GENERAL CANONICAL CORRELATIONS WITH APPLICATIONS TO GROUP SYMMETRY MODELS

STEEN A. ANDERSSON AND JESSE B. CRAWFORD
INDIANA UNIVERSITY AND TARLETON STATE 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. For seven of the ten group symmetry testing problems discussed in Andersson, Brøns, and Jensen (1983), these are the eigenvalues used to represent the maximal invariant statistic, and additional observations regarding the canonical correlations are made for these testing problems.

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 treatment of this subject in, for example, Chapter 12 of Anderson (2003), the canonical correlations are obtained by solving a matrix eigenvalue problem, and the canonical covariates are obtained by solving the corresponding linear equations. This testing problem is invariant, and the empirical canonical correlations are a maximal invariant statistic\(^1\), making them important in the study of invariant test statistics. A brief overview of canonical correlations is presented in Section 2.

In Andersson et al. (1983), abbreviated (ABJ), ten fundamental testing problems, including the one mentioned above, 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 Madsen (1998), Appendix A. For each of these invariant testing problems, (ABJ) found a concrete representation of the maximal invariant statistic and a representation of its central distribution in terms of a density wrt. Lebesgue measure. Each maximal invariant 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.

\(^1\) Consult Chapter 6 of Lehmann (1986) for an overview of invariant testing problems.
For the above test of independence, the eigenvalues used by (ABJ) to represent the maximal invariant statistic are the empirical canonical correlations. This raises the following question: can canonical correlations be generalized in a natural way to the other fundamental testing problems, and if so, are the general canonical correlations a maximal invariant in all ten cases?

In Section 3, general canonical correlations and covariates are defined for a class of testing problems, including the fundamental testing problems discussed above, that involve the covariance matrix of a centered multivariate normal distribution. The general canonical correlations and covariates are characterized in terms of eigenvalues and eigenvectors, and in Section 4, the general empirical canonical correlations are shown to be a maximal invariant statistic for seven of the ten fundamental testing problems. Furthermore, the general canonical correlations and covariates for these testing problems exhibit additional structure, providing the maximal invariant statistic with a statistical interpretation, as described in Remark 2.5. An alternative statistical interpretation of the maximal invariant is provided by the canonical variance distances, described in Section 5, which are analogous to general canonical correlations.

2. Classical Canonical Correlations

Let $\text{PD}(I)$ and $\text{GL}(I)$ be the sets of positive definite and invertible $I \times I$ real matrices, respectively, and suppose $X_1, \ldots, X_N$ are i.i.d. random vectors with distribution $N(\Sigma)$, the multivariate normal distribution on $\mathbb{R}^I$ with expectation zero and covariance matrix $\Sigma \in \text{PD}(I)$. Consider testing the null hypothesis that the first $I_1$ components of $X_n$ are independent of the last $I_2$ components, where $I_1 \geq I_2$, and $I_1 + I_2 = I$, that is,

$$(2.1) \quad H_0 : \Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \quad \text{vs.} \quad H : \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. $$

Partition $X_1, \ldots, X_N$ and $\Sigma$ accordingly, setting

$$ X_n = \begin{pmatrix} X_n^{(1)} \\ X_n^{(2)} \end{pmatrix}, \quad n = 1, \ldots, N, \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. $$

The idea of canonical correlations is to measure the statistical dependence between $X_n^{(1)}$ and $X_n^{(2)}$ via correlations between linear combinations of their components, $x^t X_n^{(1)}$ and $y^t X_n^{(2)}$. Given $x \in \mathbb{R}^{I_1}$ and $y \in \mathbb{R}^{I_2}$, define $\text{Cov}_\Sigma(x, y) := x^t \Sigma_{12} y$, the covariance between $x^t X_n^{(1)}$ and $y^t X_n^{(2)}$. Similarly, define $\text{Var}_\Sigma(x) := x^t \Sigma_{11} x$ and $\text{Var}_\Sigma(y) := y^t \Sigma_{22} y$, the variances of these random variables. The first canonical correlation is

$$ c_1 = \max \{ \text{Cov}_\Sigma(x, y) \mid x \in \mathbb{R}^{I_1}, y \in \mathbb{R}^{I_2}, \text{Var}_\Sigma(x) = \text{Var}_\Sigma(y) = 1 \}, $$

and a pair of vectors $(x_1, y_1)$ where the maximum is attained is called the first pair of canonical covariates.\(^2\) The second canonical correlation is the maximum value

\(^2\)Such a pair is not unique, so this terminology is somewhat misleading.
\[ c_2 = \max \{ \text{Cov}_\Sigma(x, y) \mid x \in \mathbb{R}^{I_1}, y \in \mathbb{R}^{I_2}, \forall \Sigma(x) = \forall \Sigma(y) = 1, \text{Cov}_\Sigma(x, x_1) = \text{Cov}_\Sigma(y, y_1) = 0 \}, \]

and a pair of vectors \((x_2, y_2)\) where the maximum is attained is called the second pair of canonical covariates. Further canonical correlations and covariates are defined similarly. The \(k\)th canonical correlation is the maximum value

\[ c_k = \max \{ \text{Cov}_\Sigma(x, y) \mid x \in \mathbb{R}^{I_1}, y \in \mathbb{R}^{I_2}, \forall \Sigma(x) = \forall \Sigma(y) = 1, \text{Cov}_\Sigma(x, x_{k'}) = \text{Cov}_\Sigma(y, y_{k'}) = 0, k' = 1, \ldots, k - 1 \}, \]

and a pair of vectors \((x_k, y_k)\) where the maximum is attained is called the \(k\)th pair of canonical covariates. This process is repeated until \(I_2\) canonical correlations and pairs of canonical covariates are obtained. The empirical canonical correlations and covariates are defined the same way by replacing the parameter \(\Sigma\) with the empirical covariance matrix

\[ S = N^{-1} \sum_{n=1}^{N} X_n X_n^t. \]  

Because the canonical covariates are not unique, it is more precise to call \((x_1, y_1), \ldots, (x_{I_2}, y_{I_2})\) a family of canonical covariate pairs, which is the terminology adopted in this paper. The following proposition characterizes the canonical correlations, in particular showing that they do not depend on the chosen family of canonical covariate pairs.

**Proposition 2.1.** The \(k\)th canonical correlation \(c_k\) is the \(k\)th largest root of

\[ p(c, \Sigma) = \begin{vmatrix} -c \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -c \Sigma_{22} \end{vmatrix} = 0, \]

and the canonical covariates satisfy

\[ \begin{pmatrix} -c_k \Sigma_{11} \\ \Sigma_{21} \end{pmatrix} \begin{pmatrix} \Sigma_{12} \\ -c_k \Sigma_{22} \end{pmatrix} \begin{pmatrix} x_k \\ y_k \end{pmatrix} = 0. \]

**Proof.** The first canonical correlation is the maximum value of \(x^t \Sigma_{12} y\) subject to the constraints \(x^t \Sigma_{11} x = 1\) and \(y^t \Sigma_{22} y = 1\). Applying Lagrange multipliers to this optimization problem shows that \(c_1\) is the largest root of the above polynomial, and \((x_1, y_1)\) satisfies the above linear equation. The proof is completed by inducting on \(I_2\), cf. Theorem 12.2.1 from Anderson (2003). \(\square\)

Before proceeding further, it will be helpful to reformulate this proposition in terms of eigenvalues, which requires the following definition.

**Definition 2.2.** Suppose \(R\) and \(T\) are \(I \times I\) matrices, \(\lambda \in \mathbb{R}\), and \(x\) is a nonzero vector in \(\mathbb{R}^I\) such that

\[ (R - \lambda T)x = 0. \]

Then \(\lambda\) is called an eigenvalue of \(R\) with respect to \(T\), and \(x\) is an eigenvector of \(R\) with respect to \(T\) corresponding to \(\lambda\).
Given sets of matrices $A$ and $B$, define
\[
A \oplus B := \left\{ \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \mid A \in \mathcal{A}, B \in \mathcal{B} \right\},
\]
and observe that the parameter sets for $H$ and $H_0$ are $\text{PD}(I)$ and $\text{PD}(I_1) \oplus \text{PD}(I_2)$, respectively. Let $t$ be the projection map
\[
t : \text{PD}(I) \rightarrow \text{PD}(I_1) \oplus \text{PD}(I_2)
\]
\[
\Sigma \mapsto \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}.
\]
Defining $r(\Sigma) := \Sigma - t(\Sigma)$, for $\Sigma \in \text{PD}(I)$, we see that the roots of the polynomial equation (2.3) are the eigenvalues of $r(\Sigma)$ wrt. $t(\Sigma)$. Furthermore, $(x_k, y_k)$ satisfies (2.4) if and only if $(x_k, y_k)$ is an eigenvector of $r(\Sigma)$ wrt. $t(\Sigma)$ corresponding to $c_k$, and $(x_k, -y_k)$ is an eigenvector of $r(\Sigma)$ wrt. $t(\Sigma)$ corresponding to $-c_k$. Making the identifications
\[
x_k \equiv \begin{pmatrix} x_k \\ 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} x_k \\ y_k \end{pmatrix} + \frac{1}{2} \begin{pmatrix} x_k \\ -y_k \end{pmatrix}, \quad \text{and}
\]
\[
y_k \equiv \begin{pmatrix} 0 \\ y_k \end{pmatrix} = \frac{1}{2} \begin{pmatrix} x_k \\ y_k \end{pmatrix} - \frac{1}{2} \begin{pmatrix} x_k \\ -y_k \end{pmatrix},
\]
we reformulate Proposition 2.1 as follows.

**Proposition 2.3.** The $k$th canonical correlation $c_k$ is the $k$th largest eigenvalue of $r(\Sigma)$ wrt. $t(\Sigma)$. The canonical covariates satisfy

1. $x_k = u_k + v_k$, and
2. $y_k = u_k - v_k$,

where $u_k$ and $v_k$ are eigenvectors of $r(\Sigma)$ wrt. $t(\Sigma)$ corresponding to $c_k$ and $-c_k$, respectively.

It will now be convenient to replace the sample $X_1, \ldots, X_N$ with the sufficient statistic $S$, whose distribution is $\mathcal{W}_{\Sigma, N}$, the Wishart distribution on $\text{PD}(I)$ with expectation $\Sigma$ and $N$ degrees of freedom. After this transformation, $\text{PD}(I)$ is the sample space and the parameter space for the statistical model $H$. If $A \in \text{GL}(I_1) \oplus \text{GL}(I_2)$, then $\text{ASA}^t \sim \mathcal{W}_{\Sigma \text{ASA}^t, N}$. In other words, $H$ is invariant under the following actions of $\text{GL}(I_1) \oplus \text{GL}(I_2)$ on its sample space and parameter space
\[
(A, S) \mapsto \text{ASA}^t
\]
\[
(A, \Sigma) \mapsto A \Sigma A^t.
\]
Furthermore, the second action restricts to a transitive action on $\text{PD}(I_1) \oplus \text{PD}(I_2)$, the parameter space for $H_0$, which means that the testing problem (2.1) is invariant under the above actions.

These group actions represent basis changes for the first $I_1$ components and last $I_2$ components of $\mathbb{R}^I$ and intuitively should be irrelevant to the testing problem. Therefore, it is natural to require statistical procedures for this testing problem to be invariant as well. For example, the likelihood ratio statistic $q$ is invariant, meaning that $q(\text{ASA}^t) = q(S)$, for all $A \in \text{GL}(I_1) \oplus \text{GL}(I_2)$ and $S \in \text{PD}(I)$. It is therefore desirable to find a maximal invariant statistic, an invariant statistic $\pi$ such that every invariant statistic is a function of $\pi$. To the extent that the basis
changes above are irrelevant, the maximal invariant statistic contains all relevant information about the sample data.

**Proposition 2.4.** The family of empirical canonical correlations is a maximal invariant statistic for the testing problem (2.1).

**Proof.** By Proposition 2.1, the empirical canonical correlations are the $I_2$ largest roots of $p(c,S) = 0$, which are invariant because

$$p(c,ASA^t) = |A|p(c,S)|A^t|,$$

for all $A \in \text{GL}(I_1) \oplus \text{GL}(I_2)$. Let $\Lambda = \text{Diag}(\hat{c}_1,\ldots,\hat{c}_{I_2})$ be the diagonal matrix whose entries are the empirical canonical correlations $\hat{c}_1,\ldots,\hat{c}_{I_2}$. It is possible to construct a matrix $A \in \text{GL}(I_1) \oplus \text{GL}(I_2)$ in terms of the empirical canonical covariates such that

$$ASA^t = \begin{pmatrix} 1_{I_2} & 0 & \Lambda \\ 0 & 1_{I_1-I_2} & 0 \\ \Lambda & 0 & 1_{I_2} \end{pmatrix}.$$  

Therefore, if $f$ is any invariant function, $f(S) = f(ASA^t)$, which depends only on the empirical canonical correlations, proving that they are a maximal invariant statistic, cf. Theorem 12.2.2. in Anderson (2003). □

**Remark 2.5.** For this testing problem, the null hypothesis is true if and only if the random variables $x^tX_n^{(1)}$ and $y^tX_n^{(2)}$ are uncorrelated for all $x \in \mathbb{R}^{I_1}$ and $y \in \mathbb{R}^{I_2}$. The empirical canonical correlations are maximum correlations wrt. $\hat{\Sigma}$ between such pairs, so one might think of them as measuring deviations in $\Sigma$ from the parameter set for the null hypothesis. Surprisingly, Proposition 2.4 shows that the empirical canonical correlations are a maximal invariant statistic. In other words, the maximal invariant statistic has an intuitively appealing statistical interpretation as a family of canonical correlations.

### 3. General Canonical Correlations

In this section, canonical correlations are generalized to a certain class of testing problems involving the covariance matrix of a multivariate normal distribution. Specifically, assume the following:

1. $X_1,\ldots,X_N \sim N(\Sigma)$ are i.i.d.;
2. the testing problem under consideration is $H_0 : \Sigma \in \Theta_0$ vs. $H : \Sigma \in \Theta$ where $\Theta_0 \subset \Theta \subseteq \text{PD}(I)$, and $I \geq 2$;
3. the maximum likelihood estimators $\hat{\Sigma}$ and $\hat{\Sigma}_0$ are defined a.s.;
4. there exists a mapping $t : \Theta \to \Theta_0$ such that $\hat{\Sigma}_0 = t(\hat{\Sigma})$.

Group symmetry testing problems satisfy these conditions, cf. Section 4. In particular, the test of independence discussed in Section 2 satisfies these conditions with $\Theta = \text{PD}(I)$, $\Theta_0 = \text{PD}(I_1) \oplus \text{PD}(I_2)$, and with the mapping $t$ from (2.5). Also, testing problems where $\Theta = \text{PD}(I)$ and $\Theta_0 \subseteq \text{PD}(I)$ is determined by the Markov properties of an acyclic mixed graph, in particular an acyclic directed graph or undirected graph, satisfy these conditions.

---

$1_{I}$ is the $I \times I$ identity matrix.
In the classical case, the first canonical correlation is the maximum possible correlation wrt. $\Sigma$ between random variables of the form $x^t X_n^{(1)}$ and $y^t X_n^{(2)}$, random variables that are always uncorrelated wrt. $\Sigma_0 = t(\Sigma)$. For testing problems meeting the above conditions, the first general canonical correlation is defined as the maximum possible correlation wrt. $\Sigma$ between random variables of the form $x^t X$ and $y^t X$ that are uncorrelated wrt. $\Sigma_0 = t(\Sigma)$. Additional general canonical correlations are defined recursively, as in the classical case, and the pairs $(x, y)$ where the maxima are attained are called general canonical covariate pairs. As before, define $\text{Cov}_x(x, y) = x^t \Sigma y$ and $\text{V}_x(x) = x^t \Sigma x$, for $x, y \in \mathbb{R}^I$, and let $\lfloor \frac{k}{2} \rfloor$ be the greatest integer less than or equal to $\frac{k}{2}$.

**Definition 3.1.** A family $((x_1, y_1), \ldots, (x_{\lfloor \frac{k}{2} \rfloor}, y_{\lfloor \frac{k}{2} \rfloor}))$ of pairs of vectors from $\mathbb{R}^I$ is called a family of (general) canonical covariate pairs of $\Sigma$ wrt. $\Sigma_0$ if

$$c_1 := \max \{\text{Cov}_x(x, y) \mid x, y \in \mathbb{R}^I, \text{V}_x(x) = \text{V}_x(y) = 1, \text{Cov}_{\Sigma_0}(x, y) = 0\}$$

is attained at $(x_1, y_1)$, and
$$c_k := \max \{\text{Cov}_x(x, y) \mid x, y \in \mathbb{R}^I, \text{V}_x(x) = \text{V}_x(y) = 1, \text{Cov}_{\Sigma_0}(x, y) = 0, \text{Cov}_{\Sigma_0}(x, x_{k'}) = \text{Cov}_{\Sigma_0}(y, x_{k'}) = \text{Cov}_{\Sigma_0}(y, y_{k'}) = 0, k' = 1, \ldots, k - 1\}$$

is attained at $(x_k, y_k)$, for $k = 2, \ldots, \lfloor \frac{k}{2} \rfloor$. The family $(c_1, \ldots, c_{\lfloor \frac{k}{2} \rfloor})$ is called the family of (general) canonical correlations. The empirical canonical correlations and covariate pairs are defined the same way by replacing $\Sigma$ and $\Sigma_0$ with $\Sigma$ and $\Sigma_0$, respectively.

**Remark 3.2.** For some testing problems, there is a $k$ such that $c_{k+1} = \cdots = c_{\lfloor \frac{k}{2} \rfloor} = 0$, for every $\Sigma \in \Theta$ and $\Sigma_0 = t(\Sigma)$, and the choice of $(x_{k+1}, y_{k+1}), \ldots, (x_{\lfloor \frac{k}{2} \rfloor}, y_{\lfloor \frac{k}{2} \rfloor})$ is therefore uninteresting. In these cases, the families of canonical correlations and covariate pairs are shortened to $(c_1, \ldots, c_k)$ and $((x_1, y_1), \ldots, (x_k, y_k))$, respectively.

Note that families of canonical correlations and covariate pairs exist by the extreme value theorem, and $1 > c_1 \geq \cdots \geq c_{\lfloor \frac{k}{2} \rfloor} \geq 0$. Theorem 3.4 characterizes the general canonical correlations and covariate pairs in terms of the eigenvalues and eigenvectors of $\Sigma - \Sigma_0$ wrt. $\Sigma_0$, generalizing Proposition 2.3. The proof requires the following well known diagonalization theorem.

**Theorem 3.3.** If $R$ is a symmetric $I \times I$ matrix, and $T \in \text{PD}(I)$, then there exists $M \in \text{GL}(I)$ such that $M^t RM$ is a diagonal matrix, and $M^t TM = 1_I$.

**Proof.** Let $M_1$ be a matrix whose columns form an orthonormal basis wrt. $T$, so that $M_1^t TM_1 = 1_I$. Because $M_1^t RM_1$ is symmetric, there exists an orthogonal matrix $M_2$ such that $M_2^t M_1^t RM_1 M_2$ is diagonal. Then $M = M_1 M_2$ satisfies the conditions in the theorem. \hfill $\Box$

**Theorem 3.4.** Let $\lambda_1 \geq \cdots \geq \lambda_I$ be the eigenvalues of $\Sigma - \Sigma_0$ wrt. $\Sigma_0$. A family $((x_1, y_1), \ldots, (x_{\lfloor \frac{k}{2} \rfloor}, y_{\lfloor \frac{k}{2} \rfloor}))$ is a family of canonical covariate pairs of $\Sigma$ wrt. $\Sigma_0$ if and only if there exist $u_k, v_k \in \mathbb{R}^I$, for $k = 1, \ldots, \lfloor \frac{k}{2} \rfloor$, such that

1. $u_k$ and $v_k$ are eigenvectors of $\Sigma - \Sigma_0$ wrt. $\Sigma_0$ corresponding to $\lambda_k$ and $\lambda_{I+1-k}$, respectively;
2. $u_1, \ldots, u_{\lfloor \frac{k}{2} \rfloor}, v_1, \ldots, v_{\lfloor \frac{k}{2} \rfloor}$ are orthogonal wrt. $\Sigma_0$.
(3) $u_k^t \Sigma_0 u_k = v_k^t \Sigma_0 v_k = \frac{1}{\lambda_k + \lambda_{t+1-k} + 2}$;

(4) $x_k = u_k + v_k$ and $y_k = u_k - v_k$.

The $k$th canonical correlation is

$$c_k = \mu(\lambda_k, \lambda_{t+1-k}) := \frac{\lambda_k - \lambda_{t+1-k}}{\lambda_k + \lambda_{t+1-k} + 2}.$$ 

**Proof.** We begin the proof with a change of basis. For any matrix $M \in \text{GL}(I)$, $((x_k, y_k))$ is a family of canonical covariates of $\Sigma$ wrt. $\Sigma_0$ if and only if $((M^{-1}x_k, M^{-1}y_k))$ is a family of canonical covariates of $M^t \Sigma M$ wrt. $M^t \Sigma_0 M$. Also, $u_k$ and $v_k$ satisfy the conditions in the theorem if and only if $M^{-1}u_k$ and $M^{-1}v_k$ satisfy these conditions with $((x_k, y_k))$, $\Sigma$, and $\Sigma_0$ replaced by $((M^{-1}x_k, M^{-1}y_k))$, $M^t \Sigma M$, and $M^t \Sigma_0 M$, respectively. Therefore, by Theorem 3.3, we may assume without loss of generality that $\Sigma = 1_I + \Lambda$, and $\Sigma_0 = 1_I$, where $\Lambda = \text{Diag}(\lambda_1, \ldots, \lambda_I)$. For each eigenvalue $\lambda$ of $\Lambda$, let $V_\lambda \subseteq \mathbb{R}^I$ be the corresponding eigenspace. Because the eigenvalues of $\Lambda$ are not necessarily distinct, these eigenspaces may have dimension greater than one, and it is possible that $V_\lambda = V_{\lambda'}$ for distinct $i$ and $j$.

The conditions in Definition 3.1 are now

(3.1) $c_1 = \max\{x^t \Lambda y \mid x, y \in \mathbb{R}^I, x^t (1_I + \Lambda) x = y^t (1_I + \Lambda) y = 1, x^t y = 0\}$

is attained at $(x_1, y_1)$, and

(3.2) $c_k = \max\{x^t \Lambda y \mid x, y \in \mathbb{R}^I, x^t (1_I + \Lambda) x = y^t (1_I + \Lambda) y = 1, x^t y = 0, x^t x_{k'} = y^t y_{k'} = 0, k' = 1, \ldots, k \}$

is attained at $(x_k, y_k)$, and the conditions on $u_k$ and $v_k$ are

(1) $u_k \in V_{\lambda_k}$ and $v_k \in V_{\lambda_{t+1-k}}$,

(2) $u_1, \ldots, u_{\lfloor \frac{I}{2} \rfloor}, v_1, \ldots, e_{\lfloor \frac{I}{2} \rfloor}$ are orthogonal,

(3) $|u_k|^2 = |v_k|^2 = \frac{1}{\lambda_k + \lambda_{t+1-k} + 2}$, and

(4) $x_k = u_k + v_k$ and $y_k = u_k - v_k$.

To prove the rightward implication of the theorem, we assume that $x_k$ and $y_k$ satisfy the first set of conditions, and we must show the existence of $u_k$ and $v_k$ satisfying the second set of conditions. Since $x_1$ and $y_1$ satisfy (3.1), the desired $u_1$ and $v_1$ exist by Lemma 3.5, and $c_1 = \mu(\lambda_1, \lambda_I)$. Letting $e_1, \ldots, e_I$ denote the standard basis for $\mathbb{R}^I$, we may assume by a second change of basis that $u_1 \in V_{\lambda_1}$ is a scalar multiple of $e_1$, and $v_1 \in V_{\lambda_I}$ is a scalar multiple of $e_I$.

Therefore, $x_1, y_1 \in \text{Span}\{e_1, e_I\}$ and $((x_2, y_2), \ldots, (x_{\lfloor \frac{I}{2} \rfloor}, y_{\lfloor \frac{I}{2} \rfloor}))$ is a family of canonical covariate pairs of $1_I + \tilde{\Lambda}$ wrt. $1_I - 2$, where $\tilde{\Lambda} = \text{Diag}(\lambda_2, \ldots, \lambda_{I-1})$. The existence of the desired $u_k$ and $v_k$, $k = 2, \ldots, \lfloor \frac{I}{2} \rfloor$, now follows by induction, proving the rightward implication of the theorem. The leftward implication is proved similarly.

□

**Lemma 3.5.** Suppose $\Lambda = \text{Diag}(\lambda_1, \ldots, \lambda_I)$, where $\lambda_1 \geq \cdots \geq \lambda_I > -1$, and for each eigenvalue $\lambda$ of $\Lambda$, let $V_\lambda \subseteq \mathbb{R}^I$ be the corresponding eigenspace. The maximum value
\[ c_1 := \max \{ x^t A y \mid x, y \in \mathbb{R}^I, x^t (1_I + \Lambda)x = y^t (1_I + \Lambda)y = 1, x^t y = 0 \} \]

is attained at \((x_1, y_1)\) if and only if there exist vectors \(u_1\) and \(v_1\) such that

1. \(u_1 \in V_{\lambda_1}\) and \(v_1 \in V_{\lambda_I}\),
2. \(u_1\) and \(v_1\) are orthogonal,
3. \(|u_1|^2 = |v_1|^2 = \frac{1}{\lambda_1 - \lambda_I + 2}\), and
4. \(x_1 = u_1 + v_1\) and \(y_1 = u_1 - v_1\).

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, so we assume that \(\lambda_1 \neq \lambda_I\). We will first show that \(c_1 = \mu(\lambda_1, \lambda_I)\). If \(u_1\) and \(v_1\) are vectors satisfying the above conditions, then \(u_1^t v_1 = u_1^t \Lambda v_1 = 0\) by condition (1), which implies \(x_1^t (1_I + \Lambda)x_1 = y_1^t (1_I + \Lambda)y_1 = 1\), \(x_1^t y_1 = 0\), and \(x_1^t \Lambda y_1 = \mu(\lambda_1, \lambda_I)\). This proves \(c_1 \geq \mu(\lambda_1, \lambda_I) > 0\), and it remains to show that \(c_1 \leq \mu(\lambda_1, \lambda_I)\).

We will now apply Lagrange multipliers to the maximization problem. For \(x, y \in \mathbb{R}^I\), define

\[ F(x, y) = x^t A y + l_1[x^t (1_I + \Lambda)x - 1] + l_2[y^t (1_I + \Lambda)y - 1] + l_0 x^t y, \]

and assume the maximum is attained at \((x_1, y_1)\), where \(x_1 = (x_{11}, \ldots, x_{1I})^t\), and \(y_1 = (y_{11}, \ldots, y_{1I})^t\). Then

\[
\begin{align*}
\frac{\partial F}{\partial x}(x_1, y_1) &= \Lambda y_1 + 2l_1 (1_I + \Lambda) x_1 + l_0 y_1 = 0, \\
\frac{\partial F}{\partial y}(x_1, y_1) &= \Lambda x_1 + 2l_2 (1_I + \Lambda) y_1 + l_0 x_1 = 0.
\end{align*}
\]

Multiplying (3.3) by \(x_1^t\) and (3.4) by \(y_1^t\) yields \(l_1 = l_2 = -\frac{\lambda_1 + \lambda_I}{2}\).

Let \(i_1\) and \(i_2\) be indices such that \(x_{1i_1} \neq 0\), \(x_{1i_2} \neq 0\), and \(\lambda_{i_1} > \lambda_{i_2}\). If no such indices exist, then there is an eigenvalue \(\lambda\) such that \(x_{1i} \neq 0\) implies \(\lambda_{i} = \lambda\). It follows that \(c_1 = x_{1i}^t \Lambda y_1 = \lambda x_{1i}^t y_1 = 0\), a contradiction, so the desired indices exist. By considering the \(i_1\)th and \(i_2\)th components of equations (3.3) and (3.4), one obtains two systems of linear equations with nonzero solutions. Because the corresponding determinants are zero,

\[
\begin{align*}
c_1^2(1 + \lambda_{i_1})^2 - (\lambda_{i_1} + l_0)^2 &= 0, \\
(c_1^2(1 + \lambda_{i_2})^2 - (\lambda_{i_2} + l_0)^2 &= 0.
\end{align*}
\]

Noting that \(c_1 < 1\) by the Cauchy-Schwartz inequality, these equations yield

\[ c_1 = \mu(\lambda_{i_1}, \lambda_{i_2}) \leq \mu(\lambda_1, \lambda_I), \]

with equality if and only if \(\lambda_{i_1} = \lambda_1\) and \(\lambda_{i_2} = \lambda_I\). Thus, \(c_1 = \mu(\lambda_1, \lambda_I)\), as claimed.

We have also established that \(x_{1i} = 0\) if \(\lambda_{i} \notin \{\lambda_1, \lambda_I\}\), and because of symmetry, the same must be true for \(y_1\). That is, \(x_1, y_1 \in V_{\lambda_1} \oplus V_{\lambda_I}\), and substituting the values of \(c_1\) and \(l_0\) obtained from (3.5) into (3.3) and (3.4) yields \(x_1 = u_1 + v_1\) and \(y_1 = u_1 - v_1\), where \(u_1 \in V_{\lambda_1}\) and \(v_1 \in V_{\lambda_I}\).

Because \(u_1\) and \(v_1\) are eigenvectors of \(\Lambda\) corresponding to distinct eigenvalues, they are orthogonal. The condition \(x_{1i}^t y_1 = 0\) implies that \(|u_1| = |v_1|\), and the condition \(1 = x_{1i}^t (1_I + \Lambda) x_1 = (1 + \lambda_{i}) |u_1|^2 + (1 + \lambda_I) |v_1|^2\) then implies that
This proves the rightward implication of the lemma, and the leftward implication is easily established. □

4. Canonical Correlations for Group Symmetry Models

The theory of group symmetry models can be found in Appendix A of Andersson and Madsen (1998). This section provides a brief overview of this theory, followed by an investigation of general canonical correlations and covariates for the ten fundamental testing problems treated in (ABJ).

Given a finite group $G$ of orthogonal $I \times I$ matrices, define

1. $M_G(I) = \{ M \in \mathbb{R}^{I\times I} \mid gMg^t = M, \text{ for every } g \in G \}$,
2. $\mathcal{PD}_G(I) = \mathcal{PD}(I) \cap M_G(I)$, and
3. $\mathcal{GL}_G(I) = \mathcal{GL}(I) \cap M_G(I)$.

If $\Sigma \in \mathcal{PD}(I)$ and $X_1, \ldots, X_N \sim N(\Sigma)$ are i.i.d. random vectors, the group $G$ determines the group symmetry model $H : \Sigma \in \mathcal{PD}_G(I)$. For $N$ greater than or equal to a certain integer constant depending on $G$, the maximum likelihood estimator exists with probability 1 and is equal to $\hat{\Sigma} = \Psi_G(S)$, where $S$ is the empirical covariance matrix, and $\Psi_G$ is the smoothing map

$\Psi_G : \mathcal{PD}(I) \rightarrow \mathcal{PD}_G(I)
S \mapsto \frac{1}{|G|} \sum_{g \in G} gSg^t$.

If $G_0$ is another finite group of orthogonal $I \times I$ matrices, and $G$ is a subgroup of $G_0$, then $\mathcal{PD}_{G_0}(I) \subseteq \mathcal{PD}_G(I)$, which induces the group symmetry testing problem

$H_0 : \Sigma \in \mathcal{PD}_{G_0}(I)$ vs. $H : \Sigma \in \mathcal{PD}_G(I)$.

Because $\Psi_{G_0} = \Psi_{G_0} \circ \Psi_G$, the maximum likelihood estimator under $H_0$ is $\hat{\Sigma}_0 = \Psi_{G_0}(\hat{\Sigma})$. As in Section 2, we transform the above testing problem by the sufficient statistic $\hat{\Sigma}$, and afterwards, $\mathcal{PD}_G(I)$ is the sample space and parameter space for $H$. The group $\mathcal{GL}_G(I)$ acts on these spaces as follows

$\mathcal{GL}_G(I) \times \mathcal{PD}_G(I) \rightarrow \mathcal{PD}_G(I)
(A, \Sigma) \mapsto A\Sigma A^t$,

and the testing problem is invariant under these actions.

Any group symmetry testing problem can be decomposed into a sequence of testing problems, where each problem in the sequence is one of the ten problems discussed in (ABJ). Therefore, these testing problems are fundamental to the theory of group symmetry models, and they deserve special attention.

Note that group symmetry testing problems meet the general conditions given in Section 3, where $\Theta = \mathcal{PD}_G(I)$, $\Theta_0 = \mathcal{PD}_{G_0}(I)$, and $t$ is the restriction of $\Psi_{G_0}$ to $\mathcal{PD}_G(I)$. For the ten fundamental testing problems, the ordered eigenvalues $\lambda_1 \geq \cdots \geq \lambda_I$ of $r(\Sigma) = \Sigma - t(\Sigma)$ wrt. $t(\Sigma)$ are contained in $(-1, 1)$, and they are symmetric about zero, i.e., $\lambda_i = -\lambda_{I+1-i}$. The following corollary of Theorem 3.4 characterizes the canonical correlations and covariates for the ten fundamental testing problems.
Corollary 4.1. Consider one of the ten fundamental testing problems, and let \( \lambda_1 \geq \cdots \geq \lambda_{\frac{G}{2}} \) be the largest \( \frac{G}{2} \) eigenvalues of \( r(\Sigma) = \Sigma - \Psi_{G_0}(\Sigma) \) wrt. \( t(\Sigma) = \Psi_{G_0}(\Sigma) \). Then \( ((x_1, y_1), \ldots, (x_{\left\lfloor \frac{G}{2} \right\rfloor}, y_{\left\lfloor \frac{G}{2} \right\rfloor})) \) is a family of canonical covariate pairs of \( \Sigma \) wrt. \( t(\Sigma) \) if and only if there exist \( u_k, v_k \in \mathbb{R}^G \), for \( k = 1, \ldots, \left\lfloor \frac{G}{2} \right\rfloor \), such that

1. \( u_k \) and \( v_k \) are eigenvectors of \( r(\Sigma) \) wrt. \( t(\Sigma) \) corresponding to \( \lambda_k \) and \( -\lambda_k \), respectively;
2. \( u_1, \ldots, u_{\left\lfloor \frac{G}{2} \right\rfloor}, v_1, \ldots, v_{\left\lfloor \frac{G}{2} \right\rfloor} \) are orthogonal wrt. \( t(\Sigma) \);
3. \( u_k^t t(\Sigma) u_k = v_k^t t(\Sigma) v_k = \frac{1}{2} \);
4. \( x_k = u_k + v_k \) and \( y_k = u_k - v_k \).

The family of canonical correlations is \( (\lambda_1, \ldots, \lambda_{\left\lfloor \frac{G}{2} \right\rfloor}) \).

Proposition 2.4 shows that the classical empirical canonical correlations are a maximal invariant statistic for the test of independence (2.1). The following corollary extends this result.

Corollary 4.2. For each of the fundamental testing problems, the family of empirical canonical correlations is \( (\hat{\lambda}_1, \ldots, \hat{\lambda}_{\left\lfloor \frac{G}{2} \right\rfloor}) \), where \( \hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_{\frac{G}{2}} \) are the ordered eigenvalues of \( R = \hat{\Sigma} - \Psi_{G_0}(\hat{\Sigma}) \) wrt. \( T = \Psi_{G_0}(\hat{\Sigma}) \). The family of empirical canonical correlations is a maximal invariant for the seven testing problems covered in Sections 2 through 6 of (ABJ), and it is not a maximal invariant for the three testing problems covered in Section 7.

Proof. The family of empirical canonical correlations is \( (\hat{\lambda}_1, \ldots, \hat{\lambda}_{\left\lfloor \frac{G}{2} \right\rfloor}) \) by Corollary 4.1. For each of the fundamental testing problems, the representation of the maximal invariant statistic given in (ABJ) has the form

\[
\pi : PD_G(I) \rightarrow \Lambda_p := \{ (\gamma_1, \ldots, \gamma_p) \mid 1 > \gamma_1 \geq \cdots \geq \gamma_p \geq 0 \},
\]

where \( \pi(\hat{\Sigma}) \) is some family of eigenvalues determined by \( \hat{\Sigma} \). It follows immediately from these representations that \( (\hat{\lambda}_1, \ldots, \hat{\lambda}_{\left\lfloor \frac{G}{2} \right\rfloor}) \) is also a maximal invariant for the testing problems covered in Sections 2 through 6 of (ABJ), cf. Sections 4.1 through 4.4 below, and that \( (\hat{\lambda}_1, \ldots, \hat{\lambda}_{\left\lfloor \frac{G}{2} \right\rfloor}) \) is not a maximal invariant for the other three testing problems, cf. Section 4.5. \( \square \)

For each of the fundamental testing problems, the canonical correlations and covariates have a specific form that is analogous to classical canonical correlations, providing a statistical interpretation to the maximal invariant statistic for the seven testing problems mentioned in Corollary 4.2. In the following examples, canonical correlations for five of the fundamental testing problems are explored in more detail.

4.1. Testing Independence of Two Sets of Variates, cf. Section 6 of (ABJ). Defining \( G = \{1_f\} \) and \( G_0 = \{ \pm 1_f, \pm f \} \), where \( f = \text{Diag}(1_{1_f}, -1_{1_f}) \), yields \( PD_G(I) = PD(I) \), and \( PD_{G_0}(I) = PD(I_1) \oplus PD(I_2) \). Therefore, the corresponding group symmetry testing problem is the classical test of independence (2.1) examined in Section 2.

In particular, the smoothing map \( \Psi_{G_0} : PD_G(I) \rightarrow PD_{G_0}(I) \) is the mapping (2.5), and the maximum likelihood estimators are \( \hat{\Sigma} = \Psi_G(S) = S \), and \( \hat{\Sigma}_0 = \Psi_{G_0}(S) = \text{Diag}(S_{11}, S_{22}) \). For this testing problem, the ordered eigenvalues of \( r(\Sigma) \) wrt. \( t(\Sigma) \) have the form

\[ \Rightarrow \]
Testing that a Covariance Matrix with Real Structure Has Complex Structure, cf. Section 2 of ([ABJ]). Suppose that \( I = 2J \), and let \( X_n = (X_n^{(1)} , X_n^{(2)})^t \) be the decomposition of \( X_n \) into its first \( J \) components and last \( J \) components. Consider testing the null hypothesis that the random vector \( X_n^{(1)} + iX_n^{(2)} \) taking values in \( \mathbb{C}^J \) has a complex multivariate normal distribution, which is equivalent to testing

\[
\mathrm{H}_0 : \Sigma = \begin{pmatrix} \Gamma & -F \\ F & \Gamma \end{pmatrix} \text{ vs. } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},
\]

(cf. Wooding (1956). The parameter space for \( \mathrm{H}_0 \) is \( \mathbf{PD}(I) \), the set of all \( I \times I \) positive definite matrices with complex structure.\(^5\)

\(^4\)If \( I_1 > I_2 \), zero is an eigenvalue with multiplicity at least \( I_1 - I_2 \).

\(^5\)A matrix \( M \in \mathbb{R}^{I \times I} \) has complex structure if \( M = (A - B, A) \), for matrices \( A, B \in \mathbb{R}^{J \times J} \).
This is the group symmetry testing problem obtained by setting $G = \{1_J\}$ and $G_0 = \{±1_J, ±1\}$, where $i = (\begin{smallmatrix} 0 & -1_J \\ 1_J & 0 \end{smallmatrix})$, and the maximum likelihood estimator under $H_0$ is

$$\hat{\Sigma}_0 = \Psi_{G_0}(S) = \frac{1}{2} \begin{pmatrix} S_{11} + S_{22} & S_{12} - S_{21} \\ S_{21} - S_{12} & S_{11} + S_{22} \end{pmatrix}.$$

The ordered eigenvalues of $r(\Sigma)$ wrt. $t(\Sigma)$ have the form $\lambda_1, \ldots, \lambda_J, -\lambda_J, \ldots, -\lambda_1$, where $1 > \lambda_1 \geq \cdots \geq \lambda_J \geq 0$, and we define

$$\pi : \mathbf{PD}(I) \to \Lambda_J$$

$$\pi(\Sigma) = (\lambda_1, \ldots, \lambda_J).$$

The family of canonical correlations is $\pi(\Sigma)$, and the family of empirical canonical correlations, $\pi(\hat{\Sigma})$, is a maximal invariant statistic.

By Lemma 1 in (ABJ), there exists a basis $u_1, \ldots, u_J, iu_1, \ldots, iu_J$ of $\mathbb{R}^I$ that is orthonormal wrt. $t(\Sigma)$, such that $u_k$ is an eigenvector of $r(\Sigma)$ wrt. $t(\Sigma)$ corresponding to $\lambda_k$, $k = 1, \ldots, J$.

Because $r(\Sigma)i = -ir(\Sigma)$, $-iu_k$ is an eigenvector corresponding to $-\lambda_k$. By Corollary 4.1, $x_k = \frac{1}{\sqrt{2}}(u_k - iu_k)$, $y_k = \frac{1}{\sqrt{2}}(u_k + iu_k)$, $k = 1, \ldots, J$, defines a family of canonical covariate pairs, and $y_k = \sqrt{2}x_k$. Proposition 4.3 follows from these observations.

For $x, y \in \mathbb{R}^I \setminus \{0\}$, define $\text{Corr}_\Sigma(x, y)$ to be the correlation between $x^tX$ and $y^tX$, where $X \sim N(\Sigma)$.

**Proposition 4.3.** For the testing problem (4.3), there exists a family of canonical covariate pairs of the form $((x_1, i\varepsilon_1), \ldots, (x_J, i\varepsilon_J))$. Furthermore, the canonical correlations satisfy

$$c_k = \text{Corr}_\Sigma(x_k, i\varepsilon_k)$$

$$= \max \{ \text{Corr}_\Sigma(x, i\varepsilon) \mid x \in \mathbb{R}^I \setminus \{0\}, \text{Cov}_{t(\Sigma)}(x, x_{k'}) = \text{Cov}_{t(\Sigma)}(x, i\varepsilon_{k'}) = 0, k' = 1, \ldots, k - 1 \},$$

$k = 1, \ldots, J$.

**Proof.** The existence of a family of canonical covariate pairs $((x_1, i\varepsilon_1), \ldots, (x_J, i\varepsilon_J))$ has been shown, and by Definition 3.1, $c_k = \text{Corr}_\Sigma(x_k, i\varepsilon_k)$, proving that the above maximum is at least $c_k$.

Let $x \in \mathbb{R}^I \setminus \{0\}$ satisfy the optimization constraints from the proposition, and let $K_1, K_2 > 0$ such that $\forall_{\Sigma}(K_1 x) = \forall_{\Sigma}(K_2 i\varepsilon) = 1$. Because $i$ commutes with $t(\Sigma)$, $(K_1 x, K_2 i\varepsilon)$ satisfies the optimization constraints from Definition 3.1, so

$$\text{Corr}_\Sigma(x, i\varepsilon) = \text{Cov}_\Sigma(K_1 x, K_2 i\varepsilon) \leq c_k,$$

showing that the maximum value from the proposition is at most $c_k$. Propositions 4.4 through 4.6 are proved similarly.

---

6If $\lambda_1 > \cdots > \lambda_J > 0$, such a basis is obtained by letting $u_k$ be an arbitrary $t(\Sigma)$-normalized eigenvector corresponding to $\lambda_k$, $k = 1, \ldots, J$, and similar remarks apply to the examples given in Sections 4.3 through 4.5.
Proposition 4.3 effectively provides an alternative definition of the canonical correlations for testing problem (4.3), showing that they are obtained by recursively maximizing the correlation between pairs of the form \((x, ix)\). Furthermore, \(\Sigma \in PD_G(I)\) if and only if all such pairs are uncorrelated wrt \(\Sigma\). This is analogous to the classical case, where the canonical correlations are obtained by recursively maximizing the correlation between pairs of the form (4.2), and \(\Sigma \in PD(I_1) \oplus PD(I_2)\) if and only if such pairs are uncorrelated wrt \(\Sigma\). In other words, the empirical canonical correlations for testing problem (4.3) provide a statistical interpretation of the maximally invariant eigenvalues, as in the classical case.

4.3. Testing that a Covariance Matrix with Complex Structure has Real Structure, cf. Section 3 of (ABJ). In this testing problem, we assume that \(I = 2J\), where \(J \geq 2\), and that the covariance matrix of \(X_n = (X_n^{(1)}, X_n^{(2)})^t\) has complex structure, as described in Section 4.2. Under these assumptions, it is possible that \(X_n^{(1)}\) and \(X_n^{(2)}\) are independent random vectors with the same covariance matrix, which suggests the following testing problem from Khatri (1965)

\[(4.4) \quad H_0 : \Sigma = \begin{pmatrix} \Gamma & 0 \\ 0 & \Gamma \end{pmatrix} \quad \text{vs.} \quad H : \Sigma = \begin{pmatrix} \Gamma & -F \\ F & \Gamma \end{pmatrix}.\]

This is the group symmetry testing problem obtained by setting \(G = \{\pm 1, i\}\) and \(G_0 = \{\pm 1, \pm i, \pm s\}\), where \(i = (0, 1, 0)^t\) and \(s = (0, -1, 0)^t\). The maximum likelihood estimator under the null hypothesis is

\[\hat{\Sigma}_0 = \Psi_{G_0} (S) = \frac{1}{2} \begin{pmatrix} S_{11} + S_{22} & 0 \\ 0 & S_{11} + S_{22} \end{pmatrix}.\]

For this testing problem, the eigenvalues of \(r(\Sigma)\) wrt \(t(\Sigma)\) all have even multiplicity, and we define

\[\pi : PD_G(I) \to \Lambda_{2\lfloor \frac{J}{2} \rfloor} \quad \pi(\Sigma) = (\lambda_1, \lambda_1, \ldots, \lambda_{\lfloor \frac{J}{2} \rfloor}, \lambda_{\lfloor \frac{J}{2} \rfloor}),\]

where \(\pi(\Sigma)\) is the family of the \(2\lfloor \frac{J}{2} \rfloor\) largest eigenvalues, and \(1 > \lambda_1 \geq \cdots \geq \lambda_{\lfloor \frac{J}{2} \rfloor} \geq 0\). The family of canonical correlations is \(\pi(\Sigma)\), and the family of empirical canonical correlations is \(\pi(\hat{\Sigma})\), which is a maximal invariant statistic.

By Lemma 7 in (ABJ), there exist vectors \(u_k, iu_k, su_k, siu_k \in \mathbb{R}^J, k = 1, \ldots, \lfloor \frac{J}{2} \rfloor\), that are orthonormal wrt. \(t(\Sigma)\), such that \(u_k\) is an eigenvector of \(r(\Sigma)\) wrt \(t(\Sigma)\) corresponding to \(\lambda_k\). Because \(i\) commutes with \(r(\Sigma)\), \(iu_k\) is also an eigenvector corresponding to \(\lambda_k\), and because \(r(\Sigma)s = -sr(\Sigma)\), \(su_k\) and \(siu_k\) are eigenvectors corresponding to \(-\lambda_k\). Defining \(x_k' = \frac{1}{\sqrt{2}}(u_k + su_k), y_k' = \frac{1}{\sqrt{2}}(u_k - su_k), \tilde{x}_k = \frac{1}{\sqrt{2}}(iu_k + siu_k), \tilde{y}_k = \frac{1}{\sqrt{2}}(iu_k - siu_k)\), for \(k = 1, \ldots, \lfloor \frac{J}{2} \rfloor\), a family of canonical covariate pairs is given by

\[\{(x_1, y_1), \ldots, (x_{2\lfloor \frac{J}{2} \rfloor}, y_{2\lfloor \frac{J}{2} \rfloor})\} = \{(x_1', y_1'), (\tilde{x}_1, \tilde{y}_1), \ldots, (x_{\lfloor \frac{J}{2} \rfloor}', y_{\lfloor \frac{J}{2} \rfloor}'), (\tilde{x}_{\lfloor \frac{J}{2} \rfloor}, \tilde{y}_{\lfloor \frac{J}{2} \rfloor})\}.\]

Because \(sx_k = x_k\) and \(sy_k = -y_k\), these pairs have the form

\[7\{\pm 1, i, \pm s\}\]
(4.5) \[ x_k = \left( \frac{\alpha_k}{\alpha_k} \right) \text{ and } y_k = \left( \frac{\beta_k}{-\beta_k} \right), \alpha_k, \beta_k \in \mathbb{R}^J, \]

which leads to the proposition below.

**Proposition 4.4.** For the testing problem (4.4), there exists a family of canonical covariate pairs \(((x_1, y_1), \ldots, (x_{2\lfloor J/2 \rfloor}, y_{2\lfloor J/2 \rfloor}))\) of the form (4.5), and the canonical correlations satisfy

\[
\begin{align*}
        c_k &= \text{Corr}_\Sigma(x_k, y_k) \\
        &= \max \left\{ \text{Corr}_\Sigma \left( \begin{pmatrix} \alpha \\ \alpha \end{pmatrix}, \begin{pmatrix} \beta \\ -\beta \end{pmatrix} \right) \middle| \alpha, \beta \in \mathbb{R}^J \setminus \{0\} \right\}, \\
\text{Cov}_{\Sigma_{11} + \Sigma_{22}}(\alpha, \alpha_k') = \text{Cov}_{\Sigma_{11} + \Sigma_{22}}(\beta, \beta_k') &= 0, \quad k' = 1, \ldots, k - 1, \\
\end{align*}
\]

for \(k = 1, \ldots, 2\lfloor J/2 \rfloor\).

For this testing problem, the null hypothesis is equivalent to all pairs of the form (4.5) being uncorrelated wrt. \(\Sigma\). By Proposition 4.4, the empirical canonical correlations are obtained by recursively maximizing the correlation between such pairs wrt. \(\hat{\Sigma}\), thus providing a statistical interpretation to the maximal invariant statistic.

### 4.4 Testing that a Covariance Matrix with Complex Structure Has Quaternion Structure, cf. Section 4 of (ABJ).

Assume \(I = 4J\), and define \(G_0 = \{\pm 1_I, \pm i, \pm j, \pm k\}\) and \(G = \{\pm 1_I, \pm j\}\), where

\[
i = \begin{pmatrix} 0 & 1_J & 0 & 0 \\
-1_J & 0 & 0 & 0 \\
0 & 0 & 0 & -1_J \\
0 & 0 & 1_J & 0 \end{pmatrix}, \quad j = \begin{pmatrix} 0 & 0 & 1_J & 0 \\
0 & 0 & 0 & 1_J \\
-1_J & 0 & 0 & 0 \\
0 & -1_J & 0 & 0 \end{pmatrix},
\]

and \(k = ij\). The group \(G_0\) is isomorphic to the quaternion group, and \(\text{PD}_{G_0}(I) = \text{PD}_H(I)\) is the set of all \(I \times I\) positive definite matrices with quaternion structure, i.e., having the form

\[
\Sigma = \begin{pmatrix} 
\Gamma & -E & -F & -G \\
E & \Gamma & -G & F \\
F & G & \Gamma & -E \\
G & -F & E & \Gamma 
\end{pmatrix}.
\]

Also, \(\text{PD}_{G}(I) = \text{PD}_C(I)\), and the testing problem

\[ (4.7) \quad H_0 : \Sigma \in \text{PD}_H(I) \text{ vs. } H : \Sigma \in \text{PD}_C(I) \]

is one of the ten fundamental group symmetry testing problems. For this testing problem, the ordered eigenvalues of \(r(\Sigma)\) wrt. \(t(\Sigma)\) have the form

\[
\lambda_1, \lambda_1, \ldots, \lambda_J, \lambda_J, -\lambda_J, -\lambda_J, \ldots, -\lambda_1, -\lambda_1,
\]

where \(1 > \lambda_1 \geq \cdots \geq \lambda_J \geq 0\), and we define
$$\pi : \text{PD}_C(I) \rightarrow \Lambda_{2J}$$

$$\pi(\Sigma) = (\lambda_1, \lambda_1, \ldots, \lambda_J, \lambda_J).$$

The family of canonical correlations is $\pi(\Sigma)$, and the family of empirical canonical correlations is $\pi(\hat{\Sigma})$, a maximal invariant statistic.

By Lemma 8 in (ABJ), there exists a basis $u_1, \ldots, u_J, iu_1, \ldots, iu_J, ju_1, \ldots, ju_J, ku_1, \ldots, ku_J$, of $\mathbb{R}^J$ that is orthonormal wrt. $t(\Sigma)$, such that $u_k$ is an eigenvector of $r(\Sigma)$ wrt. $t(\Sigma)$ corresponding to $\lambda_k$, $k = 1, \ldots, J$. Because $r(\Sigma) = -g r(\Sigma)$, for $g \in \{i, k\}$, $iu_k$ and $ku_k$ are eigenvectors corresponding to $-\lambda_k$, and because $j$ commutes with $r(\Sigma)$, $ju_k$ is an eigenvector corresponding to $\lambda_k$, $k = 1, \ldots, J$. Defining $x'_k = \frac{1}{\sqrt{2}}(u_k - iu_k)$, $y'_k = \frac{1}{\sqrt{2}}(u_k + iu_k)$, $\tilde{x}_k = \frac{1}{\sqrt{2}}(ju_k - ku_k)$, and $\tilde{y}_k = \frac{1}{\sqrt{2}}(ju_k + ku_k)$, a family of empirical canonical covariate pairs is

$$((x_1, y_1), \ldots, (x_{2J}, y_{2J})) := ((x'_1, y'_1), (\tilde{x}_1, \tilde{y}_1), \ldots, (x'_J, y'_J), (\tilde{x}_J, \tilde{y}_J)).$$

Noting that $y_k = \text{ix}_k$, $k = 1, \ldots, 2J$, leads to the following proposition.

**Proposition 4.5.** For the testing problem (4.7), there exists a family of canonical covariate pairs of the form $((x_1, \text{ix}_1), \ldots, (x_{2J}, \text{ix}_{2J}))$, and the canonical correlations satisfy

$$c_k = \text{Corr}_r(x_k, \text{ix}_k)$$

$$= \max \{\text{Corr}_r(x, \text{ix}) \mid x \in \mathbb{R}^J \setminus \{0\},$$

$$\text{Cov}_r(x, x_k') = \text{Cov}_r(x, \text{ix}_{k'}) = 0, k' = 1, \ldots, k - 1\},$$

$k = 1, \ldots, 2J$.

If $\Sigma \in \text{PD}_C(I)$, then $\Sigma \in \text{PD}_H(I)$ if and only if $\text{Cov}_r(x, \text{ix}) = 0$ for all $x \in \mathbb{R}^J$, i.e., the null hypothesis is equivalent to the pairs $(x, \text{ix})$, $x \in \mathbb{R}^J$, being uncorrelated. As in the previous examples, the empirical canonical correlations are obtained by recursively maximizing the correlation between such pairs wrt. $\hat{\Sigma}$, providing a statistical interpretation to the maximal invariant statistic.

### 4.5. Testing that Two Covariance Matrices Are Equal, cf. Section 7 of (ABJ).

Assuming that $I = 2J$, and the first $J$ components of $X_n$ are independent of the last $J$ components, consider testing the hypothesis that the corresponding covariance matrices are equal, i.e.,

$$H_0 : \Sigma = \begin{pmatrix} \Gamma & 0 \\ 0 & \Gamma \end{pmatrix} \text{ vs. } H : \Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}.$$  

This testing problem is seen to be a group symmetry testing problem by defining $G = \{\pm 1_J, \pm \text{f}\}$ and $G_0 = \langle \pm 1_J, \pm \text{f}, \pm \text{s} \rangle$, where $\text{f} = \text{Diag}(1_J, -1_J)$ and $\text{s} = (1_J 0)$. The maximum likelihood estimators are $\hat{\Sigma} = \text{Diag}(S_{11}, S_{22})$, and

$$\hat{\Sigma}_0 = \Psi_{G_0}(S) = \frac{1}{2} \begin{pmatrix} S_{11} + S_{22} & 0 \\ 0 & S_{11} + S_{22} \end{pmatrix}.$$ 

The ordered eigenvalues of $r(\Sigma)$ wrt. $t(\Sigma)$ have the form $\lambda_1, \ldots, \lambda_J, -\lambda_J, \ldots, -\lambda_1$, where $1 > \lambda_1 \geq \cdots \geq \lambda_J \geq 0$, and we define
\[ \pi : \text{PD}(J) \oplus \text{PD}(J) \to \Lambda_J \]

The family of canonical correlations is \( \pi(\Sigma) \), and the family of empirical canonical correlations is \( \pi(\Sigma) \), which is not a maximal invariant statistic in this case. The maximal invariant statistic is the ordered family of eigenvalues of \( S_{11} \) wrt. \( S_{11} + S_{22} \), and a statistical interpretation of these eigenvalues can be obtained by modifying the canonical variance distances presented in Section 5.

As in the previous examples, the canonical correlations for this testing problem have a special form. Let \( u_1, \ldots, u_J, su_1, \ldots, su_J \), be a basis of \( \mathbb{R}^J \) that is orthonormal wrt. \( t(\Sigma) \), such that \( u_k \) is an eigenvector of \( r(\Sigma) \) wrt. \( t(\Sigma) \) corresponding to \( \lambda_k \), \( k = 1, \ldots, J \), and note that \( su_k \) is an eigenvector corresponding to \( -\lambda_k \). A family of empirical canonical covariate pairs is therefore given by \( x_k = \frac{1}{\sqrt{2}}(u_k + su_k) \) and \( y_k = \frac{1}{\sqrt{2}}(u_k - su_k) \), \( k = 1, \ldots, J \), and observing that \( sx_k = x_k \) and \( sy_k = -y_k \) yields the following proposition.

**Proposition 4.6.** For the testing problem (4.8), there exists a family of canonical variance distances and distance variates exist by the extreme value theorem, and note that families of canonical variance distances and distance variates are characterized in terms of the eigenvalues and eigenvectors of \( \Sigma - \Sigma_0 \) wrt. \( \Sigma_0 \), respectively.

**Canonical Variance Distances**

Canonical variance distances are analogous to canonical correlations, and they are defined for testing problems meeting the general conditions from Section 3. The idea is to measure the distance between \( \Sigma \in \Theta \) and \( \Sigma_0 \in \Theta_0 \) by finding linear forms \( x \in \mathbb{R}^I \), such that \( |V_{\Sigma}(x) - V_{\Sigma_0}(x)| \) is maximized. Of course, \( |V_{\Sigma}(x) - V_{\Sigma_0}(x)| \) can be made arbitrarily large if \( x \) is unbounded, so we impose the constraint \( V_{\Sigma_0}(x) = 1 \), which leads to the following definition.

**Definition 5.1.** A family \( (x_1, \ldots, x_I) \) of vectors from \( \mathbb{R}^I \) is called a family of canonical distance variates of \( \Sigma \) wrt. \( \Sigma_0 \) if

\[
\begin{align*}
d_1 := & \max \{ |V_{\Sigma}(x) - 1| \mid x \in \mathbb{R}^I, V_{\Sigma_0}(x) = 1 \} \text{ is attained at } x_1, \text{ and} \\
d_i := & \max \{ |V_{\Sigma}(x) - 1| \mid x \in \mathbb{R}^I, V_{\Sigma_0}(x) = 1, \text{Cov}_{\Sigma_0}(x, x_j) = 0, j = 1, \ldots, i - 1 \}
\end{align*}
\]

is attained at \( x_i \), for \( i = 2, \ldots, I \). The family \( (d_1, \ldots, d_I) \) is called the family of canonical variance distances. The empirical canonical variance distances and distance variates are defined the same way by replacing \( \Sigma \) and \( \Sigma_0 \) with \( \hat{\Sigma} \) and \( \hat{\Sigma}_0 \), respectively.

Note that families of canonical variance distances and distance variates exist by the extreme value theorem, and \( d_1 \geq \cdots \geq d_I \geq 0 \). Like canonical correlations, they are characterized in terms of the eigenvalues and eigenvectors of \( \Sigma - \Sigma_0 \) wrt. \( \Sigma_0 \), as seen in the following theorem.

**Theorem 5.2.** Suppose the eigenvalues \( \lambda_1, \ldots, \lambda_I \) of \( \Sigma - \Sigma_0 \) wrt. \( \Sigma_0 \) are ordered so that \( |\lambda_1| \geq \cdots \geq |\lambda_I| \). Then \( (x_1, \ldots, x_I) \) is a family of canonical distance variates of \( \Sigma \) wrt. \( \Sigma_0 \) if and only if

\[
\begin{align*}
(1) & \text{ } (x_1, \ldots, x_I) \text{ is orthonormal wrt. } \Sigma_0, \text{ and}
\end{align*}
\]
The family of canonical variance distances is $(|\lambda_1|, \ldots, |\lambda_I|)$.

**Proof.** As in the proof of Theorem 3.4, assume without loss of generality that $\Sigma = \mathbf{1}_I + \Lambda$ and $\Sigma_0 = \mathbf{1}_I$, where $\Lambda = \text{Diag}(\lambda_1, \ldots, \lambda_I)$, and let $V_{\lambda} \subseteq \mathbb{R}^I$ be the eigenspace of $\Lambda$ corresponding to $\lambda_i$ or $-\lambda_i$, for $i = 1, \ldots, I$.

The conditions from Definition 5.1 are now

\begin{align*}
(5.1) & \quad d_1 = \max \{ |x^t \Lambda x| \mid x \in \mathbb{R}^I, |x| = 1 \} \text{ is attained at } x_1, \\
(5.2) & \quad d_i = \max \{ |x^t \Lambda x| \mid x \in \mathbb{R}^I, |x| = 1, x^t x_j = 0, j = 1, \ldots, i-1 \},
\end{align*}

and the conditions from the theorem are

1. $x_1, \ldots, x_I$ are orthonormal, and
2. $x_i \in V_{\lambda_i} \cup V_{-\lambda_i}$, for $i = 1, \ldots, I$.

Let $x_1 = (x_11, \ldots, x_{1I})^t$, and suppose $|x_1| = 1$. Then

$$
|x_1^t \Lambda x_1| \leq \sum_{i=1}^I |\lambda_i| |x_{1i}|^2 \leq \sum_{i=1}^I |\lambda_1| |x_{1i}|^2 = |\lambda_1|,
$$

with equality if and only if $x_{1i} \neq 0$ implies $\lambda_i = \lambda_1$ or $x_{1i} \neq 0$ implies $\lambda_i = -\lambda_1$. This proves that $x_1$ satisfies condition (5.1) if and only if $|x_1| = 1$ and $x_1 \in V_{\lambda_1} \cup V_{-\lambda_1}$, and the theorem follows by induction on $I$. \qed

We have already seen that, for fundamental group symmetry testing problems, the ordered eigenvalues of $r(\Sigma) = \Sigma - \Psi G_0(\Sigma)$ wrt. $t(\Sigma) = \Psi G_0(\Sigma)$ have the form

$$
\lambda_1 \geq \lambda_2 \geq \cdots \geq -\lambda_2 \geq -\lambda_1,
$$

so the family of canonical variances distances is $d(\Sigma) = (|\lambda_1|, |\lambda_1|, |\lambda_2|, |\lambda_2|, \ldots)$. The family of empirical canonical variance distances is $d(\hat{\Sigma})$, which is a maximal invariant statistic for the fundamental testing problems from Sections 2 through 6 of (ABJ). Therefore, the empirical canonical variance distances provide a second statistical interpretation to these maximally invariant eigenvalues.

**References**


Department of Statistics, Indiana University, Bloomington, IN 47405
Department of Mathematics, Tarleton State University, Stephenville, TX 76402
E-mail address: standers@indiana.edu, jcrawford@tarleton.edu