

Testing for Quaternion Propriety

P. Ginzberg and A. T. Walden, *Senior Member, IEEE*

Copyright (c) 2011 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Paul Ginzberg and Andrew Walden are both at the Department of Mathematics, Imperial College London, 180 Queen's Gate, London SW7 2BZ, UK. (e-mail: paul.ginzberg05@imperial.ac.uk and a.walden@imperial.ac.uk)

Abstract

We consider the problem of testing whether a quaternion-valued Gaussian random vector is proper. The quaternion covariance matrix fully describes the second order properties of a quaternion random vector only if the distribution is proper. The exact distribution of the likelihood ratio test under the hypothesis of propriety is derived for general sample size, N , and vector dimensionality, p . As this is in terms of Meijer's G-function, various approximations are considered, including Box-type and saddlepoint approximations. We find in particular that a new approach matching the first three cumulants is easy to implement and extremely accurate.

Keywords

quaternion random vector, structured covariance, likelihood ratio test, improperness, Meijer's G-function

I. INTRODUCTION

Recently there has been much interest in quaternion-valued signal processing [5], [10], [15], [16], [17], [20], [22], [23], [24], [28]. Many multivariate problems based on vector sensors may be readily modeled in the quaternion domain which accounts naturally for the correlated nature of the signal components. Applications of quaternion signal processing include multicomponent seismic velocity analysis [5], singular value decomposition for vector sensing [10], sequential best rotation [15] for polarized source separation in convolutive mixture models, multiple signal classification (MUSIC) spectrum estimation [17], block coding [20], [32], and adaptive filtering [22].

As is the case with complex-valued random vectors, the quaternion covariance only fully captures the second order properties of *proper* quaternion random vectors. In the more general improper case, further information is contained in three complementary covariance matrices. It is shown in [28], [29] that using quaternion *linear* processing for partial least squares, principal component analysis, multivariate linear regression or canonical correlation analysis is suboptimal for improper quaternion random vectors. For the same reason the quaternion widely linear adaptive filter is introduced in [23] and incorporated into their quaternion least mean square algorithm, allowing for a unified approach to adaptive filtering of both proper and improper quaternion signals.

It is hence timely to consider the problem of testing whether a quaternion-valued Gaussian random vector is proper. [27] studies the likelihood ratio test (LRT) from a geometric and

information-theoretic perspective. We will instead consider structured real covariance matrices and concentrate on the exact distribution of the test statistic and its accurate approximation.

In the remainder of this Introduction we define mathematically the problem we wish to solve. In Section II we firstly provide some basic definitions and introduce notation, along with the augmented covariance matrix. Following [28] we define propriety for quaternions. In Section III the testing problem is developed in terms of two nested hypotheses. The LRT statistic and its moments follow. The Box-type parameters are identified. The asymptotic and exact distributions of the statistic are derived in Section III, the latter involving Meijer's G-functions. As this is inconvenient in practice, Section V examines various approximations, including Box-type and saddlepoint approximations. Section V-C introduces a new approach matching the first three cumulants which is easy to implement and extremely accurate. Section VI considers asymptotic robustness to nonGaussianity. Section VII provides our concluding comments.

A. The problem

Let $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ be real-valued p -dimensional jointly Gaussian random variables (column vectors). Then $\mathbf{q} = \mathbf{a} + \mathbf{b}i + \mathbf{c}j + \mathbf{d}k$ follows a (possibly improper) p -dimensional quaternion Gaussian distribution. Here $1, i, j, k$ are the 4 basis elements of the quaternions (see Section II-A). Studying the distribution of $\mathbf{q}$ is equivalent to studying the distribution of the $4p$ -dimensional real Gaussian random variable $\mathbf{r} = [\mathbf{a}^T, \mathbf{b}^T, \mathbf{c}^T, \mathbf{d}^T]^T$ with covariance matrix $\mathbf{\Sigma}_{\mathbf{r}} = E\{\mathbf{r}\mathbf{r}^T\}$. We are assuming throughout that $E\{\mathbf{r}\} = \mathbf{0}$, but this can be relaxed by subtracting the sample mean and replacing N by $N - 1$ in all results.

In this paper we consider the hypothesis test H_0 : the $4p \times 4p$ covariance matrix $\mathbf{\Sigma}_{\mathbf{r}}$ from a real-valued multivariate Gaussian distribution has quaternion structure, defined in (8), against the default alternative H_1 : $\mathbf{\Sigma}_{\mathbf{r}}$ does not have quaternion structure.

Just as in the complex-valued case, where testing that an appropriate $2p \times 2p$ covariance matrix has complex structure, defined in (10), is equivalent to testing for propriety for p -dimensional complex-valued Gaussian vectors [30], so testing that the $4p \times 4p$ covariance matrix $\mathbf{\Sigma}_{\mathbf{r}}$ has quaternion structure is equivalent to testing for propriety for p -dimensional quaternion-valued Gaussian vectors $\mathbf{q}$.

II. BACKGROUND

A. Basics

The quaternions $\mathbb{H}$ are a 4-dimensional real division algebra. We denote the 4 basis elements by 1, i, j, k. In this basis, a quaternion can be written as $q = a + bi + cj + dk$ with $a, b, c, d \in \mathbb{R}$. i, j, k satisfy $i^2 = j^2 = k^2 = ijk = -1$, $ij = -ji = k$, $jk = -kj = i$, $ki = -ik = j$ and quaternion multiplication is associative, distributive but not commutative.

The conjugate of q is $q^* = a - bi - cj - dk$. The quaternion norm is defined as $|q| = \sqrt{qq^*} = [a^2 + b^2 + c^2 + d^2]^{1/2}$. i_0 is a pure unit quaternion iff $|i_0| = 1$ and its real part is zero. The inverse of a quaternion $q \neq 0$ is $q^{-1} = q^*/|q|^2$. For a quaternion matrix $\mathbf{Q}$ we denote the conjugate transpose by $\mathbf{Q}^H = (\mathbf{Q}^*)^T$. The product of two matrices $\mathbf{Q} \in \mathbb{H}^{m \times n}$ and $\mathbf{S} \in \mathbb{H}^{n \times p}$ is $\mathbf{P} = \mathbf{Q}\mathbf{S}$ with $(i, j)^{\text{th}}$ element $p_{i,j} = \sum_{\ell=1}^n q_{i,\ell} s_{\ell,j}$.

B. Covariance matrices

Analogously to improper complex random variables, the second-order properties of $\mathbf{q}$, given by the $4p \times 4p$ real covariance matrix

$$\Sigma_{\mathbf{r}} = E\{\mathbf{r}\mathbf{r}^T\} = \begin{bmatrix} \Sigma_{\mathbf{a}} & \Sigma_{\mathbf{a},\mathbf{b}} & \Sigma_{\mathbf{a},\mathbf{c}} & \Sigma_{\mathbf{a},\mathbf{d}} \\ \Sigma_{\mathbf{b},\mathbf{a}} & \Sigma_{\mathbf{b}} & \Sigma_{\mathbf{b},\mathbf{c}} & \Sigma_{\mathbf{b},\mathbf{d}} \\ \Sigma_{\mathbf{c},\mathbf{a}} & \Sigma_{\mathbf{c},\mathbf{b}} & \Sigma_{\mathbf{c}} & \Sigma_{\mathbf{c},\mathbf{d}} \\ \Sigma_{\mathbf{d},\mathbf{a}} & \Sigma_{\mathbf{d},\mathbf{b}} & \Sigma_{\mathbf{d},\mathbf{c}} & \Sigma_{\mathbf{d}} \end{bmatrix}, \quad (1)$$

($\Sigma_{\mathbf{a},\mathbf{b}} = E\{\mathbf{a}\mathbf{b}^T\}$ and $\Sigma_{\mathbf{a}} = E\{\mathbf{a}\mathbf{a}^T\}$), are not fully specified by its (quaternion) covariance matrix $\Sigma_{\mathbf{q}} = E\{\mathbf{q}\mathbf{q}^H\}$.

Given a quaternion q_0 and a pure unit quaternion i_0 define the involution $q_0^{(i_0)} = i_0 q_0 i_0^{-1} = -i_0 q_0 i_0$.

Note that i, j and k are all pure unit quaternions. We define an augmented quaternion vector as

$$\underline{\mathbf{q}} = \mathbf{A}\mathbf{r} = \begin{bmatrix} \mathbf{q} \\ \mathbf{q}^{(i)} \\ \mathbf{q}^{(j)} \\ \mathbf{q}^{(k)} \end{bmatrix} = \begin{bmatrix} \mathbf{q} \\ -i\mathbf{q}i \\ -j\mathbf{q}j \\ -k\mathbf{q}k \end{bmatrix} = \begin{bmatrix} \mathbf{a} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} \\ \mathbf{a} + b\mathbf{i} - c\mathbf{j} - d\mathbf{k} \\ \mathbf{a} - b\mathbf{i} + c\mathbf{j} - d\mathbf{k} \\ \mathbf{a} - b\mathbf{i} - c\mathbf{j} + d\mathbf{k} \end{bmatrix}, \quad (2)$$

where (e.g., [24])

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_p & \mathbf{iI}_p & \mathbf{jI}_p & \mathbf{kI}_p \\ \mathbf{I}_p & \mathbf{iI}_p & -\mathbf{jI}_p & -\mathbf{kI}_p \\ \mathbf{I}_p & -\mathbf{iI}_p & \mathbf{jI}_p & -\mathbf{kI}_p \\ \mathbf{I}_p & -\mathbf{iI}_p & -\mathbf{jI}_p & \mathbf{kI}_p \end{bmatrix}. \quad (3)$$

It is worth noting that if $\mathbf{c} = \mathbf{d} = \mathbf{0}$, so that the quaternion vector is just a complex vector, then $[\mathbf{a}^T, \mathbf{b}^T]^T$ can be taken to the augmented complex vector $[(\mathbf{a} + \mathbf{bi})^T, (\mathbf{a} - \mathbf{bi})^T]^T$ using the well-known submatrix of $\mathbf{A}$ [21, p. 31]:

$$\begin{bmatrix} \mathbf{I}_p & \mathbf{iI}_p \\ \mathbf{I}_p & -\mathbf{iI}_p \end{bmatrix}. \quad (4)$$

The second-order properties of $\mathbf{q}$ are fully specified by its augmented covariance matrix

$$\Sigma_{\underline{\mathbf{q}}} = E\{\underline{\mathbf{q}}\underline{\mathbf{q}}^H\} = \mathbf{A}\Sigma_{\mathbf{r}}\mathbf{A}^H. \quad (5)$$

Indeed, since $\mathbf{A}^{-1} = \frac{1}{4}\mathbf{A}^H$ we can easily recover the real covariance matrix:

$$\Sigma_{\mathbf{r}} = \frac{1}{16}\mathbf{A}^H\Sigma_{\underline{\mathbf{q}}}\mathbf{A}. \quad (6)$$

Expanding (5) in $p \times p$ blocks gives [28, p. 3505]

$$\Sigma_{\underline{\mathbf{q}}} = \begin{bmatrix} \Sigma_{\mathbf{q}} & \Sigma_{\mathbf{q},\mathbf{q}^{(i)}} & \Sigma_{\mathbf{q},\mathbf{q}^{(j)}} & \Sigma_{\mathbf{q},\mathbf{q}^{(k)}} \\ \Sigma_{\mathbf{q},\mathbf{q}^{(i)}}^{(i)} & \Sigma_{\mathbf{q}}^{(i)} & \Sigma_{\mathbf{q},\mathbf{q}^{(k)}}^{(i)} & \Sigma_{\mathbf{q},\mathbf{q}^{(j)}}^{(i)} \\ \Sigma_{\mathbf{q},\mathbf{q}^{(j)}}^{(j)} & \Sigma_{\mathbf{q},\mathbf{q}^{(k)}}^{(j)} & \Sigma_{\mathbf{q}}^{(j)} & \Sigma_{\mathbf{q},\mathbf{q}^{(i)}}^{(j)} \\ \Sigma_{\mathbf{q},\mathbf{q}^{(k)}}^{(k)} & \Sigma_{\mathbf{q},\mathbf{q}^{(j)}}^{(k)} & \Sigma_{\mathbf{q},\mathbf{q}^{(i)}}^{(k)} & \Sigma_{\mathbf{q}}^{(k)} \end{bmatrix} \quad (7)$$

where terms of the form $\Sigma_{\mathbf{q},\mathbf{q}^{(k)}}^{(i)}$ denote involutions, e.g., $\Sigma_{\mathbf{q},\mathbf{q}^{(k)}}^{(i)} = -\mathbf{i}\Sigma_{\mathbf{q},\mathbf{q}^{(k)}}\mathbf{i}$. Since all the blocks in $\Sigma_{\underline{\mathbf{q}}}$ can be derived from the first row of blocks, the second-order properties of $\mathbf{q}$ can be described by specifying the covariance $\Sigma_{\mathbf{q}}$ along with three complementary covariance matrices

$$\begin{aligned} \Sigma_{\mathbf{q},\mathbf{q}^{(i)}} &= -E\{\mathbf{qi}\mathbf{q}^H\mathbf{i}\} \\ \Sigma_{\mathbf{q},\mathbf{q}^{(j)}} &= -E\{\mathbf{qj}\mathbf{q}^H\mathbf{j}\} \\ \Sigma_{\mathbf{q},\mathbf{q}^{(k)}} &= -E\{\mathbf{qk}\mathbf{q}^H\mathbf{k}\}. \end{aligned}$$

C. Propriety

We now give the definitions of propriety following [28].

Definition 1: $\mathbf{q}$ is $\mathbb{H}$ -proper, (or just proper), iff $\Sigma_{\mathbf{q},\mathbf{q}^{(i)}} = 0$, $\Sigma_{\mathbf{q},\mathbf{q}^{(j)}} = 0$ and $\Sigma_{\mathbf{q},\mathbf{q}^{(k)}} = 0$.

Definition 2: $\mathbf{q}$ is $\mathbb{C}^j$ -proper iff $\Sigma_{\mathbf{q},\mathbf{q}^{(i)}} = 0$ and $\Sigma_{\mathbf{q},\mathbf{q}^{(k)}} = 0$.

Proposition 1: $\mathbf{q}$ is $\mathbb{H}$ -proper (proper) iff

$$\Sigma_{\mathbf{r}} = \begin{bmatrix} \Sigma_{\mathbf{a}} & -\Sigma_{\mathbf{b},\mathbf{a}} & -\Sigma_{\mathbf{c},\mathbf{a}} & -\Sigma_{\mathbf{d},\mathbf{a}} \\ \Sigma_{\mathbf{b},\mathbf{a}} & \Sigma_{\mathbf{a}} & -\Sigma_{\mathbf{d},\mathbf{a}} & \Sigma_{\mathbf{c},\mathbf{a}} \\ \Sigma_{\mathbf{c},\mathbf{a}} & \Sigma_{\mathbf{d},\mathbf{a}} & \Sigma_{\mathbf{a}} & -\Sigma_{\mathbf{b},\mathbf{a}} \\ \Sigma_{\mathbf{d},\mathbf{a}} & -\Sigma_{\mathbf{c},\mathbf{a}} & \Sigma_{\mathbf{b},\mathbf{a}} & \Sigma_{\mathbf{a}} \end{bmatrix}. \quad (8)$$

$\mathbf{q}$ is $\mathbb{C}^j$ -proper iff

$$\Sigma_{\mathbf{r}} = \begin{bmatrix} \Sigma_{\mathbf{a}} & \Sigma_{\mathbf{a},\mathbf{b}} & -\Sigma_{\mathbf{c},\mathbf{a}} & -\Sigma_{\mathbf{c},\mathbf{b}} \\ \Sigma_{\mathbf{b},\mathbf{a}} & \Sigma_{\mathbf{b}} & -\Sigma_{\mathbf{d},\mathbf{a}} & -\Sigma_{\mathbf{d},\mathbf{b}} \\ \Sigma_{\mathbf{c},\mathbf{a}} & \Sigma_{\mathbf{c},\mathbf{b}} & \Sigma_{\mathbf{a}} & \Sigma_{\mathbf{a},\mathbf{b}} \\ \Sigma_{\mathbf{d},\mathbf{a}} & \Sigma_{\mathbf{d},\mathbf{b}} & \Sigma_{\mathbf{b},\mathbf{a}} & \Sigma_{\mathbf{b}} \end{bmatrix} \quad (9)$$

$$= \begin{bmatrix} \Sigma_{\text{Re}(\mathbf{z})} & -\Sigma_{\text{Im}(\mathbf{z}),\text{Re}(\mathbf{z})} \\ \Sigma_{\text{Im}(\mathbf{z}),\text{Re}(\mathbf{z})} & \Sigma_{\text{Re}(\mathbf{z})} \end{bmatrix} = \ddot{\Sigma}_{\mathbf{r}}, \text{ say,} \quad (10)$$

where $\mathbf{z} = [\mathbf{a}^T, \mathbf{b}^T]^T + [\mathbf{c}^T, \mathbf{d}^T]^T \mathbf{j}$.

Proof: These follow from (6) and (7). More easily, the form of (8) is immediate from [28, lemma 8] while (9) is readily derived from [28, Table 1]. $\blacksquare$

Quaternion propriety can be interpreted geometrically as invariance of the second-order properties under the transformations (rotations) of the form $\mathbf{q} \mapsto \mathbf{q}u$ where u is a unit quaternion [28, lemma 9]. Invariance under $\mathbf{q} \mapsto u\mathbf{q}$ gives a different ‘handedness’ of quaternion propriety [26]. Studying either type of handedness is however equivalent, since taking conjugates switches the handedness of $\mathbf{q}$.

Algebraically, propriety is preserved by quaternion linear transformations $\mathbf{q} \mapsto \mathbf{L}\mathbf{q}$, $\mathbf{L} \in \mathbb{H}^{m \times p}$ [28, Lemma 1]. Thus, the hypothesis test H_0 v.s. H_1 is invariant under $\mathbf{q} \mapsto \mathbf{L}\mathbf{q}u$ where $\mathbf{L} \in \mathbb{H}^{p \times p}$ is invertible and u is a unit quaternion. In particular, propriety is preserved by all rotations $\mathbf{q} \mapsto u\mathbf{q}v$, $|u| = |v| = 1$.

$\mathbb{C}^j$ -propriety of $\mathbf{q}$ is equivalent to the complex propriety of $\mathbf{z}$ or the joint complex propriety of $\mathbf{a} + \mathbf{c}\mathbf{j}$ and $\mathbf{b} + \mathbf{d}\mathbf{j}$ [28, p. 3507].

Note that if $\mathbf{q}$ is $\mathbb{H}$ -proper then $\Sigma_{\mathbf{q}} = 4\Sigma_{\mathbf{a}} + 4\Sigma_{\mathbf{b},\mathbf{a}}\mathbf{i} + 4\Sigma_{\mathbf{c},\mathbf{a}}\mathbf{j} + 4\Sigma_{\mathbf{d},\mathbf{a}}\mathbf{k}$ and if $\mathbf{z}$ is complex proper then $\Sigma_{\mathbf{z}} = 2\Sigma_{\text{Re}(\mathbf{z})} + 2\Sigma_{\text{Im}(\mathbf{z}),\text{Re}(\mathbf{z})}\mathbf{j}$.

Finally, in the scalar case ($p = 1$) propriety of $\mathbf{q}$ is equivalent to sphericity of $\mathbf{r}$, i.e., $\mathbf{q}$ is proper iff $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ are independent and identically distributed (IID). A LRT for propriety then reduces to Mauchly's test [14].

III. TESTING FOR QUATERNION STRUCTURE

A. Covariance structures

Here we will link the real-valued covariance matrices in (8) and (9) to the work of [1].

Definition 3: Let $P_{\mathbb{R}}$ denote the set of $4p \times 4p$ symmetric positive definite matrices. Let $P_{\mathbb{C}} \subset P_{\mathbb{R}}$ and $P_{\mathbb{H}} \subset P_{\mathbb{R}}$ denote the set of symmetric positive definite matrices of the form (9) (or (10)), and (8), respectively. In the terminology of [1, pp. 394, 403–404], $P_{\mathbb{C}}$ is the set of covariance matrices with complex structure and $P_{\mathbb{H}}$ is the set of covariance matrices with quaternion structure. Clearly $P_{\mathbb{H}} \subset P_{\mathbb{C}}$.

Consider our hypothesis test stated in Section I-A. Because the quaternion structure is a special case of complex structure, we can exploit the nested nature of the implied models to build a direct test from two of the tests discussed in [1]. Consider two hypothesis tests with default alternatives and null hypotheses as follows:

1. Test 1. $H_0^{\mathbb{C}}$: the $4p \times 4p$ covariance matrix $\Sigma_{\mathbf{r}}$ from a real-valued multivariate Gaussian distribution has complex structure. [$H_0^{\mathbb{C}}$: $\Sigma_{\mathbf{r}} \in P_{\mathbb{C}}$ versus $H_1^{\mathbb{C}}$: $\Sigma_{\mathbf{r}} \in P_{\mathbb{R}} \setminus P_{\mathbb{C}}$.]
2. Test 2. $H_0^{\mathbb{H}}$: the $4p \times 4p$ covariance matrix $\Sigma_{\mathbf{r}}$ with complex structure ($\Sigma_{\mathbf{r}} = \ddot{\Sigma}_{\mathbf{r}}$) has quaternion structure. [$H_0^{\mathbb{H}}$: $\Sigma_{\mathbf{r}} \in P_{\mathbb{H}}$ versus $H_1^{\mathbb{H}}$: $\Sigma_{\mathbf{r}} \in P_{\mathbb{C}} \setminus P_{\mathbb{H}}$.]

Since $P_{\mathbb{H}} \subset P_{\mathbb{C}}$, we know $H_0^{\mathbb{H}}$ is nested within $H_0^{\mathbb{C}}$.

Now let $\mathbf{q}_1, \dots, \mathbf{q}_N$ be an IID sample of quaternion Gaussian random variables from the distribution of $\mathbf{q}$ and let $\mathbf{r}_1, \dots, \mathbf{r}_N$ be the corresponding $4p$ -dimensional real-valued random variables, with $N \geq 4p$, from a Gaussian distribution with mean vector zero and positive-definite and symmetric covariance matrix $\Sigma_{\mathbf{r}}$. We shall denote the maximum likelihood estimator of $\Sigma_{\mathbf{r}}$ by $\hat{\Sigma}_{\mathbb{R}}$, where

$$\hat{\Sigma}_{\mathbb{R}} = \frac{1}{N} \sum_{\ell=1}^N \mathbf{r}_{\ell} \mathbf{r}_{\ell}^T. \quad (11)$$

Lemma 1: Consider $\Sigma_{\mathbf{r}} \in P_{\mathbb{R}}$, defined as in (1). This can be expressed as

$$\Sigma_{\mathbf{r}} = \begin{bmatrix} \Sigma_{\text{Re}(\mathbf{z})} & \Sigma_{\text{Re}(\mathbf{z}), \text{Im}(\mathbf{z})} \\ \Sigma_{\text{Im}(\mathbf{z}), \text{Re}(\mathbf{z})} & \Sigma_{\text{Im}(\mathbf{z})} \end{bmatrix}.$$

Define

$$\mathbf{J}_1 = \begin{bmatrix} \mathbf{0}_{2p} & -\mathbf{I}_{2p} \\ \mathbf{I}_{2p} & \mathbf{0}_{2p} \end{bmatrix}. \quad (12)$$

Let $\hat{c} : P_{\mathbb{R}} \rightarrow P_{\mathbb{C}}$ be a projection such that $\hat{c}(2\Sigma_{\mathbf{r}}) =$

$$\begin{bmatrix} \Sigma_{\text{Re}(\mathbf{z})} + \Sigma_{\text{Im}(\mathbf{z})} & \Sigma_{\text{Re}(\mathbf{z}), \text{Im}(\mathbf{z})} - \Sigma_{\text{Im}(\mathbf{z}), \text{Re}(\mathbf{z})} \\ \Sigma_{\text{Im}(\mathbf{z}), \text{Re}(\mathbf{z})} - \Sigma_{\text{Re}(\mathbf{z}), \text{Im}(\mathbf{z})} & \Sigma_{\text{Re}(\mathbf{z})} + \Sigma_{\text{Im}(\mathbf{z})} \end{bmatrix} \\ = \Sigma_{\mathbf{r}} + \mathbf{J}_1 \Sigma_{\mathbf{r}} \mathbf{J}_1^T. \quad (13)$$

Then $\hat{\Sigma}_{\mathbb{C}} = \hat{c}(\hat{\Sigma}_{\mathbb{R}})$ is the maximum likelihood estimator of $\Sigma_{\mathbf{r}}$ under $H_0^{\mathbb{C}}$.

Proof: We wish to maximise the likelihood function

$$(2\pi)^{-2pN} \det \{\Sigma_{\mathbb{C}}\}^{-N/2} \exp \left(-\text{tr} \left\{ \Sigma_{\mathbb{C}}^{-1} \hat{\Sigma}_{\mathbb{R}} \right\} / 2 \right) \quad (14)$$

over $\Sigma_{\mathbb{C}} \in P_{\mathbb{C}}$. In consideration we introduce Lemma 2.

Lemma 2: For $\Sigma_{\mathbb{C}} \in P_{\mathbb{C}}$, $\text{tr} \left\{ \Sigma_{\mathbb{C}}^{-1} \hat{\Sigma}_{\mathbb{C}} \right\} = \text{tr} \left\{ \Sigma_{\mathbb{C}}^{-1} \hat{\Sigma}_{\mathbb{R}} \right\}$.

Proof: This is given in Appendix-A. ■

By Lemma 2, (14) is equal to

$$(2\pi)^{-2pN} \det \{\Sigma_{\mathbb{C}}\}^{-N/2} \exp \left(-\text{tr} \left\{ \Sigma_{\mathbb{C}}^{-1} \hat{\Sigma}_{\mathbb{C}} \right\} / 2 \right).$$

From the classical case, this is maximised over $\Sigma_{\mathbb{C}} \in P_{\mathbb{R}} \supset P_{\mathbb{C}}$ by setting $\Sigma_{\mathbb{C}} = \hat{\Sigma}_{\mathbb{C}}$. (See also [1, eqn. 12, Th. 1].) This completes the proof of Lemma 1. ■

Lemma 3: Consider $\ddot{\Sigma}_{\mathbf{r}} \in P_{\mathbb{C}}$, defined as in (10) and define

$$\mathbf{J}_2 = \begin{bmatrix} \mathbf{0}_p & \mathbf{0}_p & \mathbf{0}_p & -\mathbf{I}_p \\ \mathbf{0}_p & \mathbf{0}_p & \mathbf{I}_p & \mathbf{0}_p \\ \mathbf{0}_p & -\mathbf{I}_p & \mathbf{0}_p & \mathbf{0}_p \\ \mathbf{I}_p & \mathbf{0}_p & \mathbf{0}_p & \mathbf{0}_p \end{bmatrix}. \quad (15)$$

Let $\hat{h} : P_{\mathbb{C}} \rightarrow P_{\mathbb{H}}$ be a projection such that $\hat{h}(2\ddot{\Sigma}_{\mathbf{r}}) =$

$$\begin{bmatrix} \Sigma_{\mathbf{a}} + \Sigma_{\mathbf{b}} & \Sigma_{\mathbf{a},\mathbf{b}} - \Sigma_{\mathbf{b},\mathbf{a}} & \Sigma_{\mathbf{d},\mathbf{b}} - \Sigma_{\mathbf{c},\mathbf{a}} & -\Sigma_{\mathbf{c},\mathbf{b}} - \Sigma_{\mathbf{d},\mathbf{a}} \\ \Sigma_{\mathbf{b},\mathbf{a}} - \Sigma_{\mathbf{a},\mathbf{b}} & \Sigma_{\mathbf{a}} + \Sigma_{\mathbf{b}} & -\Sigma_{\mathbf{c},\mathbf{b}} - \Sigma_{\mathbf{d},\mathbf{a}} & \Sigma_{\mathbf{c},\mathbf{a}} - \Sigma_{\mathbf{d},\mathbf{b}} \\ \Sigma_{\mathbf{c},\mathbf{a}} - \Sigma_{\mathbf{d},\mathbf{b}} & \Sigma_{\mathbf{c},\mathbf{b}} + \Sigma_{\mathbf{d},\mathbf{a}} & \Sigma_{\mathbf{a}} + \Sigma_{\mathbf{b}} & \Sigma_{\mathbf{a},\mathbf{b}} - \Sigma_{\mathbf{b},\mathbf{a}} \\ \Sigma_{\mathbf{c},\mathbf{b}} + \Sigma_{\mathbf{d},\mathbf{a}} & \Sigma_{\mathbf{d},\mathbf{b}} - \Sigma_{\mathbf{c},\mathbf{a}} & \Sigma_{\mathbf{b},\mathbf{a}} - \Sigma_{\mathbf{a},\mathbf{b}} & \Sigma_{\mathbf{a}} + \Sigma_{\mathbf{b}} \end{bmatrix} \\ = \ddot{\Sigma}_{\mathbf{r}} + \mathbf{J}_2 \ddot{\Sigma}_{\mathbf{r}} \mathbf{J}_2^T. \quad (16)$$

Then $\hat{\Sigma}_{\mathbb{H}} = \hat{h}(\hat{\Sigma}_{\mathbb{C}}) = \hat{h}(\hat{c}(\hat{\Sigma}_{\mathbb{R}}))$ is the maximum likelihood estimator of $\Sigma_{\mathbf{r}}$ under $H_0^{\mathbb{H}}$.

Proof: We wish to maximise the likelihood function

$$(2\pi)^{-2pN} \det\{\Sigma_{\mathbb{H}}\}^{-N/2} \exp\left(-\text{tr}\left\{\Sigma_{\mathbb{H}}^{-1} \hat{\Sigma}_{\mathbb{R}}\right\}/2\right) \quad (17)$$

over $\Sigma_{\mathbb{H}} \in P_{\mathbb{H}}$. In consideration we introduce Lemma 4.

Lemma 4: For $\Sigma_{\mathbb{H}} \in P_{\mathbb{H}}$, $\text{tr}\left\{\Sigma_{\mathbb{H}}^{-1} \hat{\Sigma}_{\mathbb{H}}\right\} = \text{tr}\left\{\Sigma_{\mathbb{H}}^{-1} \hat{\Sigma}_{\mathbb{R}}\right\}$.

Proof: This is given in Appendix-A. ■

By Lemma 4, (17) is equal to

$$(2\pi)^{-2pN} \det\{\Sigma_{\mathbb{H}}\}^{-N/2} \exp\left(-\text{tr}\left\{\Sigma_{\mathbb{H}}^{-1} \hat{\Sigma}_{\mathbb{H}}\right\}/2\right).$$

This is maximised over $\Sigma_{\mathbb{H}} \in P_{\mathbb{R}} \supset P_{\mathbb{H}}$ by setting $\Sigma_{\mathbb{H}} = \hat{\Sigma}_{\mathbb{H}}$. (See also [1, eqn. 56, Th. 3].) This completes the proof of Lemma 3. ■

B. Nature of the Test

Let $T_{\mathbb{C}}$ be the LRT statistic for Test 1, and $T_{\mathbb{H}}$ be the LRT statistic for Test 2. From [1, Th. 1] $T_{\mathbb{C}}$ and $\hat{\Sigma}_{\mathbb{C}}$ are independent under $H_0^{\mathbb{C}}$. Applying Test 2 to $\hat{\Sigma}_{\mathbb{C}}$ we obtain another LRT statistic $T_{\mathbb{H}}$ which is a function of $\hat{\Sigma}_{\mathbb{C}}$ and hence is independent of $T_{\mathbb{C}}$. The LRT statistic for testing $H_0^{\mathbb{H}}$ versus $H_1^{\mathbb{C}}$ takes the simple form $T = T_{\mathbb{C}} \cdot T_{\mathbb{H}}$.

Proposition 2: The likelihood ratio (LR) for testing $H_0^{\mathbb{H}}$ versus $H_1^{\mathbb{H}}$ is given by

$$T_{\mathbb{H}} = [\det\{\hat{\Sigma}_{\mathbb{C}}\} / \det\{\hat{\Sigma}_{\mathbb{H}}\}]^{N/2}.$$

Proof: By Lemmas 1 & 3 the LR is

$$\left[\frac{\det\{\hat{\Sigma}_{\mathbb{C}}\}}{\det\{\hat{\Sigma}_{\mathbb{H}}\}} \right]^{N/2} \cdot \frac{\exp\left(-\text{tr}\left\{\hat{\Sigma}_{\mathbb{H}}^{-1} \hat{\Sigma}_{\mathbb{R}}\right\}/2\right)}{\exp\left(-\text{tr}\left\{\hat{\Sigma}_{\mathbb{C}}^{-1} \hat{\Sigma}_{\mathbb{R}}\right\}/2\right)}.$$

By Lemmas 4 & 2, $\text{tr} \left\{ \hat{\Sigma}_{\mathbb{H}}^{-1} \hat{\Sigma}_{\mathbb{R}} \right\} = \text{tr} \left\{ \hat{\Sigma}_{\mathbb{H}}^{-1} \hat{\Sigma}_{\mathbb{H}} \right\} = \text{tr} \{ \mathbf{I}_{4p} \} = \text{tr} \left\{ \hat{\Sigma}_{\mathbb{C}}^{-1} \hat{\Sigma}_{\mathbb{C}} \right\} = \text{tr} \left\{ \hat{\Sigma}_{\mathbb{C}}^{-1} \hat{\Sigma}_{\mathbb{R}} \right\}$, so the exponentials cancel. ■

Proposition 3: The LR for testing $H_0^{\mathbb{C}}$ versus $H_1^{\mathbb{C}}$ is

$$T_{\mathbb{C}} = [\det\{\hat{\Sigma}_{\mathbb{R}}\} / \det\{\hat{\Sigma}_{\mathbb{C}}\}]^{N/2}.$$

Proof: Proceeds as Proposition 2; see also [30, p. 828]. ■

Consider H_0 , H_1 , defined in Section I-A. Since $P(\hat{\Sigma}_{\mathbb{R}} \in P_{\mathbb{C}}) = 0$, the maximum likelihood under $H_1^{\mathbb{C}}$ and under H_1 are almost surely equal. Also, $H_0 = H_0^{\mathbb{H}}$. The LRT statistic for testing H_0 versus H_1 is therefore

$$T = T_{\mathbb{C}} \cdot T_{\mathbb{H}} = [\det\{\hat{\Sigma}_{\mathbb{R}}\} / \det\{\hat{\Sigma}_{\mathbb{H}}\}]^{N/2}. \quad (18)$$

T satisfies $0 \leq T \leq 1$, since it is an LR and $P_{\mathbb{H}} \subset P_{\mathbb{R}}$. We reject the null hypothesis for small values of T .

Now let us consider the moments of T . By independence we can obtain the moments $E\{T^h\} = E\{T_{\mathbb{C}}^h\}E\{T_{\mathbb{H}}^h\}$.

Proposition 4: Under H_0 the LRT statistic T for H_0 versus H_1 has moments

$$E\{T^h\} = K \prod_{j=1}^{\lceil \frac{3p}{2} \rceil} \frac{\Gamma[N(h+1) - 4p + 2j - 1]}{\Gamma\left[N(h+1) + \frac{2-j - \lfloor \frac{j-1}{3} \rfloor}{2}\right]}, \quad (19)$$

where $\lfloor x \rfloor$ is the integer part of x , $\lceil x \rceil$ is the smallest integer greater or equal to x , and K does not depend on h .

Proof: This is given in Appendix-B. ■

We now identify a general class of distributions containing the LRT statistic T .

Definition 4: A random variable $0 \leq W \leq 1$ is said to be of Box-type [2] if, for all $h \in \mathbb{N}$,

$$E\{W^h\} = K \left[\frac{\prod_{j=1}^k y_j^{y_j}}{\prod_{i=1}^m x_i^{x_i}} \right]^h \frac{\prod_{i=1}^m \Gamma[x_i(1+h) + \xi_i]}{\prod_{j=1}^k \Gamma[y_j(1+h) + \eta_j]}, \quad (20)$$

where $K = \prod_{j=1}^k \Gamma(y_j + \eta_j) / \prod_{i=1}^m \Gamma(x_i + \xi_i)$, so $E\{W^0\} = 1$, and $\sum_{i=1}^m x_i = \sum_{j=1}^k y_j$, $x_i > 0$, $y_j > 0$.

Comparing (19) and (20) we see that T is a random variable of Box-type with

$$x_i = y_j = N; m = k = \lceil 3p/2 \rceil; \quad (21)$$

$$\xi_i = -4p + 2i - 1; \eta_j = [2 - j - \lfloor (j-1)/3 \rfloor]/2, \quad (22)$$

and K in (19) is

$$K = \prod_{i=1}^{\lceil \frac{3p}{2} \rceil} \frac{\Gamma[N + \eta_i]}{\Gamma[N + \xi_i]}. \quad (23)$$

Next, we will consider the non-negative random variable $M = -2 \log T$.

IV. TEST STATISTIC AND ITS DISTRIBUTION

A. Asymptotic Distribution

From Wilks' theorem [33, p. 132] under H_0 , as $N \rightarrow \infty$,

$$M = -2 \log T \xrightarrow{d} \chi_f^2 \quad (24)$$

i.e., M tends in distribution to χ_f^2 . Here f denotes the degrees of freedom for the test and is given by [2, eqn. 73]

$$\begin{aligned} f &= -2 \left[\sum_{i=1}^m \xi_i - \sum_{j=1}^k \eta_j - \frac{1}{2}(m - k) \right] \\ &= \sum_{i=1}^{\lceil 3p/2 \rceil} [8p + 4 - 5i - \lfloor (i-1)/3 \rfloor] \\ &= 3p(2p+1). \end{aligned} \quad (25)$$

This is equal to the difference between the number of free parameters in the covariance matrix under H_1 , namely $2p(4p+1)$, and under H_0 , namely $p(2p-1)$. The asymptotic distribution (24) is given in [27] (along with asymptotic distributions of other related likelihood ratio statistics).

B. Exact Distribution

We firstly show that T can be written as a product of powers of independent Beta-distributed random variables.

Proposition 5: Assume W has moments given by (20) where $m = k$, $x_i = y_i$ for all i and $\eta_i > \xi_i > -x_i$ for all i (parameters can be reordered if necessary). Then $W \sim \prod_{i=1}^m V_i^{x_i}$, where V_i are independently distributed as $\text{Beta}(\xi_i + x_i, \eta_i - \xi_i)$ random variables.

Proof: Similarly to [13, p. 122], for all $h \in \mathbb{C}$ such that $\text{Re}(h) > -\xi_i - x_i$,

$$E\{V_i^{x_i h}\} = \frac{\Gamma(x_i + \eta_i) \Gamma(\xi_i + x_i(1+h))}{\Gamma(x_i + \xi_i) \Gamma(\eta_i + x_i(1+h))}. \quad (26)$$

Hence the moments match and so does the distribution which is bounded and hence uniquely determined by the moments. ■

Exact probability density functions (PDFs) of products of powers of independent Beta random variables are given as H -functions in [13, eq. (4.15)]. Furthermore, $T^{1/N}$ is a product of independent Beta random variables; the PDFs of such products are given as G -functions in [3, eq. (4.8)].

Definition 5: Fox's H -function is defined by the following Mellin-Barnes integral [3]:

$$H_{q,r}^{u,v} \left[z \left| \begin{matrix} (a_1, \alpha_1), \dots, (a_p, \alpha_p) \\ (b_1, \beta_1), \dots, (b_q, \beta_q) \end{matrix} \right. \right] = \frac{1}{2\pi i} \int_L \frac{\prod_{j=1}^u \Gamma(b_j - \beta_j s) \prod_{j=1}^v \Gamma(1 - a_j + \alpha_j s)}{\prod_{j=u+1}^r \Gamma(1 - b_j + \beta_j s) \prod_{j=v+1}^q \Gamma(a_j - \alpha_j s)} z^s ds. \quad (27)$$

The path of integration L is chosen such that the poles $(b_j + n)/\beta_j$, $j = 1, \dots, u$, $n \in \mathbb{N}$ are separated from the poles $(a_j - 1 - n)/\alpha_j$, $j = 1, \dots, v$, $n \in \mathbb{N}$. The parameters are $0 \leq v \leq q$, $0 \leq u \leq r$, $\alpha_i, \beta_j \in [0, \infty[$, $a_i, b_j \in \mathbb{C}$.

In all our uses, we will have $v = 0$ and the path of integration L will be a vertical line from $\gamma - i\infty$ to $\gamma + i\infty$.

Definition 6: Meijer's G -function is a special case of the H -function

$$G_{q,r}^{u,v} \left[z \left| \begin{matrix} a_1, \dots, a_q \\ b_1, \dots, b_r \end{matrix} \right. \right] = H_{q,r}^{u,v} \left[z \left| \begin{matrix} (a_1, 1), \dots, (a_q, 1) \\ (b_1, 1), \dots, (b_r, 1) \end{matrix} \right. \right]. \quad (28)$$

The moment generating function for M is given by $\phi_M(s) = E\{e^{sM}\} = E\{T^{-2s}\}$ so by Proposition 5

$$\phi_M(s) = K \prod_{j=1}^{\lceil \frac{3p}{2} \rceil} \frac{\Gamma[N(1-2s) + \xi_j]}{\Gamma[N(1-2s) + \eta_j]} \quad (29)$$

for $\text{Re}(-2s) > \max_i(-\xi_i - x_i) = 4p - 1 - N$, with ξ_i, η_i, K given in (22) and (23).

Theorem 1: With parameter values as in (21) and (22), the PDF of M is given by

$$f_M(x) = \frac{K e^{-x/2}}{2N} G_{m,m}^{m,0} \left[e^{-x/(2N)} \left| \begin{matrix} \eta_1, \dots, \eta_m \\ \xi_1, \dots, \xi_m \end{matrix} \right. \right], \quad (30)$$

where the constant K is defined in (23). The cumulative distribution function (CDF) is

$$F_M(x) = K e^{-x/2} G_{m+1,m+1}^{m+1,0} \left[e^{-x/(2N)} \left| \begin{matrix} \eta_1, \dots, \eta_m, 1 - N \\ \xi_1, \dots, \xi_m, -N \end{matrix} \right. \right]. \quad (31)$$

Proof: This is given in Appendix–D. ■

To be useful, the G -function expressions for the PDF and CDF in (30) and (31), respectively, need to be computable quantities. We examined a number of symbolic computation engines which in theory have arbitrary-precision implementations of Meijer’s G -function, and encountered problems with each implementation. We were however able to compute exact quantiles of the distribution of M , using version 13 of Maple[®], by adding an ad-hoc correction whenever (31) erroneously evaluated to $F_M(x) - 1$. These exact quantiles are used to compare approximations in section V–H.

The cumulants of M can be easily obtained from the cumulant generating function $\log \phi_M(s)$.

$$\begin{aligned} \kappa_j &= \left. \frac{d^j \log \phi_M(s)}{(ds)^j} \right|_{s=0} \\ &= [-2N]^j \sum_{i=1}^{\lceil 3p/2 \rceil} \left[\psi^{(j-1)}(N + \xi_i) - \psi^{(j-1)}(N + \eta_i) \right], \end{aligned} \quad (32)$$

where $\psi(x) = [d \log \Gamma(x)]/dx$ is the digamma function and its derivatives are polygamma functions (trigamma, tetragamma, ...). κ_1 is the mean, κ_2 is the variance, $\kappa_3/\kappa_2^{3/2}$ is the skewness and κ_4/κ_2^2 is the excess kurtosis.

V. APPROXIMATING DISTRIBUTIONS

Using a χ_f^2 distribution to approximate the distribution of M as in (24) will generally give poor results for finite N . Routine computation of the exact distribution for hypothesis testing is not straightforward. We now consider approximation methods and their performance. We briefly outline various possible approaches which we will compare later by making use of the exact distribution. We firstly note that since the distribution of T is of Box type, we may approximate the distribution of $M = -2 \log T$ using methods in [2].

A. Bartlett adjustment

One such method introduces a Bartlett adjustment C , approximating M by the distribution of $C\chi_f^2$, where $f = 3p(2p + 1)$ here, as in (25). C is chosen so that the cumulants of the approximation match those of M up to an error of order $O(N^{-2})$. Box’s method [2, p. 338] implies the Bartlett adjustment $C_{\text{Box}} = (12N)/(12N + 1 - 20p)$.

This approximation is convenient since it does not require the use of a computer. A more accurate approximation is obtained if one instead chooses the Bartlett adjustment $C_{\text{exact}} = f^{-1}\kappa_1$,

which matches the mean of M *exactly*, but we will not examine this option further since it is superseded by the gamma approximation of the next section. An analysis is given in [18] where various choices of Bartlett adjustment for tests 1 and 2 are compared.

B. Gamma approximation

As suggested in [7], we can also approximate M by a Gamma random variable with PDF $f(x) = (\lambda^r / \Gamma(r)) x^{r-1} \exp(-\lambda x)$, for $x, r, \lambda > 0$, with $r = \kappa_1^2 / \kappa_2, \lambda = \kappa_1 / \kappa_2$. Such a random variable has the same mean and variance as M . Scaled χ^2 distributions are special cases of the Gamma distribution where the first (shape) parameter is a half integer, so this approach corresponds to fitting both the scale and the degrees of freedom of a scaled χ^2 approximation.

C. A new F -approximation

Based on the Pearson system of curves [9], Box uses distributions of scaled F - and Beta-distributed random variables, $bF(f_1, f_2)$ or $b\text{Beta}(f_1, f_2)$, respectively, where the distribution and parameters are chosen so that the first four cumulants match those of M up to an error of order $O(N^{-3})$.

We suggest instead to match the first 3 cumulants exactly, leading to greatly increased accuracy. It does not appear this exact approach has been used before. We ran a numerical study of all combinations $p \leq 10000$, $N \leq 4p + 2000$ and concluded that cases requiring a Beta distribution do not arise and that it will be possible to match the first 3 cumulants using an F distribution whenever $N \geq 4.13p + 0.5$, i.e., nearly always considering that $N \geq 4p$ is a requirement. The parameters are then given by

$$\begin{aligned} f_1 &= 4\kappa_1 (\kappa_1^2 \kappa_2 - \kappa_2^2 + \kappa_1 \kappa_3) / (4\kappa_1 \kappa_2^2 - \kappa_1^2 \kappa_3 + \kappa_2 \kappa_3); \\ f_2 &= (4\kappa_1^2 \kappa_2 - 8\kappa_2^2 + 6\kappa_1 \kappa_3) / (\kappa_1 \kappa_3 - 2\kappa_2^2); \\ b &= 2\kappa_1 (\kappa_1^2 \kappa_2 - \kappa_2^2 + \kappa_1 \kappa_3) / (2\kappa_1^2 \kappa_2 - 4\kappa_2^2 + 3\kappa_1 \kappa_3). \end{aligned} \quad (33)$$

D. Box's χ^2 series

Following [2] we can expand the distribution function of M as

$$F_M(x) = K_B \sum_{j=0}^{n-1} a_j F_{\chi_{f+2j}^2}(\rho x) + O(N^{-n}), \quad (34)$$

where $K_B = K(\rho N)^{-f/2}$. Using the iterative scheme of [6], which is essentially a new utilization of the algorithm which computes moments from cumulants, the coefficients are given by

$$\begin{aligned} a_0 &= 1; & a_j &= \frac{1}{j} \sum_{i=1}^j i \omega_i a_{j-i}, \quad j = 1, \dots, n-1; \\ \omega_i &= \frac{(-1)^{i+1}}{i(i+1)(\rho N)^i} \\ &\quad \times \sum_{\ell=1}^{\lceil \frac{3p}{2} \rceil} [B_{i+1}((1-\rho)N + \xi_\ell) - B_{i+1}((1-\rho)N + \eta_\ell)], \end{aligned}$$

where $B_i(x)$ denotes the Bernoulli polynomial of degree i . The free scaling parameter ρ is chosen so that $a_1 = 0$. Using [4, eqn. 1.7], $\rho = 1 - [1/(fN)] \sum_{\ell=1}^{\lceil \frac{3p}{2} \rceil} [B_2(\xi_\ell) - B_2(\eta_\ell)] = C_{\text{Box}}^{-1}$.

It is proven in [6] that for $x < 4\pi N$, the series (34) converges to $F_M(x)$ as $n \rightarrow \infty$, meaning that an arbitrary precision (exact) value can be computed. For $x \geq 4\pi N$, the series diverges if too many terms are taken, as is generally the case with asymptotic expansions, but this is unlikely to be a problem in practice.

This approximation is often applied with the parameter choice $n = 3$, and possibly a renormalisation. Due to its poor performance, we have chosen instead $n = 5$ in our numerical comparisons.

E. Lugannani & Rice saddlepoint approximation

Consider an exponentially tilted version of M , M_s say, i.e., $f_{M_s}(x) = e^{sx} f_M(x) / \phi_M(s)$. Then since $\phi_{M_s}(t) = \phi_M(s+t) / \phi_M(s)$ the cumulants of M_s are

$$\begin{aligned} \bar{\kappa}_j(s) &= \left. \frac{d^j \log \phi_{M_s}(t)}{(dt)^j} \right|_{t=0} \\ &= (-2N)^j \sum_{i=1}^{\lceil 3p/2 \rceil} \left[\psi^{(j-1)}(N(1-2s) + \xi_i) \right. \\ &\quad \left. - \psi^{(j-1)}(N(1-2s) + \eta_i) \right]. \end{aligned}$$

Let $F_{\mathcal{N}}(\cdot)$ and $f_{\mathcal{N}}(\cdot)$ denote, respectively, the standard Gaussian CDF and PDF. The Lugannani & Rice saddlepoint approximation [12], truncated to two terms is

$$F_M(x) \approx F_{\mathcal{N}}(\tilde{x}) + f_{\mathcal{N}}(\tilde{x}) \left(\frac{1}{\tilde{x}} - \frac{1}{s_0 \bar{\kappa}_2^{1/2}(s_0)} \right) \quad (35)$$

where $\tilde{x} = \text{sign}(s_0) \{2[s_0 x - \log(\phi_M(s_0))]\}^{1/2}$ and s_0 is the unique solution of $\bar{\kappa}_1(s) = x$. Since the derivative of $\bar{\kappa}_1(s)$ is $\bar{\kappa}_2(s) > 0$, s_0 can easily be found numerically, e.g., by Newton's method.

The distribution of M is asymptotically χ_f^2 , not asymptotically normal and hence (35) does not converge to $F_M(x)$ as $N \rightarrow \infty$. To solve this problem, we considered using the approximation of [31] which generalises the Lugannani & Rice saddlepoint approximation to non-normal basis distributions. However, using a Gamma basis distribution provided no significant improvement, except obviously for large N .

F. Jensen's large deviation approximation

In [7], [8] Jensen applies the Gamma approximation to the exponentially tilted M_s , which yields the following approximation for M :

$$F_M(x) \approx \phi_M(s_0) \left(1 + \frac{s_0 \bar{\kappa}_2(s_0)}{x} \right)^{-x^2/\bar{\kappa}_2(s_0)} \\ \times \gamma \left(\frac{x^2}{\bar{\kappa}_2(s_0)}, x s_0 + \frac{x^2}{\bar{\kappa}_2(s_0)} \right)$$

where as previously s_0 solves $\bar{\kappa}_1(s) = x$, and $\gamma(a, b)$ denotes the incomplete gamma function $\gamma(a, b) = \left[\int_0^b x^{a-1} \exp(-x) dx \right] / \Gamma(a)$.

Whilst for all of the previously described approximations the relative error (36) tends to infinity if one takes $x \rightarrow \infty$, the relative error of this approximation instead goes to 0, making it particularly suitable for approximating the extreme tail.

G. Direct simulation

The moments (19), which fully determine the null distribution of T , and hence of M , do not depend on the true covariance matrix $\Sigma_{\mathbf{r}}$. Hence we can assume $\Sigma_{\mathbf{r}} = \mathbf{I}_{4p}$ when simulating M . For each sample of size N the maximum likelihood estimator $\hat{\Sigma}_{\mathbb{R}}$ in (11) is computed. $\hat{\Sigma}_{\mathbb{H}}$ is then given by (13), (15) and (16), and T by (18).

H. Computational results

For a chosen combination of p and N we define the relative error of the approximation $F_{\text{approx}}(x)$ to $F_M(x)$ as

$$\left| \frac{F_{\text{approx}}(x) - F_M(x)}{\min[F_M(x), 1 - F_M(x)]} \right|, \quad (36)$$

where $F_M(x)$ is given in (31) and $F_{\text{approx}}(x)$ is any of the approximate CDFs discussed in this Section. The effect of the divisor is to make the error relate to the corresponding tail probability, depending on whether x corresponds to a value in the left or right tail. Since we will reject the

hypothesis of propriety when M is larger than some critical value, we will be most interested in the region around $F_M(x) = 0.95$ and 0.99 . In Fig. 1 the relative error is plotted against x for our considered approximation methods for the case $p = 4, N = 32$. The various methods used have been given both plot symbols and labelled A-F in the legend of the figure according to the subsection where they are discussed, and we will use this labelling in what follows. The top axis shows the value of $F_M(x)$ (as a percentage) corresponding to x on the bottom axis. We conclude that the new scaled F -approximation (method C), the saddlepoint approximation (method E) and the large deviation method (method F), perform best. Of these the scaled F -approximation is the easiest to implement in practice.

To see whether our conclusions are generally true for impropriety testing for a variety of p and N we firstly look at the relative error of approximation at the 95th percentile of $F_M(x)$. Fig. 2 shows this relative error for fixed $p = 6$ and increasing values of N . We see that the errors for approximations A, B, C, D, E, and F, behave roughly like $O(N^{-r})$ with $r = 2, 2, 3, 5, 0, 2$ respectively. Fig. 3 shows this relative error for $p = 1, \dots, 14$ and $N = 5p$. Finally, Fig. 4 gives the relative error for varying p and $N = 16p$. In all cases our earlier conclusions are borne out, particularly that the scaled F -approximation performs extremely well. Let x_0 be the 99th percentile of $F_M(x)$ and let $F_{\text{approx}}(x)$ be one of the approximate CDFs of subsections A–G. The values given in Table I are then $100(1 - F_{\text{approx}}(x_0))\%$, a quantity which is ideally 1%. For each p the entries start at the minimum N value of $4p$ and the other entries are for $5p, 8p$ and $16p$. The scaled F -approximation (method C) requires $N \geq 4.13p + 0.5$ so that it cannot be used for $N = 4p$ and thus the corresponding entry is marked by ‘N/A.’ Again we see that methods C, E and F all perform very well. Table I also includes direct simulation results (method/subsection G) with the empirical CDF calculated from 100 000 independent simulations of $M = -2 \log T$. Even with so many independent simulations the simulated results can be relatively inaccurate compared with the simple scaled F -approximation.

VI. ASYMPTOTIC ROBUSTNESS TO NONGAUSSIANTITY

As with the LRT for complex propriety [19], our LRT for quaternion propriety will be sensitive to departures from Gaussianity which could be an issue when applying the test to physical data.

We can write the hypothesis that $\Sigma_{\mathbf{r}}$ has quaternion structure in the form $[\Sigma_{\mathbf{r}} - \hat{h}(\hat{c}(\Sigma_{\mathbf{r}}))]/\text{var}\{r_1\} = \mathbf{0}_{4p}$. Taking f non-redundant equations from this, the LRT for quaternion structure satisfies the scaling condition for [25, Corollary 1] to hold. Consequently, if instead of being multivariate

p	N	method						
		A	B	C	D	E	F	G
1	4	0.0398	0.7973	0.9806	5.2512	0.9961	1.0334	0.997
	5	0.3126	0.9021	0.9972	1.5443	0.9790	1.0078	0.950
	8	0.7688	0.9756	1.0000	1.0164	0.9893	0.9980	0.976
	16	0.9535	0.9956	1.0000	1.0002	0.9991	0.9997	0.977
2	8	0.0002	0.6318	0.9812	40.612	0.9914	1.0646	1.048
	10	0.1429	0.8876	0.9988	7.9486	0.9788	1.0086	1.009
	16	0.6790	0.9791	0.9999	1.2280	0.9979	1.0021	1.004
	32	0.9339	0.9965	1.0000	1.0029	0.9999	1.0004	0.995
4	16	0.0000	0.5115	N/A	96.174	0.9873	1.0930	0.978
	20	0.0496	0.9219	0.9982	49.217	0.9952	1.0084	1.046
	32	0.5607	0.9875	0.9999	4.1031	0.9998	1.0013	1.016
	64	0.9034	0.9980	1.0000	1.0412	1.0000	1.0002	1.004
6	24	0.0000	0.4671	N/A	99.982	0.9849	1.1058	0.985
	30	0.0167	0.9442	0.9987	89.100	0.9988	1.0059	0.966
	48	0.4616	0.9914	0.9999	15.430	0.9999	1.0008	1.023
	96	0.8737	0.9986	1.0000	1.2439	1.0000	1.0001	1.021
12	48	0.0000	0.4287	N/A	100.00	0.9808	1.1194	1.021
	60	0.0004	0.9711	0.9995	100.00	0.9999	1.0026	0.977
	96	0.2494	0.9956	1.0000	83.748	1.0000	1.0004	1.068
	192	0.7888	0.9993	1.0000	6.8623	1.0000	1.0001	1.023

TABLE I

APPROXIMATE REJECTION PROBABILITIES AT THE 1% LEVEL. ENTRIES ARE $100(1 - F_{\text{approx}}(x_0))\%$, WHERE x_0 IS THE 99TH PERCENTILE OF $F_M(x)$. METHODS A-F ARE AS IN THE LEGEND TO FIG. 1.

METHOD G IS DIRECT SIMULATION.

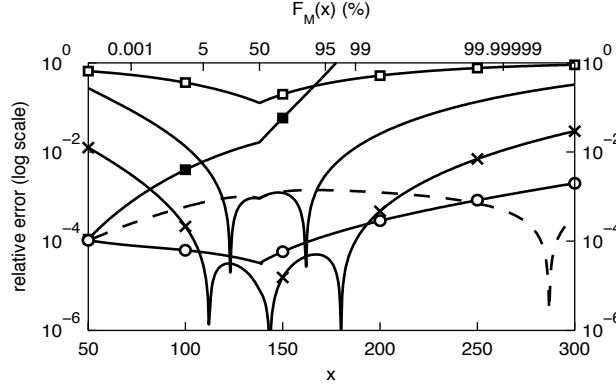


Fig. 1. Relative error defined in (36) for $p = 4, N = 32$. A, Bartlett adjustment: white boxes. B, Gamma approximation: solid line. C, New scaled F -approximation: crosses. D, Box's χ^2 series: black boxes. E, Lugannani & Rice saddlepoint approximation: white circles. F, Jensen's large deviation method: dashed line. The sharp dips correspond to points where the relative error changes sign.

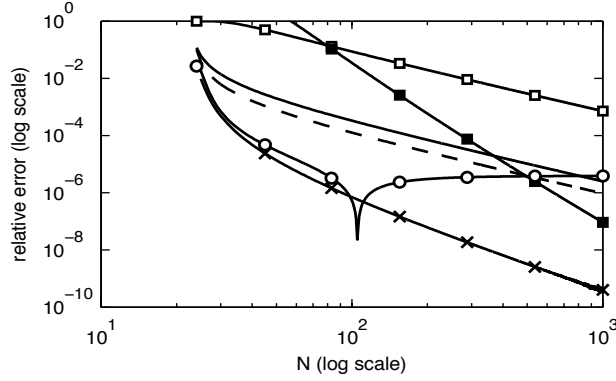


Fig. 2. Relative error of approximated CDFs at the 95th percentile, for $p = 6$. Legend: Same as Fig. 1.

Gaussian we assume that the distribution of $\mathbf{r}$ belongs to the wide class of elliptical distributions, with its excess kurtosis κ being finite, the asymptotic distribution of M becomes $(1 + \kappa/3)\chi_f^2$. Based on this theory, if we take a consistent estimator $\hat{\kappa}$ of κ , then the test statistic $(1 + \hat{\kappa}/3)^{-1}M$ will be asymptotically robust to nonGaussianity. Note however that the adjustment changes the distribution of the statistic.

VII. CONCLUDING COMMENTS

We described the LRT to test for quaternion propriety and gave its distribution under the null hypothesis. Our numerical comparison of various approximations to this distribution leads us to recommend using the scaled F -approximation for practical applications. We merely need to

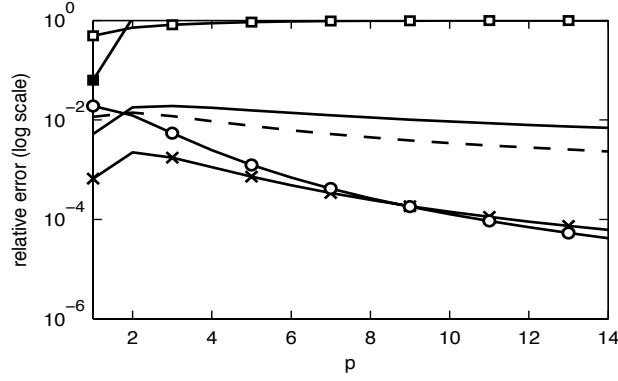


Fig. 3. Relative error of approximated CDFs at the 95th percentile, for varying p and $N = 5p$. Legend: Same as Fig. 1.

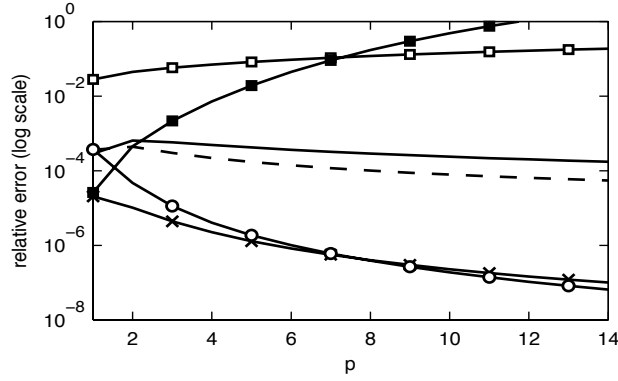


Fig. 4. Relative error of approximated CDFs at the 95th percentile, for varying p and $N = 16p$. Legend: Same as Fig. 1.

calculate the three parameters in (33), which require κ_1, κ_2 and κ_3 which follow easily from (32); the required polygamma functions of order 0, 1 and 2 are implemented in many mathematical and statistical packages and in the GNU scientific library. The approximation is conceptually straightforward and yields approximate PDF, CDF and quantiles directly.

APPENDIX

A. Proof of Lemmas 2 & 4

Consider $\mathbf{J}_1$ and $\mathbf{J}_2$ in (12), (15). The matrix $-\mathbf{J}_1 = \mathbf{J}_1^T = \mathbf{J}_1^{-1}$ commutes with all matrices in $P_{\mathbb{C}}$ [1, p. 395] and $-\mathbf{J}_2 = \mathbf{J}_2^T = \mathbf{J}_2^{-1}$ commutes with all matrices in $P_{\mathbb{H}}$ [1, p. 404]. Let $\mathbf{J}_3 = \mathbf{J}_2\mathbf{J}_1$

and $\mathbf{J}_0 = \mathbf{I}_{4p}$. Hence, for $\mathbf{\Sigma}_{\mathbb{H}} \in P_{\mathbb{H}}$ and $i = 0, 1, 2, 3$,

$$\begin{aligned} \text{tr} \left\{ \mathbf{\Sigma}_{\mathbb{H}}^{-1} \mathbf{J}_i \hat{\mathbf{\Sigma}}_{\mathbb{R}} \mathbf{J}_i^T \right\} &= \text{tr} \left\{ (\mathbf{J}_i^T \mathbf{\Sigma}_{\mathbb{H}} \mathbf{J}_i)^{-1} \hat{\mathbf{\Sigma}}_{\mathbb{R}} \right\} \\ &= \text{tr} \left\{ (\mathbf{J}_i^T \mathbf{J}_i \mathbf{\Sigma}_{\mathbb{H}})^{-1} \hat{\mathbf{\Sigma}}_{\mathbb{R}} \right\} = \text{tr} \left\{ \mathbf{\Sigma}_{\mathbb{H}}^{-1} \hat{\mathbf{\Sigma}}_{\mathbb{R}} \right\}. \end{aligned}$$

Now $4\hat{\mathbf{\Sigma}}_{\mathbb{H}} = 4\hat{h}(\hat{\mathbf{\Sigma}}_{\mathbb{C}}) = \hat{h}(2\hat{c}(2\hat{\mathbf{\Sigma}}_{\mathbb{R}}))$ and

$$\hat{h}(2\hat{c}(2\hat{\mathbf{\Sigma}}_{\mathbb{R}})) = \hat{\mathbf{\Sigma}}_{\mathbb{R}} + \mathbf{J}_1 \hat{\mathbf{\Sigma}}_{\mathbb{R}} \mathbf{J}_1^T + \mathbf{J}_2 \hat{\mathbf{\Sigma}}_{\mathbb{R}} \mathbf{J}_2^T + \mathbf{J}_2 \mathbf{J}_1 \hat{\mathbf{\Sigma}}_{\mathbb{R}} \mathbf{J}_1^T \mathbf{J}_2^T.$$

So,

$$4\text{tr} \left\{ \mathbf{\Sigma}_{\mathbb{H}}^{-1} \hat{\mathbf{\Sigma}}_{\mathbb{H}} \right\} = \sum_{i=0}^3 \text{tr} \left\{ \mathbf{\Sigma}_{\mathbb{H}}^{-1} \mathbf{J}_i \hat{\mathbf{\Sigma}}_{\mathbb{R}} \mathbf{J}_i^T \right\} = 4\text{tr} \left\{ \mathbf{\Sigma}_{\mathbb{H}}^{-1} \hat{\mathbf{\Sigma}}_{\mathbb{R}} \right\}.$$

Similarly, for $\mathbf{\Sigma}_{\mathbb{C}} \in P_{\mathbb{C}}$

$$2\text{tr} \left\{ \mathbf{\Sigma}_{\mathbb{C}}^{-1} \hat{\mathbf{\Sigma}}_{\mathbb{C}} \right\} = \sum_{i=0}^1 \text{tr} \left\{ \mathbf{\Sigma}_{\mathbb{C}}^{-1} \mathbf{J}_i \hat{\mathbf{\Sigma}}_{\mathbb{R}} \mathbf{J}_i^T \right\} = 2\text{tr} \left\{ \mathbf{\Sigma}_{\mathbb{C}}^{-1} \hat{\mathbf{\Sigma}}_{\mathbb{R}} \right\}.$$

B. Proof of Proposition 4

Test 1 corresponds to test (a) in [1]. Changing p to $2p$ in [1, eqn. 101] gives

$$E\{T_{\mathbb{C}}^h\} = K_0 \prod_{j=1}^{2p} \frac{\Gamma([N(h+1) - 2p - j + 1]/2)}{\Gamma([N(h+1) - j + 2]/2)} \quad (37)$$

where K_0 does not depend on h . Applying the duplication formula for gamma functions we can write

$$\begin{aligned} &\Gamma \left[\frac{[N(h+1) - 2p - 2k + 1]}{2} \right] \Gamma \left[\frac{[N(h+1) - 2p - 2k + 2]}{2} \right] \\ &= \sqrt{\pi} 2^{2p+2k-N(h+1)} \Gamma[N(h+1) - 2p - 2k + 1] \end{aligned}$$

and

$$\begin{aligned} &\Gamma \left[\frac{[N(h+1) - 2k + 2]}{2} \right] \Gamma \left[\frac{[N(h+1) - 2k + 3]}{2} \right] \\ &= \sqrt{\pi} 2^{2k-1-N(h+1)} \Gamma[N(h+1) - 2k + 2]. \end{aligned}$$

Hence

$$E\{T_{\mathbb{C}}^h\} = K_1 \prod_{j=1}^p \frac{\Gamma[N(h+1) - 2p - 2j + 1]}{\Gamma[N(h+1) - 2j + 2]},$$

while, from [1, eqn. 103]

$$E\{T_{\mathbb{H}}^h\} = K_2 \prod_{j=1}^p \frac{\Gamma[N(h+1) - p - j + 1]}{\Gamma[N(h+1) + \frac{3}{2} - j]},$$

where K_1 and K_2 do not depend on h . So $E\{T^h\} = E\{T_{\mathbb{C}}^h\}E\{T_{\mathbb{H}}^h\}$ takes the form

$$K \prod_{j=1}^p \frac{\Gamma[N(h+1) - 2p - 2j + 1] \Gamma[N(h+1) - p - j + 1]}{\Gamma[N(h+1) - 2j + 2] \Gamma[N(h+1) - j + \frac{3}{2}]}$$

This expression can be simplified by cancelling terms and reordering. Even and odd p can be considered separately. $\lfloor p/2 \rfloor$ of the $\Gamma[N(h+1) - 2j + 2]$ terms cancel with $\Gamma[N(h+1) - p - j + 1]$ terms. The remaining numerator terms of the product can be juxtaposed and the remaining denominator terms interspersed to form a monotone pattern. Finally by inverting the order of the product in the numerator we obtain (19).

C. H -function equivalences

The following 3 expressions are equal for arbitrary $c \in \mathbb{C}, \lambda > 0$, [3, eqs. (2.3), (2.4)]

$$H_{q,r}^{u,v} \left[z \left| \begin{matrix} (a_1, \alpha_1), \dots, (a_q, \alpha_q) \\ (b_1, \beta_1), \dots, (b_r, \beta_r) \end{matrix} \right. \right] \quad (38)$$

$$\lambda H_{q,r}^{u,v} \left[z^\lambda \left| \begin{matrix} (a_1, \lambda\alpha_1), \dots, (a_q, \lambda\alpha_q) \\ (b_1, \lambda\beta_1), \dots, (b_r, \lambda\beta_r) \end{matrix} \right. \right] \quad (39)$$

$$z^{-c} H_{q,r}^{u,v} \left[z \left| \begin{matrix} (a_1 + c\alpha_1, \alpha_1), \dots, (a_q + c\alpha_q, \alpha_q) \\ (b_1 + c\beta_1, \beta_1), \dots, (b_r + c\beta_r, \beta_r) \end{matrix} \right. \right] \quad (40)$$

D. Proof of Theorem 1

Firstly consider the PDF. The integral for inverting the moment generating function of M is

$$f_M(x) = \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} e^{-sx} \phi_M(s) ds. \quad (41)$$

If we substitute (29) into (41) and compare with the integral (27) we see that

$$f_M(x) = K H_{m,m}^{m,0} \left[e^{-x} \left| \begin{matrix} (N + \eta_1, 2N), \dots, (N + \eta_m, 2N) \\ (N + \xi_1, 2N), \dots, (N + \xi_m, 2N) \end{matrix} \right. \right],$$

with η_i and ξ_i as in (22). Then apply (39) with $\lambda = 1/(2N)$, followed by (40) with $c = N$. Then the use of (28) gives (30). It may be shown using Stirling's approximation that the tail of the characteristic function goes as

$$\phi_M(s) = O\left(|s|^{-\frac{f}{2}}\right). \quad (42)$$

We require $f > 2$ but since $f \geq 9$ ($p \geq 1$), the moment generating function will be absolutely integrable and the density will be uniformly continuous on $\mathbb{R}$.

For the CDF, choose some arbitrary $\gamma < 0$, then integrating the PDF from 0 to x or using Lévy's inversion formula [11, p. 199] gives

$$\begin{aligned}
 F_M(x) &= \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} \left[\frac{1}{s} - \frac{e^{-sx}}{s} \right] \phi_M(s) ds \\
 &= \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \left[\frac{2Ne^{-sx}}{-2Ns} - \frac{1}{-s} \right] \phi_M(s) ds \\
 &= \frac{2N}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{-sx} \frac{\Gamma(-2Ns)}{\Gamma(1-2Ns)} \phi_M(s) ds \\
 &\quad - \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \frac{1}{-s} \phi_M(s) ds.
 \end{aligned}$$

The path can be shifted by γ since no poles are crossed. Consider the first integral. It is of the form of (27) and is given by $2NK$ multiplied by

$$H_{m+1,m+1}^{m+1,0} \left[e^{-x} \left| \begin{matrix} (N + \eta_1, 2N), \dots, (N + \eta_m, 2N), (1, 2N) \\ (N + \xi_1, 2N), \dots, (N + \xi_m, 2N), (0, 2N) \end{matrix} \right. \right].$$

We now apply (39) with $\lambda = 1/(2N)$, and then (40) with $c = N$. The first integral is then $K \exp(-x/2)$ multiplied by

$$H_{m+1,m+1}^{m+1,0} \left[e^{-x/(2N)} \left| \begin{matrix} (\eta_1, 1), \dots, (\eta_m, 1), (1 - N, 1) \\ (\xi_1, 1), \dots, (\xi_m, 1), (-N, 1) \end{matrix} \right. \right].$$

Application of (28) then means that the first integral can be written as (31).

Taking $\gamma \rightarrow -\infty$ and using (42), for which we know that $f > 2$, the second integral converges to 0; the CDF will be uniformly continuous.

ACKNOWLEDGEMENTS

Paul Ginzberg thanks the EPSRC (UK) for financial support. Helpful comments by the referees were appreciated.

REFERENCES

- [1] S. A. Andersson, H. K. Brøns and S. T. Jensen, "Distribution of eigenvalues in multivariate statistical analysis," *Ann. Stat.*, vol. 11, pp. 392–415, 1983.
- [2] G. E. P. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, vol. 36, pp. 317–346, 1949.
- [3] B. D. Carter and M. D. Springer, "The distribution of products, quotients and powers of independent H -function variates," *SIAM J. Applied Math.*, vol. 33, pp. 542–58, 1977.

- [4] L. J. Gleser and I. Olkin, "A note on Box's general method of approximation for the null distributions of likelihood criteria," *Ann. Inst. Statist. Math.*, vol. 27, pp. 319–326, 1975.
- [5] A. Grandi, A. Mazzotti and E. Stucchi, "Multicomponent velocity analysis with quaternions," *Geophysical Prospecting*, vol. 55, pp. 761–777, 2007.
- [6] A. K. Gupta and J. Tang, "On a general distribution theory for a class of likelihood criteria," *Australian & New Zealand J. of Statist.*, vol. 30, pp. 359–366, 1988.
- [7] J. L. Jensen, "A large deviation-type approximation for the "Box class" of likelihood ratio criteria," *J. Amer. Statist. Assoc.*, vol. 86, pp. 437–440, 1991.
- [8] J. L. Jensen, "Correction: A large deviation-type approximation for the "Box class" of likelihood ratio criteria," *J. Amer. Statist. Assoc.*, vol. 90, p. 812, 1995.
- [9] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, London: Griffin, 1977.
- [10] N. Le Bihan and J. Mars, "Singular value decomposition of quaternion matrices: a new tool for vector-sensor signal processing," *Signal Process.*, vol. 84, pp. 1177–1199, 2004.
- [11] M. Loève, *Probability Theory, Volume 1, 4th Ed.*, New York: Springer-Verlag, 1977.
- [12] R. Lugannani and S. O. Rice, "Saddlepoint approximation for the distribution of the sum of independent random variables," *Advances Appl. Probab.*, vol. 12, pp. 475–490, 1980.
- [13] A. M. Mathai, R. K. Saxena and H. J. Haubold, *The H-Function: Theory and Applications*, Springer, 2009.
- [14] J. W. Mauchly, "Significance test for sphericity of a normal n -variate distribution," *Ann. Math. Statist.*, vol. 11, pp. 204–209, 1940.
- [15] G. M. Menanno and N. Le Bihan, "Quaternion polynomial matrix diagonalization for the separation of polarized convolutive mixture," *Signal Process.*, vol. 90, pp. 2219–2231, 2010.
- [16] G. M. Menanno, "Seismic multicomponent deconvolution and wavelet estimation by means of quaternions," Ph.D. dissertation, Dept. Earth Sci., Univ. Pisa, Italy, 2010.
- [17] S. Miron, N. Le Bihan and J. Mars, "Quaternion-MUSIC for vector-sensor array processing," *IEEE Trans. Signal Process.*, vol. 54, pp. 1218–1229, 2006.
- [18] J. Möller, "Bartlett adjustments for structured covariances," *Scand. J. Statist.*, vol. 13, pp. 1–15, 1986.
- [19] E. Olilla and V. Koivunen, "Adjusting the generalized likelihood ratio test of circularity robust to non-normality," in *IEEE 10th Int. Workshop Signal Process. Advances in Wireless Communications*, Perugia, Italy, pp. 558–562, 2009.
- [20] J. Seberry, K. Finlayson, S. S. Adams, T. A. Wysocki, T. Xia and B. J. Wysocki, "The theory of quaternion orthogonal designs," *IEEE Trans. Signal Process.*, vol. 56, pp. 256–265, 2008.
- [21] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data*. Cambridge, UK: Cambridge University Press, 2010.
- [22] C. Cheong Took and D. P. Mandic, "The quaternion LMS algorithm for adaptive filtering of hypercomplex real world processes," *IEEE Trans. Signal Process.*, vol. 57, pp. 1316–1327, 2009.
- [23] C. Cheong Took and D. P. Mandic, "A quaternion widely linear adaptive filter," *IEEE Trans. Signal Process.*, vol. 58, pp. 4427–4431, 2010.

- [24] C. Cheong Took and D. P. Mandic, "Augmented second-order statistics of quaternion random signals," *Signal Process.*, vol. 91, pp. 214–24, 2011.
- [25] D. E. Tyler, "Robustness and efficiency properties of scatter matrices," *Biometrika*, vol. 70, pp. 411–420, 1983.
- [26] N. N. Vakhania, "Random vectors with values in quaternion Hilbert spaces," *Theory Probab. Appl.*, vol. 43, pp. 99–115, 1999.
- [27] J. Vía, D. P. Palomar and L. Vielva "Generalized likelihood ratios for testing the properness of quaternion Gaussian vectors," *IEEE Trans. Signal Process.*, to be published.
- [28] J. Vía, D. Ramirez and I. Santamaria, "Properness and widely linear processing of quaternion random vectors," *IEEE Trans. Information Theory*, vol. 56, pp. 3502–15, 2010.
- [29] J. Vía, D. Ramírez, I. Santamaría and L. Vielva "Widely and semi-widely linear processing of quaternion vectors," in *Proc. 2010 IEEE Int. Conf. Acoustics Speech and Signal Process.*, Dallas, TX, pp. 3946–3949, 2010.
- [30] A. T. Walden and P. Rubin-Delanchy, "On testing for impropriety of complex-valued Gaussian vectors," *IEEE Trans. Signal Process.*, vol. 57, 825–834, 2009.
- [31] A. T. A. Wood, J. G. Booth and R. W. Butler, "Saddlepoint approximations to the CDF of some statistics with nonnormal limit distributions," *J. Amer. Statist. Assoc.*, vol. 88, pp. 680–686, 1993.
- [32] T. A. Wysocki, B. J. Wysocki and S. Spence Adams, "Correction to The theory of quaternion orthogonal designs," *IEEE Trans. Signal Process.*, vol. 57, 3298, 2009.
- [33] G. A. Young and R. L. Smith, *Essentials of Statistical Inference*. Cambridge UK: Cambridge University Press, 2005.



Paul Ginzberg received the M.Sci. degree in mathematics in 2009 from Imperial College London, U.K., where he is currently pursuing a Ph.D. in statistics. His interests include quaternion signal processing and wavelets.



Andrew T. Walden (A'86-M'07-SM'11) received the B.Sc. degree in mathematics from the University of Wales, Bangor, U.K., in 1977, and the M.Sc. and Ph.D. degrees in statistics from the University of Southampton, Southampton, U.K., in 1979 and 1982, respectively. He was a Research Scientist at BP, London, U.K., from 1981 to 1990, and then joined the Department of Mathematics at Imperial College London, London, U.K., where he is currently a Professor of statistics.