On Testing for Impropriety of Complex-Valued Gaussian Vectors

A. T. Walden* (*Member IEEE*) and P. Rubin-Delanchy

Abstract— **We consider the problem of testing whether a complex-valued random vector is proper, i.e., is uncorrelated with its complex conjugate. We formulate the testing problem in terms of real-valued Gaussian random vectors, so we can make use of some useful existing results which enable us to study the null distributions of two test statistics. The tests depend only on the sample-size, n, and the dimensionality of the vector, p. The basic behaviours of the distributions of the test statistics are derived and critical values (thresholds) are calculated and presented for certain (n, p) values. For one of these tests we derive a distributional approximation for a transform of the statistic, potentially very useful in practice for rapid and simple testing. We also study the power (detection probability) of the tests. Our results mean that testing for propriety can be a practical and undaunting procedure.**

Keywords— **Detection probability, hypothesis test, improper complex random vector, invariant statistic, threshold**

I. INTRODUCTION

A complex-valued random vector may be categorized as proper/improper, according to whether it is uncorrelated/correlated with its complex conjugate. A proper random vector has second-order statistics which are invariant to circular rotation, and the simpler statistical properties have led to the assumption of propriety being made simply for mathematical convenience.

Complex-valued signals occur naturally in areas such as quadrature Doppler ultrasound [9], seismic signal analysis [19], magnetic resonance imaging [14], and communications (where improper signals arise in binary phase shift keying and Gaussian minimum shift keying [18]).

Let $\boldsymbol{Z} = [Z_1, \ldots, Z_p]^T$ denote a complex-valued random column vector with mean zero. The covariance matrix of the complex-valued vector $\boldsymbol{Z}$ is defined as $\boldsymbol{\Gamma} = E\{\boldsymbol{Z}\boldsymbol{Z}^H\}$, and is a Hermitian positive definite matrix. Additionally, $\boldsymbol{Z}$ has a *relation matrix* [11] or *complementary covariance matrix* [16] given by $\boldsymbol{R} = E\{\boldsymbol{Z}\boldsymbol{Z}^T\}$, which is complex and symmetric.

These two matrices both appear in the so-called *augmented covariance matrix* $\boldsymbol{G}$, [16], [20]. Suppose we form the augmented vector $\boldsymbol{\check{Z}}$ by adjoining $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$, i.e.,

$$\boldsymbol{\check{Z}} = [\boldsymbol{Z}^T, \boldsymbol{Z}^H]^T = [Z_1, \ldots, Z_p, Z_1^*, \ldots, Z_p^*]^T,$$

then letting $\boldsymbol{G} = E\{\boldsymbol{\check{Z}}\boldsymbol{\check{Z}}^H\}$, we obtain

$$\boldsymbol{G} = E\left\{ \begin{bmatrix} \boldsymbol{Z} \\ \boldsymbol{Z}^* \end{bmatrix} \begin{bmatrix} \boldsymbol{Z}^H & \boldsymbol{Z}^T \end{bmatrix} \right\} = \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{R} \\ \boldsymbol{R}^* & \boldsymbol{\Gamma}^* \end{bmatrix}. \qquad (1)$$

$\boldsymbol{G}$ captures all the second-order structure of $\boldsymbol{Z}$. A complex-valued random vector for which $\boldsymbol{R} = \boldsymbol{0}$, is called a *proper complex-valued random vector*, [8]. When $\boldsymbol{R} \neq \boldsymbol{0}$ it is called an *improper complex-valued random vector* [16].

A test for propriety was developed independently by Ollila and Koivunen [10] and by Schreier, Scharf and Hanssen [17], [18], the hypothesis test being $H_0 : \boldsymbol{R} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{R} \neq \boldsymbol{0}$, and was based on n independent and identically distributed samples, $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$, of the random vector. In this paper we shall instead study propriety testing via a real-valued representation of the problem, (Section II). While one of our two test statistics is identical to the one found in the references above, our approach links readily to several important statistical inferential results already available in the statistical literature, enabling us to study the null distributions of the two test statistics, and calculate critical values, (testing thresholds), for certain combinations of sample size n and dimensionality p. For one of these tests we derive a Box approximation, potentially very useful in practice for the rapid and simple calculation of critical values. We study the power of the tests (detection probabilities) to compare the two statistics. Our results make testing for impropriety a practical procedure.

Section III develops the ideas underlying the testing problem in the real-valued Gaussian formulation, and shows that the key components, the canonical correlations, are the same as obtained in [18]. A test statistic T_1, which emerges from our formulation, is also seen to provide a generalized likelihood ratio test. Another statistic, T_2, gives a locally most powerful test, but it is pointed out in Section IV that no uniformly most powerful test exists for the problem for $p \geq 2$, answering a query posed in [18]. Section V shows that under the null hypothesis we can assume the covariance matrix is the identity, thus enabling the computation of quantiles of the distributions of the test statistics. The basic behaviours of the distributions of the test statistics as functions of (n, p) are derived in Section VI. Critical values are calculated and presented in Section VII for certain (n, p) values, and Box's approximation method is used to derive a distribution for a transformation of T_1. It is shown that Box's approximation delivers an acceptable level of accuracy for much smaller n than Wilk's asymptotic approximation. The powers of the two tests are examined in Section VIII, where the distance

of hypothesis alternatives from the null are measured in terms of the norm of a certain eigenvalue vector. Our resulting recommendations for practical testing are given in Section IX.

II. The real-valued formulation

We can also write $\boldsymbol{Z} = \boldsymbol{X} + \mathrm{i}\boldsymbol{Y}$, where $\boldsymbol{X} = \mathrm{Re}\{\boldsymbol{Z}\}, \boldsymbol{Y} = \mathrm{Im}\{\boldsymbol{Z}\}$. Defining

$$\check{\boldsymbol{X}} = [\boldsymbol{X}^T, \boldsymbol{Y}^T]^T = [X_1, \ldots, X_p, Y_1, \ldots, Y_p]^T,$$

we have [11], $\check{\boldsymbol{X}} = \boldsymbol{M}\check{\boldsymbol{Z}}$, where

$$\boldsymbol{M} = \tfrac{1}{2} \begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{I}_p \\ -\mathrm{i}\boldsymbol{I}_p & \mathrm{i}\boldsymbol{I}_p \end{bmatrix} \quad \text{and} \quad \boldsymbol{M}^{-1} = \begin{bmatrix} \boldsymbol{I}_p & \mathrm{i}\boldsymbol{I}_p \\ \boldsymbol{I}_p & -\mathrm{i}\boldsymbol{I}_p \end{bmatrix}.$$

Let $\boldsymbol{\Upsilon}$ denote the covariance matrix of the real-valued random vector $\check{\boldsymbol{X}}$ and suppose we partition up the $2p \times 2p$ covariance matrix $\boldsymbol{\Upsilon} = \mathrm{cov}\{\check{\boldsymbol{X}}\} = E\{\check{\boldsymbol{X}}\check{\boldsymbol{X}}^T\}$ into the four constituent $p \times p$ matrices,

$$\boldsymbol{\Upsilon} = \begin{bmatrix} \boldsymbol{\Upsilon}_{XX} & \boldsymbol{\Upsilon}_{XY} \\ \boldsymbol{\Upsilon}_{YX} & \boldsymbol{\Upsilon}_{YY} \end{bmatrix}. \tag{2}$$

Now let $\mathcal{H}^+$ denote the set of real $2p \times 2p$ positive definite symmetric matrices of "complex" form

$$\begin{bmatrix} \boldsymbol{A}_1 & -\boldsymbol{A}_2 \\ \boldsymbol{A}_2 & \boldsymbol{A}_1 \end{bmatrix}, \tag{3}$$

where $\boldsymbol{A}_1$ is symmetric, $\boldsymbol{A}_2$ is antisymmetric ($\boldsymbol{A}_2 = -\boldsymbol{A}_2^T$), and therefore $\boldsymbol{A}_1 + \mathrm{i}\boldsymbol{A}_2$ is Hermitian. Let us define $\mathcal{R}$ to be the set of matrices of the form

$$\begin{bmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{F} & -\boldsymbol{E} \end{bmatrix}, \tag{4}$$

where $\boldsymbol{E}$ and $\boldsymbol{F}$ are real $p \times p$ symmetric matrices. Then $\boldsymbol{\Upsilon}$ in (2) can be written as

$$\boldsymbol{\Upsilon} = \dot{\boldsymbol{\Upsilon}} + \ddot{\boldsymbol{\Upsilon}}, \tag{5}$$

where

$$\dot{\boldsymbol{\Upsilon}} = \tfrac{1}{2} \begin{bmatrix} \boldsymbol{\Upsilon}_{XX} + \boldsymbol{\Upsilon}_{YY} & \boldsymbol{\Upsilon}_{XY} - \boldsymbol{\Upsilon}_{YX} \\ \boldsymbol{\Upsilon}_{YX} - \boldsymbol{\Upsilon}_{XY} & \boldsymbol{\Upsilon}_{XX} + \boldsymbol{\Upsilon}_{YY} \end{bmatrix} \in \mathcal{H}^+,$$

and

$$\ddot{\boldsymbol{\Upsilon}} = \tfrac{1}{2} \begin{bmatrix} \boldsymbol{\Upsilon}_{XX} - \boldsymbol{\Upsilon}_{YY} & \boldsymbol{\Upsilon}_{XY} + \boldsymbol{\Upsilon}_{YX} \\ \boldsymbol{\Upsilon}_{YX} + \boldsymbol{\Upsilon}_{XY} & \boldsymbol{\Upsilon}_{YY} - \boldsymbol{\Upsilon}_{XX} \end{bmatrix} \in \mathcal{R}.$$

Now

$$\boldsymbol{\Gamma} = E\{\boldsymbol{Z}\boldsymbol{Z}^H\} = (\boldsymbol{\Upsilon}_{XX} + \boldsymbol{\Upsilon}_{YY}) + \mathrm{i}(\boldsymbol{\Upsilon}_{YX} - \boldsymbol{\Upsilon}_{XY}),$$

and

$$\boldsymbol{R} = E\{\boldsymbol{Z}\boldsymbol{Z}^T\} = (\boldsymbol{\Upsilon}_{XX} - \boldsymbol{\Upsilon}_{YY}) + \mathrm{i}(\boldsymbol{\Upsilon}_{YX} + \boldsymbol{\Upsilon}_{XY}).$$

If $\boldsymbol{R} = \boldsymbol{0}$, then $\boldsymbol{\Upsilon}_{XX} = \boldsymbol{\Upsilon}_{YY} = \mathrm{Re}\{\boldsymbol{\Gamma}\}/2$, and $\boldsymbol{\Upsilon}_{XY} = -\boldsymbol{\Upsilon}_{YX} = -\mathrm{Im}\{\boldsymbol{\Gamma}\}/2$. So a test for propriety could be

based on the hypothesis test $H_0 : \ddot{\boldsymbol{\Upsilon}} = \boldsymbol{0}$ versus $H_1 : \ddot{\boldsymbol{\Upsilon}} \neq \boldsymbol{0}$.

Under propriety, $\boldsymbol{\Upsilon}$ takes the special form

$$\boldsymbol{\Upsilon} = \tfrac{1}{2} \begin{bmatrix} \mathrm{Re}\{\boldsymbol{\Gamma}\} & -\mathrm{Im}\{\boldsymbol{\Gamma}\} \\ \mathrm{Im}\{\boldsymbol{\Gamma}\} & \mathrm{Re}\{\boldsymbol{\Gamma}\} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Upsilon}_1 & -\boldsymbol{\Upsilon}_2 \\ \boldsymbol{\Upsilon}_2 & \boldsymbol{\Upsilon}_1 \end{bmatrix} \in \mathcal{H}^+. \tag{6}$$

If we let $\mathcal{S}^+$ denote the set of real $2p \times 2p$ positive definite symmetric matrices the test can be written in the alternative form

$$H_0 : \boldsymbol{\Upsilon} \in \mathcal{H}^+ \qquad \text{versus} \qquad H_1 : \boldsymbol{\Upsilon} \in \mathcal{S}^+ \backslash \mathcal{H}^+, \tag{7}$$

exactly as studied in two technical papers by Andersson, Brøns and Jensen [1] and Andersson and Perlman [2] where the problem was called "testing the hypothesis that a covariance matrix has complex structure".

III. Invariant statistics

A. Transformation Groups

Let $\mathcal{C}$ denote the group of all nonsingular real $2p \times 2p$ matrices of the form

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_1 & -\boldsymbol{C}_2 \\ \boldsymbol{C}_2 & \boldsymbol{C}_1 \end{bmatrix}, \tag{8}$$

where $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are $p \times p$. Note that $\mathcal{H}^+ = \mathcal{S}^+ \cap \mathcal{C}$.

Now if $\boldsymbol{H} \in \mathcal{H}^+$ and $\boldsymbol{C} \in \mathcal{C}$ then we can define a group action

$$g_{\boldsymbol{C}}(\boldsymbol{H}) = \boldsymbol{C}\boldsymbol{H}\boldsymbol{C}^T \in \mathcal{H}^+. \tag{9}$$

The group $\{g_{\boldsymbol{C}} : \boldsymbol{C} \in \mathcal{C}\}$ has identity $g_{\boldsymbol{I}}$. Likewise $\mathcal{R}$ and $\mathcal{S}^+$ are closed under the action of the group, so that if $\boldsymbol{R} \in \mathcal{R}$ then

$$g_{\boldsymbol{C}}(\boldsymbol{R}) = \boldsymbol{C}\boldsymbol{R}\boldsymbol{C}^T \in \mathcal{R}. \tag{10}$$

and if $\boldsymbol{S} \in \mathcal{S}^+$ then

$$g_{\boldsymbol{C}}(\boldsymbol{S}) = \boldsymbol{C}\boldsymbol{S}\boldsymbol{C}^T \in \mathcal{S}^+. \tag{11}$$

Equations (9) and (11) imply [6, p. 213] that the test (7) is invariant under the group action $g_{\boldsymbol{C}}(\cdot)$. The principle of invariance [6, p. 11] then suggests restricting attention to test statistics T which satisfy

$$T(\boldsymbol{S}) = T(g_{\boldsymbol{C}}(\boldsymbol{S})).$$

Such a statistic is said to be invariant to the action of the group: its value does not depend on whether $\boldsymbol{S}$ or $g_{\boldsymbol{C}}(\boldsymbol{S})$ is observed. A statistic T is maximal invariant if every other invariant statistic is a function of it, or equivalently [6, p. 11], [21, p. 86]

$$T(\boldsymbol{S}_1) = T(\boldsymbol{S}_2) \Rightarrow \boldsymbol{S}_2 = g_{\boldsymbol{C}}(\boldsymbol{S}_1) \text{ for some } \boldsymbol{C} \in \mathcal{C}. \tag{12}$$

B. Invariant Parameter

It is proved in [2] that every matrix $\boldsymbol{\Upsilon} \in \mathcal{S}^+$ can be represented in the form

$$\boldsymbol{\Upsilon} = \boldsymbol{C}\boldsymbol{D}\boldsymbol{C}^T \quad \text{where} \quad \boldsymbol{D} = \begin{bmatrix} \boldsymbol{I}_p + \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_p - \boldsymbol{\Lambda} \end{bmatrix}; \quad (13)$$

where $\boldsymbol{C} \in \mathcal{C}$, $\boldsymbol{I}_p$ is the $p \times p$ identity matrix, $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, $1 > \lambda_1 \geq \cdots \geq \lambda_p \geq 0$, and $\pm\lambda_1, \ldots, \pm\lambda_p$ are the eigenvalues of $\acute{\boldsymbol{\Upsilon}}^{-1}\ddot{\boldsymbol{\Upsilon}}$. Now

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_p \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{\Lambda} \end{bmatrix},$$

where

$$\begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_p \end{bmatrix} \in \mathcal{H}^+ \text{ and } \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{\Lambda} \end{bmatrix} \in \mathcal{R}. \quad (14)$$

Using (9) and (10) we see that if $\boldsymbol{\Lambda}$ is null then $\boldsymbol{\Upsilon} = \boldsymbol{C}\boldsymbol{C}^T \in \mathcal{H}^+$ since $\boldsymbol{C}\boldsymbol{C}^T \in \mathcal{C}$ and $\boldsymbol{C}\boldsymbol{C}^T$ is positive definite; if $\boldsymbol{\Lambda}$ is non-null then $\boldsymbol{\Upsilon} \notin \mathcal{H}^+$. The testing problem can therefore be restated as

$$H_0 : \text{all } \lambda_k = 0 \quad \text{versus} \quad H_1 : \text{not all } \lambda_k = 0. \quad (15)$$

The invariance of the test to the group action is to be expected since $g_{\boldsymbol{C}}(\boldsymbol{\Upsilon}) = \boldsymbol{C}\boldsymbol{\Upsilon}\boldsymbol{C}^T$ is equivalent to transforming $\check{\boldsymbol{X}}$ to $\boldsymbol{C}\check{\boldsymbol{X}}$ which in turn is equivalent [8, p. 1296] to transforming $\boldsymbol{Z}$ to $(\boldsymbol{C}_1 + \mathrm{i}\boldsymbol{C}_2)\boldsymbol{Z}$, which is just a linear transform of $\boldsymbol{Z}$, and proper complex random vectors are still proper after such a transformation [8, p. 1295].

C. Canonical Correlations

The idea of a test for impropriety being invariant under linear (but not widely linear) transformations underpinned the approach in Schreier *et al.* [18] based on complex-valued representations. They showed that the canonical correlations between $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$ are the nonnegative square-roots of the eigenvalues of $\boldsymbol{\Gamma}^{-1}\boldsymbol{R}\boldsymbol{\Gamma}^{-*}\boldsymbol{R}^*$; the corresponding empirical or sample canonical correlations are maximal invariant and any test statistic must be a function of them.

It is straightforward to show that $\lambda_1, \ldots, \lambda_p$ are canonical correlations. $\boldsymbol{\Upsilon} \in \mathcal{S}^+$ can be represented in the form $\boldsymbol{C}\boldsymbol{D}\boldsymbol{C}^T$ for some $\boldsymbol{C} \in \mathcal{C}$, where $\boldsymbol{D}$ is given in (13). By [2, p. 36] this can be rewritten as

$$\boldsymbol{\Upsilon} = \mathrm{cov}\{\check{\boldsymbol{X}}\} = \boldsymbol{B} \begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda} & \boldsymbol{I}_p \end{bmatrix} \boldsymbol{B}^T,$$

where

$$\boldsymbol{B} = \frac{1}{\sqrt{2}}\boldsymbol{C} \begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{I}_p \\ -\boldsymbol{I}_p & \boldsymbol{I}_p \end{bmatrix} \in \mathcal{C}.$$

So,

$$\mathrm{cov}\{\boldsymbol{B}^{-1}\check{\boldsymbol{X}}\} = \begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda} & \boldsymbol{I}_p \end{bmatrix} = \boldsymbol{P},$$

say, so that $\boldsymbol{P} = \boldsymbol{B}^{-1}\boldsymbol{\Upsilon}\boldsymbol{B}^{-H}$. Let $\boldsymbol{T} = \boldsymbol{M}^{-1}$. A straightforward matrix manipulation then shows that

$$\mathrm{cov}\{\boldsymbol{T}\boldsymbol{B}^{-1}\boldsymbol{T}^{-1}\check{\boldsymbol{Z}}\} = \boldsymbol{T}\boldsymbol{P}\boldsymbol{T}^H = 2 \begin{bmatrix} \boldsymbol{I}_p & \mathrm{i}\boldsymbol{\Lambda} \\ -\mathrm{i}\boldsymbol{\Lambda} & \boldsymbol{I}_p \end{bmatrix}.$$

Let $\boldsymbol{B}^{ij}$ be the (i,j)th $p \times p$ block of $\boldsymbol{B}^{-1} \in \mathcal{C}$. Then

$$\boldsymbol{T}\boldsymbol{B}^{-1}\boldsymbol{T}^{-1} = \begin{bmatrix} \boldsymbol{B}^{11} + \mathrm{i}\boldsymbol{B}^{21} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}^{11} - \mathrm{i}\boldsymbol{B}^{21} \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}_0 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}_0^* \end{bmatrix},$$

say. Hence $\boldsymbol{T}\boldsymbol{B}^{-1}\boldsymbol{T}^{-1}$ is block diagonal and we see that

$$\mathrm{cov}\left\{ \begin{bmatrix} 2^{-1/2}\,\mathrm{e}^{-\mathrm{i}\pi/4}\boldsymbol{B}_0\boldsymbol{Z} \\ 2^{-1/2}\,\mathrm{e}^{\mathrm{i}\pi/4}\boldsymbol{B}_0^*\boldsymbol{Z}^* \end{bmatrix} \right\} = \begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda} & \boldsymbol{I}_p \end{bmatrix},$$

which means that $\lambda_1, \ldots, \lambda_p$ are the canonical correlations between $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$, (e.g., [4, p. 371]).

We note that,

$$\det(\boldsymbol{D}) = \prod_{k=1}^{p}(1 - \lambda_k^2) = \det \begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda} & \boldsymbol{I}_p \end{bmatrix},$$

so, interestingly, the determinant of the representation matrix $\boldsymbol{D}$ is identical to the determinant of the covariance matrix of the variables in canonical form. The sample version of this determinant will later provide a test statistic.

We now verify that the canonical correlations, $\lambda_1, \ldots, \lambda_p$, are identical to those in [18], by writing them in the same form. The eigenvalues of $\acute{\boldsymbol{\Upsilon}}^{-1}\ddot{\boldsymbol{\Upsilon}}$ are the solutions of the generalized eigenproblem

$$\det(\ddot{\boldsymbol{\Upsilon}} - \lambda\acute{\boldsymbol{\Upsilon}}) = 0. \quad (16)$$

Then the roots of

$$\det(\boldsymbol{T}\ddot{\boldsymbol{\Upsilon}}\boldsymbol{T}^H - \lambda\boldsymbol{T}\acute{\boldsymbol{\Upsilon}}\boldsymbol{T}^H) = 0 \quad (17)$$

are the same as those of (16) since

$$\det(\boldsymbol{T}\ddot{\boldsymbol{\Upsilon}}\boldsymbol{T}^H - \lambda\boldsymbol{T}\acute{\boldsymbol{\Upsilon}}\boldsymbol{T}^H) = \det(\boldsymbol{T}[\ddot{\boldsymbol{\Upsilon}} - \lambda\acute{\boldsymbol{\Upsilon}}]\boldsymbol{T}^H)$$
$$= \det(\boldsymbol{T}) \cdot \det(\ddot{\boldsymbol{\Upsilon}} - \lambda\acute{\boldsymbol{\Upsilon}}) \cdot \det(\boldsymbol{T}^H),$$

and $\det(\boldsymbol{T}), \det(\boldsymbol{T}^H) \neq 0$.

Hence the eigenvalues of $\acute{\boldsymbol{\Upsilon}}^{-1}\ddot{\boldsymbol{\Upsilon}}$ are also the eigenvalues of $(\boldsymbol{T}\acute{\boldsymbol{\Upsilon}}\boldsymbol{T}^H)^{-1}\boldsymbol{T}\ddot{\boldsymbol{\Upsilon}}\boldsymbol{T}^H$. Now,

$$(\boldsymbol{T}\acute{\boldsymbol{\Upsilon}}\boldsymbol{T}^H)^{-1} = \begin{bmatrix} \boldsymbol{\Gamma}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Gamma}^{-*} \end{bmatrix} \quad \text{and} \quad \boldsymbol{T}\ddot{\boldsymbol{\Upsilon}}\boldsymbol{T}^H = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{R} \\ \boldsymbol{R}^* & \boldsymbol{0} \end{bmatrix}.$$

The non-negative eigenvalues of

$$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{\Gamma}^{-1}\boldsymbol{R} \\ \boldsymbol{\Gamma}^{-*}\boldsymbol{R}^* & \boldsymbol{0} \end{bmatrix},$$

are the non-negative square-roots of the eigenvalues of $\boldsymbol{\Gamma}^{-1}\boldsymbol{R}\boldsymbol{\Gamma}^{-*}\boldsymbol{R}^*$, and so are identical to the canonical correlations derived in [18].

D. Test Statistics

Let $\check{\boldsymbol{X}}_1, \ldots, \check{\boldsymbol{X}}_n$ be n independent $2p$-dimensional random vectors from a normal distribution with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Upsilon}$. If $n \geq 2p$ the maximum likelihood estimator of $\boldsymbol{\Upsilon}$ is given by

$$\hat{\boldsymbol{\Upsilon}} = (1/n) \sum_{k=0}^{n-1} \check{\boldsymbol{X}}_k \check{\boldsymbol{X}}_k^T$$

and $\boldsymbol{W} = n\hat{\boldsymbol{\Upsilon}}$ has the real Wishart distribution on $\mathcal{S}^+$ and using (5) can be written $\boldsymbol{W} = \dot{\boldsymbol{W}} + \ddot{\boldsymbol{W}}$ with $\dot{\boldsymbol{W}} \in \mathcal{H}^+, \ddot{\boldsymbol{W}} \in \mathcal{R}$.

Note we require $n \geq 2p$. This should not be a problem in practice as we expect the dimension of the vector, p, to be small, and the sample size, n, could range from at least $2p$ up to very large, in which case results are 'asymptotic' in nature; see Sections VI–VIII.

Every matrix $\boldsymbol{W} \in \mathcal{S}^+$ can be represented in the form

$$\boldsymbol{W} = \boldsymbol{C}\hat{\boldsymbol{D}}\boldsymbol{C}^T \quad \text{where} \quad \hat{\boldsymbol{D}} = \begin{bmatrix} \boldsymbol{I}_p + \boldsymbol{L} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_p - \boldsymbol{L} \end{bmatrix}; \quad (18)$$

where $\boldsymbol{C} \in \mathcal{C}$, $\boldsymbol{L} = \mathrm{diag}(\ell_1, \ldots, \ell_p)$, $1 > \ell_1 \geq \cdots \geq \ell_p \geq 0$, and $\pm\ell_1, \ldots, \pm\ell_p$ are the eigenvalues of $\dot{\boldsymbol{W}}^{-1}\ddot{\boldsymbol{W}}$. The statistic $T(\boldsymbol{W}) \equiv (\ell_1, \ldots, \ell_p)$ is comprised of empirical or sample canonical correlations and is thus maximal invariant.

From (13) and (15) we see that, under H_0, $\det(\boldsymbol{D}) = 1$. An invariant test follows as: accept H_0 iff

$$T_1(n,p) \equiv \det(\hat{\boldsymbol{D}}) = \prod_{k=1}^p (1 - \ell_k^2) \geq c_1, \quad (19)$$

where c_1 is a constant. So if $\det(\hat{\boldsymbol{D}})$ does not depart 'too much' from unity, we accept H_0; note that $T_1(n,p)$ is a function of the maximal invariant.

The size of the test (false alarm probability), α, finds c_1 such that $\Pr\left[\prod_{k=1}^p (1 - \ell_k^2) \geq c_1 | \dot{\boldsymbol{\Upsilon}}\right] = 1 - \alpha$.

The power of a test is generally considered the gold-standard by which the quality of a test is assessed. In our notation it is defined as the probability that H_0 is rejected given the true matrix parameter $\boldsymbol{\Upsilon}$. Ideally a power function would be zero for those $\boldsymbol{\Upsilon}$ corresponding to the null hypothesis ($\boldsymbol{\Upsilon} = \dot{\boldsymbol{\Upsilon}}$), and unity for the alternative. The power of the test (19) could be written $\Pr\left[\prod_{k=1}^p (1 - \ell_k^2) < c_1 | \boldsymbol{\Upsilon}\right]$, and was shown in [2] to increase with non-zero values of the λ_k. Given a small enough value of α (and consequently small c_1) for the power of the test to approach the ideal value of unity it is sufficient (and necessary) that the largest eigenvalue approaches unity, i.e., $\lambda_1 \to 1$.

A second invariant test statistic, $T_2(n,p)$, derived from the maximal invariant provides another invariant test. We accept H_0 if

$$T_2(n,p) \equiv \tfrac{1}{2}\mathrm{tr}(\dot{\boldsymbol{W}}^{-1}\ddot{\boldsymbol{W}}\dot{\boldsymbol{W}}^{-1}\ddot{\boldsymbol{W}}) = \sum_{k=1}^p \ell_k^2 \leq c_2, \quad (20)$$

where c_2 is a constant; this is the *locally* most powerful (LMP) invariant test for (7) [2, Theorem 9.1]. Simply put this means that the test has as high a power as possible for alternatives H_1 that are close to the null hypothesis, i.e., those for which all the λ_k are small. For the LMP test of (20), the power function is $\Pr\left[\sum_{k=1}^p \ell_k^2 > c_2, | \boldsymbol{\Upsilon}\right]$, and for small enough α (large c_2) this need not approach 1 as $\lambda_1 \to 1$; it may need more than one eigenvalue close to 1. As an LMP test, it may perform poorly for alternatives far from the null. These questions are examined in Section VIII.

Since the geometric mean of $(1 - \ell_1^2), \ldots, (1 - \ell_p^2)$ is less than or equal to the arithmetic mean of the same, then, in terms of the test statistics,

$$T_1^p(n,p) \leq 1 - \tfrac{1}{p}T_2(n,p).$$

Often we shall abbreviate $T_1(n,p)$ and $T_2(n,p)$ to just T_1 and T_2, respectively.

E. Generalized Likelihood Ratio

Now $\boldsymbol{W} = n\hat{\boldsymbol{\Upsilon}}$ has the real Wishart distribution on $\mathcal{S}^+$ and hence has probability density function,

$$f(\boldsymbol{W}; n, \boldsymbol{\Upsilon}) = \tfrac{1}{C}(\det\{\boldsymbol{\Upsilon}\})^{-n/2}(\det\{\boldsymbol{W}\})^{(n-2p-1)/2} \times \exp(-\mathrm{tr}\{\boldsymbol{\Upsilon}^{-1}\boldsymbol{W}\}/2), \quad (21)$$

with

$$C = 2^{np}\pi^{p(2p-1)/2}\prod_{i=1}^{2p}\Gamma((n+1-i)/2),$$

where $\Gamma(\cdot)$ denotes the gamma function, $\mathrm{tr}\{\cdot\}$ denotes the trace of a matrix, and $n \geq 2p$. Direct calculation shows that $\mathrm{tr}\{\boldsymbol{\Upsilon}^{-1}\boldsymbol{W}\} = \mathrm{tr}\{\boldsymbol{\Upsilon}^{-1}\dot{\boldsymbol{W}}\}$, [1, p. 396], so also,

$$f(\boldsymbol{W}; n, \boldsymbol{\Upsilon}) = \tfrac{1}{C}(\det\{\boldsymbol{\Upsilon}\})^{-n/2}(\det\{\boldsymbol{W}\})^{(n-2p-1)/2} \times \exp(-\mathrm{tr}\{\boldsymbol{\Upsilon}^{-1}\dot{\boldsymbol{W}}\}/2). \quad (22)$$

The maximum likelihood estimator of $\boldsymbol{\Upsilon}$ is $\boldsymbol{W}/n$, so from (21),

$$\sup_{\boldsymbol{\Upsilon}} f(\boldsymbol{W}; n, \boldsymbol{\Upsilon}) = \tfrac{1}{C}(\det\{\boldsymbol{W}/n\})^{-n/2}(\det\{\boldsymbol{W}\})^{(n-2p-1)/2} \times \exp(-\mathrm{tr}\{n\boldsymbol{I}_{2p}\}/2),$$

Under H_0 the maximum likelihood estimator of $\dot{\boldsymbol{\Upsilon}}$ is $\dot{\boldsymbol{W}}/n$, [2, p. 38], so from (22)

$$\sup_{\dot{\boldsymbol{\Upsilon}}} f(\boldsymbol{W}; n, \boldsymbol{\Upsilon}) = \tfrac{1}{C}(\det\{\dot{\boldsymbol{W}}/n\})^{-n/2}(\det\{\boldsymbol{W}\})^{(n-2p-1)/2} \times \exp(-\mathrm{tr}\{n\boldsymbol{I}_{2p}\}/2).$$

So the generalized likelihood ratio (e.g., [7, p. 419]), L_G say, is

$$L_G = \frac{\sup_{\dot{\boldsymbol{\Upsilon}}} f(\boldsymbol{W}; n, \boldsymbol{\Upsilon})}{\sup_{\boldsymbol{\Upsilon}} f(\boldsymbol{W}; n, \boldsymbol{\Upsilon})} = \left[\frac{\det\{\boldsymbol{W}\}}{\det\{\dot{\boldsymbol{W}}\}}\right]^{n/2}.$$

Then

$$\begin{aligned} L_G^{2/n} &= \frac{\det\{\boldsymbol{W}\}}{\det\{\dot{\boldsymbol{W}}\}} & (23) \\ &= \det\{\dot{\boldsymbol{W}}^{-1}\}\det\{\boldsymbol{W}\} \\ &= \det\{\dot{\boldsymbol{W}}^{-1}\}\det\{\dot{\boldsymbol{W}} + \ddot{\boldsymbol{W}}\} \\ &= \det\{\boldsymbol{I}_{2p} + \dot{\boldsymbol{W}}^{-1}\ddot{\boldsymbol{W}}\} \\ &= \prod_{k=1}^p (1 - \ell_k)\prod_{k=1}^p (1 + \ell_k) \\ &= \prod_{k=1}^p (1 - \ell_k^2), & (24) \end{aligned}$$
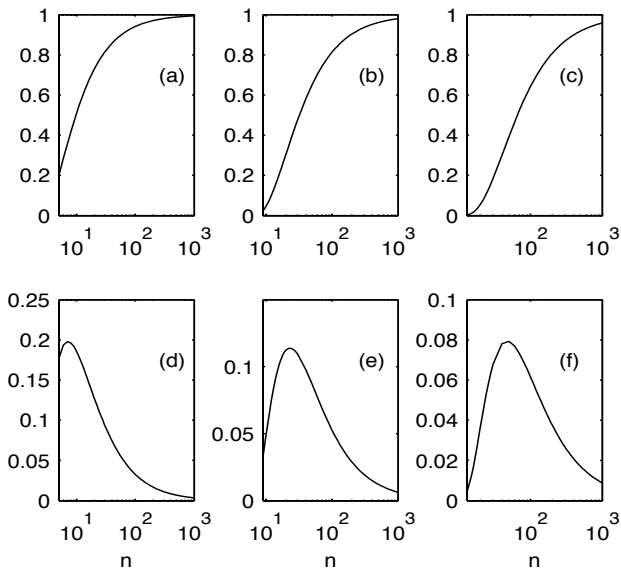
Fig. 1. The mean of the null distribution of T_1 for $p = 2, 4, 6$ in plots (a) – (c), respectively, and the standard deviation of the same in plots (d) – (f), respectively. The x-axis is on a $\log_{10}$ scale and $2p < n \leq 1000$.

Fig. 2. Empirical histograms for the null distribution of T_1 when $p = 4$ for (a) $n = 10$, (b) $n = 25$, (c) $n = 50$, and (d) $n = 400$.

since $\pm \ell_1, \ldots, \pm \ell_p$ are the eigenvalues of $\dot{\boldsymbol{W}}^{-1} \ddot{\boldsymbol{W}}$ and if $\lambda_1, \ldots, \lambda_{2p}$ are eigenvalues of $\boldsymbol{V}$ then $\det\{\boldsymbol{I}_{2p} + \boldsymbol{V}\} = \prod_{i=1}^{2p}(1 + \lambda_i)$.

So we see that the test (19) — which arose via the representation in (13) and (14) — is also a GLR test (the power $2/n$ is merely a monotone transformation). The statistic (24) was also derived in [18] via a complex-valued GLR approach.

IV. Uniformly most powerful test?

The question of the *uniformly* most powerful (UMP) invariant test for impropriety was raised in [18]. In fact no UMP invariant test exists for the problem for $p \geq 2$ [2]. The test (20) is the LMP invariant test, hence is admissible among invariant tests. But for $n > 2p$ it can be shown that (19) is admissible among all tests. Since these tests are different when $p \geq 2$, no UMP invariant test exists in this case.

When $p = 1$ the tests are the same and can be simplified: accept H_0 if $\ell_1 \leq c_3$; this is the UMP invariant test for testing for bivariate sphericity, i.e., $\boldsymbol{\Upsilon} \propto \boldsymbol{I}_2$, a problem with a long history of research [13].

V. A Key Result for Null Distributions of Test Statistics

If the distributions of T_1 and T_2 under the null hypothesis, (for which $\boldsymbol{\Upsilon} = \dot{\boldsymbol{\Upsilon}}$), were different for each known parameter pair (n, p) and unknown covariance matrix $\dot{\boldsymbol{\Upsilon}}$, then a lack of knowledge of the latter would be a major stumbling block for construction of significance levels via simulation.

A very important and useful result when considering the null distribution of both test statistics T_1 and T_2 is [1] that
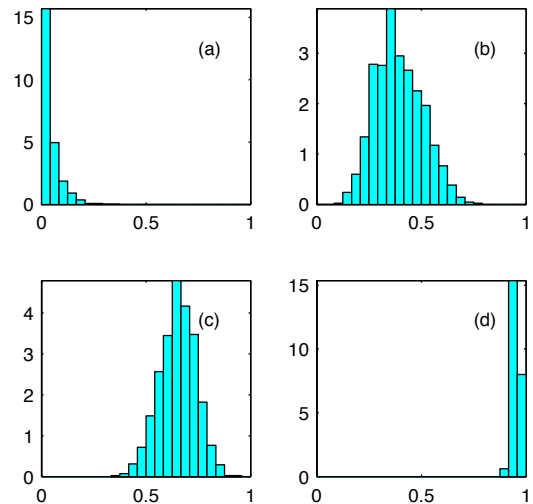
there exists a $\boldsymbol{C} \in \mathcal{C}$ such that $\boldsymbol{CHC}^T = \boldsymbol{I}_{2p}$ for $\boldsymbol{H} \in \mathcal{H}^+$. Under the null hypothesis (7), $\boldsymbol{\Upsilon} = \dot{\boldsymbol{\Upsilon}} \in \mathcal{H}^+$, and so there exists a $\boldsymbol{C} \in \mathcal{C}$ such that $\boldsymbol{C}\dot{\boldsymbol{\Upsilon}}\boldsymbol{C}^T = \boldsymbol{I}_{2p}$, and thus we can take $\dot{\boldsymbol{\Upsilon}} = \boldsymbol{I}_{2p}$ when studying the properties of the test statistics under the null hypothesis. This means that we can calculate quantiles of the null distributions of T_1 and T_2 by simulation using $\dot{\boldsymbol{\Upsilon}} = \boldsymbol{I}_{2p}$, and the calculated values will be invariant for all other $\dot{\boldsymbol{\Upsilon}}$ for the same (n, p).

VI. Distributional properties of statistics

A. Statistic T_1

Under the null hypothesis the statistic T_1 in (19) has known moments [2]. After some reduction the rth moment $E\{T_1^r\}$ may be written as

$$2^{2rp} \frac{\prod_{j=1}^{p} \Gamma(n - j + 1) \prod_{j=1}^{2p} \Gamma((2r + n - j + 1)/2)}{\prod_{j=1}^{2p} \Gamma((n - j + 1)/2) \prod_{j=1}^{p} \Gamma(2r + n - j + 1)}, \quad (25)$$

where $\Gamma(\cdot)$ is the gamma function. We notice that the rth moment depends only on (n, p), as would follow from the discussion in Section V.

Fig. 1 shows the mean and standard deviation of T_1 for $p = 2, 4, 6$ and $2p \leq n \leq 1000$. We see that for all p values the mean monotonically increases with n from 0 to 1, but the standard deviation firstly increases with n and then decreases. This suggests that the probability density function of T_1 is firstly squashed tightly close to zero for small values of n, spreads out and moves along the axis as n increases, and then finally is squashed tightly close to 1 for large values of n. This can be readily verified by simulation. With $\boldsymbol{\Upsilon} = \dot{\boldsymbol{\Upsilon}} = \boldsymbol{I}_{2p}$, 1000 independent values of T_1 were generated for $p = 4$ for n values of 10, 25, 50 and 400. The resulting empirical histograms are given in Fig. 2, verifying the postulated behaviour.
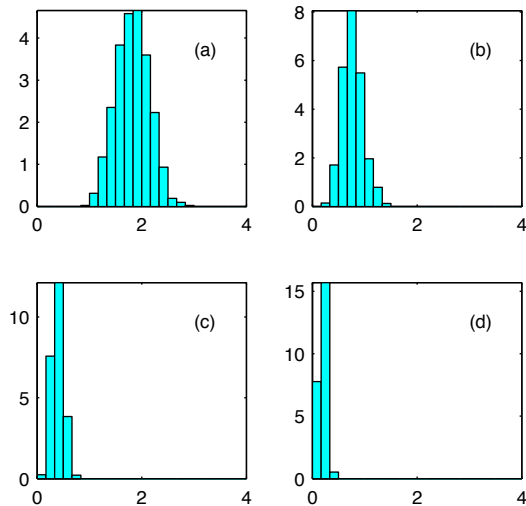
Fig. 3. Empirical histograms for the null distribution of T_2 when $p = 4$ for (a) $n = 10$, (b) $n = 25$, (c) $n = 50$, and (d) $n = 100$.

| p | α | | | n | | |
			20	50	100	1000
2	.01	c_1	0.3892	0.7004	0.8434	0.9833
		c_2	0.6806	0.3111	0.1596	0.0168
	.05	c_1	0.4939	0.7664	0.8797	0.9874
		c_2	0.5477	0.2392	0.1220	0.0125
	.1	c_1	0.5507	0.7998	0.8968	0.9894
		c_2	0.4802	0.2048	0.1042	0.0106
4	.01	c_1	0.0859	0.4355	0.6737	0.9627
		c_2	1.5260	0.6891	0.3587	0.0373
	.05	c_1	0.1282	0.4990	0.7191	0.9689
		c_2	1.3510	0.5903	0.3046	0.0313
	.1	c_1	0.1559	0.5339	0.7422	0.9719
		c_2	1.2546	0.5416	0.2771	0.0284
6	.01	c_1	0.0061	0.2135	0.4897	0.9351
		c_2	2.7021	1.2086	0.6323	0.0659
	.05	c_1	0.0113	0.2576	0.5342	0.9429
		c_2	2.4962	1.0893	0.5636	0.0582
	.1	c_1	0.0154	0.2834	0.5579	0.9469
		c_2	2.3875	1.0256	0.5263	0.0541

TABLE I

CRITICAL VALUES $c_{1,n,p,\alpha}$ AND $c_{2,n,p,\alpha}$, FOR T_1 AND T_2, RESPECTIVELY, FOR $\alpha = 0.01, 0.05$ AND 0.1.

B. Statistic T_2

An analytic form for the moments of T_2 is not known. 1000 independent values of T_2 were generated for $p = 4$ for n values of 10, 25, 50 and 100. The resulting empirical histograms are given in Fig. 3, which shows that the probability density function of T_2 is firstly spread out away from zero, and moves along the axis towards zero, with decreasing spread, as n increases, finally being squashed tightly close to 0 for large values of n. (T_2 is upper-bounded by p since each contributing eigenvalue is upper-bounded by unity.) So T_2 behaves in a more traditional way, having a decreasing variance with increasing number of samples n. This was verified for $p = 2$ and 6 also.

VII. CRITICAL REGIONS FOR TESTS

A. Simulation

Given a specified size of the test α we need to find critical regions $[0, c_{1,n,p,\alpha}]$ for $T_1(n,p)$ and $[c_{2,n,p,\alpha}, p]$ for $T_2(n,p)$ such that the null hypothesis is rejected if the sample value of the statistic falls in the critical region. So we need to find the critical values $c_{1,n,p,\alpha}$ and $c_{2,n,p,\alpha}$.

The critical values were found as follows. For a given (n,p) combination, n independent $2p$-dimensional random vectors from a normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Upsilon} = \boldsymbol{I}_{2p}$, were generated. Then $\boldsymbol{W}$, $\dot{\boldsymbol{W}}$ and $\ddot{\boldsymbol{W}} = \boldsymbol{W} - \dot{\boldsymbol{W}}$ were calculated, and $T_1(n,p) = \prod_{k=1}^{p}(1 - \ell_k^2)$ calculated from $\det\{\boldsymbol{W}\}/\det\{\dot{\boldsymbol{W}}\}$ as in (23), and $T_2(n,p) = \sum_{k=1}^{p} \ell_k^2$ from $\frac{1}{2}\mathrm{tr}(\dot{\boldsymbol{W}}^{-1}\ddot{\boldsymbol{W}}\dot{\boldsymbol{W}}^{-1}\ddot{\boldsymbol{W}})$ as in (20). This was repeated $30\,000$ times to produce smooth empirical cumulative distribution functions, $\Pr(T_1 \leq t_1)$, for T_1, and $\Pr(T_2 \leq t_2)$, for T_2. From these, the critical values $c_{1,n,p,\alpha}$ and $c_{2,n,p,\alpha}$, respectively, were "looked-up." Results for some (n,p) combinations are given in Table 1.

Because of the particular form of the moments for T_1 we can use an approximation method due to Box [3, eqn. 70] to much more easily derive the critical values for a simple transformation of T_1 for (n,p) combinations. In the following we derive Box's approximation for T_1 and examine its accuracy by reference to exact (simulated) critical values.

B. Box's χ^2 Approach For T_1

Box [3, eqn. 70] considers a statistic which has its rth moment of the form

$$C_0 \left[\frac{\prod_{j=1}^{k} y_j^{y_j}}{\prod_{i=1}^{m} x_i^{x_i}} \right]^r \frac{\prod_{i=1}^{m} \Gamma(x_i[1+r] + \xi_i)}{\prod_{j=1}^{k} \Gamma(y_j[1+r] + \eta_j)}, \qquad (26)$$

where $\sum_{i=1}^{m} x_i = \sum_{j=1}^{k} y_j$, and the constant C_0 is obtained by putting $r = 0$ and taking the reciprocal. We can make (25) equal to (26) if we set $m = 2p, k = p, x_i = 1, y_j = 2, \xi_i = (n - i - 1)/2, \eta_j = n - j - 1$, with

$$\left[\frac{\prod_{j=1}^{k} y_j^{y_j}}{\prod_{i=1}^{m} x_i^{x_i}} \right]^r = 2^{2rp} \quad \text{and} \quad c_0 = \frac{\prod_{j=1}^{p} \Gamma(n - j + 1)}{\prod_{j=1}^{2p} \Gamma((n - j + 1)/2)}.$$

We note $\sum_{i=1}^{m} x_i = \sum_{j=1}^{k} y_j = 2p$. The degrees of freedom f associated with the χ^2 approximation are given by [3,

eqn. 73]

$$f = -2\left[\sum_{i=1}^{m}\xi_i - \sum_{j=1}^{k}\eta_j - \tfrac{1}{2}(m-k)\right]$$

$$= -2\left[\sum_{i=1}^{2p}(n-i-1)/2 - \sum_{j=1}^{p}(n-j-1) - \tfrac{1}{2}(2p-p)\right]$$

$$= p(p+1). \tag{27}$$

A matrix in $\mathcal{S}^+$ is $2p \times 2p$ symmetric, and therefore has $\frac{1}{2}(4p^2 - 2p) + 2p = 2p^2 + p$ free parameters. A matrix in $\mathcal{H}^+$ of form (3) has $\frac{1}{2}(p^2 - p) + p = \frac{1}{2}p^2 + \frac{1}{2}p$ free parameters for the symmetric matrix $\boldsymbol{A}_1$ and $\frac{1}{2}(p^2 - p)$ free parameters for the skew-symmetric matrix $\boldsymbol{A}_2$ (since the diagonal must be zeros), a total of p^2 free parameters. The difference is thus $2p^2 + p - p^2 = p(p+1)$. Hence f is the difference between the number of free parameters under H_0 and H_1, as would be expected.

Box approximates $T_1' = -2\log T_1/c_B$ as χ_f^2, where c_B is a constant determined as follows. We compute $a_1 = 2\omega_1/f$ and $a_2 = 4\omega_2/f$ where

$$\omega_l = \frac{(-1)^{l+1}}{l(l+1)}\Big[\sum_{i=1}^{2p}B_{l+1}((n-i-1)/2)$$

$$- \frac{1}{2^l}\sum_{j=1}^{p}B_{l+1}(n-j-1)\Big], \tag{28}$$

where $B_l(x)$ is the Bernoulli polynomial of degree l and order unity, with

$$B_2(x) = x^2 - x + \tfrac{1}{6}; \quad B_3(x) = x^3 - \tfrac{3}{2}x^2 + \tfrac{1}{2}x.$$

For our problem we find $a_2 \geq a_1^2$ for all combinations of (n, p) so we set $c_B = 1/(1 - a_1)$, as determined by [3, p. 328].

Some tedious algebra shows that ω_1 in (28) can be rewritten in the much simpler form

$$\omega_1 = \tfrac{1}{4}[p^3 - (n-3)p^2 - (n-2)p].$$

Then with $f = p(p+1)$, as in (27),

$$c_B = \frac{f}{f - 2\omega_1} = \frac{2}{n-p},$$

so that Box's χ_f^2 approximation takes the form

$$T_1' = -2\log T_1/c_B = -(n-p)\log T_1 \stackrel{d}{=} \chi_f^2, \tag{29}$$

i.e., $-(n-p)\log T_1$ has the distribution of a chi-square random variable with f degrees of freedom. So the test may be carried out simply as: accept H_0 iff

$$-(n-p)\log T_1 < \chi_f^2(1-\alpha), \tag{30}$$

where $\chi_f^2(1-\alpha)$ is the $100(1-\alpha)\%$ point of the chi-square distribution with f degrees of freedom.
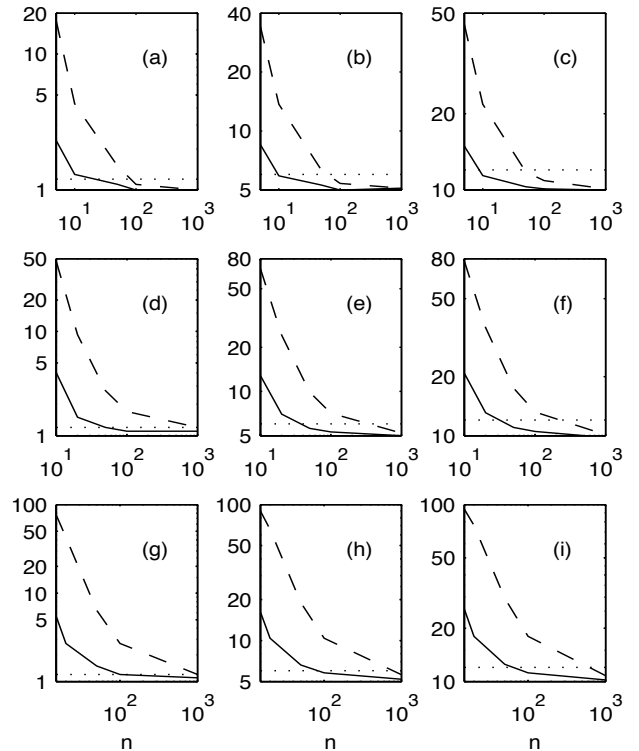

Fig. 4. Comparison of χ_f^2 approximations for transformed T_1, on a $\log n$ versus $\log 100P$ scale, where P is either P_B (Box), shown as solid line, or P_W (Wilks), shown as dashed line. (a)-(c) are for $p = 2$ and $\alpha = 0.01, 0.05$ and 0.1, respectively. (d)-(f) are likewise for $p = 4$, and plots (g)-(i) are for $p = 6$. The horizontal dotted line delineates a +20% error on the nominal level (the latter corresponding to the x-axis).

Returning to (24) we see that $L_G = T_1^{n/2}$ so that $-2\log L_G = -n\log T_1$. Now Wilk's theorem [21, p. 132] says that under H_0, as $n \to \infty$,

$$-2\log L_G = -n\log T_1 \stackrel{d}{\to} \chi_f^2 \tag{31}$$

i.e., $-n\log T_1$ tends in distribution to χ_f^2.

Comparing (29) and (31) we see that both agree as $n \to \infty$ with p fixed. However, our derived result (29) should give better results for smaller n than the purely asymptotic result (31).

The smooth empirical cumulative distribution function, $\Pr(T_1 \leq t_1)$, produced as described in Section VII-A, was used to "look-up" $P_B = \Pr(-(n-p)\log T_1 > \chi_f^2(1-\alpha))$, for (29) and $P_W = \Pr(-n\log T_1 > \chi_f^2(1-\alpha))$, for (31); the desired result is of course α. The results are given in Fig. 4. For clarity, all plots are presented on a $\log n$ versus $\log 100P$ scale, where P is either P_B or P_W. Plots (a)-(c) are for $p = 2$ and $\alpha = 0.01, 0.05$ and 0.1, respectively. Plots (d)-(f) are likewise for $p = 4$, and plots (g)-(i) are for $p = 6$. In each plot P_B is the solid line, and P_W the dashed line. While the ideal result is the horizontal x-axis, the horizontal dotted line in each plot delineates a +20% error on the nominal level. (For example, if the size of the test was 1.2% instead of 1%, or 6% instead of 5%, we

would probably find this quite acceptable for most practical purposes.)

We can see immediately that an acceptable level of accuracy is reached for Box's approximation (29) for an n about $1/10$ as large as required for Wilk's asymptotic formula (31). As p increases both approaches need larger samples for the same accuracy. For Box's method, for good accuracy, the required n increases from about 10 to 50 to 100 as p increases from 2 to 4 to 6. With reference to Fig. 1 we note that the required n seems to be on the "decreasing with n" side of the standard deviation curves.

Simulations for (n, p) combinations for $n \geq 100, p \geq 4$ were very time consuming, so Box's approximation would seem very useful in these cases. It does not work so well for small n's as p increases, but, fortunately, for these cases the critical values can be obtained by simulation in the order of minutes.

VIII. Power

To address the questions raised at the end of Section III-D regarding power of the tests, we need to be able to generate improper complex signals such that the positive eigenvalues $\lambda_1, \ldots, \lambda_p$ of $\dot{\boldsymbol{\Upsilon}}^{-1}\ddot{\boldsymbol{\Upsilon}}$ can be controlled so that we can change not only the "distance of alternatives from the null," but also the relative sizes of the eigenvalues.

Consider the simulation of a vector $[Z_1, \ldots, Z_p]^T$ with the same second-order structure (covariance and relation matrix) as a sample of length p from the complex second-order stationary autoregressive process $\{Z_t\}$ of order p, denoted CAR(p),

$$Z_t = \sum_{j=1}^{p} \phi_j Z_{t-j} + \zeta_t, \tag{32}$$

where the $\{\phi_j\}$ are complex-valued parameters and $\{\zeta_t\}$ is doubly white noise, [12]

$$\gamma_{\zeta,\tau} = \text{cov}\{\zeta_{t+\tau}, \zeta_t\} = E\{\zeta_{t+\tau}\zeta_t^*\} = \sigma_\zeta^2 \delta_{\tau,0}$$
$$r_{\zeta,\tau} = \text{rel}\{\zeta_{t+\tau}, \zeta_t\} = E\{\zeta_{t+\tau}\zeta_t\} = r_\zeta \delta_{\tau,0},$$

$\delta_{j,k}$ denotes the Kronecker delta, and rel$\{\cdot\}$ denotes the relation between $\zeta_{t+\tau}$ and ζ_t. The roots of the z-polynomial $\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ are outside the unit circle, the minimum-phase case, and values of σ_ζ^2 and r_ζ, must satisfy $|r_\zeta| < \sigma_\zeta^2$, [12]. Let $\gamma_\tau = \text{cov}\{Z_{t+\tau}, Z_t\}$ and $r_\tau = \text{rel}\{Z_{t+\tau}, Z_t\}$. Given the roots of $\Phi(z)$ we firstly find [15], $\gamma_0, \ldots, \gamma_{p-1}$, proportional to σ_ζ^2, and $r_0, \ldots, r_{p-1}$, proportional to r_ζ. Then

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_0 & \gamma_1^* & \gamma_2^* & \cdots & \gamma_{p-1}^* \\ \gamma_1 & \gamma_0 & \gamma_1^* & \cdots & \gamma_{p-2}^* \\ \vdots & \vdots & & \vdots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{bmatrix},$$

and

$$\boldsymbol{R} = \begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & r_0 & r_1 & \cdots & r_{p-2} \\ \vdots & \vdots & & \vdots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & r_0 \end{bmatrix},$$

so $\boldsymbol{G}$ follows from (1).

The covariance matrix of $\check{\boldsymbol{X}}$ is given by $\boldsymbol{\Upsilon} = \boldsymbol{M}\boldsymbol{G}\boldsymbol{M}^H$. We can calculate a Cholesky decomposition of $\boldsymbol{\Upsilon}$ such that $\boldsymbol{\Upsilon} = \boldsymbol{V}\boldsymbol{V}^H$. If we let $\boldsymbol{U} = [U_1, \ldots, U_p]^T$ where the U_k's are independent standard normal random variables, then if $\boldsymbol{Q} = \boldsymbol{V}\boldsymbol{U}$, we know that the covariance matrix of $\boldsymbol{Q}$ is given by $\boldsymbol{\Upsilon}$ and so to obtain $Z_1, \ldots, Z_p$ with the correct second-order (covariance and relation) structure we simply compute $\boldsymbol{M}^{-1}\boldsymbol{Q}$, and take the first p entries.

With (15) in mind we can measure the "distance of alternatives from the null" by the Euclidean norm $\|\lambda\|_2 = (\sum_{l=1}^{p} \lambda_l^2)^{1/2}$ of the vector $[\lambda_1, \ldots, \lambda_p]$ of positive eigenvalues of $\dot{\boldsymbol{\Upsilon}}^{-1}\ddot{\boldsymbol{\Upsilon}}$.

Consider the case $p = 2$. Let β_1 and β_2 be the roots of $\Phi(z) = 1 - \phi_1 z - \phi_2 z^2$. Then $\phi_1 = (1/\beta_1) + (1/\beta_2)$ and $\phi_2 = -1/(\beta_1\beta_2)$. Let β_1 and β_2 be $\beta_1 = |\beta_1|\exp(\text{i}\pi\theta_1)$, $\beta_2 = |\beta_2|\exp(-\text{i}\pi\theta_1)$, and let $r_\zeta = |r_\zeta|\exp(-\text{i}\pi\theta_2)$. (We took $\theta_1 = 1/5$ and $\theta_2 = 1/3$.)

We set $\sigma_\zeta^2 = 1$ and as $|r_\zeta| \to \sigma_\zeta^2$, from zero, the norm $\|\lambda\|_2$ increases. Furthermore, if we take $|\beta_1| \approx |\beta_2|$ (both greater than unity), then $\lambda_1 \approx \lambda_2$, while if we take, say, $|\beta_2| >> |\beta_1|$, then $\lambda_1 >> \lambda_2$, where as usual λ_1 is the largest eigenvalue. So we are able to control not only the "distance of alternatives from the null" by changing $|r_\zeta|$, but also examine whether disparately-sized eigenvalues affect the behaviour of the LMP test, by changing $|\beta_2|/|\beta_1|$. Model A has $|\beta_1| = 1.1$ and $|\beta_2| = 1.2$, so the eigenvalues are nearly equal, and model B has $|\beta_1| = 1.1$ and $|\beta_2| = 20$, for which $\lambda_1 \approx 6\lambda_2$. For $|r_\zeta| = 0$ we know the power is simply α, so we looked at $|r_\zeta| = 0.1, 0.2, \ldots, 1.0$. n was set to 50.

Each choice of $|\beta_1|, |\beta_2|, |r_\zeta|$ corresponds to a different $\boldsymbol{\Upsilon}$. We estimated $\Pr\left[\prod_{k=1}^{p}(1 - \ell_k^2) < c_1 | \boldsymbol{\Upsilon}\right]$, and $\Pr\left[\sum_{k=1}^{p} \ell_k^2 > c_2, | \boldsymbol{\Upsilon}\right]$, from 5000 independent repetitions.

The results for $p = 2$ are shown in Fig. 5, where the power for T_1 is shown by the solid line, and T_2 by the dotted line. Results for model A are given for 100α of 1, and 5, in Figs. 5(a) and (b), respectively, and for model B likewise in Figs. 5(c) and (d). For model A the LMP test marginally outperforms the GLR test over all alternatives (as measured by the norm $\|\lambda\|_2$ as $|r_\zeta|$ ranges from 0 to 1). For model B the LMP test is slightly superior only for small alternative distances, but for larger ones it is slightly inferior. We also looked at sample sizes of $n = 1000$; here the power plots reach unity for a norm as small as 0.2, but again the two tests are barely distinguishable in terms of power.

Finally we considered $p = 6$. Roots were found for a model C giving a range of positive eigenvalues with the smallest two being 1/5th and 1/200th of the magnitude of the largest. Again the norm $\|\lambda\|_2$ was found as $|r_\zeta|$ ranged from 0 to 1. Power results are given in Fig. 6 for 100α of 1 and 5 in Figs. 6(a) and (b), respectively, when $n = 50$, and likewise in Figs. 6(c) and (d), when $n = 1000$. When $n = 50$ the LMP test outperforms the GLR test, but the difference is again small, while when $n = 1000$ the difference in power of the two tests is negligible.
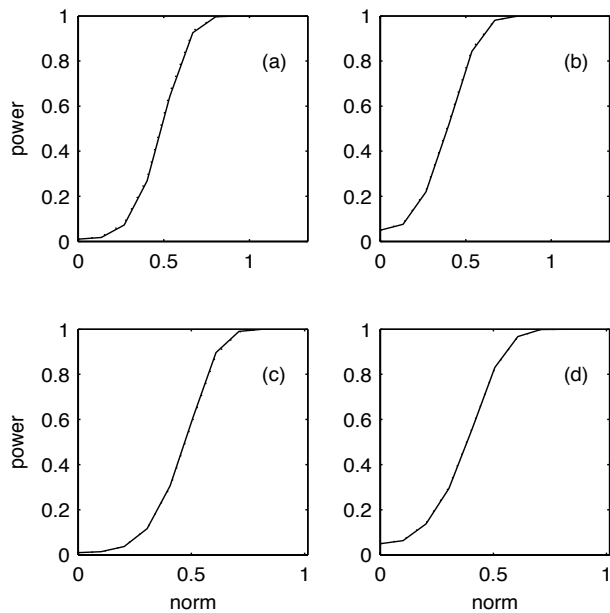
Fig. 5. Power of the test (19) shown as solid line, and (20) shown as dotted line, for $n = 50$ and $p = 2$. (The two are barely distinguishable.) (a) and (b) are for $\alpha = 0.01$ and 0.05, respectively, for model A. (c) and (d) are likewise for model B.
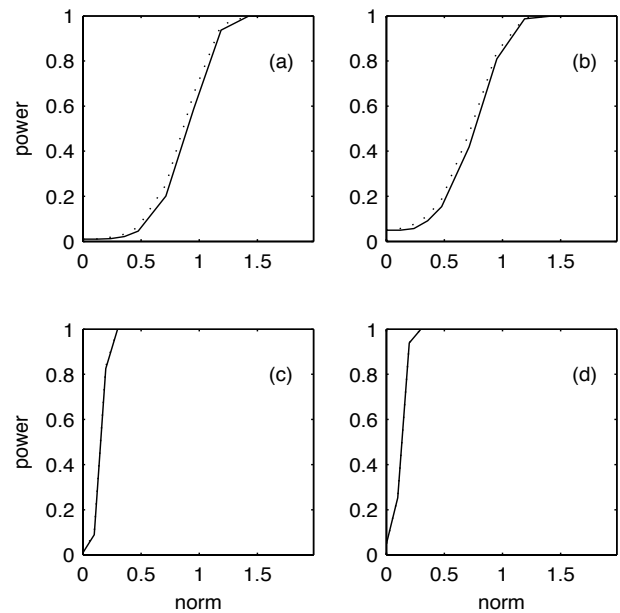

Fig. 6. Power of the test (19) shown as solid line, and (20) shown as dotted line, for $p = 6$ and model C. (a) and (b) are for $\alpha = 0.01$ and 0.05, respectively, for $n = 50$ and (c) and (d) are likewise for $n = 1000$.

IX. Concluding Comments

We have carried out a thorough study of two statistics for testing whether a complex vector $\boldsymbol{Z}$ is proper by using results based on the $2p \times 2p$ covariance matrix $\boldsymbol{\Upsilon} = \text{cov}\{\check{\boldsymbol{X}}\} = E\{\check{\boldsymbol{X}}\check{\boldsymbol{X}}^T\}$ of the real-valued random vector $\check{\boldsymbol{X}} = [\boldsymbol{X}^T, \boldsymbol{Y}^T]^T = [X_1, \ldots, X_p, Y_1, \ldots, Y_p]^T$, where $\boldsymbol{X} = \text{Re}\{\boldsymbol{Z}\}, \boldsymbol{Y} = \text{Im}\{\boldsymbol{Z}\}$. The test statistics are functions of the sample canonical correlations between $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$; it has been shown that the true canonical correlations are the same as those derived in [18].

We have been able to draw on a number of useful results in [1] and [2] enabling us to empirically study the null distributions of the test statistics, and obtain critical values. Further we were able to derive a Box approximation, namely that the distribution of $-(n-p)\log T_1$ can be taken to be a χ_f^2 distribution, for many (n, p) combinations.

Evidence provided by a power study leads us to conclude that T_1 and T_2 behave very similarly in terms of quality. In practice a good strategy seems to be to use (19) combined with the Box approximation, (for suitable (n, p) combinations as discussed in Section VII-B), enabling rapid and simple calculation of critical values. For combinations of (n, p) unsuitable for Box's approximation, the patterns of critical values in Table I may enable a conclusion to be drawn, but failing that a simulation study, along the lines described in Section VII-A will be necessary to determine critical values. In any event, the results in this paper mean that testing for propriety can be a practical and undaunting procedure.

References

[1] S. A. Andersson, H. K. Brøns and S. T. Jensen, "Distribution of eigenvalues in multivariate statistical analysis," *The Annals of Statistics*, vol. 11, pp. 392–415, 1983.

[2] S. A. Andersson and M. D. Perlman, "Two testing problems relating the real and complex multivariate normal distribution," *J. Multivariate Analysis*, vol. 15, pp. 21–51, 1984.

[3] G. E. P. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, vol. 36, 317–46, 1949.

[4] D. R. Brillinger, *Time Series: Data Analysis and Theory, Expanded Ed.* New York: McGraw-Hill, Inc., 1981.

[5] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics, Vol. 1, 4th Ed.* London: Charles Griffin, 1977.

[6] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses, 2nd Ed.* New York; Springer, 2005.

[7] A. M. Mood, F. A. Graybill and D. C. Boes, *Introduction to the Theory of Statistics, 3rd Ed.* New York: McGraw-Hill.

[8] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Information Theory*, vol. 39, pp. 1293–1302, 1993.

[9] S. C. Olhede and A. T. Walden, "Noise reduction in directional signals using multiple Morse wavelets illustrated on quadrature Doppler ultrasound," *IEEE Trans. Biomed. Eng.*, vol. 50, 51–7, 2003.

[10] E. Ollila and V. Koivunen, "Generalized complex elliptical distributions," in *Proc. Third Sensor Array and Multichannel Signal Processing Workshop*, Sitges, Spain, July, 2004, pp. 460–4.

[11] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Trans. Signal Process.*, vol. 44, pp. 2637–40, 1996.

[12] B. Picinbono and P. Bondon, "Second-order statistics of complex signals," *IEEE Trans. Signal Process.*, vol. 45, pp. 411–420, 1997.

[13] K. C. S. Pillai and B. N. Nagarsenker, "On the distribution of the sphericity test criterion in classical and complex normal pop-

ulations having unknown covariance matrices," *Annals of Mathematical Statistics*, vol. 42, pp. 764–7, 1971.

[14] D. B. Rowe, "Modeling both the magnitude and phase of complex-valued fMRI data," *Neuroimage*, vol. 25, 1310–24, 2005.

[15] P. Rubin-Delanchy and A. T. Walden, "Kinematics of complex-valued time series," *IEEE Trans. Signal Process.*, to appear, 2008.

[16] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Trans. Signal Process.*, vol. 51, pp. 714–25, 2003.

[17] P. J. Schreier, L. L. Scharf and A. Hanssen, "A statistical test for impropriety of complex random signals," in *Proc. Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Toulouse, France, May 2006, vol. III.

[18] P. J. Schreier, L. L. Scharf and A. Hanssen, "A generalized likelihood ratio test for impropriety of complex signals," *IEEE Signal Process. Lett.*, vol. 13, pp. 433–6, 2006.

[19] M. T. Taner, F. Koehler and R. E. Sheriff, "Complex seismic trace analysis," *Geophysics*, vol. 44, 1041–63, 1979.

[20] A. van den Bos, "The multivariate complex normal distribution — a generalization.," *IEEE Trans. Information Theory*, vol. 41, pp. 537–9, 1995.

[21] G. A. Young and R. L.Smith, *Essentials of Statistical Inference.* Cambridge UK: Cambridge University Press, 2005.