INTERPRETATION OF EIGENVALUES IN MULTIVARIATE STATISTICAL ANALYSIS

STEEN A. ANDERSSON[a] AND JESSE B. CRAWFORD[b]

December 07, 2007

Technical Report 07-02
Department of Statistics
Indiana University
Bloomington, IN 47405

[a] Indiana University
[b] Indiana University

INTERPRETATION OF EIGENVALUES IN MULTIVARIATE STATISTICAL ANALYSIS

STEEN A. ANDERSSON◇ AND JESSE B. CRAWFORD
INDIANA UNIVERSITY

Abstract. In this paper, we define general canonical correlations, which generalize the canonical correlations developed by Hotelling, and general canonical covariate pairs, the corresponding linear statistic. We also define canonical variance distances with corresponding canonical distance variates. In a rather broad setting, these parameters and their corresponding linear statistics are characterized in terms of certain eigenvalues and eigenvectors. Specifically, these are the eigenvalues used to represent the maximal invariant statistic for seven of the ten testing problems discussed in Andersson, Brøns, and Jensen (1983). Additional observations regarding these ten problems are made at the end of this paper.

1. Introduction.

The canonical correlations and canonical covariates were introduced by Hotelling (1936) with regards to testing independence of two jointly normal multivariate random variables.

Following the classical evaluation of this subject in f.ex. Anderson (1984), page 480-497, the canonical correlations* are first obtained as a solution to a matrix eigenvalue problem, and the canonical covariates are obtained as solutions to the corresponding linear equations. The central distribution of the empirical ordered canonical correlations is then derived in terms of a density with respect to (wrt.) the Lebesgue measure†.

This testing problem is also invariant under actions of a group on the observation space and on the parameter sets, and the ordered empirical canonical correlations are a maximal invariant statistic. All invariant test statistics must factor through the empirical maximal invariant statistic. Thus, the representations of central and non-central distributions of the ordered canonical correlations are important in the study of properties of invariant test statistics.

In Andersson *et al.* (1983), abbreviated (ABJ), ten fundamental testing problems are treated in a unified manner suggested by the general theory of normal models where the covariance matrix is invariant under a compact group, cf. Andersson and

---

Date: December, 2007

2000 *Mathematics Subject Classification.* Primary 62-09, 62H05, 62H99; Secondary 06F99, 15A30.

Key words and phrases. eigenvalues, eigenvectors, variance residual, canonical correlations, canonical covariates, interpretation.

◇Research supported by U.S. National Security Agency Grant number H98230-06-1-0054.

*theoretical as well as empirical

†The non-central distributions can also be described in terms of a multiplicative correction factor to the central density involving hypergeometric functions or zonal polynomials

Madsen (1998), Appendix A, abbreviated (AM). Each of the ten testing problems is invariant under a group of linear transformations. For each of the ten cases, (ABJ) found a concrete representation of the orbit projection, i.e., a maximal invariant statistic, and a representation of its central distribution in terms of a density wrt. the Lebesgue measure. Each orbit projection is represented in terms of ordered eigenvalues of a symmetric matrix wrt. a positive definite matrix, both with certain additional structures. More precisely, the representations involve simultaneous reduction of quadratic forms on vector spaces.
The classical test for independence mentioned above is one of the ten testing problems, and the classical ordered canonical correlations are equal to the eigenvalues used in (ABJ) to represent the maximal invariant statistic. This raises the following question: can the eigenvalues used to represent maximal invariants for the other problems also be expressed in terms of correlations? For seven of the fundamental testing problems, we show that the eigenvalues are equal to generalized canonical correlations, and the generalized canonical covariates are characterized in terms of the corresponding eigenvectors.
These *general canonical correlations* and *general canonical covariate pairs* are defined in the context of testing a covariance structure of a normal distribution against a more general covariance structure. In the same context, we will define another finite family of parameters, the *canonical variance distances*, whose corresponding linear statistics are called *canonical distance variates*.
Although not a subject for this paper, it is also interesting to find the non-central distributions by finding more explicit expressions for the correction factors, cf. f.ex. Andersson (1982), Berthelsen(1989, 1990), and the references herein.

2. The general testing problem.

We shall start with a simple situation in a classical formulation. Let I be a finite set[‡], and consider any vector $z \equiv (z_i|i \in I) \in \mathbb{R}^I$ as an I-column vector. Let $X \equiv (X_\nu|\nu \in N)$ be a family of i.i.d. random variables indexed by a finite set N, each with values in $\mathbb{R}^I$, and each following a multivariate normal distribution on $\mathbb{R}^I$ with expectation $0 \in \mathbb{R}^I$ and unknown $I \times I$ (co)variance matrix $\Sigma \in \mathbf{PD}(I)$, where $\mathbf{PD}(I)$ denotes the set of $I \times I$ positive definite matrices. These assumptions constitute a statistical hypothesis H. A subset $\Theta_0 \subset \mathbf{PD}(I)$ induces a subhypothesis H_0 and thus a testing problem: H_0: $\Sigma_0 \in \Theta_0$ vs. H: $\Sigma \in \mathbf{PD}(I)$. The ML estimator $\hat{\Sigma}$ of $\Sigma \in \mathbf{PD}(I)$ under H exists a.e. iff $N \geq I$, and it is uniquely given by $\hat{\Sigma}(X) = S := \frac{1}{N}\sum(X_\nu X_\nu^t|\nu \in N)$, the *normed empirical variance matrix*. A reasonable estimator $\hat{\Sigma}_0(X)$ for $\Sigma_0 \in \Theta_0$ under H_0 is also based on S, and we shall assume that this estimator also exists a.e. and that $\hat{\Sigma}_0 = t(\hat{\Sigma})$ for some surjective idempotent mapping $t : \mathbf{PD}(I) \to \Theta_0$, i.e., $t(\mathbf{PD}(I)) = \Theta_0$ and $t \circ t = t$.
More generally, let a random variable X represent the observable from a statistical model with parameter set $\Theta \subseteq \mathbf{PD}(I)$. Let $\Theta_0 \subset \Theta$ and consider the induced testing problem classically expressed as

$$\mathrm{H}_0 : \ \Sigma_0 \in \Theta_0 \ \text{ vs. } \ \mathrm{H} : \ \Sigma \in \Theta.$$

Suppose that the elements of $\Theta \subseteq \mathbf{PD}(I)$ have an interpretation as $I \times I$ variance matrices for distributions P_Σ, $\Sigma \in \Theta$, on the vector space $\mathbb{R}^I$. Let Z be a random

[‡]The cardinality of a set I is also denoted by I.

variable with distribution[§] P_Σ. Any matrix $\Sigma \in \mathbf{PD}(I)$ represents a positive definite form on $\mathbb{R}^I$ given by $(x,y) \to x^t\Sigma y$, $x,y \in \mathbb{R}^I$. Thus $x^t\Sigma y$ is the covariance $\mathbb{C}ov_\Sigma(x^tZ, y^tZ) =: \mathbb{C}ov_\Sigma(x,y)$, between the random variables x and y under P_Σ, $\Sigma \in \Theta$. In particular, $x^t\Sigma x$ is the variance $\mathbb{V}_\Sigma(x^tZ) =: \mathbb{V}_\Sigma(x)$ of x under P_Σ, $\Sigma \in \Theta$.
Suppose that $S := s(X)$ and $S_0 := s_0(X)$ are almost surely defined estimators for the unknown parameter $\Sigma \in \Theta$, i.e., under H, and $\Sigma_0 \in \Theta_0$, i.e., under H_0, respectively. Furthermore, assume that $S_0 = t(S)$, where $t : \Theta \to \Theta_0$ is surjective and idempotent.
In accordance with the point of view in most of the ten cases from (ABJ), we define the *(empirical) variance residual* R and the *(theoretical) variance residual* Δ as $R := S - t(S)$ and $\Delta = \Sigma - t(\Sigma)$. The variance residuals are symmetric $I \times I$ matrices, i.e., $R, \Delta \in \mathbf{S}(I)$, where $\mathbf{S}(I)$ denotes the vector space of symmetric $I \times I$ matrices. The variance residuals constitute the basis for our general interpretations of eigenvalues and canonical linear forms presented in Section 3 and Section 6 below. The interpretations are then in Section 8 applied to the ten testing problems in (ABJ).

3. Canonical variance distances and canonical distance variates.

Let $\Delta := \Sigma - t(\Sigma)$, $\Sigma \in \Theta$, be the theoretical residual from Section 2.

Definition 3.1. An ordered family $(x_1, \cdots, x_I)$ of vectors from $\mathbb{R}^I$ is called a family of *(theoretical) canonical distance variates* (CDV) (wrt. Σ) if

$$\begin{aligned}
&m_1 := |\mathbb{V}_\Sigma(x_1) - \mathbb{V}_{t(\Sigma)}(x_1)| = \\
&\max\{|\mathbb{V}_\Sigma(x) - \mathbb{V}_{t(\Sigma)}(x)| \; |x \in \mathbb{R}^I, \; \mathbb{V}_{t(\Sigma)}(x) = 1\}, \\
&\text{and} \\
&m_i := |\mathbb{V}_\Sigma(x_i) - \mathbb{V}_{t(\Sigma)}(x_i)| = \\
&\max\{|\mathbb{V}_\Sigma(x) - \mathbb{V}_{t(\Sigma)}(x)| \; |x \in \mathbb{R}^I, \; \mathbb{V}_{t(\Sigma)}(x) = 1, \\
&\mathbb{C}ov_{t(\Sigma)}(x, x_j) = 0, j = 1, \cdots, i-1\}, \\
&\text{for } i = 2, \cdots, I.
\end{aligned} \tag{3.1}$$

The ordered family $(m_1, \cdots, m_I) \in [0, \infty[^I$ is called the *corresponding (theoretical) canonical variance distances* (CVD). Note that CVD satisfies $m_1 \geq m_2 \geq \cdots \geq m_I \geq 0$. When Σ is replaced by the estimator $S = \hat{\Sigma}$ in the definition, the word *theoretical* is replaced by the word *empirical*.

When the ordered lists $(x_1, \cdots, x_I)$ and $(m_1, \cdots, m_I)$ are shortened to $(x_1, \cdots, x_k)$ and $(m_1, \cdots, m_k)$, respectively, $k < I$, it means that $m_{k+1} = \cdots = m_I = 0$. In that case the choice of $x_{k+1}, \cdots, x_I$ is uninteresting, cf. Remark 5.1.
We shall later see, cf. Corollary 5.1, that theoretical CDVs are pairwise uncorrelated even under the hypothesis H. Similarly, empirical CDVs satisfy $x_\iota^t S x_{\iota'} = 0$, $\iota, \iota' = 1, \cdots, I$, $\iota \neq \iota'$.

[§]In statistics it is not a good idea to represent a distribution with a random variable, since the dependency of the parameter becomes suppressed. Nevertheless this is the classical way.

The trivial reformulation of the conditions in Definition 3.1 into the analytic conditions

$$(3.2)\qquad \begin{array}{l} m_1 := |x_1^t \Delta x_1| = \\ \max\{|x^t \Delta x| \; |x \in \mathbb{R}^I,\; x^t t(\Sigma) x = 1\}, \\ \text{and} \\ m_i := |x_i^t \Delta x_i| = \\ \max\{|x^t \Delta x| \; |x \in \mathbb{R}^I,\; x^t t(\Sigma) x = 1,\; x^t t(\Sigma) x_j = 0, j = 1, \cdots, i-1\}, \\ \text{for } i = 2, \cdots, I, \end{array}$$

shows that a family of theoretical canonical distance variates exists by the extreme value theorem.

The connection between the above concepts and eigenvalues/eigenvectors of symmetric forms wrt. positive definite forms provides an understanding, extra properties, uniqueness descriptions, and equations to find these variates and distances. The next two sections therefore formulate the eigenvalues/eigenvector results and connect them to the above concepts.

4. On the basic well known diagonalization theorem.

The well known classical result below is the foundation for the interpretations presented in this paper. Let $V \neq \{0\}$ be a finite dimensional vector space with $\mathrm{Dim}(V) = I$. Let $\tau : V \times V \to \mathbb{R}$ be a positive definite form on V, and $\delta : V \times V \to \mathbb{R}$ be a symmetric bilinear form on V.

Theorem 4.1. *There exists a finite subset $\Lambda \subset \mathbb{R}$ and a decomposition $V = \oplus(V_\lambda | \lambda \in \Lambda)$ of V into a direct sum of subspaces $V_\lambda \subseteq V$ indexed by $\lambda \in \Lambda$, such that*

(i) : $V_\lambda \neq \{0\}$,

(ii) : For all $\lambda \in \Lambda$:
$\delta(x, x) = \lambda \tau(x, x),\; x \in V_\lambda$,

(iii) : For all $\lambda_1, \lambda_2 \in \Lambda$ with $\lambda_1 \neq \lambda_2$:
$\delta(x_1, x_2) = \tau(x_1, x_2) = 0,\; x_1 \in V_{\lambda_1}, x_2 \in V_{\lambda_2}$.

On the other hand, if $V = \oplus(U_\mu | \mu \in M)$ is a decomposition of V into a direct sum of subspaces $U_\mu \subseteq V$ indexed by a subset $M \subseteq \mathbb{R}$, and satisfying (i), (ii), and (iii) with Λ replaced by M, i.e.,

(i) : $U_\mu \neq \{0\},\; \mu \in M$

(ii) : For all $\mu \in M$:
$\delta(x, x) = \mu \tau(x, x),\; x \in U_\mu$,

(iii) : For all $\mu_1, \mu_2 \in \Lambda$ with $\mu_1 \neq \mu_2$:
$\delta(x_1, x_2) = \tau(x_1, x_2) = 0,\; x_1 \in U_{\mu_1}, x_2 \in U_{\mu_2}$,

then $M = \Lambda$ and $U_\lambda = V_\lambda$, $\lambda \in \Lambda \equiv M$.

Definition 4.1. The elements $\lambda \in \Lambda$ are called the *eigenvalues* of δ wrt. to τ and a vector different from the zero vector in V_λ is called an *eigenvector* of δ wrt. τ (with eigenvalue λ). The dimension $n_\lambda := \text{Dim}(V_\lambda)$, $\lambda \in \Lambda$, is called the *(geometric) multiplicity* of the eigenvalue λ.

Remark 4.1. Let $(v_i|i \in I)$ be an arbitrary basis for V and let

$$T := (\tau(v_i, v_j)|(i,j) \in I \times I) \quad \text{and} \quad D := (\delta(v_i, v_j)|(i,j) \in I \times I)$$

be the corresponding $I \times I$ matrices for τ and δ, respectively. The unique set $\Lambda \subset \mathbb{R}$ of eigenvalues can then be obtained as the set of solutions (roots) to the I degree polynomial equation

$$\det(D - \lambda T) = 0. \tag{4.1}$$

The (algebraic) multiplicity of a root λ to this polynomial equation is equal to the (geometric) multiplicity of the eigenvalue λ. For each $\lambda \in \Lambda$ the coordinates $x \in \mathbb{R}^I$ of the vectors in V_λ can be obtained as the solutions to the singular linear equation in $x \in \mathbb{R}^I$:

$$(D - \lambda T)x = 0. \tag{4.2}$$

Remark 4.2.
(I): Condition (iii) says that the direct sum is orthogonal wrt. τ **and** wrt. δ.

(II): The standard formulation of the existence part of this well known theorem is usually:

"There exists a basis $(e_i|i \in I)$ for V such that the $I \times I$ matrices T and D for τ and δ, respectively, take the form¶ $T = \text{Diag}(1|i \in I) =: 1_I$ and $D = \text{Diag}(\lambda_i|i \in I)$, where $\lambda_i \in \mathbb{R}$, $i \in I$." Thus $\Lambda = \{\lambda_i|i \in I\}$ and $V_\lambda = \text{Span}\{e_i|\lambda_i = \lambda\}$.

(III): Usually different versions of the uniqueness statement in the theorem are added to this standard formulation, f.ex.

(IV) If $\text{Dim}(V_\lambda) = 1$, $\lambda \in \Lambda$, the basis in (III) is unique up to independent signshifts of e_i, $i \in I$.

5. The diagonalization theorem continued. An observation.

Let $U^\perp$ denote the orthogonal complement wrt. τ of a subspace $U \subseteq V$. The invariant formulation of (3.2) is then:

Definition 5.1. An ordered family $(x_1, \cdots, x_I)$ of vectors from V is called a family of $(|\delta|, \tau)$-*maximizers* if

$$\begin{aligned} &m_1 := |\delta(x_1, x_1)| = \max\{|\delta(x,x)|\ |x \in V,\ \tau(x,x) = 1\} \\ &\text{and} \\ &m_i = |\delta(x_i, x_i)| = \max\{|\delta(x,x)|\ |x \in \text{Span}\{x_1, \cdots, x_{i-1}\}^\perp,\ \tau(x,x) = 1\} \\ &\text{for } i = 2, \cdots, I. \end{aligned} \tag{5.1}$$

The ordered family $(m_1, \cdots, m_I) \in [0, \infty[^I$ is called the *corresponding* family of $(|\delta|, \tau)$-*maxima.*

Lemma 5.1. *Assume that x_1 satisfies the first condition in Definition 5.1, set $V_1 := \text{Span}\{x_1\}^\perp$, and let δ_1 and τ_1 be restrictions of δ and τ to V_1, respectively. Then $(x_1, \cdots, x_I)$ is a family of $(|\delta|, \tau)$-maximizers with corresponding $(|\delta|, \tau)$-maxima $(m_1, \cdots, m_I)$ if and only if $(x_2, \cdots, x_I)$ is a family of $(|\delta_1|, \tau_1)$-maximizers with corresponding $(|\delta_1|, \tau_1)$-maxima $(m_2, \cdots, m_I)$.*

¶The $I \times I$ diagonal matrix with diagonal $(\alpha_i|i \in I) \in \mathbb{R}^I$ is denoted $\text{Diag}(\alpha_i|i \in I)$.

Proof. Trivial. □

Let $\gamma_1 > \cdots > \gamma_\Gamma$ be the ordering of elements of $\Gamma := \{|\lambda| \; |\lambda \in \Lambda\} \subset [0, \infty[$ and define the **sets**

$$W_j := \begin{cases} V_{\gamma_j} & \text{if} \quad \gamma_j \in \Lambda \text{ and } -\gamma_j \notin \Lambda \\ V_{-\gamma_j} & \text{if} \quad -\gamma_j \in \Lambda \text{ and } \gamma_j \notin \Lambda \\ V_{\gamma_j} \cup V_{-\gamma_j} & \text{if} \quad \gamma_j \in \Lambda \text{ and } -\gamma_j \in \Lambda \end{cases}, \; j = 1, \cdots, \Gamma.$$

Furthermore define

$$k_j := \begin{cases} n_{\gamma_j} & \text{if} \quad \gamma_j \in \Lambda \text{ and } -\gamma_j \notin \Lambda \\ n_{-\gamma_j} & \text{if} \quad -\gamma_j \in \Lambda \text{ and } \gamma_j \notin \Lambda \\ n_{\gamma_j} + n_{-\gamma_j} & \text{if} \quad \gamma_j \in \Lambda \text{ and } -\gamma_j \in \Lambda \end{cases}, \; j = 1, \cdots, \Gamma,$$

$k_{\bullet j} := \sum(k_\nu | \nu = 1, \cdots, j)$, $j = 1, \cdots, \Gamma$, and $k_{\bullet 0} = 0$.

Proposition 5.1. *Let $(x_1, \cdots, x_I)$ be an ordered family of vectors from V. The following two conditions, (a) and (b) below, are then equivalent:*

(a): The family $(x_1, \cdots, x_I)$ is a family of $(|\delta|, \tau)$ maximizers

and

(b):

$$\begin{array}{l} (x_1, \cdots, x_I) \textit{ is an orthogonal and normalized (wrt. to } \tau) \textit{ basis of } V \\ \text{and} \\ x_i \in W_j, \; j = 1, \cdots, \Gamma, \; i = k_{\bullet(j-1)} + 1, \cdots, k_{\bullet j}. \end{array} \tag{5.2}$$

Proof. Let $[a, b]$ be the smallest interval containing Λ and note that $a, b \in \Lambda$. In the case $a = b$ the claimed result is trivial. Thus assume that $a < b$. Then $a = -\gamma_1$ or $b = \gamma_1$. Let $x = \sum(x_\lambda | \lambda \in \Lambda)$ the unique decomposition of $x \in V$ into its components $x_\lambda \in V_\lambda$. Then $\delta(x, x) = \sum(\lambda \tau(x_\lambda, x_\lambda) | \lambda \in \Lambda)$. The condition $\tau(x, x) = 1$ means that $\sum(\tau(x_\lambda, x_\lambda) | \lambda \in \Lambda) = 1$. It thus follows that $a \leq \delta(x, x) \equiv \sum(\lambda \tau(x_\lambda, x_\lambda) | \lambda \in \Lambda) \leq b$ and in particular that $|\delta(x, x)| \leq \max\{|a|, |b|\} = \gamma_1$ with equality if and only if $\delta(x, x) = -\gamma_1$ or $\delta(x, x) = \gamma_1$. In the first case the convex linear combination of the λs is degenerate, i.e., $\tau(x_{-\gamma_1}, x_{-\gamma_1}) = 1$, which is true iff $x \in V_{-\gamma_1}$. The second case is similar.
The proposition now follows from Lemma 5.1 and Theorem 4.1. □

Corollary 5.1. *Let $(x_1, \cdots, x_I)$ be a family of $(|\delta|, \tau)$ maximizers then $x_1, \cdots, x_I$ is also pairwise orthogonal wrt. $\sigma := \tau + \delta$.*

Proof. Trivial □

Remark 5.1. Let $(x_1, \cdots, x_I)$ be a family of $(|\delta|, \tau)$-maximizers. Then in particular $|\delta(x_i, x_i)| \equiv m_i = \gamma_j$, $j = 1, \cdots, \Gamma$, $i = k_{\bullet j-1} + 1, \cdots, k_{\bullet j}$. The corresponding family $(m_1, \cdots, m_I)$ of $(|\delta|, \tau)$-maxima family is thus uniquely determined by δ and τ (and not of the choice of a family of $(|\delta|, \tau)$-maximizers).

6. On general canonical correlations and covariate pairs.

In this section, our initial starting point is again a variation of (3.1).
Let $[\frac{I}{2}]$ denote the greatest integer less than or equal to $\frac{I}{2}$, and assume that $I \geq 2$.

Definition 6.1. An ordered family $((u_1,v_1),\cdots,(u_{[\frac{I}{2}]},v_{[\frac{I}{2}]}))$ of pairs of vectors from $\mathbb{R}^I$ is called a family of *(theoretical) general canonical covariate pairs* (GCCP) (wrt. Σ) if

$$
\begin{aligned}
(6.1)\quad & c_1 := \mathbb{C}ov_\Sigma(u_1,v_1) = \\
& \max\{\mathbb{C}ov_\Sigma(u,v)\ |u,v\in\mathbb{R}^I,\ \mathbb{V}_\Sigma(u)=\mathbb{V}_\Sigma(v)=1, \mathbb{C}ov_{t(\Sigma)}(u,v)=0\},\\
& \text{and}\\
& c_\kappa := \mathbb{C}ov_\Sigma(u_\kappa,v_\kappa) = \\
& \max\{\mathbb{C}ov_\Sigma(u,v)\ |u,v\in\mathbb{R}^I,\ \mathbb{V}_\Sigma(u)=\mathbb{V}_\Sigma(v)=1, \mathbb{C}ov_{t(\Sigma)}(u,v)=0,\\
& \mathbb{C}ov_{t(\Sigma)}(u,u_{\kappa'})=\mathbb{C}ov_{t(\Sigma)}(u,v_{\kappa'})=\mathbb{C}ov_{t(\Sigma)}(u_{\kappa'},v)=\mathbb{C}ov_{t(\Sigma)}(v_{\kappa'},v)=0,\\
& \kappa'=1,\cdots,\kappa-1\},\\
& \text{for } \kappa=2,\cdots,[\tfrac{I}{2}].
\end{aligned}
$$

The ordered family $(c_1,\cdots,c_I)\in[0,1]^{[\frac{I}{2}]}$ is called the *corresponding (theoretical) general canonical correlations* (GCC). Note that GCC satisfy $1>c_1\geq c_2\geq\cdots\geq c_{[\frac{I}{2}]}\geq 0$. When Σ is replaced by the estimator $S=\hat{\Sigma}(X)$ in the definition, the word *theoretical* is replaced by the word *empirical.*

When the ordered lists $((u_1,v_1),\cdots,(u_{[\frac{I}{2}]},v_{[\frac{I}{2}]}))$ and $(c_{[\frac{I}{2}]},\cdots,c_{[\frac{I}{2}]})$ are shortened to the families $((u_1,v_1),\cdots,(u_k,v_k))$ and $(c_1,\cdots,c_k)$, respectively, $k<[\frac{I}{2}]$, it means that $c_{k+1}=\cdots=c_{[\frac{I}{2}]}=0$. In that case the choice of $((u_{k+1},v_{k+1}),\cdots,(u_{[\frac{I}{2}]},v_{[\frac{I}{2}]})$ is uninteresting, cf. Remark 6.1.
The invariant formulation of Definition 6.1, skipping a step similar to the intermediate step (3.2), becomes:

Definition 6.2. Let τ and σ be positive definite forms on V and let $\delta=\sigma-\tau$. An ordered family $((u_1,v_1),\cdots,(u_{[\frac{I}{2}]},u_{[\frac{I}{2}]})$ of pairs of vectors from V is called a family of pairs of $[\delta,\tau]$-*maximizers* if
(6.2)

$$
\begin{aligned}
& c_1 := \delta(u_1,v_1)=\max\{\delta(u,v)\ |u,v\in V,\ \sigma(u,u)=\sigma(v,v)=1,\ \tau(u,v)=0\},\\
& \text{and}\\
& c_\kappa=\delta(u_\kappa,v_\kappa)=\\
& \max\{\delta(u,v)|u,v\in\operatorname{span}(u_1,v_1,\cdots,u_{\kappa-1},v_{\kappa-1})^\perp,\sigma(u,u)=\sigma(v,v)=1,\tau(u,v)=0\},\\
& \text{for } \kappa=2,\cdots,[\tfrac{I}{2}].
\end{aligned}
$$

The family $(c_1,\cdots,c_{[\frac{I}{2}]})\in[0,1[^{[\frac{I}{2}]}$ is called the *corresponding* family of $[\delta,\tau]$-*maxima.*

By the extreme value theorem, it is obvious that such a family of pairs of maximizers exists.

Lemma 6.1. *Assume that (u_1,v_1) satisfies the first condition in Definition 6.2, set $V_1:=\operatorname{span}(u_1,v_1)^\perp$, and let δ_1 and τ_1 be restrictions of δ and τ to V_1, respectively. Then $((u_\kappa,v_\kappa)\in V\times V|\kappa=1,\cdots,[\frac{I}{2}]))$ is a family of pairs of $[\delta,\tau]$-maximizers with corresponding family of maxima $(c_1,\cdots,c_{[\frac{I}{2}]})$ if and only if $((u_\kappa,v_\kappa)\in V\times V|\kappa=2,\cdots,[\frac{I}{2}]))$ is a family of pairs of $[\delta_1,\tau_1]$-maximizers with $(c_2,\cdots,c_{[\frac{I}{2}]})$ as the corresponding family of maxima.*

Proof. Trivial. □

Proposition 6.1. *Let $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_I(\geq -1)$ be the ordering of the elements of Λ, where each element occurs according to its multiplicity.*

The family $((u_\kappa, v_\kappa) \in V \times V | \kappa = 1, \cdots, [\frac{I}{2}])$ *is a family of* $[\delta, \tau]$*-pairs maximizers if and only if*
there exists τ*-orthogonal vectors* $x_1, \cdots, x_{[\frac{I}{2}]}, y_1, \cdots, y_{[\frac{I}{2}]} \in V$ *such that*

(i) $x_\kappa \in V_{\lambda_\kappa}$*,* $y_\kappa \in V_{\lambda_{I+1-\kappa}}$*,*

(ii) $\tau(x_\kappa, x_\kappa) = \tau(y_\kappa, y_\kappa) = \frac{1}{\lambda_\kappa + \lambda_{I+1-\kappa} + 2}$*,*

(iii) $u_\kappa = x_\kappa + y_\kappa$ *and* $v_\kappa = x_\kappa - y_\kappa$*,*

for $\kappa = 1, \cdots, [\frac{I}{2}]$*.*

In that case, $c_\kappa = \mu(\lambda_\kappa, \lambda_{I+1-\kappa}) := \frac{\lambda_\kappa - \lambda_{I+1-\kappa}}{\lambda_\kappa + \lambda_{I+1-\kappa} + 2}$*,* $\kappa = 1, \cdots, [\frac{I}{2}]$*.*

Proof. This follows from the following lemma and Lemma 6.1 by inducting on $[\frac{I}{2}]$. □

Lemma 6.2. *The maximum value*

$$c_1 := \max\{\delta(u, v) \; | u, v \in V, \; \sigma(u, u) = \sigma(v, v) = 1, \; \tau(u, v) = 0\}$$

is attained at (u_1, v_1) *if and only if there exists* τ*-orthogonal vectors* $x_1 \in V_{\lambda_1}$ *and* $y_1 \in V_{\lambda_I}$ *such that*

(i) $\tau(x_1, x_1) = \tau(y_1, y_1) = \frac{1}{\lambda_1 + \lambda_I + 2}$*, and*

(ii) $u_1 = x_1 + y_1$ *and* $v_1 = x_1 - y_1$*.*

The maximum value is $c_1 = \mu(\lambda_1, \lambda_I) := \frac{\lambda_1 - \lambda_I}{\lambda_1 + \lambda_I + 2}$*.*

Proof. If $\lambda_1 = \lambda_I$ the proof is trivial (and $c_1 = 0$). We can thus assume that $\lambda_1 \neq \lambda_I$. Note that $c_1 < 1$ by the Cauchy-Schwartz inequality.
We will first show $c_1 = \mu(\lambda_1, \lambda_I)$. If u_1 and v_1 are vectors satisfying the above conditions, it is easy to check that $\sigma(u_1, u_1) = \sigma(v_1, v_1) = 1$, $\tau(u_1, v_1) = 0$, and $\delta(u_1, v_1) = \mu(\lambda_1, \lambda_I)$. This proves that $c_1 \geq \mu(\lambda_1, \lambda_I) > 0$, and it remains to show that $c_1 \leq \mu(\lambda_1, \lambda_I)$.
Let $u_1 \equiv \sum(u_\lambda | \lambda \in \Lambda)$ and $v_1 \equiv \sum(v_\lambda | \lambda \in \Lambda)$ be the decompositions of u_1 and v_1, respectively, according to the decomposition $V = \oplus(V_\lambda | \lambda \in \Lambda)$, cf. Theorem 4.1.
For all $\lambda \in \Lambda$ with $u_\lambda \neq 0$ there exists $\lambda' \in \Lambda$ with $\lambda' \neq \lambda$ and $u_{\lambda'} \neq 0$. Suppose otherwise that there exists $\lambda \in \Lambda$ such that $u_\lambda \neq 0$ and $u_{\lambda'} = 0$ for all $\lambda' \in \Lambda \backslash \{\lambda\}$. Then $c_1 = \delta(u_1, v_1) = \delta(u_\lambda, v_\lambda) = \lambda\tau(u_\lambda, v_\lambda) = \lambda\tau(u_1, v_1) = 0$, contradicting that $c_1 > 0$.
We shall now apply Lagrange multipliers. Consider the function

$$F(u, v) := \delta(u, v) + l_1(\sigma(u, u) - 1) + l_2(\sigma(v, v) - 1) + l_0\tau(u, v), \; u, v \in V.$$

Then

$$\frac{\partial F}{\partial u}(u_1, v_1) = \delta(\bullet, v_1) + 2l_1\sigma(u_1, \bullet) + l_0\tau(\bullet, v_1) = 0,$$
$$\frac{\partial F}{\partial v}(u_1, v_1) = \delta(\bullet, u_1) + 2l_2\sigma(v_1, \bullet) + l_0\tau(\bullet, u_1) = 0. \tag{6.3}$$

Furthermore u_1 and v_1 satisfy the constraints

$$\sigma(u_1, u_1) = 1, \; \sigma(v_1, v_1) = 1, \text{ and } \; \tau(u_1, v_1) = 0.$$

Replacing $\bullet$ with u_1 and v_1 in the first and second equation, respectively, yields $l_1 = l_2 = -\frac{c_1}{2}$.

Let $\lambda, \lambda' \in \Lambda$ with $\lambda > \lambda'$, $u_\lambda \neq 0$, and $u_{\lambda'} \neq 0$. The equations in (6.3) then yield

$$\begin{aligned} &\lambda v_\lambda - c_1(1+\lambda)u_\lambda + l_0 v_\lambda = 0, \\ &\lambda u_\lambda - c_1(1+\lambda)v_\lambda + l_0 u_\lambda = 0, \end{aligned} \tag{6.4}$$

and a similar system of 2 linear homogeneous equation with λ replaced by λ'. Since both of these systems have non-zero solutions, it follows that

$$\begin{aligned} c_1^2(1+\lambda)^2 - (\lambda + l_0)^2 = 0 \\ c_1^2(1+\lambda')^2 - (\lambda' + l_0)^2 = 0. \end{aligned} \tag{6.5}$$

After some trivial but tedious calculations, (6.5) yields that

$$c_1 = \frac{\lambda - \lambda'}{\lambda + \lambda' + 2} \quad \text{and} \quad l_0 = -\frac{\lambda + \lambda' + 2\lambda\lambda'}{\lambda + \lambda' + 2}. \tag{6.6}$$

We then conclude that

$$c_1 = \frac{\lambda - \lambda'}{\lambda + \lambda' + 2} \leq \frac{\lambda_1 - \lambda_I}{\lambda_1 + \lambda_I + 2}, \tag{6.7}$$

with equality if and only if $\lambda = \lambda_1$, and $\lambda' = \lambda_I$.
Thus $c_1 = \mu(\lambda_1, \lambda_I)$ as claimed. We have also established that $u_\lambda = 0$ for all $\lambda \in \Lambda \backslash \{\lambda_1, \lambda_I\}$. By reason of symmetry the same is valid for v_1. Substituting (6.6) into the second equation of (6.4) with λ replaced by λ_1 and with λ replaced by λ_I yields $v_{\lambda_1} = u_{\lambda_1}$ and $v_{\lambda_I} = -u_{\lambda_I}$, respectively. Therefore, $u_1 = x_1 + y_1$ and $v_1 = x_1 - y_1$, where $x_1 := u_{\lambda_1} \in V_{\lambda_1}$ and $y_1 := u_{\lambda_I} \in V_{\lambda_I}$.
Because x_1 and y_1 are eigenvectors corresponding to distinct eigenvalues, $\tau(x_1, y_1) = 0$ and $\sigma(x_1, y_1) = 0$. The condition $\tau(u_1, v_1) = 0$ implies that $\tau(x_1, x_1) = \tau(y_1, y_1)$ and the condition $1 = \sigma(u_1, u_1) = (1+\lambda_1)\tau(x_1, x_1) + (1+\lambda_I)\tau(y_1, y_1)$ then implies that

$$\tau(x_1, x_1) = \tau(y_1, y_1) = \frac{1}{\lambda_1 + \lambda_I + 2}.$$

□

Remark 6.1. Let $((u_1, v_1), \cdots, (u_{[\frac{I}{2}]}, v_{[\frac{I}{2}]}))$ be a family of $[\delta, \tau]$-maximizers. Then in particular $c_\kappa := \delta(u_\kappa, v_\kappa)$, $\kappa = 1, \cdots, [\frac{I}{2}]$ only depends on δ and τ (and not on the choice of a family of $[\delta, \tau]$-maximizers).

Remark 6.2. If Λ is symmetric around 0 the two expressions $\frac{1}{\lambda_\kappa + \lambda_{I+1-\kappa} + 2}$ and $\frac{\lambda_\kappa - \lambda_{I+1-\kappa}}{\lambda_\kappa + \lambda_{I+1-\kappa} + 2}$ in Proposition 6.1 reduce to the simple expressions $\frac{1}{2}$ and λ_κ, respectively.

7. The group symmetry models.

The theory of group symmetry models can be found in (AM). This theory is a small but central part of an unpublished general algebraic theory of normal models developed by Andersson, Brøns, and Tolver Jensen in the years 1972-1985, cf. the list of references.
A group symmetry model can be stated simply as follows, cf. (AM), cf. Section A6: Let $(X_1, \cdots, X_N) \equiv X \in (\mathbb{R}^I)^N \equiv \mathbb{R}^{I \times N}$ be the family of N i.i.d. observables from a normal distribution N_Σ on $\mathbb{R}^I$ with expectation $0 \in \mathbb{R}^I$ and unknown $I \times I$

variance matrix $\Sigma \in \mathbf{PD}_G(I) \subseteq \mathbf{PD}(I)$, where G is a finite[‖] subgroup of the othogonal group of $I \times I$ matrices and

$$\mathbf{PD}_G(I) := \{\Sigma \in \mathbf{PD}(I) | g\Sigma g^t = \Sigma,\ g \in G\}.$$

More generally we define

$$\mathbf{M}_G(I) := \{M \in \mathbf{M}(I) | gM = Mg,\ g \in G\} \text{ and } \mathbf{S}_G(I) := \{A \in \mathbf{M}_G(I) | A = A^t\},$$

where $\mathbf{M}(I) := \mathbb{R}^{I\times I}$ denote the vector space of all $I \times I$ matrices. For N greater than or equal to a certain integer constant depending on G the ML-estimator $\hat{\Sigma}(X_1, \cdots, X_N)$ exists with probability 1, cf. AM (A.9)[**], and is given by

$$\hat{\Sigma}(X) := \frac{1}{G} \sum (g(\frac{1}{N} XX^t)g^t | g \in G),\ X \equiv (X_1, \cdots, X_N) \in \mathbb{R}^{I\times N}.$$

Let $G \subseteq G_0$ be a finite subgroup of another finite subgroup of the othogonal group. Then the relation $\mathbf{PD}_{G_0}(I) \subseteq \mathbf{PD}_G(I)$ is obvious and induces the *symmetry testing problem*, cf. (AM) Section A7.

$$\text{H}_0:\ \Sigma \in \mathbf{PD}_{G_0}(I) \text{ vs. } \text{H}:\ \Sigma \in \mathbf{PD}_G(I). \tag{7.1}$$

This testing problem is invariant under the group $\mathbf{GL}_{G_0}(I)$ of non-singular elements of $\mathbf{M}_{G_0}(I)$ and is a special case of the very general testing problem in Section 2 with $\Theta := \mathbf{PD}_G(I)$, $\Theta_0 := \mathbf{PD}_{G_0}(I)$, $S = \hat{\Sigma}(X)$, $T := \hat{\Sigma}_0(X)$, $P_\Sigma = N_\Sigma$, and $t : \mathbf{PD}_G(I) \to \mathbf{PD}_{G_0}(I)$ given by

$$T := t(\Sigma) = \frac{1}{G_0} \sum (g_0 S g_0^t | g_0 \in G_0), \tag{7.2}$$

an *average map*. From this it follows that the variance residuals $R := S - t(S) \in \mathbf{S}_G(I)$ and $\Delta := \Sigma - t(\Sigma) \in \mathbf{S}_G(I)$ have the further property

$$\sum (g_0 R g_0^t | g_0 \in G_0) = 0 \text{ and } \sum (g_0 \Delta g_0^t | g_0 \in G_0) = 0, \tag{7.3}$$

respectively.

The ten special symmetry testing problems considered in (ABJ) are the building stones of the general symmetry testing problem (7.1), cf. (ABJ), p392, the second paragraph. Thus they deserves special attention. In all ten problems (ABJ) obtain a representation of a maximal invariant statistic and find its central distribution in terms of a density wrt. to the standard Lebesque measure. In 7 out the ten problems the representation is based directly on the eigenvalues of the residuals R wrt. the estimator T under H_0. This fact uses the property (7.3) which reduces in all ten cases to a sum of only two different terms.

In the next section we will use the general results from the previous sections to obtained interpretations of these maximal invariants in terms of canonical variates. The classical case of canonical correlations and variates of course becomes a special case.

[‖]In fact just a compact subgroup, but any compact subgroup G can be replaced by a finite subgroup $G^{\text{finite}} \subseteq G$ such that $\mathbf{PD}_G(I) = \mathbf{PD}_{G^{\text{finite}}}(I)$

[**]Otherwise it will not exists for any observation, cf. (AM), Section A1.

8. Interpretations of the representation of the maximal invariant statistic in the ten fundamental symmetry testing problems.

In the ten cases below we comment on the **empirical** CDVs, CVDs, GCCPs, and GCCs. The similar comments on the same theoretical families are of course obtained by replacing the estimator $S := \hat{\Sigma}(X)$ under H by the parameter Σ. Thus T and R will be replaced by $t(\Sigma)$ and $\Delta := S - t(\Sigma)$, respectively.

1. Testing that a covariance matrix with real structure has complex structure, cf. Section 2 of (ABJ). In this testing problem, the index set I is replaced by the disjoint union $I^{\dot{\cup}2} := I \dot{\cup} I$ and $I \geq 2$. All $I^{\dot{\cup}2} \times I^{\dot{\cup}2}$ matrices are thus partitioned into 2×2 block matrices according to the decomposition $I^{\dot{\cup}2}$ of the index set. We can choose $G = \{\pm 1_{I^{\dot{\cup}2}}\}$ and $G_0 = \{\pm 1_{I^{\dot{\cup}2}}, \pm \mathbf{i}\}$, where 1_I denotes the $I \times I$ unit matrix for an index set I, and

$$\mathbf{i} := \begin{pmatrix} 0 & -1_I \\ 1_I & 0 \end{pmatrix}$$

is an imaginary unit in the complex field. Thus $\Theta := \mathbf{PD}_G(I^{\dot{\cup}2}) = \mathbf{PD}(I^{\dot{\cup}2}) = \mathcal{H}^+(I^{\dot{\cup}2}, \mathbb{R})$ and $\Theta_0 := \mathbf{PD}_{G_0}(I^{\dot{\cup}2}) = \mathcal{H}^+(I, \mathbb{C})$, using the notation[††] from (ABJ). Roughly speaking, we consider the testing problem that an arbitrary non-singular $I^{\dot{\cup}2} \times I^{\dot{\cup}2}$ variance matrix Σ, an $\mathbb{R}$-*structure*, has the form

$$\Sigma = \begin{pmatrix} \Gamma & -F \\ F & \Gamma \end{pmatrix}, \tag{8.1}$$

a $\mathbb{C}$-*structure*. The average mapping t given in (7.2) reduces to (12) in (ABJ) and the residual R is given by (13) in (ABJ). The condition (7.3) has only two different terms and reduces to the condition that the residual has the form (7) from (ABJ). The eigenvalues of R wrt. T, i.e., the solution set Λ to the $2I$-degree polynomial (4.1) is symmetric (around 0), i.e., $\Lambda = -\Lambda$, and $|\lambda| < 1$, $\lambda \in \Lambda$. The ordered family $(\lambda_1, \cdots, \lambda_I)$ of non-negative eigenvalues, each repeated according to its multiplicity, i.e., $1 > \lambda_1 \geq \cdots \geq \lambda_I \geq 0$, becomes a maximal invariant statistic[‡‡], cf. (17) of (ABJ).
For the sake of simplicity in the description of interpretations, we shall assume that $\lambda_1 > \cdots > \lambda_I > 0$, i.e., all the $2I$ eigenvalues are different[*].
To find a family of CDV we first find for each λ_ι, $\iota = 1, \cdots, I$, a normalized[†] (wrt. T) solution $x_\iota \in \mathbb{R}^{I^{\cup 2}}$ to (4.2) when $\lambda = \lambda_\iota$. Then $\mathbf{i}x_\iota$ is also normalized and solves (4.2) with $\lambda = -\lambda_\iota$, $\iota = 1, \cdots, I$, and we thus have all eigenvalues and eigenvectors. Thus $(x_1, \mathbf{i}x_1, \cdots, x_I, \mathbf{i}x_I)$ is a family of CDV with the corresponding family $(\lambda_1, \lambda_1, \cdots, \lambda_I, \lambda_I)$ of CVD. The CDV and the corresponding CVD provide through Definition 3.1 then the first interpretation of the maximal invariant statistic.
It then follows from Proposition 6.1 that $((u_1, v_1), \cdots, (u_I, v_I))$, where $u_\iota := \frac{1}{\sqrt{2}}(x_\iota + \mathbf{i}x_\iota)$, and $v_\iota := \frac{1}{\sqrt{2}}(x_\iota - \mathbf{i}x_\iota)$, $\iota = 1, \cdots, I$, is a family of GCCP with $(\lambda_1, \cdots, \lambda_I)$ as the corresponding family of GCC. Noting that $u_\iota = \mathbf{i}v_\iota$ the GCCP becomes $((\mathbf{i}v_1, v_1), \cdots, (\mathbf{i}v_I, v_I))$. Furthermore, the condition $(\mathbf{i}x)^t \Sigma x = 0$, $x \in \mathbb{R}^{I^{\dot{\cup}2}}$ also

[††] We have replaced and will also below replace p from (ABJ) with the index set I.
[‡‡] More precisely, the set Λ_p from (ABJ) together with the mapping π is a topological faithful representation of the orbit projection.
[*] For the empirical case this is true a.e. under H.
[†] The vector space of solutions has dimension 1.

defines H$_0$: $\Sigma \in \mathcal{H}^+(I,\mathbb{C})$ within H: $\mathcal{H}^+(I^{\dot\cup 2},\mathbb{R})$. Therefore, the GCCP and the corresponding GCC provide through Definition 6.1 an interpretation of the maximal invariant statistic in a similar way that classical canonical correlations and covariates provide an interpretation of the maximal invariant statistic for testing independence of two sets of variates.

2. Testing that a covariance matrix with complex structure has real structure, cf. Section 3 of (ABJ). In this testing problem the index set I is also replaced by the disjoint union $I^{\dot\cup 2}$ and $I \geq 2$. We can choose $G = \{\pm 1, \pm \mathbf{i}\}$ and G_0 generated by $\{\pm 1, \pm \mathbf{i}\ \pm \mathbf{s}\}$, where

$$\mathbf{s} = \begin{pmatrix} 0 & 1_I \\ 1_I & 0 \end{pmatrix}, \tag{8.2}$$

the matrix that *interchanges* the two I-components. Thus $\Theta := \mathbf{PD}_G(I^{\dot\cup 2}) = \mathcal{H}^+(I,\mathbb{C})$ and $\Theta_0 := \mathbf{PD}_{G_0}(I^{\dot\cup 2}) = \mathcal{H}^+(I,\mathbb{R}) \otimes I_2$, again using the notation from (ABJ).
Roughly speaking, we consider the testing problem that an arbitrary non-singular $I^{\dot\cup 2} \times I^{\dot\cup 2}$ variance matrix Σ of the form (8.1), a $\mathbb{C}$-*structure*, has the form

$$\Sigma = \begin{pmatrix} \Gamma & 0 \\ 0 & \Gamma \end{pmatrix},$$

an $\mathbb{R}$-*structure*. The average mapping t given in (7.2) reduces to the mapping t in (ABJ), page 400, and the residual R is given on the same page in (ABJ). The condition (7.3) has only two different terms and reduces to the condition that the residual belongs to $\mathcal{A}(I,\mathbb{R}) \otimes J_1$ defined in (ABJ) on page 400.
The set of eigenvalues of R wrt. T is again symmetric (around 0), i.e., $\Lambda = -\Lambda$, $|\lambda| < 1$, $\lambda \in \Lambda$, and each eigenvalue has even multiplicity and at least multiplicity 2. (When I is odd zero is always an eigenvalue.) The ordered family of the $[\frac{I}{2}]$ non-negative eigenvalues $(\lambda_1, \cdots, \lambda_{[\frac{I}{2}]})$ of R wrt. T each one repeated with half of its multiplicity[‡], i.e., $1 > \lambda_1 \geq \cdots \geq \lambda_{[\frac{I}{2}]} \geq 0$ becomes a the maximal invariant statistic, cf. (42) of (ABJ).
For the sake of simplicity in the description of interpretations we shall assume that $\lambda_1 > \cdots > \lambda_{[\frac{I}{2}]} > 0$, i.e., all the $2I$ eigenvalues are different.
To find a family of CDV we find for each λ_ι, $\iota = 1, \cdots, [\frac{I}{2}]$, one normalized (wrt. T) solution[§] $x_\iota \in \mathbb{R}^{I^{\cup 2}}$ to (4.2) when $\lambda = \lambda_\iota$. Another normalized and orthogonal (wrt. T) solution is then $\mathbf{i}x_\iota$. The vectors $\mathbf{s}x_\iota$ and $\mathbf{si}x_\iota$ are also normalized and orthogonal and solve (4.2) with $\lambda = -\lambda_\iota$, $\iota = 1, \cdots, [\frac{I}{2}]$. We thus have all non-zero eigenvalues and corresponding eigenvectors. Thus $(x_1, \mathbf{s}x_1, \mathbf{i}x_1, \mathbf{si}x_1, \cdots, x_{[\frac{I}{2}]}, \mathbf{s}x_{[\frac{I}{2}]}, \mathbf{i}x_{[\frac{I}{2}]}, \mathbf{si}x_{[\frac{I}{2}]})$ is a family of CDV with the corresponding family

$$(\lambda_1, \lambda_1, \lambda_1, \lambda_1, \lambda_2, \lambda_2, \lambda_2, \lambda_2, \cdots, \lambda_{[\frac{I}{2}]}, \lambda_{[\frac{I}{2}]}, \lambda_{[\frac{I}{2}]}, \lambda_{[\frac{I}{2}]})$$

of CVD. The CDV and the corresponding CVD provide through Definition 3.1 the first interpretation of the maximal invariant statistic.
It then follows from Proposition 6.1 that

$$((u_{1\mathbf{1}}, v_{1\mathbf{1}}), (u_{1\mathbf{i}}, v_{1\mathbf{i}}), (u_{2\mathbf{1}}, v_{2\mathbf{1}}), (u_{2\mathbf{i}}, v_{2\mathbf{i}}), \cdots, (u_{[\frac{I}{2}]\mathbf{1}}, v_{[\frac{I}{2}]\mathbf{1}}), (u_{[\frac{I}{2}]\mathbf{i}}, v_{[\frac{I}{2}]\mathbf{i}})),$$

[‡]When I is odd a zero eigenvalue is only repeated with half of its multiplicity minus 1.
[§]The vector space of solutions has dimension 2.

where $u_{\iota\mathbf{1}} := \frac{1}{\sqrt{2}}(x_\iota + \mathbf{s}x_\iota)$, $v_{\iota\mathbf{1}} := \frac{1}{\sqrt{2}}(x_\iota - \mathbf{s}x_\iota)$, $u_{\iota\mathbf{i}} := \frac{1}{\sqrt{2}}(\mathbf{i}x_\iota + \mathbf{si}x_\iota)$, and $v_{\iota\mathbf{i}} := \frac{1}{\sqrt{2}}(\mathbf{i}x_\iota - \mathbf{si}x_\iota)$, $\iota = 1, \cdots, I$, is a family of GCCP with $(\lambda_1, \lambda_1, \lambda_2, \lambda_2, \cdots, \lambda_{[\frac{I}{2}]}, \lambda_{[\frac{I}{2}]})$ as the corresponding family of GCC. Noting that all of these pairs (u, v) satisfy $\mathbf{s}u = u$ and $\mathbf{s}v = -v$, these vectors must have the form

$$u = \begin{pmatrix} \alpha \\ \alpha \end{pmatrix} \quad \text{and} \quad v = \begin{pmatrix} \beta \\ -\beta \end{pmatrix}, \ \alpha, \beta \in \mathbb{R}^I.$$

The condition

$$\begin{pmatrix} \alpha \\ \alpha \end{pmatrix}^t \Sigma \begin{pmatrix} \beta \\ -\beta \end{pmatrix} = 0, \ \alpha, \beta \in \mathbb{R}^I,$$

also defines H$_0$: $\Sigma \in \mathcal{H}^+(I, \mathbb{R}) \otimes I_2$ within H: $\Sigma \in \mathcal{H}^+(I, \mathbb{C})$. Therefore, the GCCP and the corresponding GCC provide through Definition 6.1 an interpretation of the maximal invariant statistic in a similar way that classical canonical correlations and covariates provide an interpretation of the maximal invariant statistic for testing independence of two sets of variates.

3. Testing that a covariance matrix with complex structure has quaternion structure, cf. Section 4 of (ABJ). In this testing problem, the index set I is replaced by the disjoint union $I^{\dot\cup 4}$ and $I \geq 2$. All $I^{\dot\cup 4} \times I^{\dot\cup 4}$ matrices are thus partitioned into 4×4 block matrices according to the decomposition $I^{\dot\cup 4}$ of the index set. We can choose $G = \{\pm 1, \pm\mathbf{j}\}$ and $G_0 = \{\pm 1, \pm\mathbf{i}, \pm\mathbf{j}, \pm\mathbf{k}\}$, where

$$\mathbf{i} := \begin{pmatrix} 0 & 0 & 0 & -1_I \\ 0 & 0 & 1_I & 0 \\ 0 & -1_I & 0 & 0 \\ 1_I & 0 & 0 & 0 \end{pmatrix}, \ \mathbf{j} := \begin{pmatrix} 0 & 0 & -1_I & 0 \\ 0 & 0 & 0 & -1_I \\ 1_I & 0 & 0 & 0 \\ 0 & 1_I & 0 & 0 \end{pmatrix}, \text{ and}$$

$$\mathbf{k} := \begin{pmatrix} 0 & -1_I & 0 & 0 \\ 1_I & 0 & 0 & 0 \\ 0 & 0 & 0 & 1_I \\ 0 & 0 & -1_I & 0 \end{pmatrix}$$

are the quaternion units. Thus $\Theta := \mathbf{PD}_G(I^{\dot\cup 4}) = \mathcal{H}^+(I^{\dot\cup 2}, \mathbb{C})$ and $\Theta_0 := \mathbf{PD}_{G_0}(I^{\dot\cup 4}) = \mathcal{H}^+(I, \mathbb{H})$, again using the notation from (ABJ). Roughly speaking, we consider the testing problem that an arbitrary non-singular $I^{\dot\cup 4} \times I^{\dot\cup 4}$ variance matrix with $\mathbb{C}$-structure

(8.3)

$$\Sigma = \begin{pmatrix} \Gamma_1 & E & -G & -F \\ E^t & \Gamma_2 & F^t & -H \\ G & F & \Gamma_1 & E \\ -F^t & H & E^t & \Gamma_2 \end{pmatrix} \text{ has the form } \begin{pmatrix} \Sigma_1 & E_0 & -G_0 & -F_0 \\ -E_0 & \Sigma_2 & -F_0 & G_0 \\ G_0 & F_0 & \Sigma_1 & E_0 \\ F_0 & -G_0 & -E_0 & \Sigma_2 \end{pmatrix}.$$

i.e., the $\mathbb{H}$-*structure*. The average mapping t given in (7.2) reduces to a sum with only two terms, cf. (56) in (ABJ). The condition (7.3) has only two different terms and reduces to the condition that the residual has the form (52) from (ABJ).

The eigenvalues of R wrt. T, i.e., the solution set Λ to the $4I$-degree polynomial equation (4.1) in $\lambda \in \mathbb{R}$, is symmetric (around 0), i.e., $\Lambda = -\Lambda$, and $|\lambda| < 1$, $\lambda \in \Lambda$, and each eigenvalue has even multiplicity of at least 2. The ordered family $(\lambda_1, \cdots, \lambda_I)$ of non-negative eigenvalues (of R wrt. T), each one repeated with half of its multiplicity, i.e., $1 > \lambda_1 \geq \cdots \geq \lambda_I \geq 0$, becomes a maximal invariant statistic, cf. (61) of (ABJ).

For the sake of simplicity in the description of interpretations, we shall assume that $\lambda_1 > \cdots > \lambda_I > 0$, i.e., all the $2I$ eigenvalues each with multiplicity 2 are different and non-zero.
To find a family of CDV we find for each λ_ι, $\iota = 1, \cdots, I$, one normalized¶ (wrt. T) solution $x_\iota \in \mathbb{R}^{I^{\dot\cup 4}}$ to (4.2) when $\lambda = \lambda_\iota$. Another normalized and orthogonal (wrt. T) solution is then $\mathbf{j}x_\iota$. The vectors $\mathbf{i}x_\iota$ and $\mathbf{k}x_\iota$ are also normalized and orthogonal and solve (4.2) with $\lambda = -\lambda_\iota$, $\iota = 1, \cdots, I$, and we thus have all eigenvalues and corresponding eigenvectors. Thus $(x_1, \mathbf{i}x_1, \mathbf{j}x_1, \mathbf{k}x_1, \cdots, x_I, \mathbf{i}x_I, \mathbf{j}x_I, \mathbf{k}x_I)$ is a family of CDV with the corresponding CVD family

$$(\lambda_1, \lambda_1, \lambda_1, \lambda_1, \lambda_2, \lambda_2, \lambda_2, \lambda_2, \cdots, \lambda_I, \lambda_I, \lambda_I, \lambda_I).$$

The CDV and the corresponding CVD provide through Definition 3.1 the first interpretation of the maximal invariant statistic.
It then follows from Proposition 6.1 that

$$((u_{1\mathbf{1}}, v_{1\mathbf{1}}), (u_{1\mathbf{j}}, v_{1\mathbf{j}}) \cdots, (u_{I\mathbf{1}}, v_{I\mathbf{1}}), (u_{I\mathbf{j}}, v_{I\mathbf{j}})),$$

where $u_{\iota\mathbf{1}} := \frac{1}{\sqrt{2}}(x_\iota + \mathbf{i}x_\iota)$, $v_{\iota\mathbf{1}} := \frac{1}{\sqrt{2}}(x_\iota - \mathbf{i}x_\iota)$, $u_{\iota\mathbf{j}} := \frac{1}{\sqrt{2}}(\mathbf{j}x_\iota + \mathbf{k}x_\iota)$, and $v_{\iota\mathbf{j}} := \frac{1}{\sqrt{2}}(\mathbf{j}x_\iota - \mathbf{k}x_\iota)$, $\iota = 1, \cdots, I$, is a family of GCCP with $(\lambda_1, \lambda_1, \lambda_2, \lambda_2, \cdots, \lambda_I, \lambda_I)$ as the corresponding family of GCC. All pairs (u, v) in GCCP satisfy $u = \mathbf{i}v$ and thus have the form $(\mathbf{i}v, v)$. Furthermore the condition $(\mathbf{i}u)^t\Sigma u = 0$, $u \in \mathbb{R}^{I^{\dot\cup 4}}$ also defines H$_0$: $\Sigma \in \mathcal{H}^+(I, \mathbb{H})$ within H: $\mathcal{H}^+(I^{\dot\cup 2}, \mathbb{C})$. Therefore, the GCCP and the corresponding GCC provide through Definition 6.1 an interpretation of the maximal invariant statistic in a similar way that classical canonical correlations and covariates provide an interpretation of the maximal invariant statistic for testing independence of two sets of variates.

4. Testing that a covariance matrix with quaternion structure has complex structure, cf. Section 5 of (ABJ). In this testing problem, the index set I is replaced by the disjoint union $I^{\dot\cup 4}$ and $I \geq 2$. We can choose $G = \{\pm 1, \pm\mathbf{i}, \pm\mathbf{j}, \pm\mathbf{k}\}$ and G_0 is generated by $\{\pm 1, \pm\mathbf{i}, \pm\mathbf{j}, \pm\mathbf{k}, \pm\mathbf{s}\}$, where

$$\mathbf{s} = \begin{pmatrix} 0 & 0 & 1_I & 0 \\ 0 & 0 & 0 & 1_I \\ 1_I & 0 & 0 & 0 \\ 0 & 1_I & 0 & 0 \end{pmatrix}, \tag{8.4}$$

the matrix that interchanges two $I^{\dot\cup 2}$-components. Thus $\Theta := \mathbf{PD}_G(I^{\dot\cup 4}) = \mathcal{H}^+(I, \mathbb{H})$ and $\Theta_0 := \mathbf{PD}_{G_0}(I^{\dot\cup 4}) = \mathcal{H}^+(I, \mathbb{C}) \otimes I_2$, again using the notation from (ABJ).
Roughly speaking, we consider the testing problem that an arbitrary non-singular $I^{\dot\cup 4} \times I^{\dot\cup 4}$ variance matrix Γ with $\mathbb{H}$-*structure*, i.e.,

$$\Sigma = \begin{pmatrix} \Sigma_1 & E & -G & -F \\ -E & \Sigma_2 & -F & G \\ G & F & \Sigma_1 & E \\ F & -G & -E & \Sigma_2 \end{pmatrix} \text{ has the form } \begin{pmatrix} \Sigma_1 & E & 0 & 0 \\ -E & \Sigma_2 & 0 & 0 \\ 0 & 0 & \Sigma_1 & E \\ 0 & 0 & -E & \Sigma_2 \end{pmatrix},$$

¶The vector space of solutions has dimension 2.

i.e., a $\mathbb{C}$-*structure*. The average mapping t given in (7.2) reduces to the mapping t in (ABJ), page 406 and the residual R is given on next page in (ABJ). The condition (7.3) has only two different terms and reduces to the condition that the residual belongs to $\mathcal{A}(I,\mathbb{C}) \otimes J_1$ defined in (ABJ) page 407.

The set of eigenvalues of R wrt. T is again symmetric (around 0), i.e., $\Lambda = -\Lambda$, $|\lambda| < 1$, $\lambda \in \Lambda$, and each eigenvalue has multiplicity divisible by 4. (When I is odd zero is always an eigenvalue with multiplicity at least 4.) The ordered family of the first $[\frac{I}{2}]$ non-negative eigenvalues $(\lambda_1, \cdots, \lambda_{[\frac{I}{2}]})$ of R wrt. T each one repeated with a quarter of its multiplicity[‖], i.e., $1 > \lambda_1 \geq \cdots \geq \lambda_{[\frac{I}{2}]} \geq 0$, becomes a maximal invariant statistic, cf. (71) of (ABJ).

For the sake of simplicity in the description of interpretations, we shall assume that $\lambda_1 > \cdots > \lambda_{[\frac{I}{2}]} > 0$.

For each $\iota = 1, \cdots, [\frac{I}{2}]$ we find one $x_\iota \in \mathbb{R}^{I^{\dot\cup 4}}$ normalized (wrt. T) solution [**] to (4.2) with $\lambda = \lambda_\iota$. Then $\mathbf{i}x_i, \mathbf{j}x_i$, and $\mathbf{k}x_i$ are 3 other mutually orthogonal and normalized solutions all orthogonal to x_i. The 4 vectors $\mathbf{s}x_\iota, \mathbf{si}x_\iota, \mathbf{sj}x_\iota, \mathbf{sk}x_\iota$ are normalized, mutually orthogonal and solve (4.2) for $\lambda = -\lambda_\iota$. We have thus found all eigenvectors corresponding to all non-zero eigenvalues. Define

$$\mathbf{x}_\iota := (x_\iota, \mathbf{s}x_\iota, \mathbf{i}x_\iota, \mathbf{si}x_\iota, \mathbf{j}x_\iota, \mathbf{sj}x_\iota, \mathbf{k}x_\iota, \mathbf{sk}x_\iota) \in (\mathbb{R}^{I^{\dot\cup 4}})^8,$$

$\iota = 1, \cdots, [\frac{I}{2}]$. Then the vector $(\mathbf{x}_1, \cdots, \mathbf{x}_{[\frac{I}{2}]})$ is a family of CDV with the corresponding family

$$((\lambda_1 | \iota = 1, \cdots, 8), \cdots, (\lambda_{[\frac{I}{2}]} | \iota = 1, \cdots, 8))$$

of CVD. The CDV and the corresponding CVD provide through Definition 3.1 the first interpretation of the maximal invariant statistic.

It then follows from Proposition 6.1 that

$$\begin{array}{l} ((u_{1\mathbf{1}}, v_{1\mathbf{1}}), (u_{1\mathbf{i}}, v_{1\mathbf{i}}), (u_{1\mathbf{j}}, v_{1\mathbf{j}}), (u_{1\mathbf{k}}, v_{1\mathbf{k}}), \\ (u_{2\mathbf{1}}, v_{2\mathbf{1}}), (u_{2\mathbf{i}}, v_{2\mathbf{i}}), (u_{2\mathbf{j}}, v_{2\mathbf{j}}), (u_{2\mathbf{k}}, v_{1\mathbf{k}}), \cdots, \\ (u_{[\frac{I}{2}]\mathbf{1}}, v_{[\frac{I}{2}]\mathbf{1}}), (u_{[\frac{I}{2}]\mathbf{i}}, v_{[\frac{I}{2}]\mathbf{i}}), (u_{[\frac{I}{2}]\mathbf{j}}, v_{[\frac{I}{2}]\mathbf{j}}), (u_{[\frac{I}{2}]\mathbf{k}}, v_{[\frac{I}{2}]\mathbf{k}})) \end{array}$$

where $u_{\iota\mathbf{1}} := \frac{1}{\sqrt{2}}(x_\iota + \mathbf{s}x_\iota)$, $v_{\iota\mathbf{1}} := \frac{1}{\sqrt{2}}(x_\iota - \mathbf{s}x_\iota)$, $u_{\iota\mathbf{i}} := \frac{1}{\sqrt{2}}(\mathbf{i}x_\iota + \mathbf{si}x_\iota)$, $v_{\iota\mathbf{i}} := \frac{1}{\sqrt{2}}(\mathbf{i}x_\iota - \mathbf{si}x_\iota)$, $u_{\iota\mathbf{j}} := \frac{1}{\sqrt{2}}(\mathbf{j}x_\iota + \mathbf{sj}x_\iota)$, $v_{\iota\mathbf{j}} := \frac{1}{\sqrt{2}}(\mathbf{j}x_\iota - \mathbf{sj}x_\iota)$, $u_{\iota\mathbf{k}} := \frac{1}{\sqrt{2}}(\mathbf{k}x_\iota + \mathbf{sk}x_\iota)$, $v_{\iota\mathbf{k}} := \frac{1}{\sqrt{2}}(\mathbf{k}x_\iota - \mathbf{sk}x_\iota)$, $\iota = 1, \cdots, [\frac{I}{2}]$, is a family of GCCP with $((\lambda_1 | \iota = 1, \cdots, 4), \cdots, (\lambda_{[\frac{I}{2}]} | \iota = 1, \cdots, 4))$ as the corresponding family of GCC. Noting that all of these pairs (u, v) satisfy $\mathbf{s}u = u$ and $\mathbf{s}v = -v$, these vectors must have the form

$$u = \begin{pmatrix} \alpha \\ \alpha \end{pmatrix} \quad \text{and} \quad v = \begin{pmatrix} \beta \\ -\beta \end{pmatrix}, \; \alpha, \beta \in \mathbb{R}^{I^{\dot\cup 2}}.$$

The condition

$$\begin{pmatrix} \alpha \\ \alpha \end{pmatrix}^t \Sigma \begin{pmatrix} \beta \\ -\beta \end{pmatrix} = 0, \; \alpha, \beta \in \mathbb{R}^{I^{\dot\cup 2}},$$

also defines H_0: $\Sigma \in \mathcal{H}^+(I,\mathbb{C}) \otimes I_2$ within H: $\Sigma \in \mathcal{H}^+(I,\mathbb{H})$. Therefore, the GCCP and the corresponding GCC provide through Definition 6.1 an interpretation of the maximal invariant statistic in a similar way that classical canonical correlations and

[‖]When I is odd a zero eigenvalue is only repeated with a quarter of its multiplicity minus 1.

[**]The vector space of solutions has dimension 4.

covariates provide an interpretation of the maximal invariant statistic for testing independence of two sets of variates.

5. Testing independence of two families of variates where the simultaneous variance matrix has real structure, cf. Section 6 of (ABJ). In this testing problem the index set I is replaced by the disjoint union $I\dot{\cup}J$ assuming without loss of generality that $I \geq J$. All $(I\dot{\cup}J) \times (I\dot{\cup}J)$ matrices are thus partitioned into 2×2 block matrices according to the decomposition $I\dot{\cup}J$ of the index set. We can choose $G = \{\pm 1_{I\dot{\cup}J}\}$ and $G_0 = \{\pm 1_{I\dot{\cup}J}, \pm \mathbf{f}\}$ where $\mathbf{f} := \mathrm{Diag}(1_I, -1_J)$, a *signshift matrix*.
Thus $\Theta := \mathbf{PD}_G(I\dot{\cup}J) = \mathbf{PD}(I\dot{\cup}J) = \mathcal{H}^+(I\dot{\cup}J, \mathbb{R})$ and $\Theta_0 := \mathbf{PD}_{G_0}(I\dot{\cup}J) = \mathcal{H}^+(I,\mathbb{R}) \oplus \mathcal{H}^+(J,\mathbb{R})$, again using the notation from (ABJ).
Roughly speaking, we consider the testing problem that an arbitrary non-singular $(I\dot{\cup}J) \times (I\dot{\cup}J)$ variance matrix

$$\Sigma = \begin{pmatrix} \Sigma_1 & E \\ E^t & \Sigma_2 \end{pmatrix} \text{ has the form } \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}, \tag{8.5}$$

i.e., testing independence between the I and J components of a normally distributed random variable with values in $\mathbb{R}^{I\dot{\cup}J}$. The average mapping t given in (7.2) reduces to the mapping t in (ABJ), page 409 ($\mathbb{D} = \mathbb{R}$) and the residual R is given by

$$R = \begin{pmatrix} 0 & E \\ E^t & 0 \end{pmatrix}. \tag{8.6}$$

The condition (7.3) has again only two different terms and reduces to the condition that the residual takes the form (8.6).
The eigenvalues of R wrt. T, i.e., the solution set Λ to the $(I+J)$-degree polynomial equation (4.1) in $\lambda \in \mathbb{R}$, is again symmetric (around 0), i.e., $\Lambda = -\Lambda$, and $|\lambda| < 1$, $\lambda \in \Lambda$. If $I > J$, zero is always an eigenvalue with at least multiplicity $I - J$. The ordered family $(\lambda_1, \cdots, \lambda_J)$ of the first J non-negative eigenvalues (of R wrt. T), each one repeated according to its multiplicity[††], i.e., $1 > \lambda_1 \geq \cdots \geq \lambda_J \geq 0$, becomes a maximal invariant statistic.
For the sake of simplicity in the description of interpretations, we shall again assume that $\lambda_1 > \cdots > \lambda_J > 0$, i.e., there is $2J$ eigenvalues different from zero, all with multiplicity 1.
To find a family of CDV, we find for each $\iota = 1, \cdots, J$ one normalized (wrt. T) solution[‡‡] x_ι to (4.2) when $\lambda = \lambda_\iota$. Then $\mathbf{f}x_\iota$ is also normalized and solves (4.2) for $\lambda = -\lambda_\iota$, and we have all eigenvectors corresponding to all non-zero eigenvalues. The family $(x_1, \mathbf{f}x_1, \cdots, x_J, \mathbf{f}x_J)$ is then a family of CDV with the corresponding family $(\lambda_1, \lambda_1, \cdots, \lambda_J, \lambda_J)$ of CVD. The CDV and the corresponding CVD provide through Definition 3.1 the first interpretation of the maximal invariant statistic.
It then follows from Proposition 6.1 that $((u_1, v_1), \cdots, (u_J, v_J))$, where $u_\iota := \frac{1}{\sqrt{2}}(x_\iota + \mathbf{f}x_\iota)$, and $v_\iota := \frac{1}{\sqrt{2}}(x_\iota - \mathbf{f}x_\iota)$, $\iota = 1, \cdots, J$, is a family of GCCP with $(\lambda_1, \cdots, \lambda_J)$ as the corresponding family of GCC. Noting that all of these pairs (u, v) satisfy

[††]If $I > J$ the eigenvalue zero is repeated with its multiplicity minus $I - J$.
[‡‡]The vector space of solutions has dimension 1.

$\mathbf{f}u = u$ and $\mathbf{f}v = -v$, these vectors must have the form

$$u = \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \text{ and } v = \begin{pmatrix} 0 \\ \beta \end{pmatrix}, \ \alpha \in \mathbb{R}^I, \ \beta \in \mathbb{R}^J.$$

The condition

$$\begin{pmatrix} \alpha \\ 0 \end{pmatrix}^t \Sigma \begin{pmatrix} 0 \\ \beta \end{pmatrix} = 0, \ \alpha \in \mathbb{R}^I, \ \beta \in \mathbb{R}^J,$$

also defines H_0: $\Sigma \in \mathcal{H}^+(I, \mathbb{R}) \otimes I_2$ within H: $\Sigma \in \mathcal{H}^+(I, \mathbb{C})$. Therefore, the GCCP and the corresponding GCC provide the second interpretation of the maximal invariant statistic, namely that of classical canonical correlations and covariates.

6. Testing independence of two families of variates where the simultaneous variance matrix has complex structure, cf. Section 6 of (ABJ). In this testing problem, the index set I from case 2 above is replaced by the disjoint union $(I\dot{\cup}J)$, assuming without loss of generality that $I \geq J$. All $(I\dot{\cup}J)^{\dot{\cup}2} \times (I\dot{\cup}J)^{\dot{\cup}2}$ matrices are thus partitioned into 4×4 block matrices according to the decomposition of the index set, $(I\dot{\cup}J)^{\dot{\cup}2}$. We can choose $G = \{\pm 1, \pm \mathbf{i}\}$ and $G_0 := \{\pm 1, \pm \mathbf{i}, \pm \mathbf{g}\}$, $\mathbf{g} := \mathrm{Diag}(1, -1, 1, -1)$, a *signshift matrix*.
Thus $\Theta := \mathbf{PD}_G((I\dot{\cup}J)^{\dot{\cup}2}) = \mathcal{H}^+((I\dot{\cup}J), \mathbb{C})$ and
$\Theta_0 := \mathbf{PD}_{G_0}((I\dot{\cup}J)^{\dot{\cup}2}) = \mathcal{H}^+(I, \mathbb{C}) \oplus \mathcal{H}^+(J, \mathbb{C})$, again using the notation from (ABJ).
Roughly speaking, we consider the testing problem that a non-singular $(I\dot{\cup}J) \times (I\dot{\cup}J)^{\dot{\cup}2}$ variance matrix

$$(8.7) \quad \Sigma = \begin{pmatrix} \Sigma_1 & E & -F & -G \\ E^t & \Sigma_2 & G^t & -H \\ F & G & \Sigma_1 & E \\ -G^t & H & E^t & \Sigma_2 \end{pmatrix} \text{ has the form } \begin{pmatrix} \Sigma_1 & 0 & -F & 0 \\ 0 & \Sigma_2 & 0 & -H \\ F & 0 & \Sigma_1 & 0 \\ 0 & H & 0 & \Sigma_2 \end{pmatrix},$$

i.e., testing independence between the I components and the J components of a normally distributed random variable with values in $\mathbb{R}^{(I\dot{\cup}J)^{\dot{\cup}2}}$. The average mapping t given in (7.2) reduces to the mapping t in (ABJ), page 409 ($\mathbb{D} = \mathbb{C}$) and the residual R is given by

$$(8.8) \qquad R = \begin{pmatrix} 0 & E & 0 & -G \\ E^t & 0 & G^t & 0 \\ 0 & G & 0 & E \\ -G^t & 0 & E^t & 0 \end{pmatrix}.$$

The condition (7.3) has again only two different terms and reduces to the condition that the residual takes the form (8.8).
The eigenvalues of R wrt. T, i.e., the solution set Λ to the $2(I+J)$-degree polynomial equation (4.1) in $\lambda \in \mathbb{R}$, is again symmetric (around 0), i.e., $\Lambda = -\Lambda$, $|\lambda| < 1$, $\lambda \in \Lambda$, and each solution has even multiplicity of at least 2. If $I > J$, zero is always an eigenvalue with at least multiplicity $2(I - J)$. The ordered family $(\lambda_1, \cdots, \lambda_J)$ of non-negative eigenvalues (of R wrt. T), each one repeated according to half of its multiplicity*, i.e., $1 > \lambda_1 \geq \cdots \geq \lambda_J \geq 0$, becomes a maximal invariant statistic.

*If $I > J$ the eigenvalue zero is repeated with its multiplicity minus $2(I - J)$.

For the sake of simplicity in the description of interpretations, we shall again assume that $\lambda_1 > \cdots > \lambda_J > 0$, i.e., there are $2J$ distinct eigenvalues different from zero all with multiplicity 2.
To find a family of CDV we find for each $\iota = 1, \cdots, J$ one normalized (wrt. T) solution[†] x_ι to (4.2) when $\lambda = \lambda_\iota$. Then $\mathbf{i}x_\iota$ is also a normalized solution to the same equation orthogonal to x_ι. The vectors $\mathbf{g}x_\iota$ and $\mathbf{gi}x_\iota$ are orthogonal and normalized vectors and they both solve (4.2) when $\lambda = -\lambda_\iota$. Thus

$$(x_1, \mathbf{g}x_1, \mathbf{i}x_1, \mathbf{gi}x_1, \cdots, x_1, \mathbf{g}x_J, \mathbf{i}x_J, \mathbf{gi}x_J)$$

is a family of CDV with the corresponding family

$$(\lambda_1, \lambda_1, \lambda_1, \lambda_1, \lambda_2, \lambda_2, \lambda_2, \lambda_2, \cdots, \lambda_J, \lambda_J, \lambda_J, \lambda_J)$$

of CVD. The CDV and the corresponding CVD provide through Definition 3.1 the first interpretation of the maximal invariant statistic.
It then follows from Proposition 6.1 that

$$((u_{1\mathbf{1}}, v_{1\mathbf{1}}), (u_{1\mathbf{i}}, v_{1\mathbf{i}}), \cdots, (u_{J\mathbf{1}}, v_{J\mathbf{1}}), (u_{J\mathbf{i}}, v_{J\mathbf{i}})),$$

where $u_{\iota\mathbf{1}} := \frac{1}{\sqrt{2}}(x_\iota + \mathbf{g}x_\iota)$, $v_{\iota\mathbf{1}} := \frac{1}{\sqrt{2}}(x_\iota - \mathbf{g}x_\iota)$, $u_{\iota\mathbf{i}} := \frac{1}{\sqrt{2}}(\mathbf{i}x_\iota + \mathbf{gi}x_\iota)$, $v_{\iota\mathbf{i}} := \frac{1}{\sqrt{2}}(\mathbf{i}x_\iota - \mathbf{gi}x_\iota)$, $\iota = 1, \cdots, J$, is a family of GCCP with $(\lambda_1, \lambda_1, \lambda_2, \lambda_2, \cdots, \lambda_J, \lambda_J)$ as the corresponding family of GCC. Noting that all of these pairs (u, v) satisfy $\mathbf{g}u = u$ and $\mathbf{g}v = -v$, these vectors must have the form

$$u = \begin{pmatrix} \alpha_1 \\ 0 \\ \alpha_2 \\ 0 \end{pmatrix} \quad \text{and} \quad v = \begin{pmatrix} 0 \\ \beta_1 \\ 0 \\ \beta_2 \end{pmatrix}, \ \alpha_1, \alpha_2 \in \mathbb{R}^I, \ \beta_1, \beta_2 \in \mathbb{R}^J.$$

The condition

$$\begin{pmatrix} \alpha_1 \\ 0 \\ \alpha_2 \\ 0 \end{pmatrix}^t \Sigma \begin{pmatrix} 0 \\ \beta_1 \\ 0 \\ \beta_2 \end{pmatrix} = 0, \ \alpha_1, \alpha_2 \in \mathbb{R}^I, \ \beta_1, \beta_2 \in \mathbb{R}^J,$$

also defines H_0: $\Sigma \in \mathcal{H}^+(I, \mathbb{C}) \oplus \mathcal{H}^+(J, \mathbb{C})$ within H: $\Sigma \in \mathcal{H}^+(I \dot\cup J, \mathbb{C})$. Therefore, the GCCP and the corresponding GCC provide through Definition 6.1 the second interpretation of the maximal invariant statistic, namely that of classical canonical correlations and covariates in the $\mathbb{C}$-case.

7. Testing independence of two families of variates where the simultaneous variance matrix has quaternion structure, cf. Section 6 of (ABJ). This case is very similar to case 6 above, and we shall not give the details. The index set is $(I \dot\cup J)^{\dot\cup 4}$, thus matrices are partitioned into 8×8 block matrices. Also, Λ is symmetric, each eigenvalue has multiplicity at least 4, and zero is an eigenvalue with multiplicity at least $4(I - J)$. The ordered family of the first J eigenvalues is a maximal invariant. The space of solutions to (4.2) with $\lambda = \lambda_\iota$, $\iota = 1, \cdots, J$, is 4 dimensional, and the structure of CDV and CVD are similar[*] to case 6. The GCCP and GCC are also similar[†]. They are the canonical correlations and covariates in the $\mathbb{H}$-case.

[†]The vector space of solutions has dimension 2.
[*]Each λ_i, $i = 1, \cdots, J$, is repeated 8 times.
[†]Each λ_i, $i = 1, \cdots, J$, is repeated 4 times.

8. Testing that two variance matrices with real structure are identical, cf. Section 7 of (ABJ). In this testing problem, the index set I is replaced by the disjoint union $I^{\dot\cup 2}$ and $I \geq 2$. All $I^{\dot\cup 2} \times I^{\dot\cup 2}$ matrices are thus partitioned into 2×2 equal size block matrices according to the decomposition $I^{\dot\cup 2}$ of the index set. We can choose $G = \{\pm 1, \pm \mathbf{f}\}$ and G_0 as the group generated by $\{\pm 1, \pm \mathbf{f}, \pm \mathbf{s}\}$.
Thus $\Theta := \mathbf{PD}_G(I^{\dot\cup 2}) = \mathcal{H}^+(I, \mathbb{R}) \oplus \mathcal{H}^+(I, \mathbb{R})$ and $\Theta_0 := \mathbf{PD}_{G_0}(I^{\dot\cup 2}) = \mathcal{H}^+(I, \mathbb{R}) \otimes I_2$, using the notation from (ABJ).
Roughly speaking, we consider the testing problem[‡] that an arbitrary non-singular $I^{\dot\cup 2} \times I^{\dot\cup 2}$ variance matrix of the form

$$\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \text{ has the form } \begin{pmatrix} \Gamma & 0 \\ 0 & \Gamma \end{pmatrix}. \tag{8.9}$$

The average mapping t given in (7.2) is then given by

$$t\left(\begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix}\right) := \frac{1}{2} \begin{pmatrix} S_1 + S_2 & 0 \\ 0 & S_1 + S_2 \end{pmatrix} =: \begin{pmatrix} S_0 & 0 \\ 0 & S_0 \end{pmatrix},$$

with the residual given by $R := \frac{1}{2}\mathrm{Diag}((S_2 - S_1), (S_1 - S_2))$. The condition (7.3) has only two different terms and reduces to the condition that the residual has the form $R = \mathrm{Diag}(-R_0, R_0)$, where R_0 is a symmetric $I \times I$ matrix[§].
The eigenvalues of R wrt. T, i.e., the solution set Λ to the $2I$-degree polynomial (4.1), is symmetric (around 0), i.e., $\Lambda = -\Lambda$, and $|\lambda| < 1$, $\lambda \in \Lambda$. The ordered family $(\lambda_1, \cdots, \lambda_I)$ of non-negative eigenvalues each one repeated according to its multiplicity, i.e., $1 > \lambda_1 \geq \cdots \geq \lambda_I \geq 0$, is in this case **not** a maximal invariant statistic, but only an invariant statistic.
For the sake of simplicity in the interpretations, we shall assume that $\lambda_1 > \cdots > \lambda_I > 0$, i.e., all the $2I$ eigenvalues are distinct and different from zero.
To find a family of CDV we find for each λ_ι, $\iota = 1, \cdots, I$, a normalized[*] (wrt. T) solution $x_\iota \in \mathbb{R}^{I^{\dot\cup 2}}$ to (4.2) when $\lambda = \lambda_\iota$. Then $\mathbf{s}x_\iota$ solves (4.2) with $\lambda = -\lambda_\iota$, $\iota = 1, \cdots, I$. Thus $(x_1, \mathbf{s}x_1, \cdots, x_I, \mathbf{s}x_I)$ is a family of CDV with the corresponding family $(\lambda_1, \lambda_1, \cdots, \lambda_I, \lambda_I)$ of CVD. The CDV and the corresponding CVD provide through Definition 3.1 the first interpretation of the eigenvalues of R wrt. T, although they do not constitute a maximal invariant.
It then follows from Proposition 6.1 that $((u_1, v_1), \cdots, (u_I, v_I))$, where $u_\iota := \frac{1}{\sqrt{2}}(x_\iota + \mathbf{s}x_\iota)$, and $v_\iota := \frac{1}{\sqrt{2}}(x_\iota - \mathbf{s}x_\iota)$, $\iota = 1, \cdots, I$, is a family of GCCP with $(\lambda_1, \cdots, \lambda_I)$ as the corresponding family of GCC. Noting that any of these pairs (u, v) satisfies $\mathbf{s}u = u$ and $\mathbf{s}v = -v$ these vectors have the form

$$u = \begin{pmatrix} \alpha \\ \alpha \end{pmatrix} \text{ and } v = \begin{pmatrix} \beta \\ -\beta \end{pmatrix}, \quad \alpha, \beta \in \mathbb{R}^I.$$

Furthermore, the condition

$$\begin{pmatrix} \alpha \\ \alpha \end{pmatrix}^t \Sigma \begin{pmatrix} \beta \\ -\beta \end{pmatrix} = 0 \quad \alpha, \beta \in \mathbb{R}^I,$$

[‡]The testing problem in (ABJ) Section 7 with $\mathbb{D} = \mathbb{R}$ is somewhat more general since we only consider the case $N = N_1 = N_2$ from (ABJ). Nevertheless, the present treatment of the equal size sample case only differs from the more general case in an insignificant manner.

[§]In the more general case where the sample sizes are not necessary equal $\frac{1}{2}(\Sigma_1 + \Sigma_2)$ is replaced by a convex linear combination of $c_1\Sigma_1 + c_2\Sigma_2$ and thus $R = \mathrm{Diag}(c_2(\Sigma_1 - \Sigma_2), c_1(\Sigma_2 - \Sigma_1))$.

[*]The vector space of solutions has dimension 1.

also defines H$_0$: $\Sigma \in \mathcal{H}^+(I,\mathbb{R}) \otimes I_2$ within H: $\Sigma \in \mathcal{H}^+(I,\mathbb{R}) \oplus \mathcal{H}^+(I,\mathbb{R})$. Therefore, the GCCP and the corresponding GCC provide through Definition 6.1 a second interpretation of the eigenvalues of R wrt. T.

As it is easily obtained from (ABJ) page 411, the maximal invariant statistic is the ordered family $(\lambda_{01}, \cdots, \lambda_{0I})$ of eigenvalues of R_0 wrt. S_0, i.e., of $\frac{1}{2}(S_1 - S_2)$ wrt. $\frac{1}{2}(S_1 + S_2)$, each one repeated according to its multiplicity, i.e., $1 > \lambda_{01} \geq \lambda_{02} \geq \cdots \geq \lambda_{0I} > -1$.

For the sake of simplicity in the formulation of the interpretation below, we shall assume that $\lambda_{01} > \cdots > \lambda_{0I}$, i.e., all the I eigenvalues are distinct and that they are the *theoretical eigenvalues*, i.e., the eigenvalues of $\Delta_0 := \frac{1}{2}(\Sigma_1 - \Sigma_2)$ wrt $\Sigma_0 := \frac{1}{2}(\Sigma_1 + \Sigma_2)$. To obtain these I eigenvalues, one solves the I degree polynomial equation (4.1) with R and T replaced by Δ_0 and Σ_0, respectively. Let $x_{0\kappa} \in \mathbb{R}^I$, for $\kappa = 1, \cdots, I$, be Σ_0-normalized solutions of (4.2) with λ, R and T, replaced by $\lambda_{0\kappa}$, Δ_0, and Σ_0. Note that these solutions are unique up to sign changes.

The $2I$ ordered vectors

$$\begin{pmatrix} x_{01} \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} x_{0I} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ x_{01} \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ x_{0I} \end{pmatrix} \in \mathbb{R}^{I \dot{\cup} 2}$$

are then eigenvectors of $\Delta = \mathrm{Diag}(\Delta_0, -\Delta_0)$ wrt. $t(\Sigma) = \mathrm{Diag}(\Sigma_0, \Sigma_0)$ with corresponding eigenvalues

$$(\lambda_{01}, \cdots, \lambda_{0I}, -\lambda_{01}, \cdots, -\lambda_{0I}).$$

$$\lambda_{0\kappa} = \mathbb{V}_\Sigma(\begin{pmatrix} x_{0\kappa} \\ 0 \end{pmatrix}) - \mathbb{V}_{t(\Sigma)}(\begin{pmatrix} x_{0\kappa} \\ 0 \end{pmatrix}) = \mathbb{V}_{\Sigma_1}(x_{0\kappa}) - \mathbb{V}_{\Sigma_0}(x_{0\kappa}),$$

$$-\lambda_{0\kappa} = \mathbb{V}_\Sigma(\begin{pmatrix} 0 \\ x_{0\kappa} \end{pmatrix}) - \mathbb{V}_{t(\Sigma)}(\begin{pmatrix} 0 \\ x_{0\kappa} \end{pmatrix}) = \mathbb{V}_{\Sigma_2}(x_{0\kappa}) - \mathbb{V}_{\Sigma_0}(x_{0\kappa}).$$

This provides one possible interpretation of the (theoretical) maximal invariant statistic, namely differences (not numerical) between variances of certain independent normalized (wrt. Σ_0) linear forms on $\mathbb{R}^I$.

9-10. Testing that two variance matrices with complex or quaternion structure are identical, cf. Section 7 of (ABJ). These cases are very similar to case 8 above and will be left to the reader.

References

Andersson, S.A. (1982). Distributions of maximal invariants using quotient measures. *Ann. Statist.* **10** 955-961.

Andersson, S.A., Brøns, H.K., and Tolver Jensen, S. (1983). Distribution of Eigenvalues in multivariate statistical analysis. *Ann. Statist.* **11** 392-415.

Andersson, S.A. and Madsen, J. (1998). Symmetry and lattice conditional independence in a multivariate normal distribution. *Ann. Statist.* **26** 525-572.

Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis, 2nd ed.* John Wiley & Sons, New York.

Bertelsen, A. (1989). On non-null distributions connected with testing reality of a complex normal distribution. *Ann. Statist.* **17** 929-936.

Bertelsen, A. 1990. On non-null distributions connected with testing that a real normal distribution is complex. *J. Mult. Anal.* **32** 282-289.

Hotelling, Harold (1936). Relations Between Two Sets of Variates. *Biometrika* **28** 321-377.

Department of Statistics, Indiana University, Bloomington, IN 47405
Department of Mathematics, Indiana University, Bloomington, IN 47405
E-mail address: standers@indiana.edu, jescrawf@indiana.edu