A MIMO Version of the Reed-Yu Detector and its Connection to the Wilks Lambda and Hotelling T^2 Statistics

Ronald W. Butler[1], Pooria Pakrooh[2], Member IEEE, and Louis L. Scharf[3], Life Fellow IEEE

[1]Department of Statistical Sciences, Southern Methodist University, Dallas, TX, USA

[2]Qualcomm, Inc, San Diego, CA, USA

[3]Departments of Mathematics and Statistics, Colorado State University, Fort Collins, CO, USA

Abstract

In this paper we study the problem of detecting a known signal transmitted over a MIMO channel of unknown complex gains and additive noise of unknown covariance. The problem arises in many contexts, including transmit-receiver synchronization. We derive the exact probability distribution for a generalized likelihood ratio (GLR) statistic, and establish the connection between this statistic and the Wilks Lambda and Hotelling T^2 statistics. We give alternatives to the GLR statistic, which include the Bartlett-Nanda-Pillai trace, the Lawley-Hotelling trace, and the Roy maximum eigenvalue statistics, each of which is favored under special conditions on the actual MIMO channel. For example, if the channel is an incoherent scattering channel, then the competition for greatest power is among the Bartlett-Nanda-Pillai, Lawley-Hotelling, and GLR statistics. If it is a coherent channel that supports a propagating wave, then Roy's test is more powerful. We discuss the null distribution theory of the GLR at length to show how it may be used to accurately predict false alarm probabilities.

Keywords— signal detection, likelihood ratio, MIMO communication, RX detector, Hotelling's T^2, Wilks Lambda, Bartlett-Nanda-Pillai trace, Lawley-Hotelling trace, Roy's maximum eigenvalue, matrix beta distribution, large random matrices, saddlepoint approximation

This work is supported in part by NSF under grant CCF-1712788, and by the Air Force Office of Scientific Research under award number FA9550-14-1-0185. Consequently the U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

I. Introduction

In this paper we re-visit the problem of detecting a known signal transmitted over a MIMO channel of unknown complex gains and additive noise of unknown covariance. The problem arises in many contexts, including transmit-receiver synchronization. The first contribution in the engineering literature to this general problem was made by Reed and Yu [1], who derived the probability distribution for a generalized likelihood ratio test in the case of a SIMO channel. Their detector was generalized to the MIMO case by Bliss and Parker [2]. Hiltunen, Loubaton, and Chevalier [3], [4] used large random matrix theory to derive asymptotic approximations for the distribution of the MIMO detector.

We derive the generalized likelihood ratio (GLR) for the Reed-Yu problem, as generalized by Bliss and Parker, and establish its connection to the Wilks Lambda and Hotelling T^2 statistics. We establish its exact probability distribution by deriving its moment generating function. Our results are based on the exact distribution theory for finite beta-distributed random matrices, whereas the results of [3], [4] are asymptotic and based on large random matrix theory. With these exact distributions, it is now known what distribution is being approximated by the asymptotic approximations of [3], [4]. It is now known that the MIMO version of the Reed-Yu detector is, in fact, a Wilks Lambda statistic, and that it generalizes the Hotelling T^2 statistic.

Under the null hypothesis of no signal in a measurement matrix, we show that the GLR is CFAR, and equal in distribution to the product of independent beta-distributed random variables, a result previously reported in [5]. This stochastic representation offers a simple and numerically-stable method for predicting false alarm probability, without computing determinants of large random matrices. Moreover, we show that under the null the moment generating function is a rational function, which can be exactly inverted for the density of the GLR in finite parameter regimes where asymptotic approximations may not accurately predict performance. Then, to demonstrate the importance of saddlepoint approximations for inverting moment generating functions, we compute saddlepoint approximations to the null survival function, which provides the probability of false alarm versus threshold setting. These approximations are then compared with those suggested by large random matrix theory.

The GLR detector statistic is derived from maximum likelihood reasoning. Moreover, it is derived under the assumption that the unknown channel matrix has no constraining structure. When it does, then there are alternatives that are more powerful, which is to say their ROC curves would lie above the ROC curve of the Wilks Lamba. For example, if the channel is a coherent channel that supports a propagating wave, then the channel matrix is near to rank-one, and Roy's maximum eigenvalue test would be preferred over the Wilks Lambda. If the channel is an incoherent scattering channel, then the competition for greatest power is among the Bartlett-Nanda-Pillai, Lawley-Hotelling, and GLR statistics. If the eigenvalues of the channel matrix are equal, then the Bartlett-Nanda-Pillai is more powerful, and with a mix of small and large eigenvalues either Lawley-Hotelling or GLR will be most powerful.

Contributions. In summary, this paper complements the papers of Bliss and Parker [2], and Hiltunen, et al. [3], [4], in the following ways :

1) it is now known that the MIMO version of the Reed-Yu detector statistic, derived by Bliss and Parker [2], is in fact a Wilks Lambda statistic, and a generalized version of the Hotelling T^2 statistic;
2) it is now known what distribution is approximated by the large random matrix results of [3], [4];
3) under the null, the GLR statistic is equal in distribution to a product of independent beta-distributed random variables, thereby providing a preferred alternative to the direct simulation of the detector statistic for Monte Carlo prediction of false alarm probability;
4) under the null the moment generating function of the GLR statistic is a rational function, which means it may be inverted exactly, and this inversion is practical in parameter regimes where asymptotic results may not be accurate;
5) under the null, the moment generating function may be inverted by saddlepoint methods to return highly accurate approximations throughout the support of the distribution. Such approximations preserve relative error in the extreme right tail of the distribution in large-deviations regions where simulations and normal approximations fail to provide adequate relative accuracy.

Our results are presented in the style of the literature on multivariate analysis, where the distributions of many important statistics are known by their moment generating functions, even though inversion of these functions may lie beyond the power of current methods, such as saddlepoints, to detemine them for their probability density functions (pdfs) and cumulative distribution functions (cdfs). Throughout the paper we direct the reader to continuing research on the inversion of moment generating functions by the method of saddlepoints.

Under the null our results are to be preferred over asymptotic results, because they can be used with saddlepoint approximations to efficiently and stably predict false alarm probabilities with very high accuracy, even in the extreme right tail where small false alarm probabilities are to be computed. This is demonstrated with several numerical examples. Under the alternative, the inversion of the moment generating function (MGF) of the GLR is a work in progress, which builds on the non-cental distribution theory developed herein.

II. Problem Statement

Consider a subspace signal-plus-noise model $\mathbf{x}_m = \mathbf{H}\mathbf{s}_m + \mathbf{n}_m$, for $m = 1, \ldots, M$. The signal component lies in an unknown p-dimensional subspace $\langle \mathbf{H} \rangle$, with unknown basis $\mathbf{H} \in \mathbb{C}^{L \times p}$. For each time sample, its location in this subspace is determined by the vector of signals $\mathbf{s}_m \in \mathbb{C}^p$. The noise snapshots $\mathbf{n}_m$ are proper and independent for $m = 1, \ldots, M$, and distributed as $\mathbf{n}_m \sim \mathcal{CN}_L[\mathbf{0}, \mathbf{\Sigma}]$, with $\mathbf{\Sigma}$ a positive definite covariance matrix. Thus, the measurements $\mathbf{x}_m$ are independent and distributed as $\mathbf{x}_m \sim \mathcal{CN}_L[\mathbf{H}\mathbf{s}_m, \boldsymbol{\Sigma}]$, with $\mathbf{H}$ and $\boldsymbol{\Sigma}$ unknown; the signal sequence $\{\mathbf{s}_m, m = 1, 2, \ldots, M\}$ is known. We assume $M > L + p$, which is to say the number of measurements, ML, exceeds the number of unknowns, L^2, in the noise covariance matrix $\boldsymbol{\Sigma}$ plus the number of unknowns, $2Lp$, in the unknown channel matrix $\mathbf{H}$.

This data model corresponds to one channel use of a multiple input multiple output (MIMO) transmission system with p transmitting antennas, L receiving antennas, and M symbol transmissions, when the transmitting and receiving antennas are perfectly synchronized [2]. We are interested in testing

$$H_0 : \mathbf{S} = \mathbf{0} \quad \text{versus} \quad H_1 : \text{ known full row rank } \mathbf{S} \neq \mathbf{0} \tag{1}$$

Define $\mathbf{S}^H = [\mathbf{s}_1, \ldots, \mathbf{s}_M] \in \mathcal{C}^{p \times M}$, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_M] \in \mathcal{C}^{L \times M}$, and $\mathbf{N} = [\mathbf{n}_1, \ldots, \mathbf{n}_M] \in \mathcal{C}^{L \times M}$. Then, the data matrix $\mathbf{X} = \mathbf{H}\mathbf{S}^H + \mathbf{N}$ is distributed as

$$\begin{aligned} f(\mathbf{X}, \mathbf{H}, \mathbf{S}, \boldsymbol{\Sigma}) = & \frac{1}{\pi^{LM} (\det \boldsymbol{\Sigma})^M} \times \\ & \exp\{-\mathrm{tr}[(\mathbf{X} - \mathbf{H}\mathbf{S}^H)^H \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \mathbf{H}\mathbf{S}^H)]\} \end{aligned} \tag{2}$$

The notation $\mathbf{S}^H$ for the $p \times M$ symbol matrix is merely a convenience to make formulas familiar.

In this paper, we assume $\mathbf{S}$ is known, but the channel gains $\mathbf{H}$ and the noise covariance $\boldsymbol{\Sigma}$ are unknown. We make no assumptions on the channel matrix $\mathbf{H}$, but once the detector is derived, we discuss alternative detectors that would provide more power than others, when the true $\mathbf{H}$ is structured. This way of constructing an unconstrained detector, followed by approximation for structured $\mathbf{H}$, stands in contrast to the problem of structuring $\mathbf{H}$ up-front, and then trying to solve an optimization problem for $\mathbf{H}$, so structured.

For $p = 1$, the Generalized Likelihood Ratio Test (GLRT) for this measurement model has been derived in [1], for a problem of optical pattern detection with unknown spectral distribution. For $p \geq 1$ Bliss and Parker [2] generalized this result for synchronization in a MIMO channel. The analysis in [1] considers the special case of a real measurement model for the rank one signal model $p = 1$. The analysis in [2] assumes complex measurements and a rank-p signal model. We follow the approaches of [1], [2] to identify Maximum Likelihood (ML) estimates of the unknown parameters, and use these estimates to form the GLR for the hypothesis test (1). Then we extend the distribution results of [1], [5] for the detector statistic.

III. Generalized Likelihood Ratio Test and Its Distribution

To find the GLR for this problem, we need the ML estimate of $\boldsymbol{\Sigma}$ under H_0 and H_1, and the ML estimate of $\mathbf{H}$ under H_1. Under H_0 the ML estimate of the covariance matrix $\boldsymbol{\Sigma}$ is

$$\hat{\boldsymbol{\Sigma}}_0 = \frac{1}{M}\mathbf{X}\mathbf{X}^H \tag{3}$$

Similarly, under H_1, the ML estimates of $\mathbf{\Sigma}$ and $\mathbf{H}$ are

$$\hat{\mathbf{\Sigma}}_1 = \frac{1}{M}(\mathbf{X} - \hat{\mathbf{H}}\mathbf{S}^H)(\mathbf{X} - \hat{\mathbf{H}}\mathbf{S}^H)^H, \tag{4}$$

and

$$\hat{\mathbf{H}} = \mathbf{X}\mathbf{S}(\mathbf{S}^H\mathbf{S})^{-1}. \tag{5}$$

The GLR $\ell = f_{H_0}(\mathbf{X}, \mathbf{HS} = \mathbf{0}, \hat{\mathbf{\Sigma}}_0)/f_{H_1}(\mathbf{X}, \hat{\mathbf{H}}, \mathbf{S}, \hat{\mathbf{\Sigma}}_1)$ is then

$$\ell^{1/M} = \frac{\det(\mathbf{X}(\mathbf{I}_M - \mathbf{P}_\mathbf{S})\mathbf{X}^H)}{\det(\mathbf{X}\mathbf{X}^H)}. \tag{6}$$

where $\mathbf{P}_\mathbf{S} = \mathbf{S}(\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H$ is the rank-p projection onto the subspace $\langle\mathbf{S}\rangle$, spanned by the linearly independent columns of the $M \times p$ matrix $\mathbf{S}$. The columns of $\mathbf{X}$ are assumed to be statistically independent, so that $\mathbf{X}\mathbf{X}^H$ is full rank with probability one.

The distribution of (6) has been derived for finite L and M in [1] for the case $p = 1$. Hiltunen, et al. [3] show that in the case where the number of receiving antennas L and the number of snapshots M are large and of the same order of magnitude, but the number of transmitting antennas p remains fixed, a standardized version of -log(ℓ) converges to a normal distribution under H_0 and H_1. Then, pragmatic approximations for the distribution are derived for large p. In [4], it is shown that asymptotically in L, M, and p, log(ℓ) is approximated as a normal random variable (See Theorem 1 of [3]). In [5], we outlined a derivation of the exact distribution of (6) under H_0 and H_1, in the non-asymptotic case. We refine this derivation here, with detail added to the original proof, and establish the connection between the GLR, the Wilks Lambda, and Hotelling T^2 statistics.

The distribution of ℓ is invariant to the transformation $\mathbf{X} \rightarrow \mathbf{\Sigma}^{-1/2}\mathbf{X}$. Thus we may assume $\mathbf{X}$ to be proper, and distributed as the proper complex normal, $\mathcal{CN}_{L\times M}(\mathbf{\Sigma}^{-1/2}\mathbf{H}\mathbf{S}^H, \mathbf{I}_L \otimes \mathbf{I}_M)$. Let $\mathbf{U} = [\mathbf{U}_\mathbf{S}, \mathbf{U}_\mathbf{S}^\perp] \in \mathcal{C}^{M\times M}$, where $\mathbf{U}_\mathbf{S} \in \mathcal{C}^{M\times p}$ is a unitary basis for $\mathbf{S}$, and $\mathbf{P}_\mathbf{S} = \mathbf{U}_\mathbf{S}\mathbf{U}_\mathbf{S}^H$; $\mathbf{U}_\mathbf{S}^\perp \in \mathcal{C}^{M\times(M-p)}$ is a unitary basis for its orthogonal space. The matrix $\mathbf{U}$ is unitary. Define $L \times p$ $\mathbf{Y}_1 = \mathbf{X}\mathbf{U}_\mathbf{S}$, and $L \times (M-p)$ $\mathbf{Y}_2 = \mathbf{X}\mathbf{U}_{\mathbf{S}^\perp}$, so that

$$W := \ell^{1/M} = \frac{\det(\mathbf{Y}_2\mathbf{Y}_2^H)}{\det(\mathbf{Y}_1\mathbf{Y}_1^H + \mathbf{Y}_2\mathbf{Y}_2^H)}, \tag{7}$$

which is the same as the Wilks Lambda [6]. Here, $\mathbf{Y}_1 \sim \mathcal{CN}_{L\times p}(\mathbf{M}_1, \mathbf{I}_L \otimes \mathbf{I}_p)$, $\mathbf{Y}_2 \sim \mathcal{CN}_{L\times(M-p)}(\mathbf{0}, \mathbf{I}_L \otimes \mathbf{I}_{M-p})$, are independent, $\mathbf{M}_1 = \mathbf{\Sigma}^{-1/2}\mathbf{H}\mathbf{S}^H\mathbf{U}_\mathbf{S}$, and $\mathbf{Y}_2\mathbf{Y}_2^H \sim \mathcal{CW}(L, M-p, \mathbf{I}_L)$, a complex Wishart distribution for the sample covariance matrix of $M-p$ draws of an L-dimensional complex Gaussian vector whose covariance is $\mathbf{I}_L$. For $M - p \geq L$, as assumed, the matrix $\mathbf{Y}_2\mathbf{Y}_2^H$ is non-singular, with probability one.

A. More antennas than sources, $L \geq p$.

For $L \geq p$, the statistic in (7) may be written as

$$W = \det(\mathbf{Y}_2\mathbf{Y}_2^H)/\det(\mathbf{Y}_1\mathbf{Y}_1^H + \mathbf{Y}_2\mathbf{Y}_2^H) = 1/\det\big[\mathbf{I}_L + (\mathbf{Y}_2\mathbf{Y}_2^H)^{-1/2}\mathbf{Y}_1\mathbf{Y}_1^H(\mathbf{Y}_2\mathbf{Y}_2^H)^{-1/2}\big]$$

$$= 1/\det\big[\mathbf{I}_p + \mathbf{Y}_1^H(\mathbf{Y}_2\mathbf{Y}_2^H)^{-1}\mathbf{Y}_1\big] = \det\big[(\mathbf{I}_p + \mathbf{F})^{-1}\big] \quad (8)$$

$$= \det(\mathbf{B}). \quad (9)$$

where $\mathbf{B} = (\mathbf{I}_p + \mathbf{F})^{-1}$. The pdf of $\mathbf{F} = \mathbf{Y}_1^H(\mathbf{Y}_2\mathbf{Y}_2^H)^{-1}\mathbf{Y}_1$ is given in [7] as

$$f(\mathbf{F}) = e^{-\text{tr}(\mathbf{M}_1^H\mathbf{M}_1)}{}_1\tilde{F}_1(M; L; \mathbf{M}_1^H\mathbf{M}_1(\mathbf{I} + \mathbf{F}^{-1})^{-1})\times \frac{\tilde{\Gamma}_p(M)}{\tilde{\Gamma}_p(M-L)\tilde{\Gamma}_p(L)}\frac{\det(\mathbf{F})^{L-p}}{\det(\mathbf{I}+\mathbf{F})^M} \quad (10)$$

which is a *complex* noncentral matrix $\mathbf{F}$ distribution. The Jacobian for the transformation $\mathbf{F} \to \mathbf{B}$ is $(\det \mathbf{B})^{-2p}$, so the pdf of $\mathbf{B}$ may be written as

$$f(\mathbf{B}) = e^{-\text{tr}(\mathbf{M}_1^H\mathbf{M}_1)}{}_1\tilde{F}_1(M; L; \mathbf{M}_1^H\mathbf{M}_1(\mathbf{I} - \mathbf{B}))\frac{\tilde{\Gamma}_p(M)}{\tilde{\Gamma}_p(M-L)\tilde{\Gamma}_p(L)}\times \det(\mathbf{B})^{M-L-p}\det(\mathbf{I}-\mathbf{B})^{L-p}, \quad (11)$$

which may be considered a complex noncentral matrix variate Beta distribution, and denoted $\mathcal{CB}_p(M-L, L, \mathbf{M}_1^H\mathbf{M}_1)$.

Under H_0 ($\mathbf{S} = \mathbf{0}$), $W = \det(\mathbf{B})$ is distributed as the product of independent beta random variables. That is,

$$W = \det(\mathbf{B}) \sim \prod_{i=1}^{p} b_i; \quad b_i \sim \beta(M - L - i + 1, L). \quad (12)$$

This is a stochastic representation of the GLR under the null. It offers numerically-stable stochastic simulation of the GLR, without simulation of the detector statistic itself, which may involve the computation of determinants of large matrices. Moreover, from this representation, the MGF of the GLR may be derived. The pdf may be computed to a high degree of accuracy using saddlepoint inversion of the MGF as shown in [8], Sections 2.4.1 and 11.1.1, for the real case. To this end, the MGF of $Z = \log\det(\mathbf{B})$ and its saddlepoint inversions are treated in Appendix, Section A, where numerical results are given for parameter choices made in [3],[4].

Under H_1 we derive the MGF for $Z = \log W = \log\det(B)$ in Appendix, Section B. Of course the result specializes to the MGF under the null. The proof follows the approach used in Muirhead (1982, Thm. 10.5.1) [9], extended to apply to complex matrix beta distributions. The result as expected doubles the degrees of freedom.

THEOREM 1. *Assume $p \leq L$. Under H_1, $\mathcal{M}_Z(s) = E(W^s)$ is*

$$\mathcal{M}_Z(s) = \frac{\tilde{\Gamma}_p(M-L+s)\tilde{\Gamma}_p(M)}{\tilde{\Gamma}_p(M-L)\tilde{\Gamma}_p(M+s)}\,{}_1\tilde{F}_1(s; M+s; -\mathbf{M}_1^H\mathbf{M}_1) \quad (13)$$

for $\text{Re}(s) > p + L - M - 1 < 0$.

PROOF. See Appendix, Section B.

In the noncentral setting, saddlepoint approximations can be used to invert the MGF in (13) but this is complicated by the difficulty in computing the ${}_1\tilde{F}_1$ hypergeometric function. To get around this in the real case using Muirhead's (1982, Thm. 10.5.1) [9] expression for the MGF, Butler and Wood (2005) [10] replaced the true value of ${}_1F_1$ with an explicit Laplace approximation to get an approximation to $E(e^{sZ})$, which could be subsequently inverted using saddlepoint approximations; see [8], Section 11.3.1, for details. The same approach applied to (13) using Laplace approximation for ${}_1\tilde{F}_1$ allows for approximation to the non-null distribution and density functions of $Z = \log W$ and W. Such a Laplace approximation for ${}_1\tilde{F}_1$ is proposed in Butler and Wood (2020) [11]. In the real case, these saddlepoint approximations achieve a high degree of accuracy when the eigenvalues of $\mathbf{M}_1^H\mathbf{M}_1$ are not exceptionally large. Similar results should apply in the complex case.

B. More sources than antennas, $p > L$

For $p > L$, there is a parallel development that produces the $L \times L$ matrix $\mathbf{B} = (\mathbf{Y}_1\mathbf{Y}_1^H + \mathbf{Y}_2\mathbf{Y}_2^H)^{-1/2}\mathbf{Y}_2\mathbf{Y}_2^H(\mathbf{Y}_1\mathbf{Y}_1^H + \mathbf{Y}_2\mathbf{Y}_2^H)^{-H/2}$.

From arguments given in Appendix, Section B, the distribution of W under H_0 is

$$W = \det(\mathbf{B}) \sim \prod_{i=1}^{L} b_i;\ \ b_i \sim \beta(M - p - i + 1, p) \tag{14}$$

which is (12) with the roles of p and L interchanged. This interchange is also apparent in the noncentral MGF of $Z = \log W$ given in the next corollary, which is proved in Appendix C.

COROLLARY 1. *Assume $p > L$. Under H_1, $\mathcal{M}_Z(s) = E(W^s)$* is

$$\mathcal{M}_Z(s) = \frac{\tilde{\Gamma}_L(M - p + s)\tilde{\Gamma}_L(M)}{\tilde{\Gamma}_L(M - p)\tilde{\Gamma}_L(M + s)}\, {}_1\tilde{F}_1(s; M + s; -\mathbf{M}_1\mathbf{M}_1^H) \tag{15}$$

for $\mathrm{Re}(s) > p + L - M - 1 < 0$.

PROOF. See Appendix C.

IV. CONNECTION TO HOTELLING'S T^2 STATISTIC.

In the special case $p = 1$ then $F = \mathbf{Y}_1^H(\mathbf{Y}_2\mathbf{Y}_2^H)^{-1}\mathbf{Y}_1$ is a real scalar. Upon rescaling by the degrees of freedom ratio $(M - L)/L$, its noncentral density given in (10) reduces to that of a Noncentral $F\{2L; 2(M - L); 2\mathbf{M}_1^H\mathbf{M}_1\}$ distribution with noncentrality parameter $2\mathbf{M}_1^H\mathbf{M}_1$. This reproduces the distribution determined in Srivastava and Khatri [12].

For the symbol sequence $\mathbf{s}^H = \mathbf{1}^H$, $\mathbf{Y}_1 = \mathbf{X}\mathbf{1}/\sqrt{M} = \sqrt{M}\bar{\mathbf{x}}$ and $\mathbf{Y}_2\mathbf{Y}_2^H = \mathbf{X}\mathbf{X}^H - M\bar{\mathbf{x}}\bar{\mathbf{x}}^H$ so that

$$F = \sqrt{M}\bar{\mathbf{x}}\{\mathbf{X}\mathbf{X}^H - M\bar{\mathbf{x}}\bar{\mathbf{x}}^H)^{-1}\sqrt{M}\bar{\mathbf{x}} = T_0/(1 - T_0)$$

where $T_0 = M\bar{\mathbf{x}}^H(\mathbf{X}\mathbf{X}^H)^{-1}\bar{\mathbf{x}}$. Thus, F and W are monotone increasing in Hotelling's T^2 statistic. For further emphasis, the argument for $W = \det(\mathbf{B})$ proceeds as follows. When $p = 1$, meaning $\mathbf{H} = \mathbf{h} \in \mathcal{C}^L$, and the symbol sequence $\mathbf{s}^H = \mathbf{1}^T$, the likelihood may be written

$$W = \frac{\det\left[\mathbf{X}(\mathbf{I}_M - \mathbf{P_1})\mathbf{X}^H\right]}{\det\left[\mathbf{X}\mathbf{X}^H)\right]} = \frac{\det(\mathbf{X}\mathbf{X}^H - M(\sum_1^M \mathbf{x}[m])(\sum_1^M \mathbf{x}[m])^H}{\det(\mathbf{X}\mathbf{X}^H)}$$
$$= 1 - \sqrt{M}(\sum_1^M \mathbf{x}[m])^H(\mathbf{X}\mathbf{X}^H)^{-1}(\sum_1^M \mathbf{x}[m])\sqrt{M} \quad (16)$$

The monotone function $(1/M)(1-W) = (\sum_1^M \mathbf{x}[m])^H(\mathbf{X}\mathbf{X}^H)^{-1}(\sum_1^M \mathbf{x}[m])$ is Hotelling's T^2 statistic.

So the multi-rank version of the RX problem is a generalization of the Hotelling problem, where Hotelling's constant but unknown $\mathbf{h}$ is replaced by a sequence of unknown $\mathbf{H}\mathbf{s}[m]$, with the linear combining weights $\mathbf{s}[m]$ known, but the common channel matrix $\mathbf{H}$ unknown.

For $p = 1$, the test based on F is the uniformly most powerful (UMP) invariant test among tests for H_0 versus H_1 at fixed level α. The uniformity is over all non-zero values of the $M \times 1$ symbol matrix $\mathbf{S}$. The invariance uses the group of transformations $\mathbf{X} \to \mathbf{N}\mathbf{X}$ where $\mathbf{N}$ is any $L \times L$ nonsingular complex matrix. Starting with sufficient statistics $\mathbf{Y}_1$ and $\mathbf{Y}_2\mathbf{Y}_2^H$, it is easily shown that F is the maximal invariant and $\mathbf{M}_1^H\mathbf{M}_1$ is the associated maximal invariant parameter. Since the noncentral F distribution is known to possess the monotone likelihood ratio property (Lehmann and Romano, 2006, p. 307, problem 7.4) [13], we conclude that the test which rejects for large F is UMP invariant. For the i.i.d. case in which $\mathbf{s}_m = 1$ for all m, this result was shown in Giri (1965, Thm. 3.2) [14].

V. Related Tests: Bartlett-Nanda-Pillai, Lawley-Hotelling, and Roy

The hypothesis test may also be addressed by using three other competing test statistics as alternatives to the likelihood ratio test. For the case $p = 1$, all four tests reduce to the use of the Hotelling T^2 test statistic, which is UMP invariant. For the case $p > 1$, however, no single test can be expected to dominate the others in terms of power. The test which achieves the greatest power depends upon the configuration of eigenvalues for the noncentrality matrix $\mathbf{M}^H\mathbf{M}$.

The three other tests use the Bartlett-Nanda-Pillai trace statistic $V = \text{tr}(\mathbf{B})$, the Lawley-Hotelling trace statistic $U = \text{tr}(\mathbf{F}) = \text{tr}\{\mathbf{B}^{-1}(\mathbf{I}_p - \mathbf{B})\}$, and Roy's maximum root λ_1, the largest eigenvalue of $\mathbf{B}$.

The tests are broadly divided into two groups: $W, V,$ and U which often share similar power and λ_1 with a quite different power level. Roy's test using λ_1 should be more powerful than the other three when the noncentrality matrix $\mathbf{M}_1^H\mathbf{M}_1$ admits a single non-zero eigenvalue. This dominance diminishes quickly when $\mathbf{M}_1^H\mathbf{M}_1$ admits further non-zero eigenvalues. When the eigenvalues of $\mathbf{M}_1^H\mathbf{M}_1$ are very unequal, then typically U is more powerful than W which is more powerful than V. If the eigenvalues are close, then this ordering of power gets reversed. For further discussion see Muirhead [9], Section 10.5.6, and the references therein.

Apart from W, the non-null distribution theory for the other three statistics is quite difficult, theoretically and computationally, in both the real and complex cases. However, null distributions in the complex case are tractable and we consider each of the three null distributions below.

The MGF for V is computed directly from the integral representation of ${}_1\tilde{F}_1$ as

$$\mathcal{M}_V(s) = E(e^{sV}) = {}_1\tilde{F}_1(M-L; M; s\mathbf{I}_p)$$

for all $s \in \mathbb{C}$, and $M - L \geq p$. When $p > L$ the MGF is ${}_1\tilde{F}_1(M-p; M; s\mathbf{I}_L)$. The null distribution and density function of V can be computed using a saddlepoint approximation based on $\mathcal{M}_V$ after replacing ${}_1\tilde{F}_1$ with its Laplace approximation as described in Butler and Wood [11].

The MGF for U is computed from the integral representation of $\tilde{\Psi}$, the confluent hypergeometric function of second type given in Chikuse (1975, Eqn. 4.13), [15]. For $p \leq L$, this leads to

$$\mathcal{M}_U(s) = E(e^{sU}) = \frac{\tilde{\Gamma}_p(M)}{\tilde{\Gamma}_p(M-L)}\,\tilde{\Psi}(L; L+p-M; -s\mathbf{I}_p), \qquad \mathrm{Re}(s) \leq 0.$$

For $p > L$, the MGF is

$$\frac{\tilde{\Gamma}_L(M)}{\tilde{\Gamma}_L(M-p)}\tilde{\boldsymbol{\Psi}}(p; p+L-M; -s\mathbf{I}_L), \qquad \mathrm{Re}(s) \leq 0.$$

Saddlepoint methods cannot be used with $\tilde{\Psi}$ or when it is replaced with a Laplace approximation since both are not "steep" at $s = 0$ as discussed in [8], Section 16.2.2. However, numerical inversion along $\mathrm{Re}(s) = -\varepsilon < 0$ for some $\varepsilon > 0$ can be used either with $\tilde{\Psi}$ or its Laplace approximation.

The null distribution of the maximum root test for $p \leq L$ was derived in Khatri [16] and takes the explicit form

$$F(t) = P(\lambda_1 \leq t) = c\det\{\mathbf{A}(t)\},$$

where

$$c = \pi^{p(p-1)}\tilde{\Gamma}_p(M)/\{\tilde{\Gamma}_p(p)\tilde{\Gamma}_p(M-L)\tilde{\Gamma}_p(L)\}$$

and $\mathbf{A}(t) = \{a_{ij}(t)\}$ is a $p \times p$ symmetric matrix with $a_{ij}(t) = \mathcal{B}_t(M-L-p+i+j-1, L-p+1)$ and $\mathcal{B}_t(\alpha, \beta) = \int_0^t u^{\alpha-1}(1-u)^{\beta-1}du$ as the incomplete beta function. The computations for filling out the matrix $\mathbf{A}(t)$ only require the single incomplete beta integral for $a_{11}(t)$. All other entries in $\mathbf{A}(t)$ result from recursive computation using the recursion for $\mathcal{B}_t$ given in (8.17.20) of [17] as

$$\mathcal{B}_t(\alpha+1, \beta) = \frac{1}{\alpha+\beta}\left\{\alpha\mathcal{B}_t(\alpha, \beta) - t^\alpha(1-t)^\beta\right\}.$$

This allows for fast efficient computation, although the computation of $\det[\mathbf{A}(t)]$ for large p is sensitive to numerical error in the computation of the matrix entries. For larger values of p, the Tracy-Widom limiting distribution (of order 2) for a transformed value of λ_1 may be accurately used as given in Theorem 2 of Johnstone (2008) [18].

The case $p > L$ is handled by interchanging the roles of p and L in the expressions above so that $\mathbf{A}(t)$ is now an $L \times L$ matrix.

VI. Numerical Results

Under the null, the distribution of $W = \det(\mathbf{B})$ may be approximated with a histogram computed from Monte-Carlo simulation of a product of independent beta-distributed random variables. The analytical alternative is to invert the moment generating function of W, using the method of saddlepoints or exact inversion of its rational MGF. These methods may be used to predict the probability of false alarm (or equivalently the survival function) with precision, without asymptotic approximation or direct simulation of W.

In this section we use Monte-Carlo simulation of beta-distributed random variables and saddlepoint approximations to compute what may be called exact false alarm probabilities, and compare them with the asymptotic approximations of [3], [4]. The purpose is not to call into question asymptotic results, which for some parameter choices and false alarm probabilities may be quite accurate. Rather it is to show that asymptotic approximations are just that: approximations that are to be used with caution in non-asymptotic regimes.

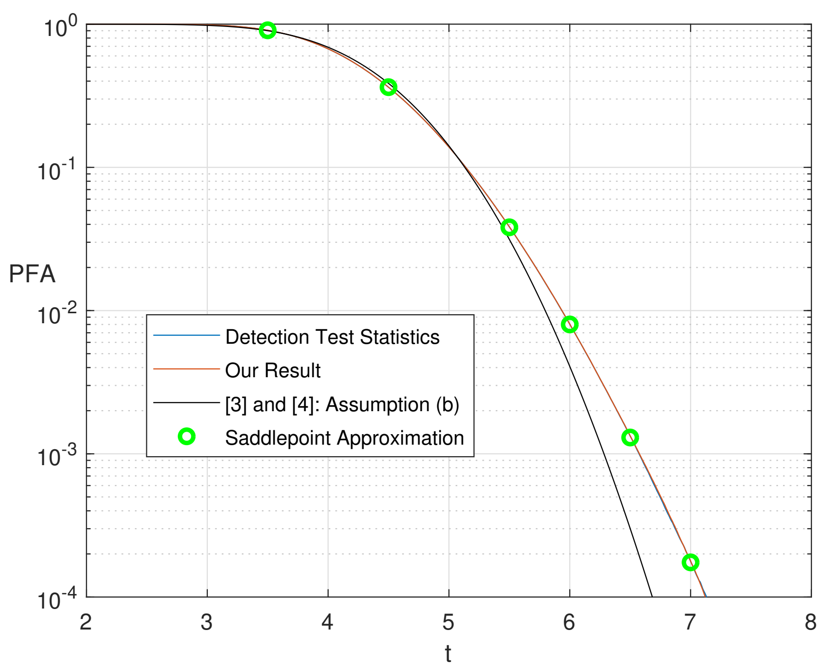


Fig. 1: Probability of false alarm (PFA) on log-scale, $p = 5$ sources, $L = 10$ antenna elements, $M = 20$ snapshots. The blue curve is generated from tail probabilities of a histogram computed from 10^7 realizations of the detector statistic W; the red curve is generated from tail probabilities of a histogram computed from 10^7 realizations of the product of p independent beta-distributed random variables; the circles are computed from saddlepoint inversion of the moment generating function for W; the black curve is computed from the asymptotic approximation of [3], [4], using their approximation (b).

The results are illustrated in Figures 1-3, for parameter choices made in [3], [4]. In each figure false alarm probabilities are predicted from stochastic simulation of a product of betas, from saddlepoint approximation of the null distribution of W, and from the large random matrix approximation. These are compared with the false alarm probabilities predicted from simulation of W itself. These latter are invisible, as they lie exactly under the predictions from the stochastic simulation of a product of betas and from saddlepoint inversion of the moment generating function. The conclusions from these figures is what might be expected: for numbers of sensors less than about 100, number of measurements less than about 200, and number of sources in the range of 5 to 50, and for false alarm probabilities in the range $10^{-4} - 10^{-6}$, the false alarm probabilities predicted from asymptotic approximations can be too low by a factor of 0.1. For example, in Fig. 1 it is demonstrated that when the asymptotic approximation to the probability of false alarm is predicted to be 10^{-4}, the actual probability of false alarm is 10^{-3}. For applications in radar and communications, this has consequences. For other applications it may not.

Details concerning the computation of saddlepoint approximations and their preservation of relative error uniformly in the right tail are given in Appendix A. Exact inversion using the method of partial fraction expansion is also described in Appendix A.

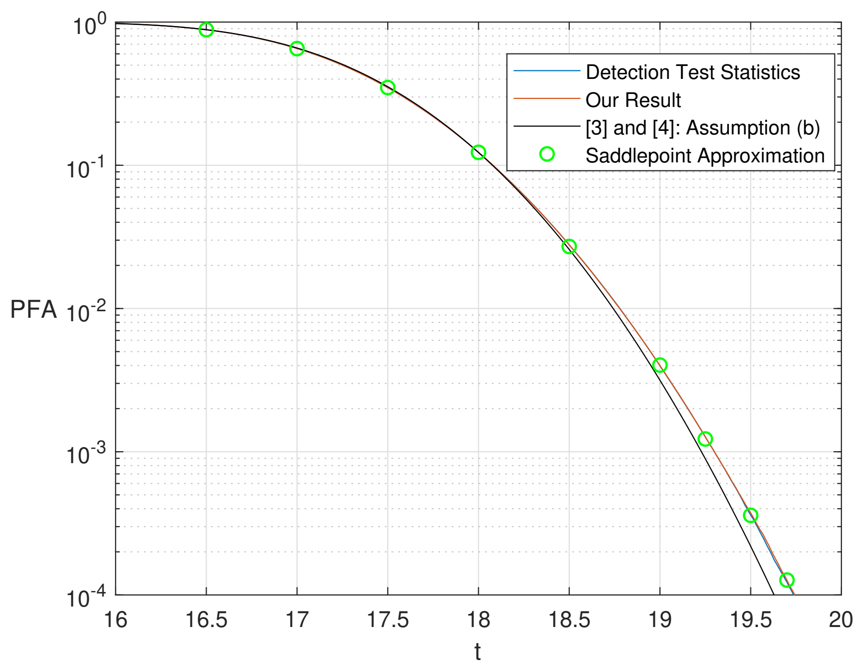


Fig. 2: Probability of false alarm (PFA) on log-scale, $p = 20$ sources, $L = 40$ antenna elements, $M = 80$ snapshots. See the caption of Fig 1 for a description of the plot features.

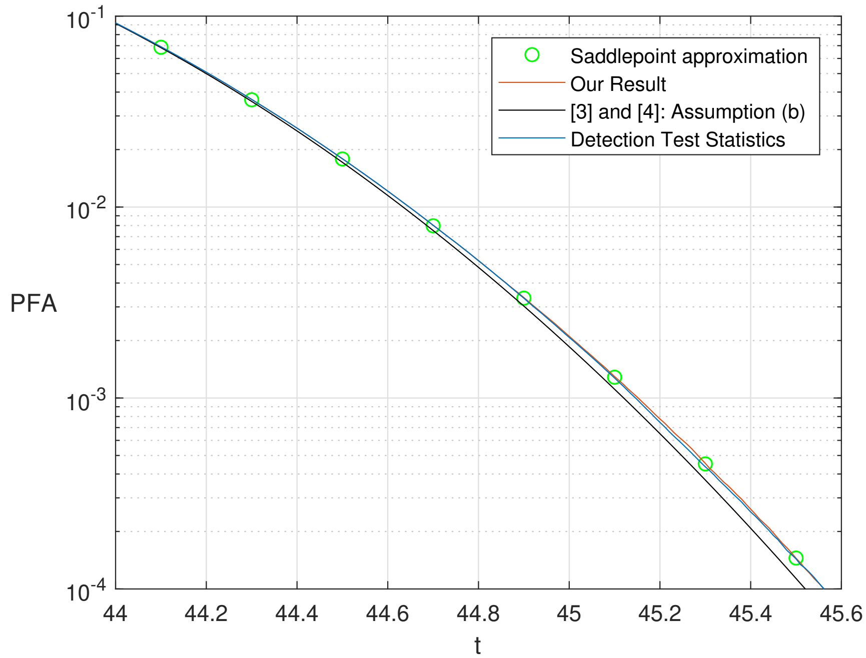


Fig. 3: Probability of false alarm (PFA) on log-scale, $p = 50$ sources, $L = 100$ antenna elements, $M = 200$ snapshots. See the caption of Fig 1 for a description of the plot features.

VII. Conclusions

In this paper we have revisited a detector first derived by Reed and Yu [1], generalized by Bliss and Parker [2], and recently studied by Hiltunen, Loubaton, and Chevalier [3], [4]. The problem is to detect a known signal transmitted over a MIMO channel of unknown complex gains and additive noise of unknown covariance. We have shown that the generalized likelihood ratio (GLR) is a Wilks Lambda statistic, and a generalization of the Hotelling T^2 statistic.

The probability distribution for the GLR was first derived for the SIMO channel in [1]. We have generalized this distribution for the case of a MIMO channel, and shown that the GLR is distributed as the determinant of a complex Beta-distributed matrix, which may be written as a product of independent scalar beta random variables under the null. The moment generating function of the GLR is derived, and saddlepoint techniques for inverting it under the null are presented. Our results hold for M symbols transmitted from p transmitters and received at L receivers. They contrast with the asymptotic results of [3], [4], based on large random matrix theory, which assume L and M to be large.

The GLR may be modified to arrive at several related statistics, including the Bartlett-Nanda-Pillai, Lawley-Hotelling, and Roy tests. We indicate how the null distributions for each statistic may be computed. For the case $p = 1$, all four tests reduce to Hotelling's T^2 test which is UMP invariant. For $p > 1$, however, we contend that no single test will dominate, and that the dominant test is determined by the nature of the MIMO channel as expressed through the arrangement of eigenvalues of the noncentrality matrix $\mathbf{M}_1^H \mathbf{M}_1$.

This paper is not to be construed as a competition between exact distribution results and asymptotic results. Each has its virtues. Asymptotic results do not claim to be exact, nor even accurate, in non-asymptotic regimes, but can claim accuracy and ease of computation in their asymptotic regimes when small tail probabilities are not involved. Our exact results under the null are to be preferred over asymptotic results because they can be used to efficiently and stably compute false alarm probabilities with high precision. In false alarm ranges of 10^{-2} to 10^{-6} the proposed saddlepoint approximations provide high accuracy and preserve relative error, unlike Monte-Carlo simulations or asymptotic normal results. With more effort, exact inversion of a rational MGF can be achieved, as described in Appendix A.

Under the alternative hypothesis, the inversion of the MGF of the GLR is a work in progress. At the time of this writing we offer no computationally feasible way to invert it to determine the pdf and to compute detection probabilities and ROC curves, although promising methods are under development.

VIII. Appendices

A. Moment Generating Function and Saddlepoint Inversion for the Distribution of $W = \det(\mathbf{B})$ Under the Null.

The null distribution of W, characterized in (12) as a product of independent beta random variables, leads to exact and saddlepoint computations for the density of $Z = \log W$, and for false alarm probabilities. The scalar $Z = \log \det(\mathbf{B})$ is negative with probability one, so for convenience, we derive the MGF of $-Z$, which is the MGF of Z, evaluated at $-s$:

$$\mathcal{M}_{-Z}(s) = \mathcal{M}_Z(-s) = c\prod_{i=1}^{p} \frac{\Gamma(N-i+1-s)}{\Gamma(M-i+1-s)} \qquad c = \prod_{i=1}^{p} \frac{\Gamma(M-i+1)}{\Gamma(N-i+1)}, \tag{17}$$

with $N = M - L$. The arguments of the gamma function ratios differ by the integer $M - N = L$ so the expression in (17) is a rational function in s which can be expanded into partial fractions and inverted term-by-term for exact computation. Using $\Gamma(x+1) = x\Gamma(x)$ successively in each, this becomes

$$\begin{aligned}\mathcal{M}_{-Z}(s) &= c \prod_{j=N-p+1}^{N} \frac{\Gamma(j-s)}{\Gamma(j+L-s)} = c \prod_{j=N-p+1}^{N} \prod_{k=j}^{j+L-1} \frac{1}{k-s} \\ &= c \prod_{k=N-p+1}^{N} \left(\frac{1}{k-s}\right)^{k-(M-L-p)} \times \prod_{k=N+1}^{M-p-1} \left(\frac{1}{k-s}\right)^{p} \times \prod_{k=M-p}^{M-1} \left(\frac{1}{k-s}\right)^{M-k}. \end{aligned} \tag{18}$$

The expression in (18) shows that this MGF has poles from $N-p+1$ to $M-1$. Poles of order 1 to p run from $k = N-p+1$ to N, respectively, fixed order p-poles run from $N+1$ to $M-p+1$, and poles of order p to 1 run from $M-p$ to $M-1$, respectively. Altogether there are $L+p-1$ poles with orders running from 1 to p.

Exact inversion of (18) proceeds by expanding $\mathcal{M}_{-Z}$ into a partial fraction sum using the symbolic computational capabilities of Maple or Mathematica. This is followed by direct symbolic inversion to yield exact expressions for the density function of $-Z$ as well as its cumulative distribution function. This method will be limited to settings in which the value of p is not large since higher-order p-poles substantially increase the size of the partial fraction expansions involved.

Approximate saddlepoint inversion is based upon the cumulant generating function (CGF) $\mathcal{K}(s) = \log\{\mathcal{M}_{-Z}(s)\}$ and its derivatives. Using (6.3.2) of Abramowitz and Stegun (1972) [19], its first derivative can be expressed as

$$\begin{aligned}\mathcal{K}'(s) \;&=\; (N-p-s)\psi\left(1+N-p-s\right)+(M-s)\psi\left(M-s\right)\\&\quad-(M-p-s)\psi\left(M-p-s\right)-(N-s)\psi\left(N+1-s\right),\end{aligned}$$

where $\psi(s) = d\log\Gamma(s)/ds$ is the digamma function. Higher-order derivatives follow from Leibnitz's formula of the derivative of a product to give closed form expressions in terms of polygamma functions given as

$$\begin{aligned}\mathcal{K}^{(k)}(s) \;&=\; (-1)^{k-1}\Big[(N-p-s)\psi^{(k-1)}\left(1+N-p-s\right)+(M-s)\psi^{(k-1)}\left(M-s\right)\\&\quad+(k-1)\Big\{\psi^{(k-2)}\left(1+N-p-s\right)+\psi^{(k-2)}\left(M-s\right)\Big\}\Big]\\&\quad+(-1)^{k}\Big[(M-p-s)\psi^{(k-1)}\left(M-p-s\right)+(N-s)\psi^{(k-1)}\left(N+1-s\right)\\&\quad+(k-1)\Big\{\psi^{(k-2)}\left(M-p-s\right)+\psi^{(k-2)}\left(N+1-s\right)\Big\}\Big],\end{aligned} \tag{19}$$

for $k \geq 1$. The exact mean and variance of $-Z$ are $\mu = \mathcal{K}'(0)$ and $\sigma^2 = \mathcal{K}^{(2)}(0)$.

The CGF of the standardized $Z' = (-Z-\mu)/\sigma$ is $\mathcal{K}_{Z'}(s) = \mathcal{K}(s/\sigma) - s\mu/\sigma$. Sufficient conditions for the convergence of Z' to a standard normal are that the third and higher-order cumulants $\mathcal{K}_{Z'}^{(k)}(0)$ for $k \geq 3$ all convergence to 0. Given the benign nature of these CGFs, the sufficient conditions are likely to also be necessary. Thus convergence of the third and higher-order cumulants most likely characterizes the asymptotic regimes for which there is a weak normal limit. Therefore it is the asymptotic behavior of the polygamma functions involved which determines regimes leading to a normal limit.

1) Example $p = 5, L = 10, M = 20$: The first 4 cumulants of $-Z$ are

$$\mathcal{K}'(0) = 4.308\,3 \qquad \mathcal{K}^{(2)}(0) = 0.4032 \qquad \mathcal{K}^{(3)}(0) = 0.08222 \qquad \mathcal{K}^{(4)}(s) = 0.02734$$

and the standardized skewness and kurtosis in terms of the 3rd and 4th cumulants are

$$\mathcal{K}^{(3)}\left(0\right)/\{\mathcal{K}^{(2)}(0)\}^{3/2} = 0.3212 \qquad \mathcal{K}^{(4)}\left(0\right)/\{\mathcal{K}^{(2)}(0)\}^{2} = 0.1682. \tag{20}$$

The standardized cumulants in (20) suggest the distribution is close to normal, with the caution that small tail probabilities may still be poorly approximated.

We compute the exact density of $-Z$, denoted $f(t)$, its saddlepoint approximation $\hat{f}(t)$ as given in (1.4) of Section 1.1.2 of Butler (2007) [8], and the normal approximation of [4]. In Fig. 4, the saddlepoint approximation is seen to be graphically indistinguishable from the exact density.

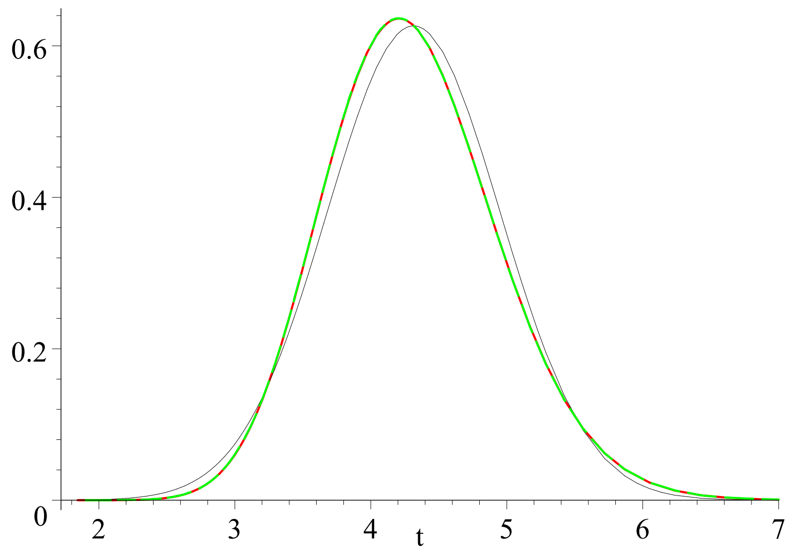


Fig. 4: The exact density of $-Z$ (solid red), the saddlepoint density approximation (solid line green), and the normal approximation (solid black) from [4] (solid black).

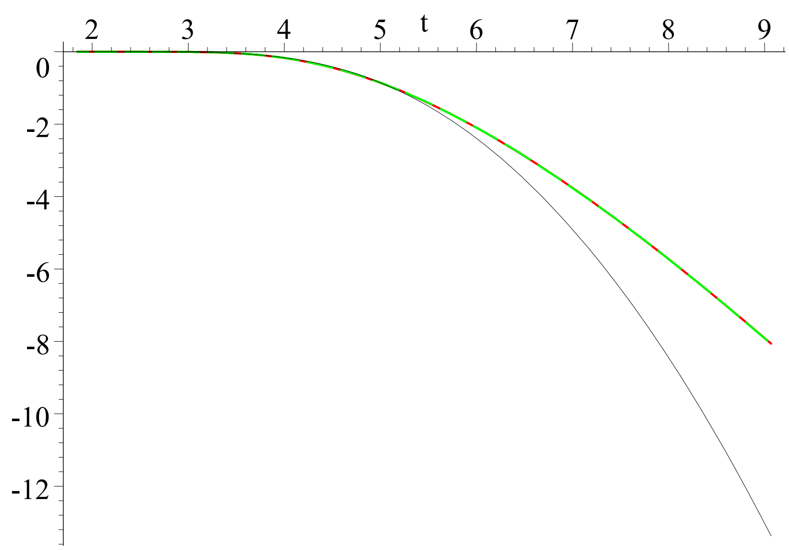


Fig. 5: Plot of the exact $\log_{10}$(PFA) (solid red), its $\log_{10}$-saddlepoint approximation (solid lime green), and the $\log_{10}$-normal approximation from [4] (solid black).

Fig. 5 compares $\log_{10}\{P_{FA}(t)\}$, the true $\log_{10}$-false alarm rate (solid red), with $\log_{10}\{\hat{P}_{FA}(t)\}$, the $\log_{10}$-Lugannani-Rice saddlepoint approximation (solid lime green) given in (1.21) of Section 1.2 of Butler (2007) [8], and the

$\log_{10}$-normal approximation used in [4]. The saddlepoint approximation is graphically indistinguishable from the exact $\log_{10}$-false alarm curve and the $\log_{10}$-normal approximation begins to deviate substantially at $t = 5.5$ where the tail probability is 0.038 and still quite substantial.

The percentage relative errors of $\hat{f}(t)$ and the normal approximation are shown in the left panel of Fig. 6, where this error is plotted as $100\{\hat{f}(t)/f(t) - 1\}$ versus t. At $t = 7$, $\mathbb{P}\{-Z > 7\} = 0.000\,175$, $\hat{f}(7)$ has a relative error -0.469% and the normal density approximation has diverged to -100% the lowest possible valued that can be plotted for the percentage relative error computation. The saddlepoint density satisfies the large deviation limit

$$\lim_{t\to\infty} 100 \left\{ \frac{\hat{f}(t)}{f(t)} - 1 \right\} = 100 \left(\frac{1}{\sqrt{2\pi} e^{-1}} - 1 \right) = 8.44\%, \tag{21}$$

as may be deduced from Theorem 3 and Corollary 4 in Butler and Wood (2019) [20]. Thus its relative error is uniform in the right tail and such uniformity and the limit in (21) apply for all values of p, L, and M. The normal approximation in [4] does not preserve such uniformity since the tail of the normal density approximation $\to 0$ at a faster rate than $f(t) = O(e^{-6t})$ as $t \to \infty$ so that the corresponding ratio in (21) converges to -100%.

The percentage relative errors $100\{\hat{P}_{FA}(t)/P_{FA}(t) - 1\}$ versus t are plotted in the right panel of Fig. 6. At $t = 6.6$ and 6.9, the normal approximation differs from the exact right tail probabilities of 8.9×10^{-4} and 2.9×10^{-4} by factors of $1/5$ and $1/10$ respectively. Unlike the normal CDF approximation, the Lugannani and Rice saddlepoint approximation $\hat{P}_{FA}(t)$ preserves uniform relative in the right tail and its limiting percentage relative error is the same as that in (21) for all values of p, L, and M. This again follows from Theorem 3 and Corollary 4 in Butler and Wood (2019) [20].

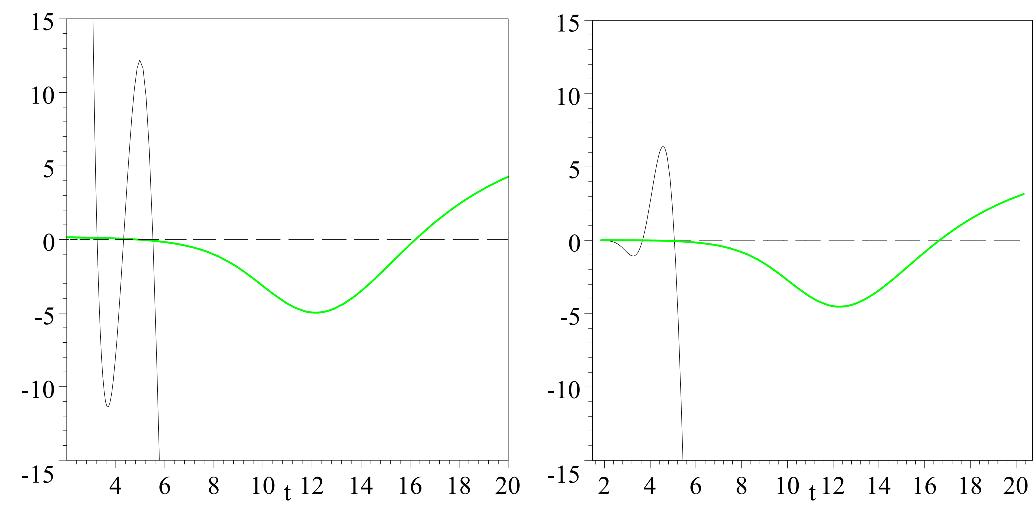


Fig. 6: (Left panel). Plots of the percentage relative error for the saddlepoint density approximation $\hat{f}(t)$ (solid lime green) and the normal approximation (solid black). (Right panel). Plots of the percentage relative error for the saddlepoint PFA approximation $\hat{P}_{FA}(t)$ (solid lime green) and the normal CDF approximation (solid black).

2) Example $p = 20, L = 40, M = 80$: Standardized skewness and kurtosis, as computed from (19), provide a means for determining whether normal approximations will provide adequate accuracy. Both values are zero for a normal distribution and their closeness to zero provides information about the accuracy in using the normal approximation. For this example these standardized cumulants are

$$\mathcal{K}^{(3)}(0)/\{\mathcal{K}^{(2)}(0)\}^{3/2} = \quad 8.067 \times 10^{-2} \qquad \mathcal{K}^{(4)}(0)/\{\mathcal{K}^{(2)}(0)\}^{2} = \quad 1.066 \times 10^{-2}$$

suggesting that the exact distribution of $-Z$ is quite close to a normal distribution, with the caution that small tail probabilities may still be poorly approximated.

3) Example $p = 50, L = 100, M = 200$*:* For this example the standardized skewness and kurtosis are

$$\mathcal{K}^{(3)}(0)/\{\mathcal{K}^{(2)}(0)\}^{3/2} = \quad 3.228 \times 10^{-2} \qquad \mathcal{K}^{(4)}(0)/\{\mathcal{K}^{(2)}(0)\}^{2} = \quad 1.706 \times 10^{-3},$$

suggesting that the exact distribution of $-Z$ is close to a normal distribution, with the caution that small tail probabilities may still be poorly approximated.

B. *Proof of Theorem 1*

Starting with the noncentral matrix beta density in (10), let

$$c = e^{-\operatorname{tr}(\mathbf{M}_1^H\mathbf{M}_1)}\tilde{\Gamma}_p(M)/\{\tilde{\Gamma}_p(M-L)\tilde{\Gamma}_p(L)\}$$

so that

$$E(W^s) = c\int_{\mathbf{0}<\mathbf{B}=\mathbf{B}^H<\mathbf{I}_p} {}_1\tilde{F}_1\{M; L; \mathbf{M}_1^H\mathbf{M}_1(\mathbf{I}_p - \mathbf{B})\}\det(\mathbf{B})^{M-L-p+s}\det(\mathbf{I}_p-\mathbf{B})^{L-p}(d\mathbf{B}.)$$

Change variables $\mathbf{B} \to \mathbf{I}_p - \mathbf{B}$ so

$$E(W^s) = c\int_{\mathbf{0}<\mathbf{B}=\mathbf{B}^H<\mathbf{I}_p} {}_1\tilde{F}_1\{M; L; \mathbf{M}_1^H\mathbf{M}_1\mathbf{B})\}\det(\mathbf{B})^{L-p}\det(\mathbf{I}_p-\mathbf{B})^{M-L-p+s}(d\mathbf{B}.)$$

Using the zonal polynomial expansion for ${}_1\tilde{F}_1$ given in James (1964, Section 8) [7] this becomes

$$E(W^s) = c\int_{\mathbf{0}<\mathbf{B}=\mathbf{B}^H<\mathbf{I}_p} \sum_{k=0}^{\infty}\sum_{\kappa}\frac{[M]_\kappa}{[L]_\kappa}\frac{\tilde{C}_\kappa(\mathbf{M}_1^H\mathbf{M}_1\mathbf{B})}{k!}\det(\mathbf{B})^{L-p}\det(\mathbf{I}_p-\mathbf{B})^{M-L-p+s}(d\mathbf{B}.)$$

The integration and double summation may be interchanged. Using the reproductive property of zonal polynomial $\tilde{C}_\kappa$ as given in Khatri (1966, Section 5) [21], then

$$E(W^s) = c\sum_{k=0}^{\infty}\sum_{\kappa}\frac{[M]_\kappa}{[L]_\kappa}\frac{1}{k!}\frac{\tilde{\Gamma}_p(L,\kappa)\tilde{\Gamma}_p(M-L+s)}{\tilde{\Gamma}_p(M+s;\kappa)}\tilde{C}_\kappa(\mathbf{M}_1^H\mathbf{M}_1),$$

where $\tilde{\Gamma}_p(L,\kappa) = [L]_\kappa\,\tilde{\Gamma}_p(L)$. Cancelling c with values $\tilde{\Gamma}_p(L,\kappa)$ and $\tilde{\Gamma}_p(M+s;\kappa)$ gives

$$\begin{aligned}
E(W^s) &= e^{-\operatorname{tr}(\mathbf{M}_1^H\mathbf{M}_1)}\frac{\tilde{\Gamma}_p(M-L+s)\tilde{\Gamma}_p(M)}{\tilde{\Gamma}_p(M-L)\tilde{\Gamma}_p(M+s)}\sum_{k=0}^{\infty}\sum_{\kappa}\frac{[M]_\kappa}{[M+s]_\kappa}\frac{\tilde{C}_\kappa(\mathbf{M}_1^H\mathbf{M}_1)}{k!} \\
&= e^{-\operatorname{tr}(\mathbf{M}_1^H\mathbf{M}_1)}\frac{\tilde{\Gamma}_p(M-L+s)\tilde{\Gamma}_p(M)}{\tilde{\Gamma}_p(M-L)\tilde{\Gamma}_p(M+s)}\,{}_1\tilde{F}_1(M; M+s; \mathbf{M}_1^H\mathbf{M}_1) \qquad (22)\\
&= \frac{\tilde{\Gamma}_p(M-L+s)\tilde{\Gamma}_p(M)}{\tilde{\Gamma}_p(M-L)\tilde{\Gamma}_p(M+s)}\,{}_1\tilde{F}_1(s; M+s; -\mathbf{M}_1^H\mathbf{M}_1), \qquad (23)
\end{aligned}$$

where (23) follows from (22) using the Euler relation for ${}_1\tilde{F}_1$ as given in Herz (1955) [22].

C. *Proof of Corollary 1*

The proof follows the approach used in Theorem 10.4.1 in Muirhead (1982) [9] but applied to Hermitian matrices rather than real symmetric matrices. Starting with $\mathbf{W}_1 = \mathbf{Y}_1\mathbf{Y}_1^H \sim$ noncentral $\mathcal{CW}_L(p, \mathbf{M}_1\mathbf{M}_1^H)$ and $\mathbf{W}_2 = \mathbf{Y}_2\mathbf{Y}_2^H \sim \mathbb{C}\mathbf{W}_L(M-p, \mathbf{I}_L)$, the joint density of $\mathbf{W}_1$ and $\mathbf{W}_2$ from James (1964, Section 8) [7] is

$$f(\mathbf{W}_1, \mathbf{W}_2) = \frac{1}{\tilde{\Gamma}_L(p)} \det(\mathbf{W}_1)^{p-L} e^{-\operatorname{tr}(\mathbf{M}_1\mathbf{M}_1^H + \mathbf{W}_1)} \, {}_0\tilde{F}_1(p; \mathbf{M}_1\mathbf{M}_1^H\mathbf{W}_1) \\ \times \frac{1}{\tilde{\Gamma}_L(M-p)} \det(\mathbf{W}_2)^{M-p-L} e^{-\operatorname{tr}\mathbf{W}_2}.$$

Transform $\mathbf{W}_1$, $\mathbf{W}_2 \to \mathbf{W}_1$, $\mathbf{F} = \mathbf{W}_1^{1/2}\mathbf{W}_2^{-1}\mathbf{W}_1^{1/2}$ with Jacobian

$$(d\mathbf{W}_1)(d\mathbf{W}_2) = \det(\mathbf{W}_1)^L \det(\mathbf{F})^{-2L} (d\mathbf{W}_1)(d\mathbf{F})$$

to get the joint density of $\mathbf{W}_1$ and $\mathbf{F}$. From this, the marginal density of $\mathbf{F}$ is

$$f(\mathbf{F}) = \frac{\det(\mathbf{F})^{p-L-M}}{\tilde{\Gamma}_L(M-p)\tilde{\Gamma}_L(p)} e^{-\operatorname{tr}(\mathbf{M}_1\mathbf{M}_1^H)} \times \\ \int_{\mathbf{W}_1 = \mathbf{W}_1^H > \mathbf{0}} {}_0\tilde{F}_1(p; \mathbf{M}_1\mathbf{M}_1^H\mathbf{W}_1) e^{-\operatorname{tr}\{\mathbf{W}_1(\mathbf{I}_L + \mathbf{F}^{-1})\}} \det(\mathbf{W}_1)^{M-L} (d\mathbf{W}_1).$$

Using Lemma 1 below, the integral is

$$f(\mathbf{F}) = \frac{\tilde{\Gamma}_L(M)}{\tilde{\Gamma}_L(M-p)\tilde{\Gamma}_L(p)} e^{-\operatorname{tr}(\mathbf{M}_1\mathbf{M}_1^H)} \times \\ {}_1\tilde{F}_1\{M; p; \mathbf{M}_1\mathbf{M}_1^H(\mathbf{I}_L + \mathbf{F}^{-1})^{-1}\} \frac{\det(\mathbf{F})^{p-L}}{\det(\mathbf{I}_L + \mathbf{F})^M}.$$

Transforming $\mathbf{F} \to \mathbf{B} = (\mathbf{I}_L + \mathbf{F})^{-1}$ with Jacobian $\det(\mathbf{B})^{-2L}$ leads to

$$f(\mathbf{B}) = \frac{\tilde{\Gamma}_L(M)}{\tilde{\Gamma}_L(M-p)\tilde{\Gamma}_L(p)} e^{-\operatorname{tr}(\mathbf{M}_1\mathbf{M}_1^H)} \times \\ {}_1\tilde{F}_1\{M; p; \mathbf{M}_1\mathbf{M}_1^H(\mathbf{I}_L - \mathbf{B})\} \det(\mathbf{B})^{M-p-L} \det(\mathbf{I}_L - \mathbf{B})^{p-L}.$$

This is a noncentral complex matrix variate beta $\mathcal{CB}_L(M-p, p, \mathbf{M}_1\mathbf{M}_1^H)$. This compares to $\mathcal{CB}_p(M-L, L, \mathbf{M}_1^H\mathbf{M}_1)$ for the case in which $p \geq L$. The results in (14) and (15) follow directly by interchanging the roles of p and L and applying the same arguments used in Theorem 1. □

LEMMA 1. *Let* $\mathbf{Y}$ be $L \times L$ *Hermitian positive definite and let* $\mathbf{Z}$ be $L \times L$ Hermitian. *For* $\operatorname{Re}(a) > L - 1$,

$$\int_{\mathbf{W} = \mathbf{W}^H > \mathbf{0}} {}_0\tilde{F}_1(p; \mathbf{Y}\mathbf{W}) e^{-\operatorname{tr}(\mathbf{W}\mathbf{Z})} \det(\mathbf{W})^{M-L} (d\mathbf{W}) \\ = \tilde{\Gamma}_L(M) \det(\mathbf{Z})^{-M} \, {}_1\tilde{F}_1\{M; p; \mathbf{Y}\mathbf{Z}^{-1}\} \tag{24}$$

Proof. Take the zonal polynomial expansion of ${}_0\tilde{F}_1$ given in James (1964, Section 8, Eqn. 87) [7], integrate term-by-term using the reproductive property of the zonal polynomials in Eqn. 85, and then recognize that this leads to the zonal polynomial expansion for ${}_1\tilde{F}_1$ on the right side of 24. □

References

[1] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1760–1770, Oct 1990.

[2] D. W. Bliss and P. A. Parker, "Temporal synchronization of MIMO wireless communication in the presence of interference," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1794–1806, March 2010.

[3] S. Hiltunen, P. Loubaton, and P. Chevalier, "Large system analysis of a GLRT for detection with large sensor arrays in temporally white noise," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5409–5423, Oct 2015.

[4] S. Hiltunen and P. Loubaton, "Asymptotic analysis of a GLR test for detection with large sensor arrays: New results," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 4506–4510.

[5] P. Pakrooh, L. L. Scharf, and R. W. Butler, "Distribution results for a multi-rank version of the Reed-Yu detector," in *2017 51st Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2017, pp. 783–784.

[6] S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika*, vol. 24, no. 3/4, pp. 471–494, 1932. [Online]. Available: http://www.jstor.org/stable/2331979

[7] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *The Annals of Mathematical Statistics*, vol. 35, no. 2, pp. 475–501, 06 1964. [Online]. Available: http://dx.doi.org/10.1214/aoms/1177703550

[8] R. W. Butler, *Saddlepoint approximations with applications*. Cambridge University Press, 2007, vol. 22.

[9] R. J. Muirhead, *Aspects of multivariate statistical theory*. John Wiley & Sons, 2009, vol. 197.

[10] R. W. Butler and A. T. A. Wood, "Approximation of power in multivariate analysis," *Statistics and Computing*, vol. 15, no. 4, pp. 281–287, 2005.

[11] R. W. Butler and A. T. Wood, "Laplace approximations for hypergeometric functions with Hermitian matrix argument," *In preparation*, 2020.

[12] M. Srivistava and C. Khatri, *Introduction to multivariate analysis*. North-Holland, NY, 1979.

[13] E. L. Lehmann and J. P. Romano, *Testing statistical hypotheses*. Springer, New York, 2006.

[14] N. Giri, "On the complex analogues of T^2 and R^2 tests," *The Annals of Mathematical Statistics*, vol. 36, pp. 664–670, 1965.

[15] Y. Chikuse, "Partial differential equations for hypergeometric functions of complex argument matrices and their applications," *Annals of the Institute of Statistical Mathematics*, vol. 28, no. 1, pp. 187–199, 1976.

[16] C. Khatri, "Distribution of the largest or the smallest characteristic root under null hypothesis concerning complex multivariate normal populations," *The Annals of Mathematical Statistics*, vol. 35, no. 4, pp. 1807–1810, 1964.

[17] "NIST digital library of mathematical functions," *http://dlmf.nist.gov, Release 1.0.24*, 2019.

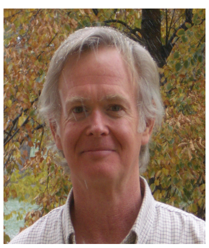
[18] I. M. Johnstone, "Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence," *Annals of Statistics*, vol. 36, no. 6, pp. 2638–2716, 2008.

[19] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions*. New York: Dover, 1972.

[20] R. W. Butler and A. T. A. Wood, "Limiting saddlepoint relative errors in large deviation regions under purely Tauberian conditions," *Bernoulli*, vol. 25, no. 4B, pp. 3379–3399, 2019.

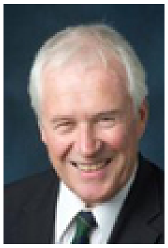
[21] C. Khatri, "On certain distribution problems based on positive definite quadratic functions in normal vectors," *The Annals of Mathematical Statistics*, vol. 37, pp. 468–479, 1966.

[22] C. S. Herz, "Bessel functions of matrix argument," *The Annals of Mathematics*, vol. 61, pp. 474–523, 1955.



Ronald Butler received his Ph.D. from the University of Michigan, Ann Arbor. He is currently C.F. Frensley Professor of Mathematics and Professor of Statistics at Southern Methodist University. Previously he has served on the faculties at University of Texas, Austin and Colorado State University. He has held visiting positions at University of Michigan, University of Nottingham, Oxford University, and University of Warwick. Prof. Butler's research interests are in applied probability, transform inversion, saddlepoint approximation, and complex variables. His book "Saddlepoint Approximations with Applications" uses simple mathematics to explain the theory and widespread applications of saddlepoint approximations.

PooriaPakrooh (S'13-M'15) received the B.Sc. degree in electrical engineering from Iran University of Science and Technology, Tehran, Iran, in 2008, and the M.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2011. He received his Ph.D. degree in electrical engineering from Colorado State University, Fort Collins, Colorado, in 2015. From June 2015 to August 2017 he was a postdoctoral research associate with the Department of Mathematics at Colorado State University. He is currently a senior system engineer at Qualcomm Inc, San Diego, California. His current research interests include statistical signal processing, wireless communication and compressed sensing.



Louis L. Scharf (S'67-M'69-SM'77-F'86-LF'07) received his Ph.D. from the University of Washington, Seattle. He has served on the faculties at Colorado State University, University of Colorado, and University of Rhode Island. He is currently Research Professor of Mathematics and Emeritus Professor of Electrical and Computer Engineering, with a joint appointment in Statistics at Colorado State University. He has held visiting positions here and abroad, including Duke University, University of Wisconsin, Ecole Superieure d'Electricite, Ecole Superieure des Telecommunications, EURECOM, University of La Plata, University of Tromso, University of Newcastle, and University of Paderborn. Prof. Scharf 's research interests are in detection, estimation, and space-time series analysis for statistical signal processing and machine learning. He has authored three books: L.L. Scharf, "Statistical Signal Processing: Detection, Estimation, and Time Series Analysis," Addison-Wesley, 1991; L.L. Scharf, "A First Course in Electrical and Computer Engineering," Addison-Wesley, 1998 (available on OpenStax); and P.J. Schreier and L.L. Scharf, "Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals," Cambridge University Press, 2010.

Prof. Scharf has received several awards for his work and professional service, including an IEEE Third Millennium Medal, the Technical Achievement and Society Awards from the IEEE Signal Processing Society (SPS), the Donald W. Tufts Award for Underwater Acoustic Signal Processing, a 2016 Diamond Award from the University of Washington (Seattle), and the 2016 IEEE Jack S. Kilby Signal Processing Medal, endowed by Texas Instruments. He is a Life Fellow of IEEE.