1.22 \\ -1.30 & -1.36 & 2.44 & -0.34 & 0.11 & 1.05 \\ -0.52 & -0.41 & -0.34 & 2.44 & 1.54 & -0.72 \\ -0.86 & -0.78 & 0.11 & 1.54 & 2.44 & -0.19 \\ -1.16 & -1.22 & 1.05 & -0.72 & -0.19 & 2.44 \end{bmatrix},$$

showing the first two CC having the largest covariance. They are shown in Table 2 as CC1b and CC2b and they explain 34% of the total variance within the subspace. CC1b is dominated by genes 5, 8, 15, 16 and 26 (ANPEP, CD9, CD37, CD22 and CD19) and CC2b by genes 5, 12, 15, 18, 23 and 29 (ANPEP, TFRC, CD37, ME491, CD24 and GB DEF).

4.2. Parallel coordinates plots for nonmetric scaling

Parallel coordinates plots help visualize high dimensional multivariate data. Variables are represented by a sequence of vertical parallel axes. For a particular observation, the variable values are marked as points on the axes which are then joined across all the axes giving a "profile" for the observation. This is done for all the observations. Patterns in the data can then be noticed and investigated. Some statistical packages allows for interactive manipulation of the parallel coordinates plot. One problem to be overcome with parallel coordinates plots is the choice of order of the variables. Here parallel coordinates plots are used in conjunction with nonmetric multidimensional scaling. Multidimensional scaling covers a variety of methods for reducing and summarizing multivariate data, see for example Cox & Cox (2001).

Nonmetric scaling is a technique for representing objects or observations as a configuration of points in a low dimensional space, usually Euclidean, where distances between pairs of points match, as well as possible, the original dissimilarities measured between corresponding pairs of objects. The p dimensional configuration of points obtained can be rotated, reflected and translated to give an equivalent configuration, that is, it is only the relative positions of the points that is important. Thus for the purposes of showing the positions of the points in the multidimensional scaling configuration via a parallel coordinates plot, the points can be rotated to different axes and a more interpretable plot. Figure 2 illustrates this where points in a multidimensional scaling configuration obtained from some consumer questionnaire data for deodorants are plotted. Here it is the variables that are of interest with dissimilarities based on the correlation coefficient. The 49 questions in the questionnaire covered areas such as (i) application, with typical questions, which have been abbreviated: ease of application, how wet whilst applying; (ii) fragrance: strength of fragrance, fragrance lasted long enough and (iii) overall opinion: overall opinion of effectiveness, overall opinion of fragrance. Responses were on 5-point Likert scales which were then deemed to be interval-scale data. It is not necessary to give further details regarding the data here.

The upper plot in Figure 2 shows the parallel coordinates plot for the axes obtained from a 5-dimensional multidimensional scaling analysis and the lower plot the same points but after a rotation of the axes, where the sum of the covariances has been maximized and the solution chosen that had largest sum of off-diagonal covariances maximized over ten thousand runs of the solution generating algorithm. The rotated view shows a group of variables at the top of the plot separated from the remaining variables (heavy dashed lines). These were negative aspects of the deodorants: felt wet during application, felt sticky while drying, left visible deposits, cold on application, marked clothes, waited longer than usual to dry and felt greasy. This pattern is not clear in the unrotated plot and would have been missed.

5. Discussion

The authors believe that the idea of finding highly correlated components, possibly alongside the principal components, is novel and merits further research. Efficient algorithms, perhaps based on genetic algorithms, need to be established for optimizing criteria used to choose particular sets of correlated components when there is no unique set. The aim here has been to establish the theory for correlated components with a further non-mathematical publication being planned to demonstrate the practical use of correlated components on consumer and medical data, with the intention of reaching a wider audience. Some R code will be made available for practitioners of principal components analysis and factor analysis to explore the use of correlated components within their fields of research.

6. Appendix

6.1. Finding q $(q < p)$ correlated components using Criterion 1

Suppose instead of finding a full set of p correlated components, only q $(q < p)$ correlated components are to be found that have maximum sum of pairwise covariances. Again, the covariance matrix $\mathbf{V}$ is sought. Let the q correlated components be labelled 1 to q and have covariance matrix $\mathbf{V}^{11}$ and let $\mathbf{V}$ be partitioned as

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}^{11} & \mathbf{V}^{12} \\ \mathbf{V}^{21} & \mathbf{V}^{22} \end{bmatrix},$$

$\mathbf{V}^{22}$ being the covariance matrix of the remaining components and $\mathbf{V}^{12} = (\mathbf{V}^{21})^T$ the covariances between the q correlated components and the rest of the components. Let $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_p)$.

Using Lagrange multipliers for the various constraints, the quantity to be maximised is

$$M = \frac{1}{2}\sum_{\substack{i=1 \\ i\neq j}}^{q}\sum_{j=1}^{q} \mathbf{a}_i^T \mathbf{\Lambda} \mathbf{a}_j - \frac{1}{2}\sum_{i=1}^{p} \mu_i(\mathbf{a}_i^T\mathbf{a}_i - 1) - \frac{1}{2}\sum_{\substack{i=1 \\ i\neq j}}^{p}\sum_{j=1}^{p} \mu_{ij}(\mathbf{a}_i^T\mathbf{a}_j).$$

Differentiating M with respect to $\mathbf{a}_i$ and equating to $\mathbf{0}$, gives the equations

$$I_{\{i\leq q\}} \sum_{j\neq i}^{q} \mathbf{\Lambda}\mathbf{a}_j - \mu_i \mathbf{a}_i - \sum_{j\neq i}^{p} \mu_{ij}\mathbf{a}_j = \mathbf{0} \quad (i = 1, \ldots, p), \tag{4}$$

where $I_{\{.\}}$ is the indicator function. For $i > q$, pre-multiplying each equation in (4) by $\mathbf{a}_k^T$, gives $\mu_i = 0$ and $\mu_{ik} = 0$. For $i \leq q$, pre-multiplying by $\mathbf{a}_k^T$, leads to

$$\sum_{j\neq i}^{q} V_{kj} - \mu_{ik} = 0 \quad (k \neq i), \qquad \sum_{j\neq i}^{q} V_{ij} - \mu_i = 0 \quad (k = i).$$

For $i, k \leq q$, equating μ_{ik} and μ_{ki} gives $\sum_{j\neq i}^{q} V_{kj} = \sum_{j\neq k}^{q} V_{ij}$. Adding V_{ki} and V_{ik} (these are equal) to either side of this equation shows that the row sums of V^{11} are all equal. For $i \leq q, k > q$, $\sum_{j\neq i}^{q} V_{kj} = 0$ since $\mu_{ik} = 0$. Adding these equations over i from 1 to q gives $(q-1)\sum_{j=1}^{q} V_{ij} = 0$. Considering $\sum_{j\neq i}^{q} V_{kj} - \sum_{j\neq i'}^{q} V_{kj} = V_{ki'} - V_{ki} = 0$, shows all V_{ki}'s are equal and as their sum is equal to zero, $V_{ki} = 0$ $(i \leq q, k > q)$. Thus $\mathbf{V}^{12}$ contains all zero elements and so the correlated components are uncorrelated with the remaining components and are found in a similar manner to finding the full set, but using the q eigenvalues chosen from the set of p eigenvalues that give the maximum value of

M. From (2) it is clear that for a maximum, the largest eigenvalue and the $q-1$ smallest eigenvalues are to be used. The matrix $\mathbf{V}^{22}$ is arbitrary as long as it has eigenvalues equal to the $p-q$ remaining ones. In practice this can be left as the diagonal matrix of these eigenvalues, that is, the principal components.

6.2. Finding q ($q < p$) correlated components using Criterion 2

In a similar manner to that for Criterion 1, the quantity to be maximised is

$$M = \frac{1}{2}\sum_{\substack{i=1\\ i\neq j}}^{q}\sum_{j=1}^{q}(\mathbf{a}_i^T\mathbf{\Lambda}\mathbf{a}_j)^2 - \frac{1}{2}\sum_{i=1}^{p}\mu_i(\mathbf{a}_i^T\mathbf{a}_i - 1) - \frac{1}{2}\sum_{\substack{i=1\\ i\neq j}}^{p}\sum_{j=1}^{p}\mu_{ij}(\mathbf{a}_i^T\mathbf{a}_j).$$

Differentiating M with respect to $\mathbf{a}_i$ and equating to $\mathbf{0}$ gives the equations

$$I_{\{i\leq q\}}\sum_{j\neq i}^{q}(\mathbf{a}_i^T\mathbf{\Lambda}\mathbf{a}_j)\mathbf{\Lambda}\mathbf{a}_j - \mu_i\mathbf{a}_i - \sum_{j\neq i}^{p}\mu_{ij}\mathbf{a}_j = \mathbf{0} \quad (i = 1,\ldots,p).$$

Pre-multiply these equations by $\mathbf{a}_k^T$ to give

$$I_{\{i\leq q\}}\sum_{j\neq i}^{q}V_{ij}V_{jk} - \mu_i\mathbf{a}_i^T\mathbf{a}_k - \sum_{j\neq i}^{p}\mu_{ij}\mathbf{a}_j^T\mathbf{a}_k \quad (i = 1,\ldots,p).$$

Thus for $i > q$, $\mu_i = 0$ and $\mu_{ij} = 0$. For $i, k \leq q$, considering $i = k$ and $i \neq k$,

$$\sum_{j\neq i}^{q}V_{ij}^2 - \mu_i = 0, \qquad \sum_{j\neq i}^{q}V_{ij}V_{jk} - \mu_{ik} = 0.$$

Now as $\mu_{ik} = \mu_{ki}$, $\sum_{j\neq i}^{q}V_{ij}V_{jk} = \sum_{j\neq k}^{q}V_{jk}V_{ij}$, leading to $V_{ik}(V_{kk} - V_{ii}) = 0$. Thus all variances for the q correlated components are equal. Let this value be w.

For $i \leq q, k > q$, $\sum_{j\neq i}^{q}V_{ij}V_{kj} = 0$ and for an arbitrary k these equations can be written as $(\mathbf{V}^{11} - w\mathbf{I})\mathbf{v} = \mathbf{0}$, where $\mathbf{v} = (V_{k1},\ldots,V_{km})^T$. Solving this equation gives $\mathbf{v} = \mathbf{0}$ unless w were to be an eigenvalue of $\mathbf{V}^{11}$. However, this cannot be as the characteristic equation for $\mathbf{V}^{11}$ is a qth order polynomial in $\lambda - w$, where λ is an eigenvalue and as the polynomial has a constant term, w cannot be an eigenvalue. Thus the set of correlated components are uncorrelated with the rest. From (3), the eigenvalues chosen for $\mathbf{V}^{11}$ are those which maximize M calculated for q of the eigenvalues.

References

Basilevsky, A. (1994) *Statistical Factor Analysis and Related Methods*, London: Griffin.

Browne, M. W. (2001) An overview of analytic rotation in exploratory factor analysis. *Multivar. Behav. Res.*, 36, 111–50. DOI: 10.1207/S15327906MBR3601_05.

Cadima, J. F. C. L. & Jolliffe, I. T. (2001) Variable selection and the interpretation of principal subspaces. *J. Agr. Biol. Envir. St.*, 6, 62–79. DOI: 10.1198/108571101300325256.

Cox, T. F. (2005) *An Introduction to Multivariate Data Analysis*, London: Hodder Arnold.

Cox, T. F. & Cox, M. A. A. (2001) *Multidimensional Scaling*, Boca Raton: Chapman and Hall/CRC Press.

Everitt, B & Hothorn, T. (2011) *An Introduction to Applied Multivariate Analysis with R*, New York: Springer.

Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C. et al. (1999) Molecular classifcation of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286, 531–537. DOI: 10.1126/science.286.5439.531.

Härdle, W. & Simar, L. (2007) *Applied Multivariate Statistical Analysis*, Heidelberg: Springer.

Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.*, 24, 417–41, 498–520. DOI: 10.1037/h0070888.

Jolliffe, I. T. (2002) *Principal Components Analysis*, 2nd edn., New York: Springer Verlag.

Kaiser, H. F. (1958) The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187–200. DOI: 10.1007/BF02289233.

Krijnen, W. P. (2009) Applied Statistics for Bioinformatics using R. (Available from https://cran.r-project.org/doc/contrib/Krijnen-IntroBioInfStatistics.pdf.)

Mestas-Nunez, A. M. (2000) Orthogonality properties of rotated empirical modes. *Int. J. Climatol.*, 20, 1509–16. DOI: 10.1002/1097-0088(200010)20:12<1509::AID-JOC553>3.0.CO;2-Q

Pearson, K. (1901) On lines and planes of closest fit to systems of points in space. *Phil. Mag.*, 2, 559–72. DOI: 10.1098/rspl.1901.0094.

Richman, M. B. (1986) Rotation of principal components. *J. of Climatol.*, 6, 293–335. DOI: 10.1002/joc.3370060305.

Table 1. Genes used in the analysis of gene expression data for Leukemia

Gene no.	Gene name	Description
173	D10495_at	PRKCD Protein kinase C, delta
398	D38524_at	NT5 5' nucleotidase (CD73)
634	D84276_at	CD38 CD38 antigen (p45)
1106	J04132_at	CD3Z CD3Z antigen, zeta polypeptide (TiT3 complex)
1817	M22324_at	ANPEP Alanyl (membrane) aminopeptidase (aminopeptidase N, aminopeptidase M, microsomal aminopeptidase, CD13)
1834	M23197_at	CD33 CD33 antigen (differentiation antigen)
1995	M37033_at	CD53 CD53 antigen
2010	M38690_at	CD9 CD9 antigen
2014	M54992_at	CD72 CD72 antigen
2475	S72008_at	CDC10 Cell division cycle 10 (homologous to CDC10 of S. cerevisiae
3722	U77948_at	KAI1 Kangai 1 (suppression of tumorigenicity 6, prostate; CD82 antigen (R2 leukocyte antigen, antigen detected by monoclonal and antibody IA4))
4024	X01060_at	TFRC Transferrin receptor (p90, CD71)
4052	X04085_rna1_at	Catalase (EC 1.11.1.6) 5'flank and exon 1 mapping to chromosome 11, band p13 (and joined CDS)
4055	X04145_at	CD3G CD3G antigen, gamma polypeptide (TiT3 complex)
4142	X14046_at	CD37 CD37 antigen
4324	X59350_at	CD22 CD22 antigen
4376	X62573_at	FCGR2B Fc fragment of IgG, low affinity IIb, receptor for (CD32)
4377	X62654_rna1_at	ME491 gene extracted from H.sapiens gene for Me491/CD63 antigen
5122	Z32765_at	GB DEF = CD36 gene exon 15
5336	M24283_at	ICAM1 Intercellular adhesion molecule 1 (CD54), human rhinovirus receptor
5407	U36341_rna1_at	SLC6A8 gene (creatine transporter) extracted from Human Xq28 cosmid, creatine transporter (SLC6A8) gene, and CDM gene, partial cds
5543	D00749_s_at	T-CELL ANTIGEN CD7 PRECURSOR
5688	L33930_s_at	CD24 signal transducer mRNA and 3' region
6079	U59632_s_at	Cell division control related protein (hCDCrel-1) mRNA
6180	M16336_s_at	CD2 CD2 antigen (p50), sheep red blood cell receptor
6225	M84371_rna1_s_at	CD19 gene
6373	M81695_s_at	ITGAX Integrin, alpha X (antigen CD11C (p150), alpha polypeptide)
6405	M98399_s_at	CD36 CD36 antigen (collagen type I receptor, thrombospondin receptor)
6685	X89101_s_at	GB DEF = Fas (Apo-1, CD95)
964	HG4018-HT4288_at	Opioid-Binding Cell Adhesion Molecule
1928	M31303_rna1_at	Oncoprotein 18 (Op18) gene
640	D84557_at	P105MCM mRNA
5254	D38073_at	MCM3 Minichromosome maintenance deficient (S. cerevisiae) 3
5625	X62153_s_at	MCM3 Minichromosome maintenance deficient (S. cerevisiae) 3

Table 2. Principal components and correlated components for the gene expression data

Gene No.	Gene	PC1	PC2	PC3	PC4	PC5	PC6	CC1a	CC2a	CC1b	CC2b
1	D10495_at	-0.18	-0.17	0.12	-0.13	0.19	-0.21	-0.27	0.02	0.15	-0.15
2	D38524_at	0.10	0.08	0.10	0.33	0.14	0.02	0.09	0.06	-0.04	0.01
3	D84276_at	0.09	0.03	0.08	0.12	0.10	0.02	0.08	0.05	0.02	0.05
4	J04132_at	-0.03	0.30	-0.10	-0.11	0.10	0.12	0.06	-0.11	-0.04	0.14
5	M22324_at	-0.15	-0.16	-0.22	0.28	-0.31	0.06	-0.06	-0.15	-0.30	-0.21
6	M23197_at	-0.16	-0.17	-0.04	-0.01	0.16	0.04	-0.09	-0.14	-0.05	0.00
7	M37033_at	0.08	-0.03	0.09	-0.04	-0.06	0.05	0.09	0.02	0.00	0.07
8	M38690_at	0.42	-0.29	-0.50	0.07	0.06	-0.31	0.08	0.52	0.48	0.02
9	M54992_at	0.20	-0.10	0.60	-0.02	-0.11	-0.13	0.05	0.23	0.14	-0.04
10	S72008_at	0.08	0.08	0.03	0.09	0.16	0.01	0.06	0.05	0.05	0.06
11	U77948_at	0.11	0.07	0.20	0.14	-0.07	0.15	0.18	-0.02	-0.12	0.10
12	X01060_at	-0.19	-0.06	-0.03	0.36	-0.21	-0.25	-0.31	0.04	-0.12	-0.46
13	X04085_rna1_at	-0.12	-0.23	0.34	0.38	0.17	-0.08	-0.15	-0.03	-0.10	-0.20
14	X04145_at	-0.04	0.28	-0.09	-0.08	0.16	-0.08	-0.09	0.03	0.10	-0.01
15	X14046_at	-0.11	-0.08	0.01	-0.15	-0.21	0.57	0.33	-0.48	-0.46	0.33
16	X59350_at	0.29	0.03	0.09	-0.34	0.08	-0.17	‘0.09	‘0.33	0.41	0.16
17	X62573_at	-0.07	-0.05	-0.08	-0.05	-0.05	0.00	-0.05	-0.05	-0.03	-0.03
18	X62654_rna1_at	-0.13	-0.29	0.02	-0.03	0.42	0.23	0.07	-0.25	-0.07	0.24
19	Z32765_at	-0.10	-0.03	-0.03	-0.08	-0.04	-0.10	-0.14	‘0.00	0.04	-0.11
20	M24283_at	-0.14	-0.12	-0.08	-0.27	0.21	-0.09	-0.16	-0.04	0.16	0.01
21	U36341_rna1_at	-0.09	-0.05	-0.03	-0.01	-0.33	-0.04	-0.09	-0.04	-0.12	-0.17
22	D00749_s_at	-0.03	0.27	-0.04	-0.11	0.05	0.05	0.01	-0.05	0.00	0.07
23	L33930_s_at	0.54	-0.20	-0.11	0.17	0.08	0.43	0.68	0.08	-0.03	0.58
24	U59632_s_at	-0.06	-0.05	0.03	0.05	-0.08	0.11	0.04	-0.12	-0.15	0.01
25	M16336_s_at	-0.04	0.49	-0.15	0.35	0.16	0.03	0.00	-0.05	-0.13	-0.06
26	M84371_rna1_s_at	0.24	-0.07	0.02	-0.13	-0.22	-0.16	0.05	0.28	0.22	-0.01
27	M81695_s_at	-0.17	-0.04	-0.06	-0.15	-0.10	0.07	-0.07	-0.17	-0.11	-0.03
28	M98399_s_at	-0.17	-0.14	-0.07	0.05	0.20	-0.10	-0.19	-0.05	0.04	-0.11
29	X89101_s_at	-0.08	-0.04	-0.07	0.02	-0.27	-0.12	-0.14	0.03	-0.04	-0.22
30	HG4018-HT4288_at	-0.09	-0.07	-0.18	0.02	-0.01	0.10	0.01	-0.14	-0.12	0.02
31	M31303_rna1_at	0.09	0.14	0.00	-0.01	-0.07	-0.12	-0.02	0.15	0.11	-0.06
32	D84557_at	0.02	0.09	-0.06	0.03	0.05	-0.08	-0.04	0.07	0.07	-0.04
33	D38073_at	0.13	0.17	0.06	0.07	0.03	-0.10	0.02	0.16	0.11	-0.02
34	X62153_s_at	0.05	0.15	0.11	-0.11	-0.22	0.03	0.05	0.01	-0.04	0.00
	Variance	6.62	4.39	1.35	0.91	0.69	0.64	3.63	3.63	2.44	2.44
	% variance explained	36	24	7	5	4	4	25	25	17	17

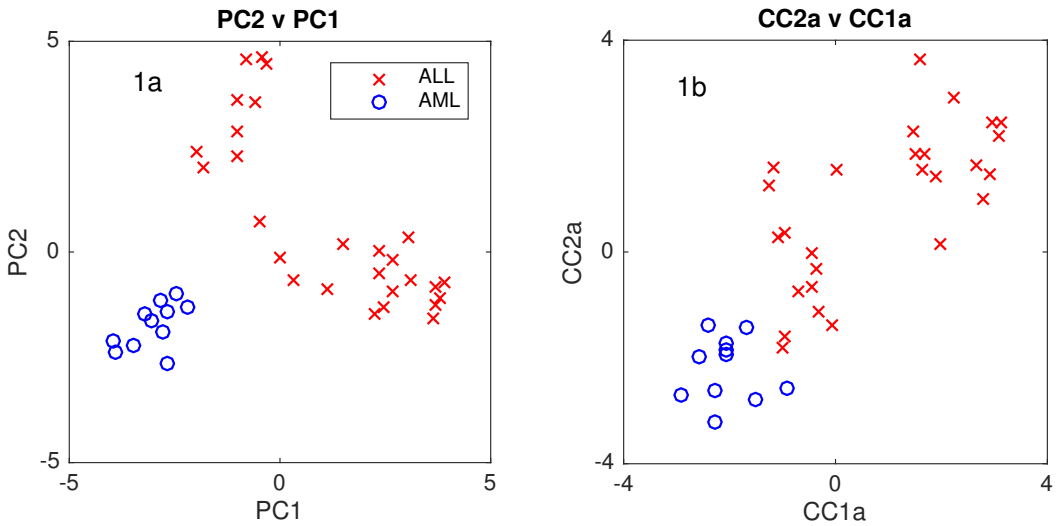

Figure 1. Principal components plot (a) and correlated components plot (b)

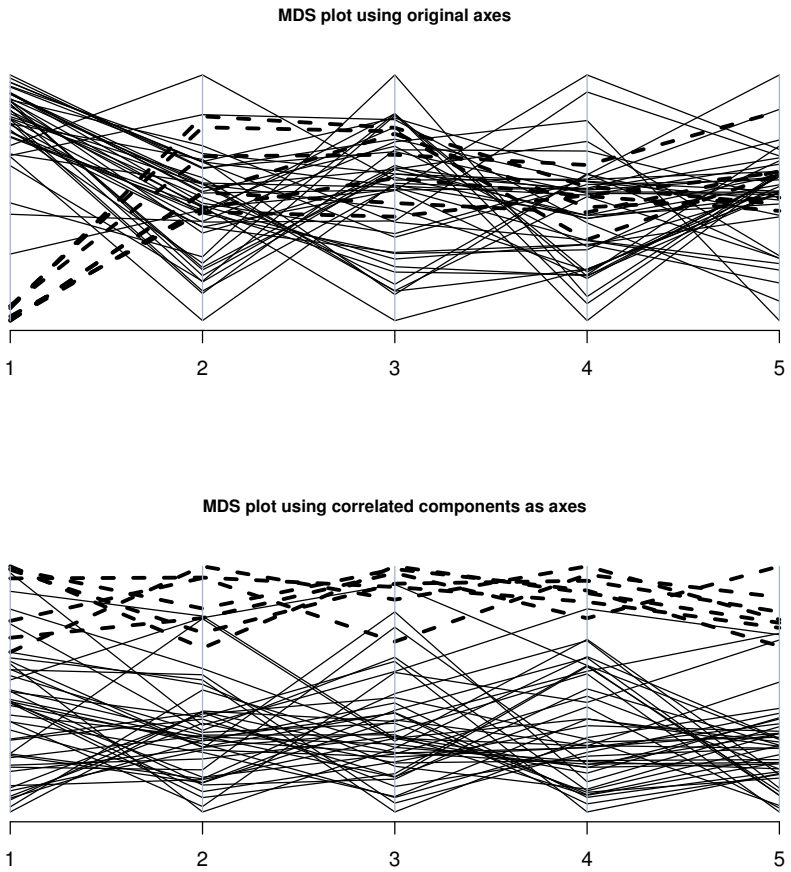

Figure 2. Parallel coordinates plots for nonmetric scaling of consumer data with and without rotation