Submitted to the Annals of Applied Statistics

ORTHOGONAL SIMPLE COMPONENT ANALYSIS

By Karim Anaya-Izquierdo, Frank Critchley and Karen Vines

The Open University

A new methodology to aid interpretation of a principal component analysis is presented. While preserving orthogonality, each eigenvector is replaced by a vector, close to it in angle terms, whose entries are small integers. We call such vectors simple. The approach is exploratory, a range of sets of pairwise orthogonal simple components being systematically obtained, from which the user may choose. Examples and simulations show that this approach effectively combines simplicity, retention of optimality and computational efficiency. Further, when the population eigenvectors have simple structure, this can be recovered using only the information from the sample. The methodology can also be used to simplify a given subset of components. An efficient algorithm for the computationally challenging problem of producing orthogonal components is provided.

1. Introduction. Principal components are optimal linear combinations of a set of variables with coefficients given by the eigenvectors of the covariance or correlation matrix ((8) chapters 2 and 3). To be useful in practice, such components often need interpretation in the context of the data studied. Unfortunately, the optimality of these principal components does not ensure they have a simple interpretation. They may possess optimal theoretical properties, but be of limited practical interest. This motivates replacing them by simpler, more interpretable, components, at the expense of a certain degree of optimality.

In a broad sense, simplicity means the appearance of nice structures in the loadings matrix $\mathbf{Q} = (\mathbf{q}_1|...|\mathbf{q}_k)$. Often, the scientist in charge of the study, would like to see clear-cut patterns in $\mathbf{Q}$ which help him or her to better understand the meaning of the components $\mathbf{q}_r^\top \mathbf{x}$ $(r = 1, ..., k)$ which it generates. Examples of nice structures include sparseness, the presence of simple weighted averages, contrasts and groups of variables. But simplicity inevitably implies loss of optimality, and it is the scientist in charge of the study who needs to calibrate the trade-off between simplicity and optimality.

The oldest approach to simplifying principal components is rotation (see (8), chapter 11 for an excellent review). Assuming normality, (12) recently proposed a penalised profile likelihood method, using varimax as the penalty function, which favours rotation of ill-defined components (those whose eigenvalues are close). In general, rotation methods lead to orthogonal simplified components defining new coordinate axes on which the data can be displayed, while total variance is preserved. However, the loadings involved are usually real numbers, which means that interpretation can still be difficult.

Another approach to simplification is to target sparsity. The presence of many zeros in $\mathbf{Q}$ can be useful for interpretation, for example when dealing with many variables. See, for example, (2), (3), (5), (11) and the references therein. One class of methods which targets sparseness is that based on the Least Absolute Shrinkage Selection Operator (LASSO). See, for example, (9) (15), (18), and (20). Although most of these methods lead to orthogonal simplified components,

AMS 2000 subject classifications: Primary 62H25
Keywords and phrases: simplified principal components, orthogonal integer loadings

combined with the presence of exact zero loadings, the remaining loadings are still real numbers, impeding interpretation.

A more explicitly modelling approach to simplification has recently been suggested by (14). Intrinsically restricted to principal component analysis of a correlation matrix, it assumes a particular form of that matrix inducing a pattern in its eigenstructure in which groups and contrasts of variables are forced to appear. For example, when all variables are positively correlated, the first eigenvector is always close to the average of the variables and consequently the remaining eigenvectors are basically contrasts. The loadings obtained are all proportional to integers, aiding interpretation. However, the components obtained need not be orthogonal.

(6), (16) and (19) suggest similar methods of simplification, in the sense that all three give orthogonal components with loading vectors proportional to integers. Hausman's method only allows the loadings to take the values -1, 0 or 1 and therefore is not always able to find a complete set of orthogonal vectors. In contrast, Vines' method produces loading vectors that are proportional to integers via a sequence of pairwise 'simplicity-preserving' transformations which ensure that orthogonality is maintained. However, although always proportional to integers, the size of the integers is not bounded and may at times be very large. A fuller discussion of the method, and its properties, can be found in (16).

The methodology we introduce in this paper produces the same clear structure as (19) for a data set with four different measurements of an index of resistance to flow in blood-vessels (17), as shown in Table 1:

Variable	$\mathbf{q}_1$	$\hat{\mathbf{z}}_1$	$\mathbf{q}_2$	$\hat{\mathbf{z}}_2$	$\mathbf{q}_3$	$\hat{\mathbf{z}}_3$	$\mathbf{q}_4$	$\hat{\mathbf{z}}_4$
Right Doppler	0.42	1	−0.32	−1	−0.58	−1	−0.62	−1
Left Doppler	0.43	1	0.30	1	−0.55	−1	0.65	1
Right CVI	0.55	1	−0.65	−1	0.43	1	0.30	1
Left CVI	0.58	1	0.63	1	0.42	1	−0.31	−1
% variance	58	57	25.9	23.8	9.5	10.5	6.5	8.6

TABLE 1

Principal component loadings for the resistance index data $\mathbf{q}_1, \ldots, \mathbf{q}_4$ *and the corresponding simplified loading vectors* $\hat{\mathbf{z}}_1, \ldots, \hat{\mathbf{z}}_4$

The simplified loadings are easier to understand than the continuous ones, so much so in fact that it looks like we have uncovered nature's design: a main effect, plus three orthogonal contrasts. The simplified components being orthogonal, the total variance is retained, being redistributed among the components so as to enhance interpretability. In particular, there is just a little loss in the variance explained by the first two components.

In this paper, we propose a methodology to obtain simple approximations to the vectors of loadings of the principal components with the following characteristics:

1. Proportionality to vectors of small integers.
2. Individual angle-closeness.
3. Pairwise orthogonality.

As examples and simulations will show, we can obtain good approximations in the sense that only small integers are used, while retaining closeness to the original components and exact orthogonality. The rationale behind the choice of these three characteristics is as follows:

1. Integers aid interpretation, vectors proportional to small integers being typically easier to interpret than vectors of real entries. Again, exact zeros and simple averages appear naturally.

2. By keeping each approximating vector close to its exact counterpart, we ensure that we do not lose potentially meaningful individual eigenvectors and that overall optimality is maximally retained.
3. Orthogonality also aids interpretation. Contrasts, simple relations between components and groups of variables may all appear as a direct consequence of using orthogonality combined with integer coefficients. Again, total variance is preserved, while simple components now define orthogonal axes, relative to which data can be plotted and interpreted.

However, interpretability is neither guaranteed nor amenable to precise mathematical formulation. This has two, key, methodological consequences. Whereas, as we will show by example, our approach can help in the vital step of interpretation, we do not expect any method to lead to interpretable results in all cases. Again, rather than attempt to find a unique optimal simplification in any pre-defined sense, we follow an exploratory approach which systematically produces a range of sets of pairwise orthogonal simple components, each close to its parent, amongst which the user can choose. Factors that can guide this choice include: (a) interpretability, simplicity and accuracy, as above, (b) subject matter considerations, particular to the context of the data set under analysis, and (c) suboptimality with respect to exact principal component analysis, including lower explanatory power in terms of the proportion of total variance, loss of focus on potential scientific laws (near constant relationships between the variables) and correlation. Only principal component analysis itself can give orthogonal loadings *and* uncorrelated components. So, as we require our simple loading vectors to be orthogonal, our components will always show some degree of correlation.

The space of all integer orthogonal matrices is huge, the combinatorial complexity growing both with the number of variables and the maximum size of integer allowed. However, the nature of our method allows efficient exploration of this vast space without restriction to any of its particular subsets, such as those determined by modelling assumptions. Although other approaches are possible, such as the one described in (1), the methodology proposed here relies on the fact that interpretation is possible using only the matrix of loadings.

The plan of the paper is as follows. Section 2 describes our methodology, discussing the ideas involved in detail and illustrating them with the help of a running example. Section 3 discusses two other examples. Section 4 presents a simulation study that shows the exact recovery properties of our method, while Section 5 discusses some possible extensions and other issues. Technical and computational details are given in an appendix.

2. Approximating eigenvectors.

2.1. *General setup.* Throughout this paper, the set of all integers and real numbers will be denoted by $\mathbb{Z}$ and $\mathbb{R}$ respectively, while $\mathbb{Z}^{(k)}$ and $\mathbb{R}^{(k)}$ denote the punctured product spaces $\mathbb{Z}^k\backslash\{(0,\ldots,0)^\top\}$ and $\mathbb{R}^k\backslash\{(0,\ldots,0)^\top\}$ respectively. For any $\mathbf{u} \in \mathbb{R}^{(k)}$, $\ell(\mathbf{u}) := \{c\,\mathbf{u} : c \in \mathbb{R}\}$ will be called the axis generated by $\mathbf{u}$.

Let $\mathbf{S}$ be a given $k \times k$ sample covariance or correlation matrix for a set of k commensurable variables. We assume $\mathbf{S}$ is nonsingular with k distinct eigenvalues, denoted $\lambda_1 > \ldots > \lambda_k > 0$. The overall sign of the unit-length eigenvector associated with each λ_i being intrinsically indeterminate, we denote *either* by $\mathbf{q}_i$. The eigenspace corresponding to each λ_i is then the axis $\alpha_i = \ell(\mathbf{q}_i)$, each of which we approximate by some other axis $\hat{\alpha}_i$. For any axis $\alpha = \ell(\mathbf{u})$ approximated by $\hat{\alpha} = \ell(\widehat{\mathbf{u}})$, the angle discrepancy

$$d(\alpha, \hat{\alpha}) := \arccos\left(\frac{|\mathbf{u}^\top \hat{\mathbf{u}}|}{\|\mathbf{u}\|\|\hat{\mathbf{u}}\|}\right)$$

defines a distance between them. The choice of this discrepancy is partly motivated by the fact that it does not depend on the choice of the nonzero vectors $\mathbf{u}$ and $\hat{\mathbf{u}}$ within their spans. For reporting purposes, we use the accuracy measure defined by $accu(\alpha, \hat{\alpha}) := \cos(d(\alpha, \hat{\alpha}))$ which takes values in $[0, 1]$.

	$\mathbf{q}_1$	$\mathbf{q}_2$	$\mathbf{q}_3$	$\mathbf{q}_4$	$\mathbf{q}_5$
Mechanics (Closed)	0.40	-0.65	0.62	0.15	-0.13
Vectors (Closed)	0.43	-0.44	-0.71	-0.30	-0.18
Algebra (Open)	0.50	0.13	-0.04	0.11	0.85
Analysis (Open)	0.46	0.39	-0.14	0.67	-0.42
Statistics (Open)	0.40	0.47	0.31	-0.66	-0.23
% variance	63.6	14.8	8.9	7.8	4.9

TABLE 2
Principal component loadings for the exams data.

To illustrate our methodology we use, as a running example, a data set consisting of the scores achieved by 88 students in 5 tests, a combination of open- and closed-book exams (10). Table 2 shows the unit length eigenvectors (rounded to 2 decimal places) of the sample correlation matrix for this data set. Thus, the first principal component is a weighted average of all the different subject scores and the other principal components can be interpreted as contrasts. However, more detailed interpretation of the principal components, particularly those other than the first, is not easy.

This section is organised as follows. Sections 2.2 and 2.3 describe how to obtain approximations to all the eigenvectors, while Section 2.4 shows how the approach can be adapted to situations in which only a subset of eigenvectors are of interest. Section 2.5 introduces three variants of our methodology, and some of their important features. Finally, in Section 2.6, we discuss how to proceed in practice and interpret the exams data based on our systematic, exploratory analysis.

2.2. *Description of the approximations.* We only consider approximations $\hat{\alpha}$ which are generated by integer vectors.

DEFINITION 1. *An axis $\hat{\alpha}$ is called simple if there exists $\mathbf{z} \in \mathbb{Z}^{(k)}$ such that $\hat{\alpha} = \ell(\mathbf{z})$. If the nonzero elements of $\mathbf{z}$ are coprime, then $\mathbf{z}$ is called an integer representation of $\hat{\alpha}$ and their maximum absolute value is called the complexity of $\hat{\alpha}$, denoted here by $compl(\widehat{\alpha})$.*

We approximate the $\{\alpha_r\}_{r=1}^k$ one at a time. The order in which we do this matters, different orders providing different approximations. Subsuming a permutation of $(1, \ldots, k)$, there is no loss of generality in describing our approximations $\{\widehat{\alpha}_r\}_{r=1}^k$ for a single, fixed order. Here, we fit the axes with larger variances first – that is, we work in increasing numerical order. Other possibilities are described in Section 2.5.

For each $r \in \{1, \ldots, k\}$, we construct an appropriate set $\mathcal{M}_r$ of simple axes, within which we then find the best approximation $\hat{\alpha}_r$ to α_r. The orthogonality of our approximating eigenvectors is guaranteed by the way we construct $\mathcal{M}_r$ at each step. In the approximation of the first axis there is no orthogonality restriction, so we take $\mathcal{M}_1$ to be set of all simple axes in $\mathbb{R}^k$. For the exams data, $\mathcal{M}_1$ is the set of all axes generated by vectors in $\mathbb{Z}^{(5)}$. Thus an approximation to α_1 could be the simple axis generated by $\hat{\mathbf{z}}_1 = (1, 1, 1, 1, 1)^\top$.

For $2 \leq r \leq k-1$, we take $\mathcal{M}_r$ to be the set of all simple axes which are orthogonal to $\hat{\alpha}_1, \ldots, \hat{\alpha}_{r-1}$. Let $\mathbf{O}_{r-1}$ be a matrix whose columns are integer vectors, orthogonal to $\hat{\alpha}_1, \ldots, \hat{\alpha}_{r-1}$.

To ensure our approximating axes are simple, we use linear combinations of the columns of $\mathbf{O}_{r-1}$ with integer coefficients to generate $\mathcal{M}_r$. It is always possible to find such a matrix $\mathbf{O}_{r-1}$ (see Appendix A). For the exams data, if we have $\hat{\alpha}_1 = \ell((1,1,1,1,1)^\top)$, a suitable $\mathbf{O}_1$ is

$$\mathbf{O}_1 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Thus, for example, if we take $\mathbf{y}^\top = (-1,0,1,1)$, then

$$\hat{\mathbf{z}}_2 = \frac{1}{\mathrm{hcf}(|\mathbf{O}_1\,\mathbf{y}|)}\,\mathbf{O}_1\,\mathbf{y} = (1,1,0,-1,-1)^\top$$

is a member of $\mathcal{M}_2$, where $\mathrm{hcf}(|\mathbf{z}|)$ denotes the highest common factor of the nonzero absolute values of the elements of $\mathbf{z}$ (dividing by it ensures the corresponding values for $\hat{\mathbf{z}}_2$ are coprime).

Geometrically, it is clear that the angle-closest axis to α which is orthogonal to $\hat{\alpha}_1, \ldots, \hat{\alpha}_{r-1}$ is its projection onto the orthogonal complement of their span, restricting the maximum accuracy achievable. For, if $\mathbf{q}_r^\perp$ is the orthogonal projection of the unit vector $\mathbf{q}_r$ onto $\mathcal{M}_r$, for any approximation $\hat{\alpha} \in \mathcal{M}_r$,

$$\begin{aligned} accu(\alpha_r, \hat{\alpha}) &= accu(\alpha_r, \ell(\mathbf{q}_r^\perp))\, accu(\ell(\mathbf{q}_r^\perp), \hat{\alpha}) \\ &= \|\mathbf{q}_r^\perp\|\, accu(\ell(\mathbf{q}_r^\perp), \hat{\alpha}) \end{aligned}$$

so that $accu(\alpha_r, \hat{\alpha}) \le \|\mathbf{q}_r^\perp\|$, equality holding if and only if $\hat{\alpha} = \ell(\mathbf{q}_r^\perp)$ (which requires $\ell(\mathbf{q}_r^\perp)$ to be simple). Thus, over $\mathcal{M}_r$, not every possible accuracy is achievable for α_r, although no such upper bound applies to $accu(\ell(\mathbf{q}_r^\perp), \hat{\alpha})$.

For the exams data with $\hat{\alpha}_1 = \ell((1,1,1,1,1)^\top)$, the projection of $\mathbf{q}_2$ onto the orthogonal complement of $\hat{\alpha}_1$ is $\mathbf{q}_2^\perp = (-0.63, -0.42, 0.15, 0.41, 0.49)^\top$ (to 2 decimal places). Since $\|\mathbf{q}_2^\perp\| = 0.999$, there is no approximation to α_2 orthogonal to $\hat{\alpha}_1$ which can achieve an accuracy bigger than this. In particular, $\hat{\alpha}_2 = \ell((1,1,0,-1,-1)^\top)$ has an accuracy of 0.973 with respect to α_2, while its accuracy with respect to $\ell(\mathbf{q}_2^\perp)$ is given by $accu(\alpha_2, \hat{\alpha}_2)/\|\mathbf{q}_2^\perp\| = 0.973/0.999 \approx 0.974$. Similar information for other axes is given in Table 3.

Any given angle $\theta \in (0, \pi/2)$ defines a 'cone' around α_r in $\mathcal{M}_r$ of fixed complexity $N \in \{1, 2, \ldots\}$ via

$$C_r(\theta, N) := \{\hat{\alpha} \in \mathcal{M}_r \,:\, accu(\hat{\alpha}, \alpha_r) \ge \cos(\theta)\,,\ compl(\hat{\alpha}) = N\}.$$

If $\cos(\theta) > \|\mathbf{q}_r^\perp\|$, then $C_r(\theta, N)$ is empty for all N, so we always take $\cos(\theta) \le \|\mathbf{q}_r^\perp\|$. For given θ, we are interested in the nonempty cone with smallest complexity, namely $C_r^*(\theta) := C_r(\theta, N_r(\theta))$, where

$$N_r(\theta) := \min\{N \ge 1 \,:\, C_r(\theta, N) \ne \emptyset\}. \tag{2.1}$$

Note that $N_r(\theta) = 1 \Leftrightarrow C_r(\theta, 1) \ne \emptyset$. Thus, $C_r^*(\theta)$ comprises all those axes with minimal complexity subject to being θ-close to α_r. For given θ, we take as our approximation to α_r the closest among all the axes in $C_r^*(\theta)$.

DEFINITION 2. *Let θ be such that $0 < \cos(\theta) \leq \|\mathbf{q}_r^{\perp}\|$. The best simple θ approximation to $\alpha_r = \ell(\mathbf{q}_r)$ orthogonal to $\hat{\alpha}_1, \ldots, \hat{\alpha}_{r-1}$ is the axis $\hat{\alpha}_r(\theta) = \ell(\mathbf{O}_{r-1}\mathbf{y})$ for some $\mathbf{y} \in \mathbb{Z}^{(k-r+1)}$ which maximises $accu(\hat{\alpha}, \alpha_r)$ among all simple axes $\hat{\alpha}$ in the cone $C_r^*(\theta)$. Either of the two possible integer representations of $\hat{\alpha}_r(\theta)$ will be denoted by $\hat{\mathbf{z}}_r(\theta)$.*

In practice, we require $\theta \leq \pi/4$, as larger values of θ will clearly give poor approximations. Thus, overall, we have the following bounds on the accuracy attained:

$$\cos(\pi/4) \leq \cos(\theta) \leq accu(\alpha_r, \hat{\alpha}_r(\theta)) \leq \|\mathbf{q}_r^{\perp}\|. \tag{2.2}$$

We call $\cos(\theta)$ the *minimum accuracy required* for the approximation $\hat{\alpha}_r$. We compute all the approximations $\hat{\alpha}_1(\theta), \ldots, \hat{\alpha}_k(\theta)$ using the same value of the minimum accuracy required. We denote the complete set of approximations for a given value of θ by $\hat{A}_{1:k}(\theta) := (\hat{\alpha}_1(\theta), \ldots, \hat{\alpha}_k(\theta))$. To measure the overall closeness of $\hat{A}_{1:k}(\theta)$ to $(\alpha_1, \ldots, \alpha_k)$, we use the root mean square of the accuracies attained $\{accu(\alpha_r, \hat{\alpha}_r(\theta))\}_{r=1}^{k}$.

Finding $N_r(\theta)$ and then subsequently $\hat{\alpha}_r(\theta)$ for $r < k$ can be hard combinatorial optimisation problems, especially when the dimension is large. We therefore use approximations in such cases to reduce the combinatorial complexity. We briefly describe such approximations in Appendix B.

For the exams data, suppose we take $\hat{\alpha}_1 = \ell((1,1,1,1,1)^{\top})$. Then the cone $C_2(\pi/4, 1)$ has many elements, including $\ell((1,1,0,-1,-1)^{\top})$ and $\ell((1,0,0,0,-1)^{\top})$. Of these, we prefer the first, their accuracies being 0.973 and 0.789 respectively. In fact, it can be shown that $\ell((1,1,0,-1,-1)^{\top})$ is the best simple $\theta = \pi/4$ approximation to α_2 orthogonal to $\ell((1,1,1,1,1)^{\top})$.

Note that for the first axis we trivially have $\mathbf{q}_1^{\perp} = \mathbf{q}_1$, so that $\|\mathbf{q}_1^{\perp}\| = 1$. For the exams data we have $N_1(\pi/4) = 1$ and, furthermore, out of all axes of complexity one, $\ell((1,1,1,1,1)^{\top})$ is the closest axis to α_1. Therefore, $\hat{\mathbf{z}}_1 = (1,1,1,1,1)^{\top}$ is the integer representation of the best simple $\theta = \pi/4$ approximation to α_1.

Variable	$\hat{\mathbf{z}}_1(\theta)$	$\hat{\mathbf{z}}_2(\theta)$	$\hat{\mathbf{z}}_3(\theta)$	$\hat{\mathbf{z}}_4(\theta)$	$\hat{\mathbf{z}}_5(\theta)$
Mechanics (Closed)	1	-1	1	0	-1
Vectors (Closed)	1	-1	-1	0	-1
Algebra (Open)	1	0	0	0	4
Analysis (Open)	1	1	0	-1	-1
Statistics (Open)	1	1	0	1	-1
Accuracy	0.997	0.973	0.938	0.937	0.974
RMS	0.96				
max accuracy $\|\mathbf{q}_r^{\perp}\|$	1	0.999	0.99	0.95	0.97
% variance	63.3	14.4	8.9	7.9	5.5

TABLE 3
Integer representations for the examinations data with $\theta = \pi/4$.

When $r = k$ there is nothing to optimise, because there is a unique simple axis orthogonal to $\hat{\alpha}_1, \ldots, \hat{\alpha}_{k-1}$ (see Appendix A). We are therefore obliged to take this axis as $\hat{\alpha}_k(\theta)$, having no control over its accuracy. However, if $\hat{\alpha}_r(\theta)$ is close to to α_r for all $r < k$, then $\hat{\alpha}_k(\theta)$ is usually close to α_k. We can see an example of this in Table 3, which shows the integer representations of $\hat{A}_{1:k}(\pi/4)$ for the exams data along with the corresponding accuracies attained and their upper bounds.

In general, the complexity of $\hat{\alpha}_r(\theta)$ tends to grow as r increases. This happens simply because it becomes harder for low-complexity axes to satisfy the orthogonality restrictions regardless of the accuracy required. However, it is always possible to approximate with reasonably high accuracy a

single k-dimensional axis $\ell(\mathbf{q})$ with a simple axis of low complexity. Figure 1 shows the empirical distribution (based on 10000 independent replications) of the minimum complexities $N_1(\theta)$ for different values of k and $\cos(\theta)$ when $\ell(\mathbf{q})$ is sampled from the uniform distribution over the set of all possible axes (see (4).) Clearly, without orthogonality restrictions, accurate approximations of axes tend not to be very complex.

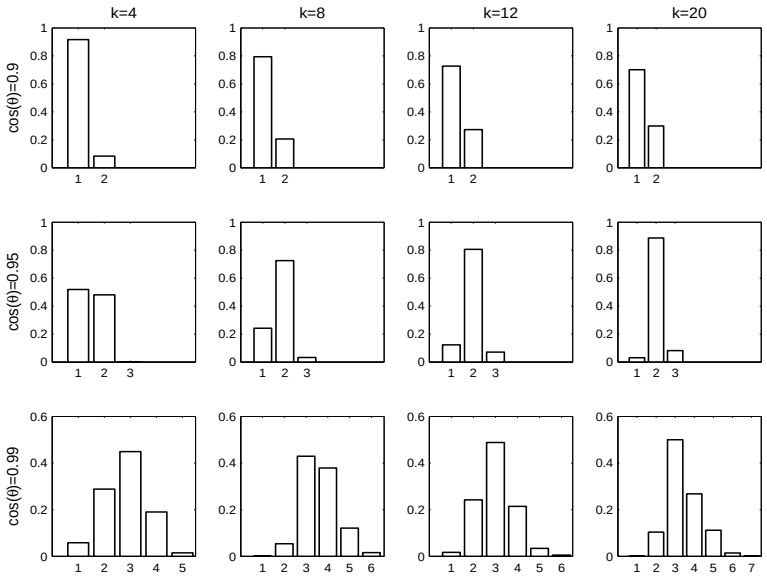

FIG 1. *Empirical distribution of minimum complexities $N_1(\theta)$ in simple approximations to k-dimensional axes.*

Nevertheless, there is a clear trade-off between accuracy and complexity. Highly accurate approximations usually have high complexity, making the interpretation of the axes more difficult. This is why we choose to control accuracy and then minimise complexity, so our method is *biased towards simplicity.*

2.3. *Effect of varying the minimum accuracy required.* When $\theta = \pi/4$ the approximations $\hat{A}_{1:k}(\theta)$ typically have low complexity overall and so can usually be interpreted. Unless all the eigenspaces are already simple, we might expect the overall complexity of the approximations to steadily increase with the minimum accuracy required. However, it turns out there is no straightforward relationship between the complexity of the approximations and $\cos(\theta)$. Further, although less complex axes tend to lead to more interpretable axes, this is not always the case. So, instead of attempting to find an optimal value of $\cos(\theta)$ under some criterion, we vary the value of θ to explore the different approximations obtained. This offers different explanations of the same data set and so gives more information for interpretation.

The good news is that it is only necessary to explore a discrete set of values of θ. This is because for a given θ', the same approximations will result for any θ such that

$$\cos(\theta') \le \cos(\theta) \le \min_{r\in\{1,\ldots,k-1\}} accu(\alpha_r, \hat{\alpha}_r(\theta')).$$

Thus to *fully* explore the range of possible approximations it is sufficient to consider the sequence of angles $\theta^{[1]}, \theta^{[2]}, \ldots$ where $\cos(\pi/4) = \cos(\theta^{[1]}) < \cos(\theta^{[2]}) < \cdots$ with

$$\cos(\theta^{[n+1]}) = \min_{r\in\{1,\ldots,k-1\}} accu(\alpha_r, \hat{\alpha}_r(\theta^{[n]})). \tag{2.3}$$

For $n = 1, 2, \ldots$, we abbreviate $\hat{\mathbf{z}}_r(\theta^{[n]})$ to $\hat{\mathbf{z}}_r^{[n]}$, denoting also by $RMS^{[n]}$ the root mean square accuracy of $\hat{A}_{1:k}(\theta^{[n]})$.

For the exams data, $\cos(\theta^{[2]}) = accu(\alpha_4, \hat{\alpha}_4(\theta^{[1]})) = 0.937$ (see Table 3). The lefthand portion of Table 4 shows the integer representations for the last three fitted axes in $\hat{A}_{1:k}(\theta^{[2]})$. Only the approximations to the fourth and fifth axes change, but they do so quite dramatically: the complexities increase substantially, losing the simple structure for the corresponding components we had in $\hat{A}_{1:k}(\theta^{[1]})$. The right-hand portion of Table 4 shows the integer representations of the fitted axes which change under the next set of approximations $\hat{A}_{1:k}(\theta^{[3]})$, where $\cos(\theta^{[3]}) = accu(\alpha_3, \hat{\alpha}_3(\theta^{[2]})) = 0.938$. The increase in minimum accuracy required is very small, but this is enough to drop down the complexities in the last two components back to a much lower level. The approximations are now almost as simple as those for $\hat{A}_{1:k}(\theta^{[1]})$, but more accurate. The next sets of approximations, starting with $\cos(\theta^{[4]}) = 0.973$ and up to $\cos(\theta) = 0.995$, are very complex with the complexities of the order of hundreds indicating we have reached a good balance between accuracy and complexity in $\hat{A}_{1:k}(\theta^{[3]})$.

Variable	$\hat{\mathbf{z}}_3^{[2]}$	$\hat{\mathbf{z}}_4^{[2]}$	$\hat{\mathbf{z}}_5^{[2]}$	$\hat{\mathbf{z}}_3^{[3]}$	$\hat{\mathbf{z}}_4^{[3]}$	$\hat{\mathbf{z}}_5^{[3]}$
Mechanics (C)	1	1	13	2	1	1
Vectors (C)	-1	1	13	-2	-1	1
Algebra (O)	0	-4	-52	0	0	-4
Analysis (O)	0	-12	23	-1	2	1
Statistics (O)	0	14	3	1	-2	1
Accuracy	0.938	0.941	0.978	0.980	0.979	0.974
RMS	0.97			0.98		
% variance	8.9	7.9	5.5	8.9	7.8	5.5

TABLE 4

Integer representations for the exams data with $\cos(\theta^{[2]}) = 0.937$ *and* $\cos(\theta^{[3]}) = 0.938$. $\hat{\mathbf{z}}_1^{[n]}$ *and* $\hat{\mathbf{z}}_2^{[n]}$ *for* $n = 2, 3$ *are the same as* $\hat{\mathbf{z}}_1$ *and* $\hat{\mathbf{z}}_2$ *in Table 3 respectively.*

This non-monotone behaviour of the approximations $\hat{A}_{1:k}(\theta)$ when $\cos(\theta)$ increases is due to the discreteness inherent in our approximations. Restricting the elements of the integer representations to be coprime is mainly responsible for this and makes it difficult to fully understand the behaviour of the sequence of approximations. However, we can clearly explain key features of the relation between consecutive approximations $\hat{A}_{1:k}(\theta^{[n]})$ and $\hat{A}_{1:k}(\theta^{[n+1]})$, as follows. Let

$$u_n := \arg\min_{i \in \{1,\ldots,k-1\}} accu(\alpha_i, \hat{\alpha}_i(\theta^{[n]})) \tag{2.4}$$

indicate the first approximation which changes from n to $n+1$. The first approximated axes do not change; specifically, $\hat{\alpha}_r(\theta^{[n+1]}) = \hat{\alpha}_r(\theta^{[n]})$ for $r = 1, \ldots, u_n - 1$. Furthermore, because of the orthogonality restrictions, we have

$$\text{span}\{\hat{\mathbf{z}}_i(\theta^{[n+1]}) \,:\, i = u_n, \ldots, k\} = \text{span}\{\hat{\mathbf{z}}_i(\theta^{[n]}) \,:\, i = u_n, \ldots, k\}.$$

That is, the subspace generated by the remaining axes is the same for $\hat{A}_{1:k}(\theta^{[n+1]})$ as for $\hat{A}_{1:k}(\theta^{[n]})$. We are just obtaining different orthogonal bases for the same subspace. In the exams data, with $\cos(\theta^{[3]}) = 0.938$, $u_2 = 3$ and thus moving from $\hat{A}_{1:k}(\theta^{[2]})$ to $\hat{A}_{1:k}(\theta^{[3]})$ only approximations to axes α_3, α_4 and α_5 change and the corresponding spanned subspace remains the same. Clearly, $\hat{A}_{1:k}(\theta^{[3]})$ provides a much simpler basis for this subspace than $\hat{A}_{1:k}(\theta^{[2]})$.

2.4. *Targeting a subset of eigenvectors.* Often, interest is focused on just a subset of the eigenvectors, typically the first few dominant ones. Again, subsuming a permutation, there is no loss of generality in considering this case.

If we want to simplify the first $k' < k$ eigenspaces $\{\alpha_1, \ldots, \alpha_{k'}\}$, we proceed as above by simplifying this subset only and varying the minimum accuracy required for this case, which we denote now by $\cos(\vartheta)$. The range of possible approximations, denoted by $\hat{A}_{1:k'}(\vartheta)$, will be determined by the sequence of angles $\vartheta^{[1]}, \vartheta^{[2]}, \ldots$ defined by

$$\vartheta^{[1]} = \pi/4,\ \cos(\vartheta^{[n+1]}) = \min_{r \in \{1,\ldots,k'\}} accu(\alpha_r, \hat{\alpha}_r(\vartheta^{[n]}))\ (n \geq 1). \tag{2.5}$$

This is a subsequence of the full sequence (2.3) where we skip the approximations $\hat{A}_{1:k}(\theta^{[n]})$ for which the target eigenvectors do not change. Note that the last approximation computed in $\hat{A}_{1:k'}(\vartheta)$ is not determined by those already fitted, as it was in the complete case $k' = k$.

For example, if we are interested only in the first $k' = 3$ eigenspaces of the exams data, then $\hat{A}_{1:3}(\vartheta^{[1]})$ is given by the first three columns in Table 3. Also from Table 3 it follows that $\cos(\vartheta^{[2]}) = 0.938$, so the approximations in $\hat{A}_{1:3}(\vartheta^{[2]})$ are given by $\ell(\hat{\mathbf{z}}_1^{[1]}), \ell(\hat{\mathbf{z}}_2^{[1]})$ and $\ell(\hat{\mathbf{z}}_3^{[3]})$. It is not necessary to calculate $\hat{A}_{1:5}(\theta^{[2]})$ to obtain $\hat{A}_{1:3}(\vartheta^{[2]})$.

2.5. *Variants of the basic approximation method.*

2.5.1. *Description of the variants.* So far, approximations to the axes $\alpha_1, \ldots, \alpha_k$ (or just $\alpha_1, \ldots, \alpha_{k'}$) have been obtained in that order. That is $\hat{\alpha}_2$ is obtained after $\hat{\alpha}_1$, $\hat{\alpha}_3$ is obtained after $\hat{\alpha}_2$, and so on. Changing the order in which these approximations are found leads to variants of the basic method. Use of different variants, apart from giving more information for interpretation, helps find strong easily interpretable signals in the data. This is because such signals would be expected to throw up the same low complexity approximation regardless of the variant used. We consider first approximation of all k axes, so that the last fitted axis is determined by its $k-1$ predecessors.

Let permutation $(p(1), \ldots, p(k))$ of $(1, \ldots, k)$ represent the order in which axes $\alpha_1, \ldots, \alpha_k$ are approximated. In total, we will consider four variants:

Forwards: $(p(1), p(2), \ldots, p(k)) = (1, 2, \ldots, k)$. This is the order of calculation we have used so far. It corresponds to following the decreasing order of the magnitudes of the eigenvalues of $\mathbf{S}$ and so prioritises approximations to the linear combinations with the larger variances.

Next-best Forwards: The next-best variant makes the accuracy upper bound in (2.2) as large as possible. In contrast with the forwards variant, this approach prioritises the potentially most accurately approximated axes. Initially, all the approximations have the same accuracy upper bound (this being 1), so we choose first to approximate the eigenspace with largest eigenvalue, i.e. $p(1) = 1$. For $2 \leq r \leq k-1$ we then set

$$p(r) = \arg \max_{j \in \{1,\ldots,k\} \setminus \{p(1),\ldots,p(r-1)\}} \|\mathbf{q}_j^{\perp}\|. \tag{2.6}$$

That is, the next eigenvector to approximate is the one which has the largest upper bound on its accuracy. (Note that there is no guarantee that the actual accuracy achieved when approximating $\alpha_{p(r)}$ will be higher than by an approximation to any other of the remaining axes, a more accurate approximation being possible for an axis with a lower maximum accuracy.)

Backwards: The order of calculation of the approximations in this variant is $(p(1), \ldots, p(k)) = (k, k-1, \ldots, 1)$, the reverse of that in the forwards variant. The aim here is to prioritise approximations to eigenvectors representing near constant linear relationships in the variables.

Next-best Backwards: This variant is the same as the next-best forwards variant but starting by simplifying the last eigenvector, that is, $p(1) = k$. For $r = 2, \ldots, k-1$ the rule for the following eigenvectors to simplify is again given by (2.6).

Overall, for many data sets, we may expect the two forwards variants to give approximations that are more similar to each other than to those produced by the two backwards variants, and conversely.

In an obvious notation, $M \in \{F, NF, B, NB\}$ denotes which variant is used. For each $n = 1, 2, \ldots$, $\theta_M^{[n]}$ denotes the angle determining the n^{th} minimum accuracy required, and $\hat{A}_{1:k}^{[n,M]}$ the corresponding set of approximations obtained (arranged back in the original order given by the indices in $\{\alpha_1, \ldots, \alpha_k\}$). Again, we abbreviate the integer representation $\hat{\mathbf{z}}_r(\theta_M^{[n]})$ to $\hat{\mathbf{z}}_r^{[n,M]}$, denoting by $RMS_M^{[n]}$ the corresponding root mean square accuracy attained.

2.5.2. *Applying the variants to a subset of eigenvectors.* When interest is focused on a subset of eigenspaces, the two backwards variants lose their 'constant linear relationship' rationale. Nevertheless, maintaining an exploratory spirit, all four variants can still be used.

If interest is focused in simplifying a subset of size $k' < k$ of $\{\alpha_1, \ldots, \alpha_k\}$ we assume the indices for the target eigenspaces have been relabeled so that the subset is $\{\alpha_1, \ldots, \alpha_{k'}\}$. Otherwise, we use the same notation as above, replacing k by k' and $\theta_M^{[n]}$ by $\vartheta_M^{[n]}$.

Variable	$\hat{\mathbf{z}}_1^{[1]}$	$\hat{\mathbf{z}}_2^{[1]}$	$\hat{\mathbf{z}}_3^{[1]}$
Mechanics (C)	1	-1	1
Vectors (C)	1	0	-1
Algebra (O)	1	0	0
Analysis (O)	1	1	0
Statistics (O)	0	1	1
Accuracy	0.896	0.868	0.946
RMS		0.904	
% variance	53.2	17.5	12

TABLE 5

Integer representations for the exams data when targeting only the first three eigenspaces for $\vartheta^{[1]} = \pi/4$ when using the backwards or next-best backwards variants.

For the exams data with $\vartheta^{[1]} = \pi/4$, if we are only interested in the first three eigenspaces $\{\alpha_1, \alpha_2, \alpha_3\}$, the forwards and next-best forwards variants give the same approximations as the first three columns in Table 3. For the backwards and next-best backwards variants, the approximations are as in Table 5. These approximations are different to approximations given by the backwards variants in the case $k' = k$. Thus, the simplification obtained depends not only on the variant used, but also on the subset of eigenvectors chosen. This is to be expected, as we are concentrating on a specific subspace which has been preselected for some reason. The problem of choosing a subset of principal components to focus on can be a difficult one (see (8) for details) and we always assume this target subspace is given.

2.6. *Exploratory analysis for the exams data.* Pursuing our exploratory approach, the four variants of our methodology systematically provide a range of possible explanations of a given data

set, corresponding to its projections onto different sets of orthogonal simple axes. As noted in the introduction, these explanations can vary on many factors, including interpretability, simplicity, accuracy, subject matter considerations and suboptimality criteria. Accordingly, different users can be lead to make different choices.

In this case, for the complete set of eigenvectors derived from the exams data, we suggest the following shortlist:

Forwards: We feel that approximation $\hat{A}_{1:5}^{[1,F]}$ (Table 3) is the simplest and most interpretable obtained using this variant. The overall accuracy attained is quite high, $RMS_F^{[1]}$ being 0.96. In particular, the percentage of variance explained by the first three components (86.6%) is close to the corresponding optimal percentage of variance explained (87.3%). An almost equally good approximation is $\hat{A}_{1:5}^{[3,F]}$ (Table 4), with $RMS_F^{[3]} = 0.98$, while the variance explained by the first three components is virtually identical under both approximations. The difference between both approximations is just the basis for the space spanned by the last 3 components.

Overall, it seems that $\hat{A}_{1:5}^{[1,F]}$ has the edge, its interpretation being particularly straightforward. Algebra turning out to behave differently from the other two open-book exams, Analysis and Statistics, it is helpful here to use 'open' to refer to this latter pair alone. The simple components explain the total variability in the data – in other words, they answer the question: 'why do students perform differently in the 5 exams?' – as follows:

$\hat{\alpha}_1$**:** Differences in overall mathematical ability.

$\hat{\alpha}_2$**:** Differences between open- and closed-book exam performances.

$\hat{\alpha}_3$**:** Differences between closed-book exam performances.

$\hat{\alpha}_4$**:** Differences between open-book exam performances.

$\hat{\alpha}_5$**:** Differences between performance in Algebra and all other subjects.

Next-best Forwards: It appears the best approximation for this variant is $\hat{A}_{1:5}^{[1,NF]}$, for which the order in which the approximations were obtained is $(p(1), \ldots, p(5)) = (1, 4, 2, 5, 3)$. Now, $\hat{A}_{1:5}^{[1,NF]}$ happens to coincide with $\hat{A}_{1:5}^{[1,F]}$ (Table 3). This is evidence of a strong signal in the data. For example, the approximations to α_4 and α_5 obtained by $\hat{A}_{1:5}^{[1,NF]}$ are the same as those obtained by $\hat{A}_{1:5}^{[1,F]}$ despite some orthogonality restrictions being dropped, similar remarks applying to α_2 and α_3 on reversing the roles of the variants.

A close contender is $\hat{A}_{1:5}^{[3,NF]}$, where $\cos(\theta_{NF}^{[3]}) = 0.959$. In fact, $\hat{A}_{1:5}^{[3,NF]}$ coincides with $\hat{A}_{1:5}^{[3,F]}$, discussed above, again signalling structure in the data. Using a minimum accuracy of $\cos(\theta_{NF}^{[2]}) = 0.937$ we obtain approximations (not shown) for which the interpretation in the approximated component 5 is quite different, and in our opinion worse, compared to $\hat{A}_{1:5}^{[1,NF]}$ and $\hat{A}_{1:5}^{[3,NF]}$. The remaining approximations under this method are rather complex overall and so are not discussed further.

Backwards: Approximations $\hat{A}_{1:5}^{[3,B]}$ and $\hat{A}_{1:5}^{[4,B]}$ (Table 6) are the simplest and more interpretable under this variant. There is a small gain in accuracy between them as $RMS_B^{[3]} = 0.95$ and $RMS_B^{[4]} = 0.96$. $\hat{A}_{1:5}^{[3,B]}$ and $\hat{A}_{1:5}^{[4,B]}$ are not as simple as the approximations given by the forwards variants, although their interpretation is similar except for the last component which, interestingly, contrasts Algebra with the other two open-book exams.

Next-best Backwards: Approximations $\hat{A}_{1:5}^{[1,NB]}$, $\hat{A}_{1:5}^{[2,NB]}$ and $\hat{A}_{1:5}^{[3,NB]}$ have a quite different and more complicated interpretation than all the previous ones. The best next-best backwards

Var	$\hat{\mathbf{z}}_1^{[3,B]}$	$\hat{\mathbf{z}}_2^{[3,B]}$	$\hat{\mathbf{z}}_1^{[4,B]}$	$\hat{\mathbf{z}}_2^{[4,B]}$	$\hat{\mathbf{z}}_3^{[4,B]}$	$\hat{\mathbf{z}}_4^{[4,B]}$	$\hat{\mathbf{z}}_5^{[4,B]}$
Mechanics	3	-1	3	-2	1	0	0
Vectors	3	-1	3	-2	-1	0	0
Algebra	2	1	4	1	0	0	2
Analysis	2	1	4	1	0	1	-1
Statistics	2	1	4	1	0	-1	-1
Accuracy	0.966	0.928	0.996	0.953	0.938	0.937	0.959
RMS	0.95		0.96				
% variance	60.2	17.3	63.2	14.3	8.9	7.9	5.7

TABLE 6
Integer representations for the examinations data with $\cos(\theta^{[3]}) = 0.898$ *and* $\cos(\theta^{[4]}) = 0.928$. $\hat{\mathbf{z}}_r^{[3,B]} = \hat{\mathbf{z}}_r^{[4,B]}$ *for* $r = 3, 4, 5$.

solution we obtained was $\hat{A}_{1:5}^{[4,NB]}$. This turned out to coincide with $\hat{A}_{1:5}^{[2,NF]}$, despite the order of calculation in those approximations being different, ($(5, 1, 4, 3, 2)$ and $(1, 4, 5, 3, 2)$ respectively), again indicating clear structure in the data.

To summarise: for this example, the forwards variants are overall better than the backwards ones, which may be due to all observed correlations being positive. There is evidence of strong signals in the data. Overall, if we were to choose a single approximation, we would suggest $\hat{A}_{1:5}^{[1,F]}$ as best combining simplicity, accuracy and subject matter interpretability, without introducing undue correlation. Figure 2 shows the scatterplot matrix for our preferred simple components. We can see visually there is no great correlation induced by this simplification – in fact the largest correlations between the simple components are about 0.2 and 0.19 between the first and the fifth and fourth respectively. The scatterplot is also visually close to the one given by the exact principal component analysis.

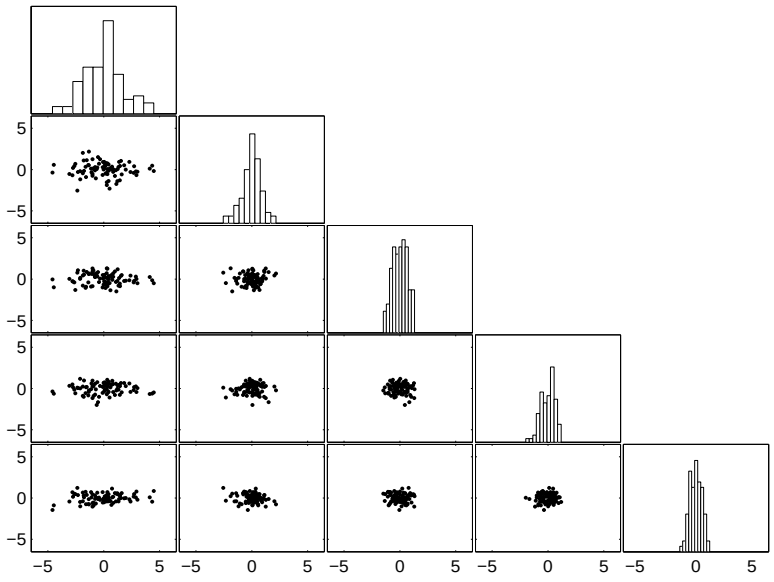

FIG 2. *Scatterplot matrix of the simplified principal components for the Exams Data.*

3. Further examples.

3.1. *Reflexes data.* The reflexes data are taken from Section 3.8.1 of (8). The data consist of reflex measurements for 143 individuals. On each individual left and right reflexes were recorded of five parts of the body: biceps, triceps, wrists, knees and ankles.

Variable	$\mathbf{q}_1$	$\mathbf{q}_2$	$\mathbf{q}_3$	$\mathbf{q}_4$	$\mathbf{q}_5$	$\mathbf{q}_6$	$\mathbf{q}_7$	$\mathbf{q}_8$	$\mathbf{q}_9$	$\mathbf{q}_{10}$
triceps.R	0.35	-0.18	0.18	0.49	-0.27	-0.06	-0.05	0.00	0.10	0.69
triceps.L	0.36	-0.19	0.15	0.47	-0.27	-0.02	-0.13	0.01	-0.13	-0.70
biceps.R	0.36	-0.13	-0.14	0.04	0.71	-0.50	-0.22	-0.03	-0.19	0.04
biceps.L	0.39	-0.14	-0.09	0.05	0.41	0.70	0.35	0.02	0.19	-0.03
wrist.R	0.34	-0.24	0.14	-0.51	-0.16	-0.21	-0.13	-0.01	0.67	-0.10
wrist.L	0.34	-0.22	0.17	-0.52	-0.23	0.11	0.08	0.03	-0.67	0.12
knee.R	0.30	0.29	-0.50	0.02	-0.24	-0.35	0.62	-0.02	0.01	-0.04
knee.L	0.27	0.35	-0.54	-0.07	-0.18	0.28	-0.63	0.02	-0.02	0.06
ankle.R	0.20	0.53	0.41	-0.03	0.07	0.03	0.00	-0.71	-0.01	-0.02
ankle.L	0.19	0.54	0.40	-0.02	0.10	-0.04	0.01	0.70	0.03	-0.01
% variance	52.23	20.36	10.94	8.57	4.96	1.08	0.86	0.59	0.23	0.19

TABLE 7
Exact PCA loadings (rounded to 2 decimal places) for the reflexes data.

A principal component analysis of the correlation matrix is reported in Table 7. This brings out some of the structure in the data. The dominant component is an average, while components two to five contrast reflexes in different parts of the body, whereas principal components 6 to 10 represent contrasts between reflexes on the left and right sides of the body. The substantially smaller variances associated with principal components 6 to 10 also suggest there is a near constant relationship between measurements on the left and right sides of the body. However, more detailed interpretation of the principal components is difficult. For example, although the first principal component is a weighted average of all the measurements it is not clear how to interpret the loadings given to the ankle measurements relative to the rest.

3.1.1. *Results of our analyses.* All four variants of our methodology were able to produce approximate loading vectors with a good balance between complexity, accuracy and interpretation. Any of their candidate best solutions clearly highlight the structure in the data, giving more interpretable refinements of the exact loading vectors.

Arguably, the next-best forwards solution $\hat{A}_{1:10}^{[3,NF]}$ shown in Table 8 is preferable overall. The minimum accuracy required is $\cos(\theta_{NF}^{[3]}) = 0.91$ and the root mean square of the accuracies attained is 0.95, the vectors of approximate loadings produced being very close individually to their 'parents'. A close competitor is $\hat{A}_{1:10}^{[5,NF]}$ in which only $\hat{\alpha}_6$ and $\hat{\alpha}_7$ change with respect to $\hat{A}_{1:10}^{[3,NF]}$, their accuracies improving from 0.91 to 0.95 and 0.97 respectively. The overall root mean square accuracy is up slightly at 0.96 at the expense of increasing the complexity by one, but without changing the interpretation (zeroes remain in the same place). Due to the orthogonality restrictions, $\hat{\alpha}_6$ and $\hat{\alpha}_7$ for $\hat{A}_{1:10}^{[3,NF]}$ and $\hat{A}_{1:10}^{[5,NF]}$ generate the same two dimensional subspace. This illustrates that rotation of some of the approximated loading vectors can increase accuracy without changing overall interpretation.

We have the following interpretations for the simple components in $\hat{A}_{1:10}^{[3,NF]}$. The first component is just the simple average of all the reflexes. The second component is a simple contrast between upper limbs and lower limbs. The third and fifth components are contrasts within lower and upper limbs respectively. The fourth component is a contrast between reflexes in wrists and triceps. Working back from component 10 to component 8 these can be interpreted as meaning

variable	$\hat{\mathbf{z}}_1^{[3]}$	$\hat{\mathbf{z}}_2^{[3]}$	$\hat{\mathbf{z}}_3^{[3]}$	$\hat{\mathbf{z}}_4^{[3]}$	$\hat{\mathbf{z}}_5^{[3]}$	$\hat{\mathbf{z}}_6^{[3]}$	$\hat{\mathbf{z}}_7^{[3]}$	$\hat{\mathbf{z}}_8^{[3]}$	$\hat{\mathbf{z}}_9^{[3]}$	$\hat{\mathbf{z}}_{10}^{[3]}$
triceps.R	1	2		1	1					-1
triceps.L	1	2		1	1					1
biceps.R	1	2			-2	-1	1			
biceps.L	1	2			-2	1	-1			
wrist.R	1	2		-1	1				-1	
wrist.L	1	2		-1	1				1	
knee.R	1	-3	-1			-1	-1			
knee.L	1	-3	-1			1	1			
ankle.R	1	-3	1					-1		
ankle.L	1	-3	1					1		
Accuracy	0.98	0.95	0.92	0.99	0.91	0.91	0.91	1.00	0.95	0.98
% variance	50.8	20.6	11.2	8.6	5.6	1.1	1.1	0.6	0.3	0.2
order fit	1	8	5	3	10	9	7	4	2	6

TABLE 8

Integer representations for the reflexes data with $\cos(\theta_{NF}^{[3]}) = 0.91$ *using the nextbest forwards method* ($RMS_{NF}^{[3]} = 0.95$). *Empty entries mean zeroes.*

the triceps, wrist and ankle reflexes have the least left-right asymmetry. This asymmetry for the biceps and knees is averaged in component 6 and contrasted in component 7.

The variance explained by the first five approximated components is 96.7% as opposed to 97.1% for the exact principal components, the variances explained by each approximate component being close to that of its optimal counterpart. The last three simple components are nearly constant, so that they are strong candidates for further investigation as possible scientific laws reflecting triceps, wrist and ankle symmetries respectively. The simplified components 6 and 7 are scarcely more variable, suggesting that biceps and knee symmetries may also be present.

The highest absolute correlation between the components derived from our preferred solution is -0.32 between simple components 5 and 7. Other high correlations are between approximate component 9 with 6 and 7, both which are about -0.31, other correlations being significantly smaller.

Interestingly, the basis matrix $\mathbf{O}_5$ (described in Section 2.2) for the forwards and backwards methods revealed features which quickly anticipate the simple and interpretable solutions obtained. In the backwards variant, the columns of the matrix $\mathbf{O}_5$ are the basis vectors $v_1 = (1,1,0,0,0,0,0,0,0,0)^\top$, $v_2 = (0,0,1,1,0,0,0,0,0,0)^\top$ and so on until v_5. Thus, the subsequently approximated components, being generated by linear combinations of them, must average reflexes over different parts of the body. Similarly, for the forwards method, the columns of $\mathbf{O}_5$ are the contrasts $c_1 = (1,-1,0,0,0,0,0,0,0,0)^\top$, $c_2 = (0,0,1,-1,0,0,0,0,0,0)^\top$ and so on until c_5, so that any subsequently approximated components must be linear combinations of left-right reflex contrasts within body parts.

3.1.2. *Comparison with other approaches.* We also applied other simplification methods to the reflexes data set. Table 9 shows the approximate components obtained using the (19) method with corresponding parameter $c = 0$. Compared to the original principal component analysis (Table 7), this gives a substantially simpler, more interpretable solution. Vines' solution has a different interpretation to ours, especially for the middle components, but interestingly picks up the same approximated first and last components. This might be because Vines' method is able to seek simplifications of components in a non-sequential fashion. Approximate components 8 and 9 under Vines' method have virtually the same accuracy as ours, but their complexity is beyond

interpretable limits, indicating that giving preference to simplicity does not necessarily sacrifice accuracy.

variable	$\hat{\mathbf{z}}_1$	$\hat{\mathbf{z}}_2$	$\hat{\mathbf{z}}_3$	$\hat{\mathbf{z}}_4$	$\hat{\mathbf{z}}_5$	$\hat{\mathbf{z}}_6$	$\hat{\mathbf{z}}_7$	$\hat{\mathbf{z}}_8$	$\hat{\mathbf{z}}_9$	$\hat{\mathbf{z}}_{10}$
triceps.R	1	1	2	1	19	-19	19	19	-19	-1
triceps.L	1	1	2	1	19	-19	19	19	-19	1
biceps.R	1	1	-1		-42	-2479	-42	-42	42	
biceps.L	1	1	-1		-40	2561	-40	-40	40	
wrist.R	1	1	2	-1	18	-19	19	18	-2541	
wrist.L	1	1	2	-1	20	-19	19	20	2503	
knee.R	1	-1	-9		10	-9	-2512	10	-10	
knee.L	1	-1	-9		8	-9	2530	8	-8	
ankle.R	1	-2	6		-5	6	-6	-2529	6	
ankle.L	1	-2	6		-7	6	-6	2517	6	
Accuracy	0.98	0.97	0.99	0.99	0.97	0.85	0.89	1.00	0.95	0.98
% variance	50.8	21.5	11	8.6	5	1.2	0.99	0.60	0.30	0.20

TABLE 9

Integer representations for the reflexes data using Vines' method with $c = 0$. Empty entries mean zeroes.

Rousson and Gasser's (2004) default solution with 1 block is identical to our best solution and with 2 blocks (which appears in (13)) only the basis for the subspace generated by the first two approximated components differs, as shown in Table 10. Their method obtains a different basis for the subspace generated by the first two approximated components, and also for the sixth and seventh components, so the attained accuracies and corresponding variances are lower than those produced by our method.

variable	$\hat{\mathbf{z}}_1$	$\hat{\mathbf{z}}_2$	$\hat{\mathbf{z}}_3$	$\hat{\mathbf{z}}_4$	$\hat{\mathbf{z}}_5$	$\hat{\mathbf{z}}_6$	$\hat{\mathbf{z}}_7$	$\hat{\mathbf{z}}_8$	$\hat{\mathbf{z}}_9$	$\hat{\mathbf{z}}_{10}$
triceps.R	1			1	1					-1
triceps.L	1			1	1					1
biceps.R	1				-2	-1				
biceps.L	1				-2	1				
wrist.R	1			-1	1				-1	
wrist.L	1			-1	1				1	
knee.R		1	-1				-1			
knee.L		1	-1				1			
ankle.R		1	1					-1		
ankle.L		1	1					1		
Accuracy	0.87	0.86	0.92	0.99	0.91	0.91	0.91	1.00	0.95	0.98
% variance	44.2	27.2	11.2	8.6	5.6	1.1	1.1	0.6	0.3	0.2

TABLE 10

Integer representations for the reflexes data using Rousson and Gasser's default method with $b = 2$ blocks. Empty entries mean zeroes.

The fact that our exploratory approach produces the same solution as their default solution with 1 block is empirical evidence in favour of this implicit Rousson-Gasser model.

3.2. *Alate adelges data.* The data considered in this section consist of 19 anatomical measurements taken from 40 *alate adelges* (winged aphids) as reported in (7). The measurements taken on each aphid are length and width of the aphid, 3 leg bone measurements, 5 antennal length segments, fore-wing and hind-wing lengths, anal fold, measurements of the rostrum and the ovipositor, and counts of the number of spiracles, antennal spines, ovipositor spines and hind-wing hooks.

Variable	$\mathbf{q}_1$	$\mathbf{q}_2$	$\mathbf{q}_3$	$\mathbf{q}_4$
Length	0.25	-0.03	-0.02	0.07
Width	0.26	-0.07	-0.01	0.10
Fore-wing	0.26	-0.03	0.05	0.07
Hind-wing	0.26	-0.09	-0.03	0.00
Spiracles	0.16	0.41	0.19	-0.62
Antseg 1	0.24	0.18	-0.04	-0.01
Antseg 2	0.25	0.16	0.00	0.02
Antseg 3	0.23	-0.24	-0.05	0.11
Antseg 4	0.24	-0.04	-0.16	0.01
Antseg 5	0.25	0.03	-0.10	-0.02
Ant-spines	-0.13	0.20	-0.93	-0.17
Tarsus 3	0.26	-0.01	-0.03	0.18
Tibia 3	0.26	-0.03	-0.08	0.20
Femur 3	0.26	-0.07	-0.12	0.19
Rostrum	0.25	0.01	-0.07	0.04
Ovipositor	0.20	0.40	0.02	0.06
Ov-spines	0.11	0.55	0.15	0.04
Fold	-0.19	0.35	-0.04	0.49
Hooks	0.20	-0.28	-0.05	-0.45
% variance	73	12.5	3.9	2.6

TABLE 11

Exact principal component analysis loadings (rounded to 2 decimal places) for the alate adelges data.

Jeffers concentrates attention on the four dominant eigenvectors of the correlation matrix, shown in Table 11, these eigenvectors accounting for 92% of the total variability in the data. The first component is a general index of the size of the aphid. Interpretation for the other components looks difficult. Jeffers suggests scaling each eigenvector by dividing by its largest element in absolute value and effectively only take into account for interpretation the largest loadings. Following that rule, the rest of the components are essentially a measure of the number of ovipositor spines, the number of antennal spines and the number of spiracles respectively.

We applied our exploratory methodology to approximate this subset of four eigenvectors. Unsurprisingly, the results using the backwards variants were not so good overall as those for the forwards ones.

For the forwards variant, $A_{1:4}^{[F,4]}$ gives arguably the best balance between simplicity and accuracy. Similarly, $A_{1:4}^{[NF,4]}$ is arguably the best solution obtained using the next-best forwards variant. In these solutions, the first approximate loading vectors were the same and have complexity of just 1. Only one variable, the number of ovipositor spines, was given zero weighting, and only two variables, the anal fold measurement and the number of antennal spines, were given a loading of -1. So this first component can be interpreted as a type of average. The approximate loadings for the other three components have 10 or more zero loadings, assisting interpretations which refine those of Jeffers' original suggestion. However, perhaps unsurprisingly given the arguably non-commensurate nature of the variables here, the overall interpretation is not as clear as with the other examples above.

Table 12 shows $A_{1:4}^{[F,4]}$ and also the solution under the method of (14), this again being an example where only one block is appropriate in their method, all the correlations having the same sign (by changing signs in the number of antennal spines and anal fold variables). We feel that our approximations are better, in as much as they have lower complexity, are more accurate and even slightly more optimal in terms of Rousson and Gasser's criterion of optimality (row labeled

Variable	$\hat{\mathbf{z}}_1$	$\hat{\mathbf{z}}_2$	$\hat{\mathbf{z}}_3$	$\hat{\mathbf{z}}_4$	$\hat{\mathbf{z}}_1$	$\hat{\mathbf{z}}_2$	$\hat{\mathbf{z}}_3$	$\hat{\mathbf{z}}_4$
Length	1			1	1		-3	
Width	1			1	1		-3	1
Forwing	1				1			1
Hinwing	1				1		-3	
Spirac	1	1	1	-3	1	4	11	-3
Antseg 1	1	1			1			
Antseg 2	1	1			1			
Antseg 3	1	-1	-1	1	1	-3	-3	1
Antseg 4	1		-1		1		-3	
Antseg 5	1		-1		1		-3	
Antspin	-1	1	-4	-1	-1	3	-11	-1
Tarsus 3	1			1	1		-3	1
Tibia 3	1			1	1		-3	1
Femur 3	1		-1	2	1		-3	1
Rostrum	1				1		-3	
Ovipos	1	1			1	4		1
Ovspin		1	1		1	4	11	1
Fold	-1	1		3	-1	3		3
Hooks	1	-1	-1	-2	1	-3	-3	-3
Accuracy	0.98	0.92	0.95	0.96	0.98	0.94	0.75	0.96
RMS	0.95				0.91			
% variance	70.3	11.1	4.4	3.0	70.2	11.3	7.8	3.0
Optim (%)	95.2				94.5			
Max Corr	0.40				0.63			

TABLE 12

Left: Integer representations for the alate adelges data using our forwards method with $\cos(\theta_F^{[4]}) = 0.899$. *Right: Rousson and Gasser's method with* $b = 1$ *block. Empty entries mean zeroes.*

Optim(%) in Table 12). Again, Rousson and Gasser's components 2 and 3 are highly correlated (0.63), while the largest correlation we obtain is 0.40 between simple components 1 and 4. Their vectors of loadings are also non-orthogonal, only the first one being orthogonal to the other three. Vines' method is not capable to produce any answer here, mainly due to the complexity of some of the approximate loading vectors growing far too big.

4. Exact recovery of simple structure: a simulation study. In this section we show how our methodology performs when there is a simple underlying structure in the population. For small dimensions it is shown, perhaps surprisingly, that the methodology is able to reproduce the population structure exactly, using only information from the sample. Such behaviour has been observed in one other simplification method (see (16)). We focus on principal component analysis based on the covariance matrix.

In the usual continuous case, whatever the underlying data generation process, it can only ever be *exactly* recovered asymptotically – there is zero probability of recovering the population eigenspaces $\{\ell(\mathbf{q}_r^*)\}_{r=1}^k$ exactly. However, as our simulations show, there is a *positive probability* of doing so if they are all simple. Our algorithm is biased towards simplicity; that is, it moves away from the continuous empirical estimates $\ell(\mathbf{q}_r)$ towards a set of simple angle-close axes. This set may include $\ell(\mathbf{q}_r^*)$ or it may contain axes *simpler* than $\ell(\mathbf{q}_r^*)$, in which case our algorithm will naturally be biased towards finding them instead. Accordingly, we focus on exact recovery of *irreducibly simple* eigenvector structures; that is, on eigenvector structures of complexity one.

We consider the events A_h and S_h, $h = 1, \ldots, k$, defined as follows. The event A_h occurs if and only if $\ell(\hat{\mathbf{z}}_r(\theta_M)) = \ell(\mathbf{q}_r^*)$ for $r = 1, \ldots, h$, for some $\theta \in (0, \pi/4)$ and some method M. That is,

the first h population eigenspaces are recovered exactly by any of the four variants for any of its minimum accuracies. This is consistent with the analysis of the examples where all the variants were applied to the data and the best looking results selected. The event S_h occurs if and only if $\ell(\mathbf{q}_r)$ is closer to $\ell(\mathbf{q}_r^*)$ than any other eigenspace $\ell(\mathbf{q}_s^*)$ $s = 1, \ldots, k$, $s \neq r$, this holding for $r = 1, \ldots, h$. That is, if the first h sample eigenspaces appear in the same order as their population counterparts.

For exact recovery, we focus on the conditional probabilities $P(A_h \mid S_h)$ for $h = 1, \ldots, k$. We condition on S_h as it would be inappropriate to check if A_h has occurred when S_h has not. As h increases, these probabilities evaluate the progressive effectiveness of our methodology in recovering the population eigenvectors.

We performed the following simulation study. For a range of matrices $\mathbf{Z}^* = [\mathbf{z}_1^*, \ldots, \mathbf{z}_k^*]$, where $\mathbf{z}_1^*, \ldots, \mathbf{z}_k^*$ are integer orthogonal vectors of dimension k, and of spectra $\mathbf{\Lambda}^* := \mathrm{diag}(\lambda_1^*, \ldots, \lambda_k^*)$, where $\lambda_1^* > \lambda_2^* > \ldots > \lambda_k^* > 0$, we constructed the corresponding population covariance matrix $\mathbf{\Sigma}^* = \mathbf{Q}^*\mathbf{\Lambda}^*(\mathbf{Q}^*)^T$ where $\mathbf{Q}^*$ is just $\mathbf{Z}^*$ with its columns normalised to have unit length. We simulated n observations from a k-variate normal distribution with mean zero and covariance $\mathbf{\Sigma}^*$. We computed the eigenvalues $\mathbf{\Lambda} := \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$ where $\lambda_1 > \lambda_2 > \ldots > \lambda_k > 0$ and eigenvectors $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_k]$ of the usual empirical covariance matrix $\mathbf{S}$. We calculated $\mathbf{Z}(\theta_M) = [\hat{\mathbf{z}}_1, \ldots, \hat{\mathbf{z}}_k]$ for all four variants and for θ in the accuracy range of $[\cos(\pi/4), \cos(\pi/10)]$. Experience showed this is a good enough range to explore. We replicated independently 1000 times.

For illustration purposes, we focus only on the case $k = 8$. Three representative structures of irreducibly simple eigenvectors were studied: the identity $\mathbf{Z}_{id}^*$ structure, the Hadamard structure defined by

$$\mathbf{Z}_{h,8}^* = \begin{pmatrix} \mathbf{Z}_{h,4}^* & \mathbf{Z}_{h,4}^* \\ \mathbf{Z}_{h,4}^* & -\mathbf{Z}_{h,4}^* \end{pmatrix},$$

where

$$\mathbf{Z}_{h,4}^* = \begin{pmatrix} \mathbf{Z}_{h,2}^* & \mathbf{Z}_{h,2}^* \\ \mathbf{Z}_{h,2}^* & -\mathbf{Z}_{h,2}^* \end{pmatrix} \quad \text{and} \quad \mathbf{Z}_{h,2}^* = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

and a mixed structure defined by

$$\mathbf{Z}_{mix,8}^* = \begin{pmatrix} \mathbf{Z}_{h,4}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{h,4}^* \end{pmatrix}.$$

The aim is to see how our methodology performs when, with high probability, sample eigenspaces appear in the same order as their population counterparts. To this end, we use well-separated eigenvalues – specifically, those occurring in geometrically increasing spectra, that is $(\lambda_1^*, \ldots, \lambda_k^*) = (r^{k-1}, \ldots, r, 1)$ for $r = 1.5$ or 2 – and $n = 100$, for which the estimated value of any $P(S_h)$ was never below 0.9.

The estimated conditional probabilities $\widehat{P}(A_h \mid S_h)$ are shown in Figure 3 together with corresponding pointwise asymptotic 95% confidence intervals. First, we note these probabilities are surprisingly large, at least for the first few components. As expected, there is a better recovery when the ratio between consecutive eigenvalues increases. When $r = 1.5$, there is better overall recovery of the identity structure, the mixed and Hadamard structures being progressively a little more difficult to recover. This maybe because a sparse structure is a bit easier to recover exactly when the eigenvectors are not well separated. However, recovery is essentially the same – and around 90% or better – for all three structures when $r = 2$.

By nestedness, the conditional probabilities $\widehat{P}(A_h \mid S_h)$ decrease with h. To remove this monotone effect we consider also the conditional probabilities $P(A_h \mid A_{h-1}, S_h)$ of recovering the hth

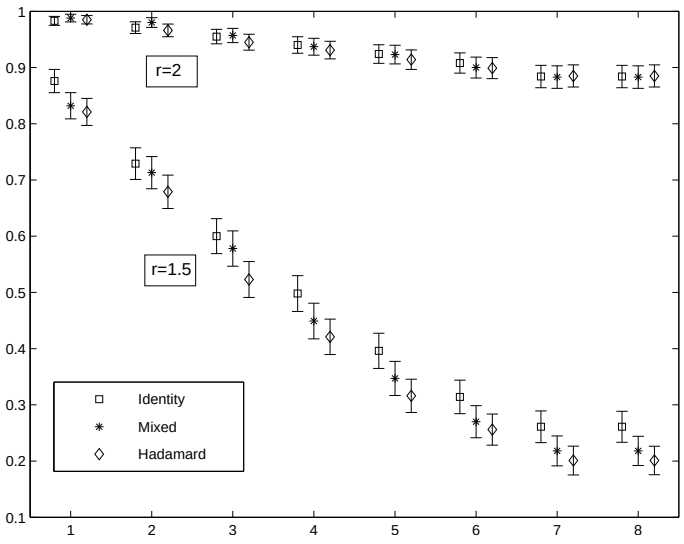

FIG 3. *Estimated conditional probabilities* $\widehat{P}(A_h \mid S_h)$ *with 95% confidence intervals.*

population eigenvector given that the first $(h-1)$ population eigenspaces have been exactly recovered and the first h sample eigenvectors appeared in the same order as their population counterparts, for $h = 1, \ldots, k$. By orthogonality, $P(A_k \mid A_{k-1}, S_k) = 1$. The estimated conditional probabilities $\widehat{P}(A_h \mid A_{h-1}, S_h)$ for $h < k$ are shown in Figure 4.

These probabilities are very large, which is a surprising result given that the number of possible simple orthogonal matrices of dimension $k \times k$ with complexities no larger than N is of order N^{k^2}. With $r = 1.5$, except for the first eigenvector, all three eigenstructures are equally easy to recover, at around 80% or better. When $r = 2$, the recovery is essentially identical – and well over 95% – for all three structures. A more detailed simulation study appears in a separate paper.

5. Discussion. This paper provides a methodology to systematically find a family of integer orthogonal matrices, each of whose members approximate the matrix of eigenvectors of a given covariance or correlation matrix, amongst which the scientist or expert in the field can choose the most interpretable ones. Each approximate eigenvector is constructed to be close in angle to its exact counterpart, while keeping the size of its maximum element as low as possible.

We showed by example that the approximations obtained are useful in practice, and by simulation that the method can have very good theoretical properties such as recovering simple population structure exactly, when such exists. The exploratory nature of the method, as compared to other similar methods, offers more possibilities for interpretation, while not compromising the usefulness of certain modelling assumptions like groups or contrasts. It is also shown that the method performs well with respect to other simplification methods that produce integer vectors.

Variations of our methodology for future study include:

- varying the minimum accuracy required across axes
- using other orders of computation, apart from the four considered in this paper, to explore more extensively the vast space of integer orthogonal matrices
- in the complete case $k' = k$ each approximation is a rotation, so we can explore more deeply the relation with other well-known rotation techniques

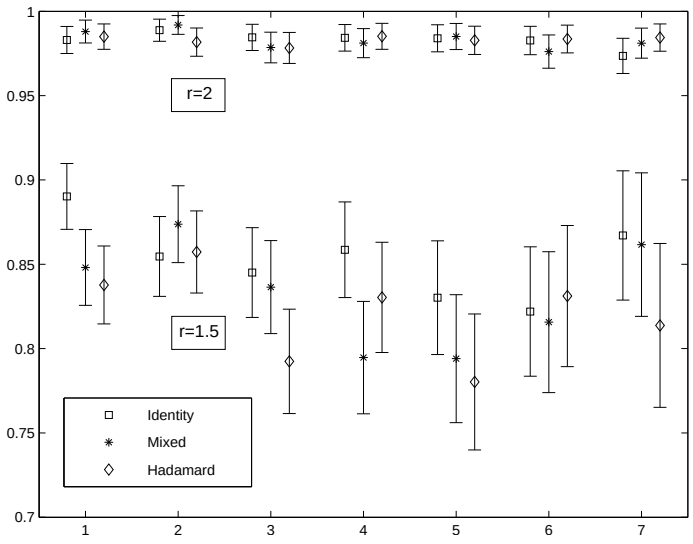

FIG 4. *Estimated conditional probabilities* $\widehat{P}(A_h \mid A_{h-1}, S_h)$ *with 95% confidence intervals*

- instead of simplifying one dimensional sample eigenspaces one at a time, it is possible to sequentially simplify higher dimensional spaces formed by their direct sums. The idea is to partition the set of sample eigenspaces and approximate each corresponding direct sum subspace by finding a simple basis within it. For example, sample eigenspaces with close eigenvalues are natural candidates to form these subspaces.

Finally, the same methodology developed here can be applied, after appropriate modifications, to other multivariate methods including Linear Discriminant Analysis and Canonical Correlation Analysis, and this is also the subject of future work.

APPENDIX A: ORTHOGONALITY

For $2 \leq r \leq k-1$, having found the approximations $\hat{\alpha}_1, \ldots, \hat{\alpha}_{r-1}$, we denote by $\mathbf{Z}_{r-1}$ any $k \times (r-1)$ matrix whose columns are integer representations of those axes. The orthogonality and integer restrictions imply that we need to search for integer vectors in $\mathcal{N}(\mathbf{Z}_{r-1}^\top)$, the null space of $\mathbf{Z}_{r-1}^\top$. Such vectors may be found as follows.

Subsuming, if required, a permutation of the rows of $\mathbf{Z}_{r-1}$, there is no loss in assuming that $\mathbf{Z}_a$ in $\mathbf{Z}_{r-1}^\top = [\mathbf{Z}_a^\top \ \mathbf{Z}_b^\top]$ is nonsingular. (In practice, we have created an algorithm to select $\mathbf{Z}_a$.) Partitioning $\mathbf{u} \in \mathbb{R}^k$ as $\mathbf{u}^\top = [\mathbf{u}_a^\top \ \mathbf{u}_b^\top]$, where $\mathbf{u}_a \in \mathbb{R}^{r-1}$ and $\mathbf{u}_b \in \mathbb{R}^{k-r+1}$, an integer basis for $\mathcal{N}(\mathbf{Z}_{r-1}^\top)$ is obtained using $\mathbf{Z}_{r-1}^\top \mathbf{u} \equiv \mathbf{Z}_a^\top \mathbf{u}_a + \mathbf{Z}_b^\top \mathbf{u}_b = (0, \ldots, 0)^\top$. This gives $\det(\mathbf{Z}_a)\,\mathbf{u}_a = -\mathrm{cof}(\mathbf{Z}_a)\mathbf{Z}_b^\top \mathbf{u}_b$, in which $\mathrm{cof}(\mathbf{Z}_a)$ is the matrix of cofactors of $\mathbf{Z}_a$. Since $\det(\mathbf{Z}_a)$ is an integer and $\mathrm{cof}(\mathbf{Z}_a)$ is an integer matrix, the $k-r+1$ columns of the integer matrix

$$\mathbf{O}_{r-1} := \begin{bmatrix} -\mathrm{cof}(\mathbf{Z}_a)\mathbf{Z}_b^\top \\ \det(\mathbf{Z}_a)\,\mathbf{I}_{k-r+1} \end{bmatrix}$$

form a basis for $\mathcal{N}(\mathbf{Z}_r^\top)$. We take $\mathcal{M}_r := \{\ell(\mathbf{O}_{r-1}\mathbf{y}) : \mathbf{y} \in \mathbb{Z}^{(k-r+1)}\}$.

In the case $r = k$, $\mathcal{N}(\mathbf{Z}_{k-1}^\top)$ is one dimensional and therefore a simple axis, since $\mathcal{N}(\mathbf{Z}_{k-1}^\top) = \ell(\mathbf{O}_{k-1})$ implies $\hat{\mathbf{z}}_k(\theta) = \mathbf{O}_{k-1}/\mathrm{hcf}(|\mathbf{O}_{k-1}|)$ and $\hat{\alpha}_k(\theta) = \ell(\mathbf{O}_{k-1})$ is constant as a function of θ,

(hcf($|\mathbf{u}|$) denoting again the highest common factor of the absolute values of the nonzero entries of $\mathbf{u}$).

APPENDIX B: IMPLEMENTATION

Numerically, we do not necessarily obtain exactly the best simple θ approximation to α_r orthogonal to $\hat{\alpha}_1, \ldots, \hat{\alpha}_{r-1}$. Instead, we compute the following approximations. We restrict attention to axes generated by vectors of the form $\mathbf{O}_{r-1}\mathbf{y}$ with $\mathbf{O}_{r-1}$ a $k \times (k-r+1)$ orthogonal matrix and $\mathbf{y} \in \mathbb{Z}^{(k-r+1)}$ (see Section 2.2). We obtain an approximation $\tilde{\alpha}_r(\theta)$ to $\hat{\alpha}_r(\theta)$ as follows:

1. Obtain a set of candidate vectors $\mathcal{Y}$ for $\mathbf{y}$.
2. Obtain $\ell(\mathbf{O}_{r-1}\mathcal{Y})$ which is the set of axes generated by vectors in $\mathbf{O}_{r-1}\mathcal{Y}$ (the image of $\mathcal{Y}$ under the linear mapping $\mathbf{O}_{r-1}$) and find the minimum complexity $\tilde{N}_r(\theta)$ of the axes in $\ell(\mathbf{O}_{r-1}\mathcal{Y})$.
3. $\tilde{\alpha}_r(\theta)$ is given by the most accurate vector of the axes in $\ell(\mathbf{O}_{r-1}\mathcal{Y})$ with complexity $\tilde{N}_r(\theta)$.

To obtain $\mathcal{Y}$ we focus attention on $\mathbf{y}^* := (\mathbf{O}_{r-1}^\top \mathbf{O}_{r-1})^{-1}\mathbf{O}_{r-1}^\top \mathbf{q}_r$ (note that $\mathbf{O}_{r-1}\mathbf{y}^* = \mathbf{q}_r^\perp$). For a user-chosen $N_{max} \geq 1$, we put $\mathcal{Y} := \cup_{N \leq N_{max}} \mathcal{Y}_N$ where, for each $N \leq N_{max}$, $\mathcal{Y}_N$ is constructed, as detailed below, with a set of candidate vectors in $\mathbb{Z}^{(k-r+1)}$ used to minimise the angle with $\mathbf{y}^*$. There is no loss in reversing the signs of any negative entries in $\mathbf{y}^*$ and then placing its entries in non-increasing order, (the corresponding inverse permutation and sign changes being applied to all candidate vectors to recover the original signs and order). We compute first the number of iterations

$$I_N = \sum_{i=1}^{\min\{N,k-r\}+1} \binom{k-r-1}{i-1}\binom{N+1}{i}$$

required for an exhaustive search for fixed N and compare this with a user-chosen limit I_{max}. If $I_N \leq I_{max}$, we take $\mathcal{Y}_N$ as the set of all non-increasing nonnegative integer vectors with complexity N. If $I_N > I_{max}$ we start with the vector $(N, 0, \ldots, 0)$, then look over all $(N, i, 0, \ldots, 0)$ for $i = 0, \ldots, N$ and keep the one which minimises the angle with $\mathbf{y}$. If the minimum is attained at zero then stop, otherwise proceed in the same way for the next entry. $\mathcal{Y}_N$ is the set all vectors encountered in this process. The case $r = 1$ is covered for the first approximation by taking $\mathbf{O}_{r-1}$ as the identity matrix.

Experience has shown that values of $N_{max} = 9$ and $I_N = 20,000$ give accurate results, while keeping the computations fast, and that the algorithm appears to be stable so long as the integer entries do not become huge. Details of the calculations appear in the corresponding R routine `simplify.r` which is available from the authors upon request.

ACKNOWLEDGEMENTS

We are grateful to Paddy Farrington and Chris Jones for useful comments on earlier versions of this manuscript, and to Nickolay Trendafilov for helpful discussions.

REFERENCES

[1] Cadima, J. and Jolliffe, I. T.(1995). Loadings and correlations in the interpretation of principal components. *Journal of Applied Statistics* **22**(2) 203–241.

[2] Chipman, H. A. and Gu, H.(2005). Interpretable dimension reduction. *Journal of Applied Statistics* **32**(9) 969–987.

[3] D'Aspremont, A. and El Ghaoui, L. and Jordan, M. I. and Lanckriet, G. R. G.(2007). A direct formulation for sparse PCA using semidefinite programming. *SIAM Review* **49**(3) 434–448.

[4] Fang, K.and Li, R.(1997). Some methods for generating both an NT-net and the uniform distribution on a Stiefel manifold and their applications. *Computational Statistics & Data Analysis* **24** 29–46.
[5] Farcomeni, A.(2008). An exact approach to sparse principal component analysis. *Computational Statistics, to appear*.
[6] Hausman, R. E.(1982). Constrained Multivariate Analysis. *Studies in the Managment Sciences* **19** 137–151.
[7] Jeffers, J.N.R.(1967). Two case studies in the application of principal component analysis. *Applied Statistics* **16** 225–236.
[8] Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer-Verlag New York, Inc.
[9] Jolliffe, I. T. and Trendafilov, N. T. and Uddin, M.(2003). A Modified Principal Component Technique Based on the LASSO. *Journal of Computational and Graphical Statistics* **12**(3) 531–547.
[10] Mardia, K. V. and Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press.
[11] Moghaddam, B. and Weiss, Y. and Avidan, S.(2006). Spectral bounds for sparse PCA: exact and greedy algorithms. *Advances in Neural Information Processing Systems*. **18** 915–922.
[12] Park, T.(2005). A penalized likelihood approach to rotation of principal components. *Journal of computational and graphical statistics* **14**(4) 867–888.
[13] Rousson, V. and Gasser, T.(2003). Some case studies of simple component analysis. *Unpublished manuscript*.
[14] Rousson, V. and Gasser, T.(2004). Simple component analysis. *Applied Statistics* **53**(4) 539–555.
[15] Sjöstrand, K. and Stegmann, M. B. and Larsen, R.(2006). Sparse principal component analysis in medical shape modeling. *International Symposium on Medical Imaging 2006*. 1579-1590.
[16] Sun, L. (2006). *Simple Principal Components*. Ph.D. Thesis. The Open University.
[17] Thompson, M.O. and Vines, S.K. and Harrington, K. (1999). Assessment of blood volume flow in the uterine artery: the influence of arterial distensibility and waveform abnormality. *Ultrasound in Obstetrics and Gynecology* **14**(1) 71.
[18] Trendafilov, N. T. and Jolliffe, I. T.(2007). dalass: Variable selection in discriminant analysis via the LASSO. *Computational Statistics and Data Analysis* **51**(8) 3718–3736.
[19] Vines, S. K.(2000). Simple principal components. *Applied Statistics* **49**(4) 441–451.
[20] Zou, H. and Hastie, T. and Tibshirani, R. (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics* **15**(2) 265–286.

Department of Mathematics and Statistics, The Open University,
Walton Hall, Milton Keynes, MK7 6AA, United Kingdom
E-mail: k.anaya@open.ac.uk