On ANOVA-Like Matrix Decompositions

Giuseppe Bove, Frank Critchley, Radka Sabolova and Germain Van Bever

Abstract The analysis of variance plays a fundamental role in statistical theory and practice, the standard Euclidean geometric form being particularly well-established. The geometry and associated linear algebra underlying such standard analysis of variance methods permit, essentially direct, generalisation to other settings. Specifically, as jointly developed here: (a) to minimum distance estimation problems associated with subsets of pairwise orthogonal subspaces; (b) to matrix, rather than vector, contexts; and (c) to general, not just standard Euclidean, inner products, and their induced distance functions. To this end, we characterise inner products rendering pairwise orthogonal a given set of nontrivial subspaces of a linear space any two of which meet only at the origin. Applications in a variety of areas are highlighted, including: (i) the analysis of asymmetry, and (ii) asymptotic comparisons in Invariant Coordinate Selection and Independent Component Analysis. A variety of possible further generalisations and applications are noted.

Giuseppe Bove
Dipartimento di Scienze della Formazione, Università degli Studi Roma Tre, Italy e-mail: giuseppe.bove@uniroma3.it

Frank Critchley
Department of Mathematics and Statistics, The Open University, United Kingdom e-mail: F.Critchley@open.ac.uk

Radka Sabolova
Department of Mathematics and Statistics, The Open University, United Kingdom e-mail: radka.sabolova@open.ac.uk

Germain Van Bever
Department of Mathematics and Statistics, The Open University, United Kingdom e-mail: germain.van-bever@open.ac.uk

1 Introduction

Geometry is a rich resource for statistical theory and practice. In particular, minimum distance estimation is an insightful, recurring, theme across many methodologies. Incorporating associated linear algebra, the analysis of variance (hereafter, ANOVA) is perhaps the stand-out example, estimation consisting essentially of orthogonal projection onto each of a set of pairwise orthogonal subspaces of a standard Euclidean space, whose direct sum is the space itself.

Specifically, for some given dimension d, standard ANOVA takes place within $\mathscr{E}^d = (\mathscr{R}^d, \langle\cdot,\cdot\rangle_{\mathbf{I}})$ – that is, within $\mathscr{R}^d$ endowed with the inner product $\langle\mathbf{x},\mathbf{y}\rangle_{\mathbf{I}} := \mathbf{x}^\top\mathbf{y}$, $(\mathbf{x},\mathbf{y} \in \mathscr{R}^d)$, inducing the squared Euclidean distance between $\mathbf{x} = (x_i)$ and $\mathbf{y} = (y_i)$:

$$\|\mathbf{x}-\mathbf{y}\|^2_{\mathbf{I}} = \sum_{i=1}^d (x_i - y_i)^2.$$

Further, $\mathscr{E}^d$ is decomposed as the direct sum of $2 \leq r \leq d$ subspaces $\{\mathscr{L}_h\}_{h=1}^r$, any two of which are orthogonal. Accordingly, each $\mathbf{y} \in \mathscr{E}^d$ has unique decomposition

$$\mathbf{y} = \sum_{h=1}^r \mathbf{y}_h \tag{1}$$

in which, for each $1 \leq h \leq r$, $\mathbf{y}_h$ is the orthogonal projection of $\mathbf{y}$ onto $\mathscr{L}_h$ – that is, the nearest point in $\mathscr{L}_h$ to $\mathbf{y}$. Crucially, pairwise orthogonality of the subspaces means that *separate* minimum distance estimation of $\mathbf{y}$ by each $\mathscr{L}_h$ gives $\mathbf{y}_h$ and, thereby, the *overall* decomposition (1).

Essentially the same ideas can be applied more generally. First, and of central interest in this paper, let $\mathscr{L}'_h \subseteq \mathscr{L}_h$ $(h = 1,...,r)$ be any given nonempty subsets of the original subspaces and suppose that, for given $\mathbf{y} \in \mathscr{E}^d$, we have the problem:

$$\text{minimise } \|\mathbf{y} - \sum_{h=1}^r \widehat{\mathbf{y}}_h\|^2_{\mathbf{I}} \text{ subject to } \widehat{\mathbf{y}}_h \in \mathscr{L}'_h \ (h = 1,...,r). \tag{2}$$

Then, again, pairwise orthogonality of the $\{\mathscr{L}_h\}$ entails that separate minimum distance estimation of each $\mathbf{y}_h$ by $\mathscr{L}'_h$ gives $\widehat{\mathbf{y}}_h$, these together solving the overall problem (2).

Secondly, vectors can be replaced by matrices. Section 2 reviews a variety of matrix analogues of the ANOVA decomposition (1). Each of these is based on the standard Euclidean inner product $\langle\mathbf{A},\mathbf{B}\rangle_E := \text{trace}(\mathbf{A}^\top\mathbf{B})$. The induced (Frobenius) norm $\|\cdot\|_E$ has squared distance $\|\mathbf{A}-\mathbf{B}\|^2_E = \sum_{i,j}(a_{ij} - b_{ij})^2$, the element-wise sum of squared differences. Applications in a variety of areas are highlighted, including: (i) the analysis of asymmetry (Constantine and Gower 1978; Gower 2014), and (ii) asymptotic comparisons in Invariant Coordinate Selection (Tyler et al., 2009) and Independent Component Analysis (Oja et al., 2006).

Thirdly, more general inner products can be used. In vector terms, the standard Euclidean inner product $\langle\cdot,\cdot\rangle_{\mathbf{I}}$ can be replaced by a general one:

$$\langle\mathbf{x},\mathbf{y}\rangle_{\mathbf{V}} := \mathbf{x}^\top\mathbf{V}\mathbf{y}$$

where $\mathbf{V}$ is any positive definite symmetric matrix of the appropriate order. Alternative inner products can also be used in a matrix context. For completeness' sake, Section 3 briefly reviews basic (essentially, standard) theory for these. Section 4 revisits analysis of asymmetry examples (Section 2.1) in this more general context. A variety of potential applications are noted.

The focus is on square matrices throughout. Extension to rectangular matrices, higher order arrays, and other vector space contexts is in principle straightforward.

To aid the flow of the paper, general terminology and definitions are introduced as we proceed. Its well-known or straightforward results are stated without proof.

2 Matrix decomposition examples

We briefly review here a variety of matrix decompositions based on standard Euclidean geometry. Sections 2.1 and 2.2 focus, respectively, on applications in the analysis of asymmetry and in Invariant Coordinate Selection.

2.1 The analysis of asymmetry

Constantine and Gower (1978) emphasised how important, in practice, it can be to model systematic departures from symmetry in square data matrices. This influential paper gave rise to a field described here as the analysis of asymmetry, the recent review paper by Gower (2014) providing interesting historical perspective. To quote from its opening section, the driving motivation for this work was that:

> 'In practice, many applications are concerned just as much with any asymmetry as they are with symmetry. Well-known examples are studies in social mobility between classes, international trade between countries and pecking order in hens.'

Additional technical details in Gower (2014) include, in particular, a formal proof of the form of the singular value decomposition of a skew-symmetric matrix.

Section 2.1.1 reviews a basic orthogonal decomposition of square matrices into their symmetric and skew-symmetric parts. In some contexts, a further decomposition may be called for. Section 2.1.2 treats a leading example of this. Section 2.1.3 discusses (i) a general orthogonal decomposition and (ii) other types of matrix subspace. Section 2.1.4 describes general benefits arising from using such decompositions, together with a worked example.

2.1.1 A basic decomposition

With the obvious definitions of addition and scalar multiplication of matrices, the set $\mathscr{M}_k$ of real square matrices of order k is a real linear (vector) space of dimension k^2.

Omitting the subscript k, its subsets of symmetric and skew(-symmetric) members, $\mathscr{S} := \{\mathbf{S} \in \mathscr{M} : \mathbf{S}^\top = \mathbf{S}\}$ and $\mathscr{K} := \{\mathbf{K} \in \mathscr{M} : \mathbf{K}^\top = -\mathbf{K}\}$, are *subspaces – i.e.* are closed under addition and scalar multiplication – with dimensions $\binom{k+1}{2}$ and $\binom{k}{2}$ respectively. Moreover,

$$\mathscr{S} + \mathscr{K} := \{\mathbf{S}+\mathbf{K} : \mathbf{S} \in \mathscr{S}, \mathbf{K} \in \mathscr{K}\} = \mathscr{M}, \text{ while } \mathscr{S} \cap \mathscr{K} = \{\mathbf{0}_{\mathscr{M}}\}. \tag{3}$$

That is (definition), $\mathscr{S}$ and $\mathscr{K}$ have *direct sum* $\mathscr{M}$, denoted $\mathscr{M} = \mathscr{S} \oplus \mathscr{K}$. Equivalently, every $\mathbf{M} \in \mathscr{M}$ can be *uniquely* represented in the form $\mathbf{M} = \mathbf{S_M} + \mathbf{K_M}$ where $\mathbf{S_M} \in \mathscr{S}$ and $\mathbf{K_M} \in \mathscr{K}$. Explicitly:

$$\mathbf{S_M} = \tfrac{1}{2}(\mathbf{M} + \mathbf{M}^\top) \text{ and } \mathbf{K_M} = \tfrac{1}{2}(\mathbf{M} - \mathbf{M}^\top). \tag{4}$$

If we endow $\mathscr{M}$ with the standard Euclidean inner product, reviewed in the Introduction, the direct sum relations (3) sharpen to:

Proposition 1. *$\mathscr{M} = \mathscr{S} \oplus \mathscr{K}$ in which $\mathscr{S} = \mathscr{K}^\perp$ and $\mathscr{K} = \mathscr{S}^\perp$ are orthogonal complements in the inner product space $(\mathscr{M}, \langle \cdot, \cdot \rangle_E)$. Accordingly, for every $\mathbf{M} \in \mathscr{M}$,*

$$\|\mathbf{M}\|_E^2 = \|\mathbf{S_M}\|_E^2 + \|\mathbf{K_M}\|_E^2. \tag{5}$$

In a seminal paper, Constantine and Gower (1978) exploited this result to advocate separate fitting of models to the symmetric and skew parts of a given data matrix $\mathbf{M}$. For, if $\mathscr{S}' \subseteq \mathscr{S}$ and $\mathscr{K}' \subseteq \mathscr{K}$ denote any symmetric and skew matrix model classes, minimisation of

$$\|\mathbf{M} - (\mathbf{S} + \mathbf{K})\|_E^2 = \|\mathbf{S_M} - \mathbf{S}\|_E^2 + \|\mathbf{K_M} - \mathbf{K}\|_E^2$$

over all $(\mathbf{S}, \mathbf{K}) \in \mathscr{S}' \times \mathscr{K}'$ is, palpably, accomplished by *separate* least squares fitting of $\mathbf{S_M}$ by $\mathbf{S} \in \mathscr{S}'$ and of $\mathbf{K_M}$ by $\mathbf{K} \in \mathscr{K}'$.

2.1.2 A further decomposition

Consider the following decomposition of $\mathscr{S}$. Denoting by $*$ the Hadamard (direct) product of matrices,

$$\mathscr{D} := \{\mathbf{D} \in \mathscr{S} : \mathbf{D} * \mathbf{I} = \mathbf{D}\} \text{ and } \mathscr{H} := \{\mathbf{H} \in \mathscr{S} : \mathbf{H} * \mathbf{I} = \mathbf{0}_{\mathscr{M}}\}$$

comprise all diagonal and all 'hollow' (zero diagonal) members of $\mathscr{S}$, these being subspaces of $\mathscr{S}$ of dimension k and $\binom{k}{2}$ respectively, and we have:

$$\mathscr{S} = \mathscr{D} \oplus \mathscr{H},$$

$\mathbf{S} \in \mathscr{S}$ being uniquely expressable as $\mathbf{S} = \mathbf{D_S} + \mathbf{H_S}$ with $\mathbf{D_S} \in \mathscr{D}$ and $\mathbf{H_S} \in \mathscr{H}$. Explicitly, $\mathbf{D_S} = \mathbf{S} * \mathbf{I}$ and $\mathbf{H_S} = \mathbf{S} * (\mathbf{1}\mathbf{1}^\top - \mathbf{I})$, where $\mathbf{1}$ denotes the vector of ones.

Overall, $\mathscr{M} = (\mathscr{D} \oplus \mathscr{H}) \oplus \mathscr{K}$ which, the order of summation being clearly irrelevant, we may write as $\mathscr{M} = \mathscr{D} \oplus \mathscr{H} \oplus \mathscr{K}$, each $\mathbf{M} \in \mathscr{M}$ having unique decomposition in terms of these subspaces as:

$$\mathbf{M} = \mathbf{D}_{\mathbf{S_M}} + \mathbf{H}_{\mathbf{S_M}} + \mathbf{K}_{\mathbf{M}}. \tag{6}$$

If we endow $\mathscr{M}$ with the standard Euclidean inner product, inherited by each of its subspaces, the above direct sum relations sharpen to:

Proposition 2. *$\mathscr{M} = \mathscr{D} \oplus \mathscr{H} \oplus \mathscr{K}$ in which $\mathscr{D}$, $\mathscr{H}$ and $\mathscr{K}$ are pairwise orthogonal subspaces of the inner product space $(\mathscr{M}, \langle\cdot,\cdot\rangle_E)$. Accordingly, for every $\mathbf{M} \in \mathscr{M}$,*

$$\|\mathbf{M}\|_E^2 = \left\|\mathbf{D}_{\mathbf{S_M}}\right\|_E^2 + \left\|\mathbf{H}_{\mathbf{S_M}}\right\|_E^2 + \|\mathbf{K}_{\mathbf{M}}\|_E^2.$$

Again, this result justifies separate fitting of models to the diagonal, hollow and skew parts of a given data matrix $\mathbf{M}$. For, if $\mathscr{D}' \subseteq \mathscr{D}$, $\mathscr{H}' \subseteq \mathscr{H}$ and $\mathscr{K}' \subseteq \mathscr{K}$ denote any diagonal, hollow and skew matrix model classes, minimisation of

$$\|\mathbf{M} - (\mathbf{D} + \mathbf{H} + \mathbf{K})\|_E^2 = \left\|\mathbf{D}_{\mathbf{S_M}} - \mathbf{D}\right\|_E^2 + \left\|\mathbf{H}_{\mathbf{S_M}} - \mathbf{H}\right\|_E^2 + \|\mathbf{K}_{\mathbf{M}} - \mathbf{K}\|_E^2$$

over all $(\mathbf{D}, \mathbf{H}, \mathbf{K}) \in \mathscr{D}' \times \mathscr{H}' \times \mathscr{K}'$ is clearly accomplished by separate least squares fitting of each of $(\mathbf{D}_{\mathbf{S_M}}, \mathbf{H}_{\mathbf{S_M}}, \mathbf{K}_{\mathbf{M}})$ within its model class.

2.1.3 A general decomposition and other subspace types

The methodology developed above generalises directly to decomposition of $\mathscr{M}$ into any collection of any number of pairwise orthogonal subspaces in an obvious way, analogous – indeed, formally, isomorphic – to familiar orthogonal decompositions of $\mathscr{E}^d$ used in standard ANOVA.

Replacing subspaces of vectors in $\mathscr{E}^d$ by subspaces of matrices in $\mathscr{M}$, we have

Proposition 3. *Let $\{\mathscr{L}_h\}_{h=1}^r$ $(2 \le r \le k^2)$ be a collection of nontrivial subspaces of $\mathscr{M}$, any two of which are orthogonal in $(\mathscr{M}, \langle\cdot,\cdot\rangle_E)$, and whose direct sum $\oplus\{\mathscr{L}_h\}$ is $\mathscr{M}$. Then, every $\mathbf{M} \in \mathscr{M}$ can be uniquely expressed as $\mathbf{M} = \sum_{h=1}^r \mathbf{M}_h$, in which $\mathbf{M}_h$ is the orthogonal projection of $\mathbf{M}$ onto $\mathscr{L}_h$, so that*

$$\|\mathbf{M}\|_E^2 = \textstyle\sum_{h=1}^r \|\mathbf{M}_h\|_E^2.$$

As a direct corollary, for any given model classes $\mathscr{L}_h' \subseteq \mathscr{L}_h$ $(h = 1, ..., r)$ within the 'parent' subspaces $\{\mathscr{L}_h\}$, and for any given $\mathbf{M} \in \mathscr{M}_k$, the overall problem:

$$\text{minimise } \left\|\mathbf{M} - \textstyle\sum_{h=1}^r \widehat{\mathbf{M}}_h\right\|_E^2 \text{ subject to } \widehat{\mathbf{M}}_h \in \mathscr{L}_h' \; (h = 1, ..., r) \tag{7}$$

can be solved by separate minimum distance estimation of each $\mathbf{M}_h$ by $\mathscr{L}_h'$.

There is a wide variety of types of subspace of potential interest in modelling data taking the form of a square matrix. In particular, the subspaces $\mathscr{L}_{\mathscr{Z}}$ of matrices (m_{ij}) for which m_{ij} vanishes for each $(i, j) \in \mathscr{Z}$, $\emptyset \subseteq \mathscr{Z} \subseteq (1, ..., k)^2$, the orthogonal complement $\mathscr{L}_{\mathscr{Z}}^{\perp}$ of each such subspace being given by:

$$\mathscr{L}_{\mathscr{Z}}^{\perp} = \mathscr{L}_{\mathscr{Z}^C}.$$

Subspaces of this type include matrices vanishing on a given set of rows or columns.

Other instances of $\mathscr{L}_{\mathscr{Z}}$ are of special interest when the rows and columns of $\mathbf{M}$ have the same natural order. For example – in a context where m_{ij} is modelled as vanishing when i and j belong to different members of a partition of $(1,...,k)$ into consecutive blocks – the corresponding blockdiagonal matrices. Or – in a context where $(1,...,k)$ label consecutive time points and m_{ij} is held to vanish for i and j sufficiently far apart in time – the (say) tri-diagonal matrices defined by: $|i-j| > 1 \Rightarrow m_{ij} = 0$.

Or, again, in social mobility studies – with $(1,...,k)$ labelling a classification of households from high to low social status, and m_{ij} the number of adults in a category i household whose parents formed a category j household – the strictly upper triangular, diagonal, and strictly lower triangular matrices are of interest. Specifically, with $\mathscr{L}_{\uparrow}$, $\mathscr{L}_{\rightarrow}$ and $\mathscr{L}_{\downarrow}$ denoting the subspaces $\mathscr{L}_{\mathscr{Z}}$ corresponding, respectively, to $\mathscr{Z} = \mathscr{Z}_{\uparrow} := \{(i,j) \in (1,...,k)^2 : j \leq i\}$, $\mathscr{Z} = \mathscr{Z}_{\rightarrow} := \{(i,j) \in (1,...,k)^2 : i \neq j\}$ and $\mathscr{Z} = \mathscr{Z}_{\downarrow} := \{(i,j) \in (1,...,k)^2 : i \leq j\}$, we have the orthogonal decomposition:

$$\mathscr{M} = \mathscr{L}_{\uparrow} \oplus \mathscr{L}_{\rightarrow} \oplus \mathscr{L}_{\downarrow}$$

in which $\mathscr{L}_{\uparrow}$ and $\mathscr{L}_{\downarrow}$ correspond to upward and downward inter-generational social mobility respectively, while $\mathscr{L}_{\rightarrow}$ corresponds to absence of either form of mobility. Typically, different models will be appropriate to these three processes, so that the expected value of a data matrix $\mathbf{M}$ is best modelled as the sum of matrices from each of three corresponding – often, parameterised – model classes $\mathscr{L}'_{\uparrow}$, $\mathscr{L}'_{\rightarrow}$ and $\mathscr{L}'_{\downarrow}$.

More generally than $\mathscr{L}_{\mathscr{Z}}$, the symmetric and skew-symmetric subspaces $\mathscr{S}$ and $\mathscr{K}$ are examples of subspaces on which a prescribed set of linear relations hold among the elements of $\mathbf{M}$.

Again, the set of subspaces of $\mathscr{M}$ is closed under orthogonal complementation, intersection and addition, giving 'new subspaces for old'. For example, $\mathscr{L}_{\rightarrow}^{\perp}$ comprises all matrices with zero diagonal entries, while $\mathscr{H} = \mathscr{L}_{\rightarrow}^{\perp} \cap \mathscr{S}$. Again, the upper and lower triangular matrices form the subspaces $\mathscr{L}_{\uparrow} \oplus \mathscr{L}_{\rightarrow}$ and $\mathscr{L}_{\downarrow} \oplus \mathscr{L}_{\rightarrow}$ respectively.

Finally in this section, we note general benefits flowing from use of orthogonal decompositions of $\mathscr{M}$ as $\oplus\{\mathscr{L}_h\}$, as in Proposition 3, together with a worked example using the decomposition $\mathscr{M} = \mathscr{D} \oplus \mathscr{H} \oplus \mathscr{K}$.

2.1.4 General benefits and a worked example

As illustrated by the social mobility example above, the expected value of a data matrix $\mathbf{M} \in \mathscr{M}$ can often be best modelled as a sum of r matrices $(2 \leq r \leq k^2)$, one from each of a set of model classes $\{\mathscr{L}'_h\}_{h=1}^r$ where $\{\mathbf{0}_{\mathscr{M}}\} \subset \mathscr{L}'_h \subseteq \mathscr{L}_h$ $(1 \leq h \leq r)$ and $\oplus_{h=1}^r\{\mathscr{L}_h\}$ is an orthogonal decomposition of $(\mathscr{M}, \langle\cdot,\cdot\rangle_E)$ into component subspaces. Assuming uncorrelated homoscedastic additive errors with zero mean and common variance σ^2, least-squares fitting – equivalently, assuming also Gaussianity, maximum likelihood estimation – then leads to the general constrained matrix approximation problem (7).

Exploiting orthogonality, this problem can be solved by separate minimum distance estimation of each $\mathbf{M}_h$ by $\mathscr{L}'_h$. This breaking up of an overall problem into component parts can be especially beneficial, computationally, when the model classes $\{\mathscr{L}'_h\}$ depend in complex nonlinear ways on separate underlying parameter vectors $\{\boldsymbol{\theta}_h\}$.

Again, such separate fitting breaks up underlying variability into its component parts. In an obvious notation, we have

$$\sigma^2 = \sum_{h=1}^{r} \sigma_h^2.$$

Each of the components of variance σ_h^2 is estimable, in the obvious way. Indeed, under Gaussianity, the resulting $\{\widehat{\sigma}_h^2\}$ are mutually independent, with the usual benefits for inference.

Finally, borrowing a term from generalised linear models, undefined or non-stochastic elements of a data matrix $\mathbf{M}$ can be accommodated by introducing corresponding *offset* terms.

To fix ideas, consider the problem of modelling a matrix $\mathbf{M}$ of scheduled flight times between cities in a part of the world subject to a strong prevailing wind blowing from West to East (say). A natural approach, described below, uses the orthogonal decomposition $\mathscr{M} = \mathscr{D} \oplus \mathscr{H} \oplus \mathscr{K}$.

The diagonal entries of $\mathbf{M}$ being undefined, we don't try to model them. Rather, formally, we set $\mathbf{D}_{\mathbf{S_M}} = \mathbf{0}_{\mathscr{M}}$, taking the corresponding model class as $\mathscr{D}' = \{\mathbf{0}_{\mathscr{M}}\}$. These offset terms assure perfect diagonal 'fit', effectively replacing minimisation of $||\mathbf{M} - \widehat{\mathbf{M}}||_E^2$ with that of

$$||\mathbf{M} - \widehat{\mathbf{M}}||_{E'}^2 := \sum_{i \neq j} (m_{ij} - \widehat{m}_{ij})^2.$$

Consider now the off-diagonal entries of $\mathbf{M}$. By definition, $\mathbf{H}_{\mathbf{S_M}}$ belongs to $\mathscr{H}^+ := \{(h_{ij}) \in \mathscr{H} : \text{each } h_{ij} \geq 0\}$, the set of all $k \times k$ dissimilarity matrices, and so can be fitted by a suitable multidimensional scaling method. Alternatively, as developed here, we may assume that the expected value of the average of the scheduled flight time from city i to city j and that for the opposite journey is a function of g_{ij}, the geodesic distance between them. As one parametric example, we may take $\mathscr{H}' = \mathscr{H}'(\boldsymbol{\theta}_\mathbf{H})$ in which $\boldsymbol{\theta}_\mathbf{H} = (\alpha, \beta, \kappa)^\top$, the general member $\mathbf{H}$ of $\mathscr{H}'(\boldsymbol{\theta}_\mathbf{H})$ having $(i,j)^{th}$ element $\alpha + \beta g_{ij}^{\kappa}$. The constant term α here reflects a fixed time needed for take off and landing.

We turn now to the skew symmetric part of $\mathbf{M}$. Constantine and Gower (1978) take an essentially exploratory approach, based on the singular value decomposition of $\mathbf{K_M}$ and, in particular, graphical representation of its dominant parts. Alternatively, for each $i < j$, the expected mean difference k_{ij} between the scheduled flight time from city i to city j and that for the opposite journey can be taken as an odd function of $(a_i - a_j)$, where a_i denotes the longitude of city i. As one parametric example, we may take the general member $\mathbf{K}$ of $\mathscr{K}' = \mathscr{K}'(\boldsymbol{\theta}_\mathbf{K})$ as having $(i,j)^{th}$ element $k_{ij} = \gamma(a_i - a_j)$, so that $\theta_{\mathbf{K}} = \gamma$. Introducing additional parameters, essentially the same model can be elaborated to include estimation of an *unknown* prevailing wind direction.

Separate minimum distance estimation of $\mathbf{H}_{\mathbf{S_M}}$ by $\mathscr{H}'$, and of $\mathbf{K_M}$ by $\mathscr{K}'$, speeds computation of the optimal estimates of $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\mathbf{H}}^\top, \boldsymbol{\theta}_{\mathbf{K}}^\top)^\top$. This computational gain is, in general, greater the larger the number of components being fitted and the more complicated the parametric functional forms involved.

Again, this separation into sub-problems allows (estimation of) the decomposition $\sigma^2 = \sigma^2_{\mathbf{H}} + \sigma^2_{\mathbf{K}}$. Under Gaussianity, $\widehat{\boldsymbol{\theta}}_{\mathbf{H}}$ and $\widehat{\boldsymbol{\theta}}_{\mathbf{K}}$ are independent. Accordingly, so too are $\widehat{\sigma}^2_{\mathbf{H}}$ and $\widehat{\sigma}^2_{\mathbf{K}}$.

2.2 Independent Component Analysis and Invariant Coordinate Selection

A somewhat different use of the $\langle\cdot,\cdot\rangle_E$-orthogonal decomposition $\mathscr{M} = \mathscr{S} \oplus \mathscr{K}$ arises in connection with Independent Component Analysis – abbreviated, ICA – see, for example, Oja et al. (2006); and with Invariant Coordinate Selection – abbreviated ICS – see Tyler et al. (2009).

ICA is a highly popular method within many applied areas. Its principal objective is to recover, as far as is possible, unobserved independent components from an observable random vector arising as an unknown (possibly shifted) convolution of them. Oja et al. (2006) were the first to point out that, under appropriate modelling assumptions, this recovery – or 'unmixing' – problem can be solved via what is now known as ICS.

ICS is a general method for exploring multivariate data, based on two scatter matrices. Scatter matrices – generalisations of the usual covariance operator – are matrix-valued functionals $\mathbf{S} = \mathbf{S}(F_{\mathbf{X}})$ of the distribution of a random vector $\mathbf{X}$ that are affine equivariant. That is, for any nonsingular matrix $\mathbf{C}$ and any vector $\mathbf{c}$, $\mathbf{X} \to \mathbf{X}^* := \mathbf{CX} + \mathbf{c}$ induces $\mathbf{S} \to \mathbf{S}^* = \mathbf{CSC}^\top$. One of these matrices, denoted $\mathbf{V}$ below, is required to be positive definite symmetric.

ICS works by transforming $\mathbf{X} \to \mathbf{Z} = \mathbf{M}^\top\mathbf{X}$, where $\mathbf{M} = \mathbf{UV}^{-1/2}$ with $\mathbf{U}$ an orthogonal matrix determined jointly by the two scatter matrices. The elements of $\mathbf{Z}$ are called 'invariant coordinates' to connote the fact that any nonsingular affine transformation $\mathbf{X} \to \mathbf{X}^*$ leaves $\mathbf{Z}$ unchanged (up to a shift term that vanishes if and only if $\mathbf{c} = 0$).

As Oja showed, for suitable ICA models, we have the striking result that *the invariant coordinates correspond to the independent components*. The question then naturally arises as to whether a pair of scatter matrices $(\mathbf{V}, \widetilde{\mathbf{V}})$ can be chosen that is, in some sense, optimal in this regard. Here, fixing $\mathbf{V}$ as the usual covariance operator, we note the following. Ilmonen et al. (2010) establish that, in the standard case,

$$\sqrt{n}(\mathbf{V} - \mathbf{I}) = O_P(1) \text{ and } \sqrt{n}(\mathbf{U} - \mathbf{I}) = O_P(1). \tag{8}$$

Recent correspondence with that paper's final author, Hannu Oja, notes that (8) implies

$$\sqrt{n}(\mathbf{S_M}-\mathbf{I}) = -\tfrac{1}{2}\sqrt{n}(\mathbf{V}-\mathbf{I}) + o_P(1) \text{ and } \sqrt{n}\mathbf{K_M} = \sqrt{n}(\mathbf{U}-\mathbf{I}) + o_P(1). \quad (9)$$

This may have implications for comparing two estimates obtained with the same $\mathbf{V}$. Relations (9) suggest that asymptotic comparison of estimates obtained with fixed $\mathbf{V}$ but different $\widetilde{\mathbf{V}}$ – hence, different $\mathbf{U}$ – should be based on the upper triangular part of $\mathbf{U}$ only (and not on the full information contained in $\mathbf{M}$). In contrast, currently, such comparison is usually made based on the off-diagonal elements of $\mathbf{M}$, introducing an unnecessary confounding with its symmetric part.

Overall, it will be of interest to explore potential methodological benefits flowing from the direct sum $\mathscr{M} = \mathscr{S} \oplus \mathscr{K}$ underpinning (9). In this connection, note also that the decomposition $\mathscr{M} = \mathscr{D} \oplus \mathscr{H} \oplus \mathscr{K}$ can be used to refine (9). And that the sharpening by orthogonality of these direct sum relations in Propositions 1 and 2, and their generalisations to other inner products established in Section 4.2 below, may bring additional benefits. In particular, it may be possible to helpfully link the choice of inner product to the asymptotic precision (inverse covariance) matrices of the estimators involved.

3 Alternative inner products

Orthogonality – in the standard Euclidean inner product $\langle\cdot,\cdot\rangle_E$ – is central to all the key results above and, thus, to their sphere of applicability. At the same time, the (often, implicit) assumption of uncorrelated homoscedastic errors being made is, as always, open to question.

Theoretical inquistiveness combines, then, with a desire for the widest possible realm of practical applications to point to the importance of finding answers to the following, natural, questions: for what *other* inner products do Propositions 1 and 2 hold? More generally (*cf*. Proposition 3), given a collection $\{\mathscr{L}_h\}$ of nontrivial subspaces with direct sum $\mathscr{M}$, any two of which are orthogonal in $(\mathscr{M}, \langle\cdot,\cdot\rangle_E)$, for what *other* inner products does such pairwise orthogonality hold? (Specific proposals for such alternative inner products, in a variety of contexts, are discussed in Section 4.3.)

In order to answer these questions, specific to the vector (sub)spaces introduced in Section 2, it will be sufficient – and actually straightforward – to answer the following generic question. Let $\mathscr{L}$ be a real linear (vector) space with finite dimension $d \geq 2$. Let now $\{\mathscr{L}_h\}_{h=1}^r$ $(2 \leq r \leq d)$ be a collection of nontrivial subspaces of $\mathscr{L}$ whose direct sum, $\oplus\{\mathscr{L}_h\}$, is $\mathscr{L}$. That is,

$$\mathscr{L}_h \cap \mathscr{L}_{h'} = \{\mathbf{0}_{\mathscr{L}}\}, (h \neq h'), \text{ while } \mathscr{L} = \textstyle\sum_h \mathscr{L}_h := \{\mathbf{l}_1 + ... + \mathbf{l}_m : \mathbf{l}_h \in \mathscr{L}_h\}, \quad (10)$$

(so that $d = \sum_h d_h$, where $\mathscr{L}_h$ has dimension $d_h > 0$).

The question posed is: in which inner product spaces $(\mathscr{L}, \langle\cdot,\cdot\rangle)$ are the $\{\mathscr{L}_h\}$ pairwise orthogonal? That is, for which inner products does $\langle \mathbf{l}_h, \mathbf{l}_{h'}\rangle$ vanish for each $\mathbf{l}_h \in \mathscr{L}_h$ and $\mathbf{l}_{h'} \in \mathscr{L}_{h'}$, $(h \neq h')$? As will be clear, the answer consists, essentially, in taking *ordered* bases.

Recall that, if $\mathscr{L}^*$ is also a real linear space, $T : \mathscr{L} \to \mathscr{L}^*$ is called an isomorphism if it is one-to-one, onto and linear: that is, for all α_1, α_2 in $\mathscr{R}$ and all $\mathbf{l}_1, \mathbf{l}_2$ in $\mathscr{L}$,

$$T(\alpha_1 \mathbf{l}_1 + \alpha_2 \mathbf{l}_2) = \alpha_1 T(\mathbf{l}_1) + \alpha_2 T(\mathbf{l}_2).$$

$\mathscr{L}$ and $\mathscr{L}^*$ are called isomorphic if such a map exists, and we write $\mathscr{L} \cong \mathscr{L}^*$ (via T).

Proposition 4. *Let* $\{\mathbf{b}_1, ..., \mathbf{b}_d\}$ *be a basis for* $\mathscr{L}$*. Then,* $\mathscr{L} \cong \mathscr{R}^d$ *via* $\mathbf{l} \to \boldsymbol{\xi}$ *where, to each* $\mathbf{l} = \sum_i \xi_i \mathbf{b}_i \in \mathscr{L}$*, we associate* $\boldsymbol{\xi} = (\xi_1, ..., \xi_d)^\top \in \mathscr{R}^d$.

Recall also that an inner product $\langle\cdot,\cdot\rangle$ on $\mathscr{L}$ is a function from $\mathscr{L}^2$ to $\mathscr{R}$ such that for all $\mathbf{l}, \mathbf{l}', \mathbf{l}_1, \mathbf{l}_2$ in $\mathscr{L}$ and α_1, α_2 in $\mathscr{R}$:

1. $\langle \mathbf{l}, \mathbf{l}'\rangle = \langle \mathbf{l}', \mathbf{l}\rangle$
2. $\langle \alpha_1 \mathbf{l}_1 + \alpha_2 \mathbf{l}_2, \mathbf{l}'\rangle = \alpha_1 \langle \mathbf{l}_1, \mathbf{l}'\rangle + \alpha_2 \langle \mathbf{l}_2, \mathbf{l}'\rangle$
3. $\langle \mathbf{l}, \mathbf{l}\rangle \geq 0$, equality holding if and only if $\mathbf{l} = \mathbf{0}_{\mathscr{L}}$.

Being bilinear, an inner product is completely determined by its matrix of values $(\langle \mathbf{b}_i, \mathbf{b}_j\rangle)$ on an ordered basis $(\mathbf{b}_1, \ldots, \mathbf{b}_d)$. Accordingly, denoting by $\mathscr{V}_d$ the set of all real, positive definite symmetric matrices of order d, the following result is immediate.

Proposition 5. *For any ordered basis* $\mathbf{b}(\mathscr{L}) = (\mathbf{b}_1, \ldots, \mathbf{b}_d)$ *for* $\mathscr{L}$*,* $\langle\cdot,\cdot\rangle \to (\langle \mathbf{b}_i, \mathbf{b}_j\rangle)$ *is a one-to-one correspondence between the set of all inner products on* $\mathscr{L}$ *and* $\mathscr{V}_d$.

Recall further that an inner product isomorphism between $(\mathscr{L}, \langle\cdot,\cdot\rangle)$ and $(\mathscr{L}^*, \langle\cdot,\cdot\rangle^*)$ is an isomorphism $T : \mathscr{L} \to \mathscr{L}^*$ which preserves inner products: that is, such that for all $\mathbf{l}_1, \mathbf{l}_2$ in $\mathscr{L}$: $\langle T(\mathbf{l}_1), T(\mathbf{l}_2)\rangle^* = \langle \mathbf{l}_1, \mathbf{l}_2\rangle$. $(\mathscr{L}, \langle\cdot,\cdot\rangle)$ and $(\mathscr{L}^*, \langle\cdot,\cdot\rangle^*)$ are called isomorphic if such a map exists, and we write $(\mathscr{L}, \langle\cdot,\cdot\rangle) \cong (\mathscr{L}^*, \langle\cdot,\cdot\rangle^*)$ (via T). Combining Propositions 4 and 5, we have:

Proposition 6. *Let* $\mathbf{b}(\mathscr{L}) = (\mathbf{b}_1, \ldots, \mathbf{b}_d)$ *be an ordered basis for* $\mathscr{L}$*, let* $\langle\cdot,\cdot\rangle$ *be an inner product on* $\mathscr{L}$*, and let* $\mathbf{V} = (\langle \mathbf{b}_i, \mathbf{b}_j\rangle)$*. Then* $(\mathscr{L}, \langle\cdot,\cdot\rangle) \cong (\mathscr{R}^d, \langle\cdot,\cdot\rangle_{\mathbf{V}})$ *via* $T_{\mathbf{b}(\mathscr{L})} : \mathbf{l} \to \boldsymbol{\xi}$*, where* $\langle \boldsymbol{\xi}, \boldsymbol{\eta}\rangle_{\mathbf{V}} := \boldsymbol{\xi}^\top \mathbf{V} \boldsymbol{\eta}$.

Denoting by $\mathscr{Q}_d$ (respectively: $\mathscr{D}_d^+$) the set of all orthogonal (respectively: diagonal, positive definite) matrices of order d, and by $diag(\cdot, \ldots, \cdot)$ the blockdiagonal matrix with the diagonal entries listed, the simple answer to the generic question posed is:

Theorem 1. *Let* $\mathscr{L}$ *and* $\{\mathscr{L}_h\}_{h=1}^r$ $(2 \leq r \leq d)$ *be as in (10). Let* $(\mathbf{b}_1, ..., \mathbf{b}_d)$ *be an ordered basis for* $\mathscr{L}$ *comprising, successively, ordered bases for each of* $\mathscr{L}_1, ..., \mathscr{L}_r$*. Then, the following are equivalent:*

1. *the* $\{\mathscr{L}_h\}$ *are pairwise orthogonal in* $(\mathscr{L}, \langle\cdot,\cdot\rangle)$,

2. $(\langle \mathbf{b}_i, \mathbf{b}_j \rangle) = diag(\mathbf{V}_1, ..., \mathbf{V}_r)$ *for some* $\mathbf{V}_h \in \mathscr{V}_{d_h}$, $h = 1, ..., r$,

3. $(\langle \mathbf{b}_i, \mathbf{b}_j \rangle) = \mathbf{Q}\boldsymbol{\Delta}\mathbf{Q}^T$ *for some* $\mathbf{Q} = diag(\mathbf{Q}_1, ..., \mathbf{Q}_r)$ *and* $\boldsymbol{\Delta} = diag(\boldsymbol{\Delta}_1, ..., \boldsymbol{\Delta}_r)$, *where* $\mathbf{Q}_h \in \mathscr{Q}_{d_h}$ *and* $\boldsymbol{\Delta}_h \in \mathscr{D}^+_{d_h}$, $h = 1, ..., r$.

Theorem 1 is couched, of course, in terms of a chosen basis $(\mathbf{b}_1, \ldots, \mathbf{b}_d)$. To see what form $\mathbf{V} = (\langle \mathbf{b}_i, \mathbf{b}_j \rangle)$ takes in a general basis, it suffices to use the following. Bilinearity of inner products gives at once:

Proposition 7. *For any inner product* $\langle\cdot,\cdot\rangle$ *on* $\mathscr{L}$, *and for any ordered basis* $(\mathbf{b}_1, \ldots, \mathbf{b}_d)$ *of* $\mathscr{L}$, *let* $\mathbf{V} := (\langle \mathbf{b}_i, \mathbf{b}_j \rangle)$. *For any other ordered basis* $(\mathbf{b}^*_1, \ldots, \mathbf{b}^*_d)$ *of* $\mathscr{L}$,

$$\mathbf{V}^* := (\langle \mathbf{b}^*_i, \mathbf{b}^*_j \rangle) = \mathbf{A}\mathbf{V}\mathbf{A}^\top$$

where the nonsingular matrix $\mathbf{A} = (a_{ij})$ *is given by* $\mathbf{b}^*_i = \sum_j a_{ij}\mathbf{b}_j$.

4 Matrix decomposition examples revisited

In the wider context of inner products preserving the same pairwise orthogonalities, we revisit here the analysis of asymmetry matrix decomposition examples reviewed in Section 2.1.

We first introduce (Section 4.1) two alternative ordered bases for $\mathscr{M}_k$, and then (Section 4.2) use them to generalise two key results underpinning the standard Euclidean analysis of asymmetry, viz.: Proposition 1 (Section 2.1.1) and Proposition 2 (Section 2.1.2). Section 4.3 indicates applications of these more general results.

4.1 Two alternative ordered bases for $\mathscr{M}_k$

In this asymmetry context, rather than the usual stacking-by-columns operator, it is convenient to order the elements of $\mathbf{M} \in \mathscr{M}_k$ by first listing those on the diagonal; then, those above the diagonal, in row-wise order; and, finally, those below the diagonal, in column-wise order. In this way, we define the *veck* operator by

$$\begin{aligned} veck(\mathbf{M}) = (&m_{11}, \ldots, m_{kk}; \\ &m_{12}, \ldots, m_{1k}, m_{23}, \ldots, m_{k-1,k}; \\ &m_{21}, \ldots, m_{k1}, m_{32}, \ldots, m_{k,k-1})^\top. \end{aligned}$$

We use relevant parts of this order to define ordered bases of $\mathscr{D}$, $\mathscr{H}$, $\mathscr{S}$, $\mathscr{K}$ and $\mathscr{M} = \mathscr{M}_k$ itself in terms of the canonical basis $\{\mathbf{E}_{ij} := \mathbf{e}_i\mathbf{e}_j^\top, 1 \leq i, j \leq k\}$ of $\mathscr{M}$.

Here, $(\mathbf{e}_1,\ldots,\mathbf{e}_k)$ is the usual ordered basis of $\mathscr{R}^k$, so that $\mathbf{e}_i$ is a binary vector with a single one in the i^{th} position. For notational convenience, the constant ρ below is $1/\sqrt{2}$.

Putting $\mathbf{D}_{ii} := \mathbf{E}_{ii}$, $\mathbf{b}(\mathscr{D}) := (\mathbf{D}_{11},\ldots,\mathbf{D}_{kk})$ is an ordered basis for $\mathscr{D}$. Again, with $\mathbf{H}_{ij} := \rho(\mathbf{E}_{ij}+\mathbf{E}_{ji})$, $i<j$, $\mathbf{b}(\mathscr{H}) := (\mathbf{H}_{12},\ldots,\mathbf{H}_{k-1,k})$ is an ordered basis for $\mathscr{H}$. Concatenating these, in a mild abuse of notation, $\mathbf{b}(\mathscr{S}) := (\mathbf{b}(\mathscr{D});\mathbf{b}(\mathscr{H}))$ is an ordered basis for $\mathscr{S}$. Finally, with $\mathbf{K}_{ij} := \rho(\mathbf{E}_{ij}-\mathbf{E}_{ji})$, $i<j$, $\mathbf{b}(\mathscr{K}) := (\mathbf{K}_{12},\ldots,\mathbf{K}_{k-1,k})$ is an ordered basis for $\mathscr{K}$, so that $\mathbf{b}(\mathscr{M}) := (\mathbf{b}(\mathscr{S});\mathbf{b}(\mathscr{K}))$ is an ordered basis for $\mathscr{M}$.

A natural alternative is to use the corresponding ordering of the canonical basis itself:

$$\mathbf{b}^*(\mathscr{M}) = (\mathbf{E}_{11},\ldots,\mathbf{E}_{kk};\mathbf{E}_{12},\ldots,\mathbf{E}_{k-1,k};\mathbf{E}_{21},\ldots,\mathbf{E}_{k,k-1})$$

which corresponds directly to the data we actually observe, since:

$$\text{each } \mathbf{M}\in\mathscr{M}_k \text{ has direct decomposition } \mathbf{M} = \textstyle\sum_{i,j} m_{ij}\mathbf{E}_{ij}. \tag{11}$$

By design, $\mathbf{b}(\mathscr{M})$ and $\mathbf{b}^*(\mathscr{M})$ are ordered *orthonormal* bases in $(\mathscr{M},\langle\cdot,\cdot\rangle_E)$. Since $\mathbf{E}_{ij} = \rho(\mathbf{H}_{ij}+\mathbf{K}_{ij})$ while $\mathbf{E}_{ji} = \rho(\mathbf{H}_{ij}-\mathbf{K}_{ij})$, $i<j$, the orthogonal matrix $\mathbf{A}$ linking these two bases via $\mathbf{b}_i^*(\mathscr{M}) = \sum_j a_{ij}\mathbf{b}_j(\mathscr{M})$ is given by:

$$\mathbf{A} = \begin{pmatrix} \mathbf{I}_k & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \rho\mathbf{I}_{\binom{k}{2}} & \rho\mathbf{I}_{\binom{k}{2}} \\ \mathbf{0} & \rho\mathbf{I}_{\binom{k}{2}} & -\rho\mathbf{I}_{\binom{k}{2}} \end{pmatrix} = diag(\mathbf{I}_k, \mathbf{B}\otimes\mathbf{I}_{\binom{k}{2}}), \tag{12}$$

$\otimes$ denoting the Kronecker product, so that $\mathbf{A}$ (hence, $\mathbf{B}$) is also symmetric, the matrix

$$\mathbf{B} = \rho\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \rho\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\cdot\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\cdot\rho\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

combining, for each $i<j$, a 45° rotation and a reflection in the two-dimensional subspace spanned by $(\mathbf{H}_{ij},\mathbf{K}_{ij})$; equivalently, by $(\mathbf{E}_{ij},\mathbf{E}_{ji})$. Note that, each being orthogonal and symmetric, $\mathbf{A}$ and $\mathbf{B}$ are self-inverse.

From now on, for any inner product $\langle\cdot,\cdot\rangle$ on $\mathscr{M}$, we define $\mathbf{V}\equiv\mathbf{V}_{(\mathscr{M},\langle\cdot,\cdot\rangle)}$ and $\mathbf{V}^*\equiv\mathbf{V}^*_{(\mathscr{M},\langle\cdot,\cdot\rangle)}$ thus:

$$\mathbf{V} := (\langle\mathbf{b}_i(\mathscr{M}),\mathbf{b}_j(\mathscr{M})\rangle) \text{ and } \mathbf{V}^* := (\langle\mathbf{b}_i^*(\mathscr{M}),\mathbf{b}_j^*(\mathscr{M})\rangle), \tag{13}$$

so that, by Proposition 7, $\mathbf{V}^* = \mathbf{A}\mathbf{V}\mathbf{A}^T$; equivalently, as $\mathbf{A}^{-1} = \mathbf{A}$, $\mathbf{V} = \mathbf{A}\mathbf{V}^*\mathbf{A}^T$.

Proposition 6 and display (11) give at once

Proposition 8. $\mathbf{M}\to\mathbf{m} := veck(\mathbf{M})$ *is an inner product isomorphism between* $(\mathscr{M}_k,\langle\cdot,\cdot\rangle)$ *and* $(\mathscr{R}^{k^2},\langle\cdot,\cdot\rangle_{\mathbf{V}^*})$, *where* $\langle\mathbf{m}_1,\mathbf{m}_2\rangle_{\mathbf{V}^*} := \mathbf{m}_1^\top\mathbf{V}^*\mathbf{m}_2$.

In particular, $(\mathscr{M}_k,\langle\cdot,\cdot\rangle_E)\cong(\mathscr{R}^{k^2},\langle\cdot,\cdot\rangle_{\mathbf{I}})$ *via* $\mathbf{M}\to veck(\mathbf{M})$.

The standard Euclidean inner product exploited in Propositions 1 and 2 corresponds, then, to taking $\mathbf{V}^* = \mathbf{I}$; equivalently, $\mathbf{V} = \mathbf{I}$. We show next exactly what other choices Theorem 1 allows.

4.2 Generalisations of Propositions 1 and 2

Recall the matrices $\mathbf{A}$, $\mathbf{V}$ and $\mathbf{V}^*$ defined in (12) and (13) above.

Considering first generalisations of Proposition 1, Theorem 1 gives at once that $\mathscr{S}$ and $\mathscr{K}$ are orthogonal in $(\mathscr{M}, \langle \cdot, \cdot \rangle)$ if and only if

$$\mathbf{V} = diag(\mathbf{V_S}, \mathbf{V_K}) \text{ for some } \mathbf{V_S} \in \mathscr{V}_{\binom{k+1}{2}} \text{ and } \mathbf{V_K} \in \mathscr{V}_{\binom{k}{2}}.$$

Partitioning any such $\mathbf{V}$ conformably with $\mathbf{A}$, we write it as

$$\mathbf{V} = \begin{pmatrix} \mathbf{V_D} & \mathbf{V_{DH}} & \mathbf{0} \\ \mathbf{V}_{\mathbf{DH}}^{\top} & \mathbf{V_H} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V_K} \end{pmatrix}. \tag{14}$$

Concerning generalisations of Proposition 2, Theorem 1 again gives at once that $\mathscr{D}$, $\mathscr{H}$ and $\mathscr{K}$ are pairwise orthogonal in $(\mathscr{M}, \langle \cdot, \cdot \rangle)$ if and only if $\mathbf{V}$, as in (14), has $\mathbf{V_{DH}} = \mathbf{0}$, so that $\mathbf{V_D} \in \mathscr{V}_k$ and $\mathbf{V_H} \in \mathscr{V}_{\binom{k}{2}}$. For general $\mathbf{V_{DH}}$ – see, for example, Theorem 7.7.6 in (Horn and Johnson, 1985, pp. 472) – we have that

Proposition 9. *The following are equivalent:*

1. $\mathbf{V_S} \equiv \begin{pmatrix} \mathbf{V_D} & \mathbf{V_{DH}} \\ \mathbf{V}_{\mathbf{DH}}^{\top} & \mathbf{V_H} \end{pmatrix} \in \mathscr{V}_{\binom{k+1}{2}}$
2. $\mathbf{V_D} \in \mathscr{V}_k$ *and* $[\mathbf{V_H} - \mathbf{V}_{\mathbf{DH}}^{\top} \mathbf{V}_{\mathbf{D}}^{-1} \mathbf{V_{DH}}] \in \mathscr{V}_{\binom{k}{2}}$
3. $\mathbf{V_H} \in \mathscr{V}_{\binom{k}{2}}$ *and* $[\mathbf{V_D} - \mathbf{V_{DH}} \mathbf{V}_{\mathbf{H}}^{-1} \mathbf{V}_{\mathbf{DH}}^{\top}] \in \mathscr{V}_k$.

It only remains to find the form of $\mathbf{V}^*$. Using Propositions 7 and 9, and recalling that $\rho = 1/\sqrt{2}$, (12) gives

Theorem 2. *$\mathscr{S}$ and $\mathscr{K}$ are orthogonal in $(\mathscr{M}, \langle \cdot, \cdot \rangle)$ if and only if*

$$\mathbf{V}^* = \begin{pmatrix} \mathbf{V_D} & \rho \mathbf{V_{DH}} & \rho \mathbf{V_{DH}} \\ \rho \mathbf{V}_{\mathbf{DH}}^{\top} & \rho^2 (\mathbf{V_H} + \mathbf{V_K}) & \rho^2 (\mathbf{V_H} - \mathbf{V_K}) \\ \rho \mathbf{V}_{\mathbf{DH}}^{\top} & \rho^2 (\mathbf{V_H} - \mathbf{V_K}) & \rho^2 (\mathbf{V_H} + \mathbf{V_K}) \end{pmatrix} \tag{15}$$

for some $\mathbf{V_K} \in \mathscr{V}_{\binom{k}{2}}$, $\mathbf{V_H} \in \mathscr{V}_{\binom{k}{2}}$ *and* $\mathbf{V_{DH}}$ *as in (14) with* $[\mathbf{V_D} - \mathbf{V_{DH}} \mathbf{V}_{\mathbf{H}}^{-1} \mathbf{V}_{\mathbf{DH}}^{\top}] \in \mathscr{V}_k$, *$\mathscr{D}$ and $\mathscr{H}$ also being orthogonal if and only if* $\mathbf{V_{DH}} = \mathbf{0}$.

Theorem 2 has two immediate corollaries. By design, pairwise orthogonality in $(\mathscr{M}, \langle \cdot, \cdot \rangle)$ of the specified set of subspaces, $\{\mathscr{S}, \mathscr{K}\}$ or $\{\mathscr{D}, \mathscr{H}, \mathscr{K}\}$, is reflected in the corresponding blockdiagonal form of $\mathbf{V}$. The first corollary characterises when $\mathbf{V}^*$ enjoys the *same* form of blockdiagonal structure, showing indeed that this happens if and only if $\mathbf{V}^* = \mathbf{V}$.

Corollary 1. *Let $\langle\cdot,\cdot\rangle$ be any inner product on $\mathscr{M}$ for which $\mathscr{S}$ and $\mathscr{K}$ are orthogonal. Then, equivalent are:*

1. $\mathbf{V}^* = diag(\mathbf{V}_a, \mathbf{V}_b)$ *for some* $\mathbf{V}_a \in \mathscr{V}_{\binom{k+1}{2}}$ *and* $\mathbf{V}_b \in \mathscr{V}_{\binom{k}{2}}$
2. $\mathbf{V}^* = diag(\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3)$ *for some* $\mathbf{V}_1 \in \mathscr{V}_k$ *and* $\mathbf{V}_2, \mathbf{V}_3 \in \mathscr{V}_{\binom{k}{2}}$
3. $\mathbf{V}^* = diag(\mathbf{V}_\mathbf{D}, \overline{\mathbf{V}}, \overline{\mathbf{V}})$ *for some* $\overline{\mathbf{V}} \in \mathscr{V}_{\binom{k}{2}}$
4. $\mathbf{V}_{\mathbf{DH}} = \mathbf{0}$ *and* $\mathbf{V}_\mathbf{H} = \mathbf{V}_\mathbf{K}$
5. $\mathbf{V}^* = \mathbf{V}$.

The second, relatively trivial, corollary characterises when $\mathbf{V}^*$ is fully diagonal, so that the inner product on $\mathscr{M}$ corresponds to a set of weights: $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i,j} w_{ij} a_{ij} b_{ij}$ for some positive $\{w_{ij}\}$. We have that the only requirement for orthogonality of $\mathscr{S}$ and $\mathscr{K}$ is the symmetry condition $w_{ji} = w_{ij}$, $i < j$. That is:

Corollary 2. *Let $\langle\cdot,\cdot\rangle$ be any inner product on $\mathscr{M}$ for which $\mathscr{S}$ and $\mathscr{K}$ are orthogonal. Then, equivalent are:*

1. $\mathbf{V}^* \in \mathscr{D}^+_{k^2}$
2. $\mathbf{V}^* = diag(\boldsymbol{\Delta}_1, \overline{\boldsymbol{\Delta}}, \overline{\boldsymbol{\Delta}})$ *for some* $\boldsymbol{\Delta}_1 \in \mathscr{D}^+_k$ *and* $\overline{\boldsymbol{\Delta}} \in \mathscr{D}^+_{\binom{k}{2}}$.

4.3 Applications

Theorems 1 and 2 determine all possible choices of $\mathbf{V}$ – equivalently, of $\mathbf{V}^*$ – consistent with the orthogonalities required for use in generalising Proposition 1 or 2. In practice, this choice of inner product will depend on context. We briefly indicate three ways in which it may be made.

One way is to allow for alternative forms of the covariance of the observed data. Recall first the vector case, reviewed briefly in the Introduction. Here, standard ANOVA methods – those based on Euclidean distances, induced by $\langle .,. \rangle_\mathbf{I}$ – are appropriate when $cov(\mathbf{y})$ is taken as proportional to $\mathbf{I}$. The move from ordinary to generalised least-squares ANOVA methods is signalled whenever, in contrast, error terms are taken as correlated and/or heteroscedastic. Effectively, in this case, Euclidean distances are replaced by their Mahalanobis counterparts, induced by $\langle .,. \rangle_{\mathbf{V_y}}$ where $\mathbf{V}_y = (cov(\mathbf{y}))^{-1}$ is assumed known up to a positive scalar multiple. Similarly, in the matrix case reviewed in Section 2.1.4, standard ANOVA methods – those based on distances induced by $\langle .,. \rangle_E$ – are appropriate for uncorrelated homoscedastic errors. However, a generally appropriate choice is to take $\mathbf{V}^*$, of the form given in Theorem 2, as a positive scalar multiple of the inverse of an assumed (asymptotic) covariance matrix for $veck(\mathbf{M})$. In particular, weighted least-squares estimation is appropriate in the uncorrelated but heteroscedastic case, when we may take $\mathbf{V}^*$ to have the diagonal form characterised in Corollary 2.

An alternative way of choosing an inner product focuses on $\mathbf{V}$. Recall that a non-singular covariance matrix has blockdiagonal form if and only if the same is true of its inverse, the number and size of the blocks being the same in both cases. A second choice is then to model the blockdiagonal form of $\mathbf{V}$ required by orthogonality as reflecting uncorrelatedness of the corresponding parts of $\mathbf{M}$. Corollary 1 shows that this happens if and only if $\mathbf{V_{DH}} = \mathbf{O}$ and $\mathbf{V_H} = \mathbf{V_K}$. Equivalently, when $\mathbf{V}^* = \mathbf{V}$. This modelling choice is very much in the same 'analyse separately' spirit as the seminal Constantine and Gower (1978) paper. The within-subspace dispersions may then be modelled according to context.

Finally, under multivariate normality – and so, in many asymptotic situations – zeroes in precision (inverse covariance) matrices have particular significance in terms of conditional independence. Depending on context, such conditional independence considerations may further guide appropriate choice of $\mathbf{V}$ or $\mathbf{V}^*$. In particular, there may be useful applications here to graphical model contexts.

5 Acknowledgement

The UK-based authors thank the EPSRC for their support under grant EP/L010429/1.

References

Constantine, A. G. and Gower, J. C.: Graphical Representation of Asymmetric Matrices. J. Royal Statistical Society, Series C. **27**, 297-304 (1978)

Gower, J. C.: Skew symmetry in retrospect. Advances in Data Analysis and Classification (2014) doi: 10.1007/s11634-014-0181-7

Horn, R. A. and Johnson, C. R.: Matrix Analysis. Cambridge University Press, (1985)

Ilmonen, P., Nevalainen J. and Oja, H.: Characteristics of multivariate distributions and the invariant coordinate system. Statistics and Probability Letters. **80**, 1844-1853 (2010)

Oja, H., Sirkiä, S. and Eriksson, J.: Scatter matrices and independent component analysis. Austrian Journal of Statistics. **35**, 175-189 (2006)

Tyler, D. E., Critchley, F., Dümbgen, L. and Oja, H.: Invariant Coordinate Selection (with Discussion). J. Royal Statistical Society, Series B. **71**, 549-592 (2009)