The Adaptive Coherence Estimator: a Uniformly-Most-Powerful-Invariant Adaptive Detection Statistic

Shawn Kraut*, Louis L. Scharf[1], and Ronald W. Butler[2]

*Corresponding author, Queen's University, Department of Mathematics and Statistics and Department of Electrical and Computer Engineering, Jeffery Hall, Kingston, Ontario K7L 3N6

[1]Colorado State University, Department of Electrical and Computer Engineering and Department of Statistics, Fort Collins, Colorado 80523-1373

[2]Colorado State University, Department of Statistics, Fort Collins, Colorado 80523-1877

E-mail: kraut@mast.queensu.ca, scharf@engr.colostate.edu, rbutler@stats.colostate.edu

Keywords: Adaptive Detection, Invariant Detection, Adaptive Coherence Estimator, Adaptive Radar and Sonar, Adaptive Array Processing, Sample Covariance, Uniformly Most Powerful Invariant, Constant False Alarm Rate

EDICS #s: 1-MSAM, 1-DETC

Submission to *IEEE Transactions on Signal Processing*

Submitted July 7, 2003,

Revised December 2, 2003.

Abstract

We show that the Adaptive Coherence Estimator (ACE) is a uniformly most powerful (UMP) invariant detection statistic. This statistic is relevant to a scenario appearing in adaptive array processing, in which there are auxiliary, signal-free, training-data vectors that can be used to form a sample covariance estimate for clutter and interference suppression. The result is based on earlier work by Bose and Steinhardt, who found a two dimensional maximal invariant when test and training data share the same noise covariance. Their 2-D maximal invariant is given by Kelly's Generalized Likelihood Ratio Test (GLRT) statistic and the Adaptive Matched Filter (AMF). We extend the maximal-invariant framework to the problem for which ACE is a GLRT: the test data shares the same covariance structure as the training data, but the relative power level is not constrained. In this case, the maximal invariant statistic collapses to a one-dimensional scalar, which is also the ACE statistic. Furthermore, we show that the probability density function for ACE possesses the property of "total positivity," which establishes that it has monotone likelihood ratio. Thus a threshold test on ACE is UMP-invariant. This means that it has a claim to optimality, having the largest detection probability out of the class of detectors that are also invariant to affine transformations on the data matrix that leave the hypotheses unchanged. This requires an additional invariance not imposed by Bose and Steinhardt: invariance to relative scaling of test and training data. ACE is invariant, and CFAR, with respect to such scaling, whereas Kelly's GLRT and the AMF are invariant, and CFAR, only with respect to common scaling.

I. Introduction

In this paper we will examine the optimality of the ACE (Adaptive Coherence Estimator) statistic when used as a threshold detection test:

$$\hat{\beta} = \frac{|\underline{\psi}^\dagger \mathbf{S}^{-1} \underline{y}|^2}{(\underline{\psi}^\dagger \mathbf{S}^{-1} \underline{\psi})(\underline{y}^\dagger \mathbf{S}^{-1} \underline{y})} \gtrless \eta. \tag{1}$$

The purpose of this statistic is to detect the presence of a signal vector (or steering/replica vector) $\underline{\psi}$ in a measurement test vector $\underline{y}$, by comparing the statistic with a threshold η [1]. The detection must be performed in the presence of noise, clutter, and interference. Training data vectors are used to construct a sample covariance $\mathbf{S}$ that approximates the true covariance $\mathbf{R}$. This statistic is an empirical version of the CFAR matched subspace detector, which is the optimal statistical if the noise covariance matrix is known exactly to within an unknown scaling level [2], [3], [4].

When the covariance must in fact be estimated from training data, an *ad hoc* adaptive statistic is then formed by substituting the sample covariance $\mathbf{S}$ for an exact covariance $\mathbf{R}$ in the detection statistic formula. The training data is presumed to be signal-free, with the same noise covariance as the test data. The class of detectors that rely on this data model have been particularly useful in radar systems, wherein data vectors from range cells in the vicinity of a range cell of interest are nominally target free, and can be used reliably to estimate the covariance of the clutter that is corrupting the test-cell data [5], [6], [7], [8], [9].

Here we are interested in the examination of the adaptive detection problem from first principles: given a test data vector $\underline{y}$ and a training data matrix $\mathbf{X}$, what is the optimal detection procedure? ACE was shown to be a GLRT for the adaptive detection problem in [10]. However this is not a strict optimality result, the GLRT merely being the maximum likelihood (ML) approximation to the optimal Neyman-Pearson likelihood ratio test. A true optimality result is provided by the formalism of Uniformly Most Powerful (UMP) invariant tests. A test is UMP-invariant when, for a given probability of false alarm, it has the largest probability of detection (is most powerful) out of all tests that are invariant to a group of transformations on the data matrix. The structure of the transformation group is chosen to characterize the scenario in which the detection statistic will be applied.

A key concept in the theory of UMP-invariant tests is that of the maximal invariant statistic [11], [2]. This is a reduced function of the data (similar to a sufficient statistic) that is insensitive to data transformations of a given form, but still sensitive to other transformations on the data. We want the statistic to be invariant when the data is subjected to a transformation group of interest (one that does not change the assumptions of the two hypotheses), but to vary when the data is subjected to other transformations. Equivalently, this means that the statistic is not only invariant to transformations within the group, but also that if the statistic is the same for any two data sets, then these data sets must be related by a transformation within the group of interest. The key consequence of this is that any function of the data that is invariant to the transformation depends on the data only through the maximal invariant (this is Theorem 1 in Chapter 6 of [11]).

In this paper we show that a threshold test on ACE is UMP-invariant. This result requires the statistic to have two properties: (1) The statistic $\hat{\beta}$ must be a maximal invariant with respect to a compelling transformation on the data matrix. (2) Its likelihood ratio must be a monotone function of $\hat{\beta}$. Then a threshold test on the likelihood ratio can be replaced a simple threshold test on $\hat{\beta}$. By the Karlin-Rubin Theorem (see [2], or Theorem 1 in Chapter 5 of [12]), this threshold test is most powerful out of all invariant tests, uniformly over all possible signal-to-ratios (SNRs), making it uniformly-most-powerful invariant.

To establish that ACE is UMP-invariant, we build upon the earlier work of Bose and Steinhardt. In [13], they considered the adaptive detection scenario in which there is homogeneity between test and training data. By homogeneity we mean that the training data vectors have the same statistical distribution as the noise in the data being tested. They obtained the interesting result that a two-dimensional statistic, closely related to ACE, is the maximal invariant. The 2-D statistic is composed of Kelly's GLRT [14] and the Adaptive Matched Filter (AMF) [15], [16]. This is a powerful result, but it is not quite sufficient to construct a simple threshold test, which requires data reduction down to a single scalar. Furthermore, in [17], Bose and Steinhardt showed that for this homogeneous adaptive detection problem, a UMP-invariant test cannot even exist. This is because the level-curves of the the 2-D likelihood surface vary with signal-to-noise ratio (SNR), meaning the decision region must change with SNR.

However, by requiring invariance to a more general transformation that permits scaling between test and training data, the two-dimensional maximal invariant of Bose and Steinhardt collapses to a one-dimensional invariant, the ACE statistic. The two scalar statistics of Bose and Steinhardt are not invariant to such scaling; however their ratio is invariant. This ratio is equivalent to ACE. We show in this paper that ACE is also a maximal invariant. This is the key observation we made in [18], and a key observation of this paper.

Our motivation for allowing scaling between test and training data comes from [10]. There it was shown that the ACE statistic is a GLRT when there is an additional scaling parameter σ, representing the scaling of the test-data noise covariance relative to the training data covariance. When ML estimates of σ under the two hypotheses are substituted into the likelihood ratio, the ratio reduces to the ACE statistic, rather than the GLRT of Kelly.

The question arises: why would one want to accomodate such scaling? The answer is that it adds robustness to deviations of the data from the assumed statistical model. The Adaptive Matched Fiter is more sensitive to scaling of the test vector than it is to scaling of any single training-data vector. So the extra scaling invariance of ACE makes it robust in models that allow for unknown fluctuatons in data power, as for example in radar clutter [19]. Another compelling fact is that the non-adaptive version of ACE, the CFAR Matched Subspace Detector [2], [3] is CFAR over all elliptically contoured (Gaussian and non-Gaussian) distributions [4]. There are other scenarios, outside of adaptive radar, wherein scaling inhomogeneity might arise. This might occur, for example, in a wireless communications system with fades over multiple sources of interference.

The cost of this robustness is a performance loss of about 2 dB when the data is in fact homogeneous and there is large training data support. The difference in performance of ACE relative to Kelly's GLRT and the AMF is discussed in more detail in [4]. When the training data support is low, there is in fact a performance gain for moderate SNR, though a performance loss persists at high SNR. Of course, when scaling inhomogeneity is present, Kelly's GLRT and the AMF cease to be CFAR, which defeats them and renders moot any attempts to compare detection performance at a constant false alarm rate.

We have found the proof that ACE possesses monotone likelihood ratio to be more difficult than we first anticipated. We prove it by relating monotone likelihood ratio to the property of "total positivity." The literature on totally-positive kernel functions makes extensive use of a basic composition formula, due to Polya and Szego (see [20]). We use this composition formula repeatedly, to build up the property of total positivity for ACE from that of its conditional probability density function. With the property of total positivity, ACE then has monotone likelihood ratio, and a threshold test on ACE is UMP-invariant. This is the result needed to complete the line of reasoning we first presented in [18].

II. Maximal Invariance in the Problem of Known Covariance (Non-Adaptive)

A. The Matched Subspace Detector (Matched Filter)

The general problem addressed here is the detection of a narrow-band signal (or target response) $\underline{\psi}$ in a measurement $\underline{y} \sim CN_N[\mu\underline{\psi}, \mathbf{R}]$ in proper Gaussian noise with covariance structure $\mathbf{R}$. The target-absent and target-present hypotheses are then parameterized by $H_0 : \mu = 0$ and $H_1 : \mu \neq 0$. Both signal and measurement vectors are N-dimensional, $\underline{y}, \underline{\psi} \in \mathbb{C}^N$, corresponding to, for example, data from N doppler/array-element channels in space-time adaptive processing.

In this section we introduce the relevant invariance issues in the simpler context of the non-adaptive problem, wherein the noise covariance is presumed to be known *a priori*. We use it to pre-whiten the noise in the test data vector: $\underline{z} = \mathbf{R}^{-\frac{1}{2}}\underline{y}$. Then $\underline{z} \sim CN_N[\mu\underline{\phi}, \mathbf{I}]$, where $\underline{\phi} = \mathbf{R}^{-\frac{1}{2}}\underline{\psi}$.

To simplify matters, we assume that a unitary transformation has been applied to $\underline{z}$ to rotate the signal vector $\underline{\phi}$ into the direction of the first standard basis vector:

$$\mathbf{U}_\phi = \left[\frac{\underline{\phi}}{\sqrt{\underline{\phi}^\dagger \underline{\phi}}}, \mathbf{U}_\perp \right], \tag{2}$$

where $\mathbf{U}_\perp^\dagger \underline{\phi} = \mathbf{0}$. Here $\langle \underline{\phi} \rangle$ is a one-dimensional signal subspace, and $\langle \mathbf{U}_\perp \rangle$ is its corresponding $(N-1)$-dimensional orthogonal subspace.

For generality, we also treat the more general case of multi-dimensional signal subspaces, wherein the rank-1 signal vector $\underline{\psi}$ is replaced by a signal mode matrix $\mathbf{H}$: $\underline{\psi} = \mathbf{H}\underline{\theta}$, where $\underline{\theta}$ contains unknown coefficients. Then the whitened data vector is distributed as $\underline{z} \sim CN_N[\mu\mathbf{R}^{-\frac{1}{2}}\mathbf{H}\underline{\theta}, \mathbf{I}]$, and the signal subspace is then spanned by the p columns of $\mathbf{G} = \mathbf{R}^{-\frac{1}{2}}\mathbf{H}$. The columns of $\mathbf{H}$ are chosen to accomodate various types of signal uncertainty, such as arrival angle or environmental/channel uncertainty. We apply a unitary rotation matrix so that the the signal subspace is then spanned by the first p standard basis vectors. That is, we construct

$$\mathbf{U}_\mathbf{G} = \left[\mathbf{G}(\mathbf{G}^\dagger\mathbf{G})^{-\frac{1}{2}}, \mathbf{U}_\perp \right], \tag{3}$$

where $\mathbf{U}_\perp^\dagger \mathbf{G} = \mathbf{0}$. Here $\langle \mathbf{G} \rangle$ is a p-dimensional signal subspace, and $\langle \mathbf{U}_\perp \rangle$ is its corresponding $(N-p)$-dimensional orthogonal subspace.

We can then write the test vector as

$$\mathbf{U}_{\mathbf{G}}^{\dagger}\underline{z} \rightarrow \quad \underline{z} = \begin{bmatrix} \underline{z}_1 \\ \underline{z}_2 \end{bmatrix}, \tag{4}$$

where the subscripts 1 and 2 indicate the signal and signal-free subspaces, respectively: $\underline{z}_1 \sim CN_N[\mu(\mathbf{G}^{\dagger}\mathbf{G})^{\frac{1}{2}}\underline{\theta}, \mathbf{I}]$ and $\underline{z}_2 \sim CN_N[\underline{0}, \mathbf{I}]$. As in [13], we seek a transformation on the data that we wish the test statistic to be invariant to, one that preserves (1) the Gaussianity of its distribution, (2) the known covariance matrix $\mathbf{R}$, and (3) the mean under the two hypotheses (i.e., the mean of $\underline{z}$ should be the zero vector under H_0 and a vector in the signal subspace under H_1). Such a transformation leaves the parameter sets that define the two hypotheses unchanged.

To preserve Gaussianity, we consider affine transformations of the form $g(\underline{z}) = \mathcal{A}\underline{z} + \underline{b}$. Preserving the mean under H_1 forces $\underline{b}$ to lie in the signal subspace. Assuming the rotation of Eqn. 3 has been applied, this means that $\underline{b} \in \mathbb{C}^p$ should be in the span of the first p standard basis vectors: $\underline{b} \in \langle \underline{e}_1 \ldots \underline{e}_p \rangle$, where $\underline{e}_1 = [1\ 0 \cdots 0]^{\dagger}$, etc. But preserving a zero mean under H_0 forces $\underline{b}$ to be the zero vector, $\underline{b} = \underline{0}$. Also, keeping the mean in $\langle \underline{e}_1 \ldots \underline{e}_p \rangle$ requires $\mathcal{A}$ to have the upper block triangular form

$$\mathcal{A} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^{\dagger} \\ \mathbf{0} & \mathbf{\Gamma} \end{bmatrix}, \tag{5}$$

where $\mathbf{A}$ is $p \times p$, $\mathbf{B}$ is $p \times (N-p)$ and $\mathbf{\Gamma}$ is $(N-p) \times (N-p)$. Finally, forcing the covariance of $\underline{z}$ to remain identity requires $\mathbf{A} = \mathbf{U}_1$ and $\mathbf{\Gamma} = \mathbf{U}_2$ both to be unitary, leading to

$$g(\underline{z}) = \begin{bmatrix} \mathbf{U}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2 \end{bmatrix} \underline{z}. \tag{6}$$

A *maximal invariant* $\underline{m}$ is then a reduced-dimension function of the data $\underline{z}$ that is invariant (insensitive) to transformations of the form $g(\cdot)$, but still varies under (is sensitive to) any other type of transformation.

Lemma 1: The maximal invariant to the transformation of Equation 6 is two-dimensional, consisting of the two scalar quadratic forms $m_1 = \|\underline{z}_1\|^2$ and $m_2 = \|\underline{z}_2\|^2$.

Proof: That these scalars are invariant to $g(\cdot)$ in Eqn. 6 is a simple consequence of unitarity: for m_1 we have, $m_1' = \underline{z}_1^{\dagger}\mathbf{U}_1^{\dagger}\mathbf{U}_1\underline{z}_1 = \underline{z}_1^{\dagger}\underline{z}_1 = m_1$; the argument is identical for m_2. Showing that they are a maximal invariant is more tricky. Consider the statement that if $\underline{z}$ undergoes any transformation to $\underline{z}'$ that is not in the form of $g(\cdot)$, then $(m_1', m_2') \neq (m_1, m_2)$ (i.e. the maximal invariant varies). A logically equivalent statement (the contrapositive) is that if $\underline{z}$ undergoes any transformation such that $(m_1', m_2') = (m_1, m_2)$, then the transformation must be expressible in the form of $g(\cdot)$. We can construct the following product of unitary matrices:

$$\mathbf{T}_1 = \left[\begin{array}{c|c} \frac{\underline{z}_1'}{\sqrt{\underline{z}_1'^\dagger \underline{z}_1'}} & \mathbf{U}_1' \end{array} \right] \left[\begin{array}{c} \frac{\underline{z}_1^\dagger}{\sqrt{\underline{z}^\dagger \underline{z}}} \\ \hline \mathbf{U}_1^\dagger \end{array} \right], \tag{7}$$

where $\mathbf{U}_1'^\dagger \underline{z}_1' = \mathbf{U}_1^\dagger \underline{z}_1 = \underline{0}$. Suppose $m_1' = m_1$; then we have

$$\mathbf{T}_1 \underline{z} = \frac{\sqrt{\underline{z}_1^\dagger \underline{z}_1}}{\sqrt{\underline{z}_1'^\dagger \underline{z}_1'}} \underline{z}_1' = \underline{z}_1'. \tag{8}$$

In an identical matter, a unitary matrix $\mathbf{T}_2$ can also be constructed such that if $m_2' = m_2$, $\underline{z}_2' = \mathbf{T}_2 \underline{z}_2$. Substituting these matrices for $\mathbf{U}_1$ and $\mathbf{U}_2$, we are able to construct a transformation in the form of Eqn. 6. ■

The first scalar, $\|\underline{z}_1\|^2$, is the Matched Subspace Detector of [2], [3]. In the case of a rank-1 signal, it simplifies to the form of a matched filter, and in the multi-rank cases it is equivalent to a projection of $\underline{z}$ onto $\langle \mathbf{G} \rangle = \langle \mathbf{R}^{-\frac{1}{2}} \mathbf{H} \rangle$:

$$\begin{aligned} |z_1|^2 &= \frac{|\underline{\psi}^\dagger \mathbf{R}^{-1} \underline{y}|^2}{\underline{\psi}^\dagger \mathbf{R}^{-1} \underline{\psi}} \gtrless \eta, \quad p = 1; \\ \|\underline{z}_1\|^2 &= \underline{y}^\dagger \mathbf{R}^{-1} \mathbf{H} (\mathbf{H}^\dagger \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^\dagger \mathbf{R}^{-1} \underline{y} = \|\mathbf{P}_{\mathbf{G}} \underline{z}\|^2 \gtrless \eta, \quad p > 1. \end{aligned} \tag{9}$$

B. The CFAR MSD

Now we consider a slightly less constrained detection problem, in which the noise covariance is assumed to be known only to within a scaling constant, i.e. it is $\sigma^2 \mathbf{R}$, where $\mathbf{R}$ is known and σ (an additional nuisance parameter) is not. In this case the transformation matrix of Equation 6 does not need to preserve the scale of the covariance, and generalizes to

$$g_\gamma(\underline{z}) = \gamma \begin{bmatrix} \mathbf{U}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2 \end{bmatrix} \underline{z}. \tag{10}$$

In this case, $m_1 = \|\underline{z}_1\|^2$ and $m_2 = \|\underline{z}_2\|^2$ are no longer invariant to transformations of the form $\underline{z}' = g_\gamma(\underline{z})$. In fact, $m_1' = \gamma^2 m_1$ and $m_2' = \gamma^2 m_2$. However, their ratio, $m_3 = \frac{m_1}{m_2}$ *is* invariant to this transformation, and furthermore it is a *maximal* invariant. As we show below, the ratio $\beta = \frac{m_1}{m_1+m_2} = \frac{m_3}{1+m_3}$ is a beta-distributed statistic that is the non-adaptive precursor to ACE. It too is a maximal invariant.

Proposition 1: The statistic $\beta = \frac{m_3}{1+m_3}$ is a maximal invariant statistic with respect the tranformation $g_\gamma(\cdot)$ of Eqn. 10.

Proof: Since β is a monotone (and 1-1) function of m_3, it suffices to show that m_3 is a maximal invariant. This requires showing that given any pair of data vectors $\underline{z}'$ and $\underline{z}$ such that $m_3' = m_3$, then there exists a transformation $g_\gamma(\cdot)$ of the form of Equation 10, such that $\underline{z}' = g_\gamma(\underline{z})$. Suppose $m_3' = m_3$. Then

$$\frac{m_1'}{m_2'} = \frac{m_1}{m_2} \to \frac{m_1'}{m_1} = \frac{m_2'}{m_2} = c^2, \tag{11}$$

where c is an unknown (positive) scaling constant. So we have

$$\begin{aligned} \|\underline{z}_1'\|^2 &= m_1' = c^2 m_1 = \|c\underline{z}_1\|^2 \\ \|\underline{z}_2'\|^2 &= m_2' = c^2 m_2 = \|c\underline{z}_2\|^2. \end{aligned} \tag{12}$$

Then from that fact that (m_1, m_2) is a maximal invariant with respect to $g(\cdot)$ of Equation 6 (Lemma 1), we know that there exists a transformation in the form of $g(\cdot)$ such that $\underline{z}' = g(c\underline{z}) = cg(\underline{z})$. And by setting $\gamma = c$ in Equation 10, we then know that there exists a transformation in the form of $g_\gamma(\cdot)$ such that $\underline{z}' = g_\gamma(\underline{z})$. ■

The statistic $m_3 = \frac{\|\underline{z}_1\|^2}{\|\underline{z}_2\|^2}$ is a ratio of chi-squared statistics, and is thus F-distributed. It is the CFAR (Constant False Alarm Rate) MSD of [2], [3]. The statistic β is its beta-distributed version:

$$\frac{m_1}{m_2} = m_3 = F = \frac{\beta}{1-\beta}; \quad \beta = \frac{F}{1+F} = \frac{m_3}{1+m_3} = \frac{m_1}{m_1+m_2}. \tag{13}$$

In the cases of rank-1 and multi-rank signals, β is given by

$$\begin{aligned} \beta &= \frac{|\underline{\psi}^\dagger \mathbf{R}^{-1} \underline{y}|^2}{(\underline{\psi}^\dagger \mathbf{R}^{-1} \underline{\psi})(\underline{y}^\dagger \mathbf{R}^{-1} \underline{y})} \gtrless \eta, \quad p = 1; \\ \beta &= \frac{\underline{y}^\dagger \mathbf{R}^{-1} \mathbf{H} (\mathbf{H}^\dagger \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^\dagger \mathbf{R}^{-1} \underline{y}}{\underline{y}^\dagger \mathbf{R}^{-1} \underline{y}} = \frac{\|\mathbf{P_G} \underline{z}\|^2}{\|\underline{z}\|^2} \gtrless \eta, \quad p > 1. \end{aligned} \tag{14}$$

These compare to a threshold the cosine-squared of the angle between the measurement $\underline{y}$ and the signal subspace $\langle\underline{\psi}\rangle$ or $\langle\mathbf{G}\rangle$. These cosines are invariant to transformations that leave this angle unaffected, which can be viewed geometrically as transformations that leave $\underline{z}$ on the surface of a cone.

III. Maximal Invariance in the Adaptive Problem

A. Development of the Transformation Group

In the adaptive problem the covariance matrix $\mathbf{R}$ is not assumed to be known *a priori*, meaning it must be estimated. We assume K training data vectors $\underline{x}_1 \dots \underline{x}_K$ are recorded. They are presumed to share the same noise statistics as the noise in the test data: $\underline{x}_i \sim CN_N[\underline{0}, \mathbf{R}]$. We construct a joint test and training data matrix as follows: $[\underline{y}|\mathbf{X}] = [\underline{y}|\underline{x}_1|\cdots|\underline{x}_K] \in \mathbb{C}^{N\times(K+1)}$.

Without loss of generality, we again make the simplifying assumption that a unitary transformation has been applied to the data to rotate the signal subspace into the subspace of the first p standard basis vectors. This unitary matrix will have the same form as that of Equation 3, but with $\mathbf{H}$ replacing $\mathbf{G} = \mathbf{R}^{-\frac{1}{2}}\mathbf{H}$ (note that in the adaptive problem, $\mathbf{H}$ is known but $\mathbf{R}$ is not). In the derivations that follow, we assume that this transformation has been made, and define the symbols $\underline{y}$ and $\mathbf{X}$ to be the test and training data in this coordinate system. We then have

$$\mathbf{U}_{\mathbf{H}}^{\dagger}[\underline{y}|\mathbf{X}] \rightarrow \quad [\underline{y}|\mathbf{X}] = \begin{bmatrix} \underline{y}_1 & \mathbf{X}_1 \\ \underline{y}_2 & \mathbf{X}_2 \end{bmatrix} \in \mathbb{C}^{N\times(K+1)}, \tag{15}$$

where the subscripts 1 and 2 indicate resolutions of the data onto the signal and signal-free subspaces.

When some prior knowledge about the covariance structure can be reasonably exploited, for example Toeplitz structure, or low-rank interference plus white noise, the GLRT approach to finding a detection statistic generally becomes intractable. For this reason, Bose and Steinhardt [13] considered the adaptive detection problem from the point of view of maximal invariance. They characterized transformations $G(\cdot)$ of the joint test and training data matrix, $[\underline{y}|\mathbf{X}]$ that leave unchanged the structural form of the covariance, and preserve the parameter regions that describe the hypotheses H_0 and H_1. Specifically, they required preservation of (1) Gaussianity, leading to consideration of affine transformations, (2) the covariance structure, and (3) the mean (i.e., the mean of $\underline{z}$ should be a vector in the signal subspace under H_1, and the zero vector under H_0; zero-mean data should remain so.) Such transformations leave the detection problem invariant (see [12], Chapter 4).

When no structure on the unknown covariance $\mathbf{R}$ is to be enforced, we simply require that the covariance remain positive-definite, and remain *the same* in both test and training data [13]. By concatenating the columns of the data matrix $[\underline{y}|\mathbf{X}]$ to form $\mathrm{vec}([\underline{y}|\mathbf{X}])$, the unknown joint covariance of the test and training data may be written as

$$\mathrm{cov}(\mathrm{vec}([\underline{y}|\mathbf{X}])) = \mathbf{R} \otimes \mathbf{I}_{K+1}, \tag{16}$$

where $\otimes$ indicates the Kronecker product. We thus require invariance to transformations of the joint test and training data matrix which preserve positive-definiteness, and which ensure that all the transformed test and training data vectors still share the same covariance.

For rank-1 signals, Bose and Steinhardt showed that an affine transformation satisfies these properties if and only if it has the following form (see the Appendix of [13]):

$$G([\mathbf{y}|\mathbf{X}]) = \begin{bmatrix} \alpha & \underline{\beta}^\dagger \\ \underline{0} & \mathbf{\Gamma} \end{bmatrix} \begin{bmatrix} y_1 & \underline{x}_1 \\ \underline{y}_2 & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} 1 & \underline{0}^\dagger \\ \underline{0} & \mathbf{U} \end{bmatrix}, \tag{17}$$

where, in a notational compromise, $\underline{x}_1$ denotes a *row* vector. The pre-multiplying matrix preserves the structure of the mean and the positive-definiteness of the covariance of the columns of $[\mathbf{y}|\mathbf{X}]$; thus, $\mathbf{\Gamma}$ is full rank. The post-multiplying matrix is unitary to preserve the i.i.d. (independent, identically distributed) nature of the noise in the data vectors; thus, $\mathbf{U}$ is unitary. The post-multiplying matrix has its first row equal to $\underline{e}_1$ to preserve zero mean in all data vectors but the test vector.

For multi-rank signal subspaces, this transformation group generalizes to

$$G([\underline{y}|\mathbf{X}]) = \begin{bmatrix} \mathbf{A} & \mathbf{B}^\dagger \\ \mathbf{0} & \mathbf{\Gamma} \end{bmatrix} \begin{bmatrix} \underline{y}_1 & \mathbf{X}_1 \\ \underline{y}_2 & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} 1 & \underline{0}^\dagger \\ \underline{0} & \mathbf{U} \end{bmatrix}, \tag{18}$$

where both $\mathbf{A}$ and $\mathbf{\Gamma}$ are full-rank and $\mathbf{U}$ is again unitary. This is actually a special case of the transformation in [21], wherein Bose and Steinhardt considered the signal to be spread in both the row and column space of the data matrix. We will denote the pre-multiplying and post-multiplying matrices by $\mathcal{A}$ and $\mathcal{U}$:

$$\mathcal{A} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^\dagger \\ \mathbf{0} & \mathbf{\Gamma} \end{bmatrix}; \quad \mathcal{U} = \begin{bmatrix} 1 & \underline{0}^\dagger \\ \underline{0} & \mathbf{U} \end{bmatrix}. \tag{19}$$

The pre-multiplying matrix $\mathcal{A}$ leaves the signal in the signal subspace spanned by the first p standard basis vectors, $\underline{e}_1 \ldots \underline{e}_p$.

B. A 2-D Maximal Invariant for the Homogeneous Detection Problem

For the case of rank-1 signals, Bose and Steinhardt showed that the following scalars form a two-dimensional maximal invariant with respect to $G(\cdot)$:

$$\hat{m}_1 = \frac{\|y_1 - \underline{x}_1 \mathbf{X}_2^\dagger (\mathbf{X}_2 \mathbf{X}_2^\dagger)^{-1} \underline{y}_2\|^2}{\underline{x}_1 \left(\mathbf{I} - \mathbf{X}_2^\dagger (\mathbf{X}_2 \mathbf{X}_2^\dagger)^{-1} \mathbf{X}_2 \right) \underline{x}_1^\dagger} \tag{20}$$

$$\hat{m}_2 = \underline{y}_2^\dagger (\mathbf{X}_2 \mathbf{X}_2^\dagger)^{-1} \underline{y}_2, \tag{21}$$

where the hat notation is used to indicate that these are adaptive statistics. Interestingly, these are a 1-1 function of the statistics given by the AMF and Kelly's GLRT [13], as we will soon show. The AMF is obtained by simply substituting the sample covariance matrix $\mathbf{S} = \mathbf{X}\mathbf{X}^\dagger$ for the known covariance $\mathbf{R}$ in the standard matched filter [15], [16]:

$$\widehat{r^2} = \frac{|\underline{\psi}^\dagger \mathbf{S}^{-1} \underline{y}|^2}{\underline{\psi}^\dagger \mathbf{S}^{-1} \underline{\psi}} = \hat{m}_1. \tag{22}$$

Kelly's statistic is obtained as a GLRT by inserting maximum-likelihood estimates for the complex signal amplitude μ and the covariance matrix $\mathbf{R}$ into the likelihood ratio [14]:

$$\widehat{\kappa^2} = \frac{|\underline{\psi}^\dagger \mathbf{S}^{-1} \underline{y}|^2}{\underline{\psi}^\dagger \mathbf{S}^{-1} \underline{\psi} (1 + \underline{y}^\dagger \mathbf{S}^{-1} \underline{y})} = \frac{\hat{m}_1}{1 + \hat{m}_1 + \hat{m}_2}. \tag{23}$$

To establish the one-to-one map between $(\hat{m}_1, \hat{m}_2)$ and $(\widehat{r^2}, \widehat{\kappa^2})$ requires some manipulations involving inverses for partitioned matrices. For generality, we will treat the case of multi-dimensional signal subspaces, and identify the multi-rank generalizations of $\hat{m}_1$ and $\hat{m}_2$, required for our subsequent analysis of ACE. First we partition the sample covariance matrix as follows:

$$\mathbf{S} = \mathbf{X}\mathbf{X}^\dagger = \begin{bmatrix} \mathbf{X}_1 \mathbf{X}_1^\dagger & \mathbf{X}_1 \mathbf{X}_2^\dagger \\ \mathbf{X}_2 \mathbf{X}_1^\dagger & \mathbf{X}_2 \mathbf{X}_2^\dagger \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix}, \tag{24}$$

in order to prove the following lemma.

Lemma 2: Let $\underline{y}_{1.2} = \underline{y}_1 - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\underline{y}_2$ be the error vector associated with estimating $\underline{y}_1$ from $\underline{y}_2$ using the sample estimate $\mathbf{S}_{12}\mathbf{S}_{22}^{-1}$ for the linear minimum mean-square-error filter. Denote the sample error covariance by $\mathbf{S}_{11.2} = \mathbf{S}_{11} - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}$. Then the quadratic form $\underline{y}^\dagger\mathbf{S}^{-1}\underline{y}$ can be decomposed as $\underline{y}^\dagger\mathbf{S}^{-1}\underline{y} = \underline{y}_{1.2}^\dagger\mathbf{S}_{11.2}^{-1}\underline{y}_{1.2} + \underline{y}_2^\dagger\mathbf{S}_{22}^{-1}\underline{y}_2$.

Proof: The block-Cholesky decomposition of the sample covariance and its inverse are given by

$$\begin{aligned}\mathbf{S} = \begin{bmatrix}\mathbf{S}_{11} & \mathbf{S}_{12}\\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{bmatrix} &= \begin{bmatrix}\mathbf{I} & \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{S}_{11.2} & \mathbf{0}\\ \mathbf{0} & \mathbf{S}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{S}_{22}^{-1}\mathbf{S}_{21} & \mathbf{I}\end{bmatrix} \quad (25)\\ \mathbf{S}^{-1} &= \begin{bmatrix}\mathbf{I} & \mathbf{0}\\ -\mathbf{S}_{22}^{-1}\mathbf{S}_{21} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{S}_{11.2}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{S}_{22}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{I} & -\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}. \quad (26)\end{aligned}$$

Let $\mathbf{W}$ denote the block upper-triangular filtering matrix, and $\mathbf{Q}$ denote the block diagonal pre-whitening matrix:

$$\mathbf{W} = \begin{bmatrix}\mathbf{I} & -\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}; \quad \mathbf{Q} = \begin{bmatrix}\mathbf{S}_{11.2}^{-\frac{1}{2}} & \mathbf{0}\\ \mathbf{0} & \mathbf{S}_{22}^{-\frac{1}{2}}\end{bmatrix}. \quad (27)$$

Then the quadratic form may be written as

$$\begin{aligned}\underline{y}^\dagger\mathbf{S}^{-1}\underline{y} &= \begin{bmatrix}\underline{y}_1^\dagger & \underline{y}_2^\dagger\end{bmatrix}\mathbf{W}^\dagger\mathbf{Q}^\dagger\mathbf{Q}\mathbf{W}\begin{bmatrix}\underline{y}_1\\ \underline{y}_2\end{bmatrix} = \begin{bmatrix}\underline{y}_{1.2}^\dagger & \underline{y}_2^\dagger\end{bmatrix}\mathbf{Q}^\dagger\mathbf{Q}\begin{bmatrix}\underline{y}_{1.2}\\ \underline{y}_2\end{bmatrix}\\ &= \underline{y}_{1.2}^\dagger\mathbf{S}_{11.2}^{-1}\underline{y}_{1.2} + \underline{y}_2^\dagger\mathbf{S}_{22}^{-1}\underline{y}_2. \quad \blacksquare \quad (28)\end{aligned}$$

Let us now label these two quadratric forms $\hat{m}_1$ and $\hat{m}_2$,

$$\hat{m}_1 = \underline{y}_{1.2}^\dagger\mathbf{S}_{11.2}^{-1}\underline{y}_{1.2}, \quad \hat{m}_2 = \underline{y}_2^\dagger\mathbf{S}_{22}^{-1}\underline{y}_2, \quad (29)$$

and consider whether they are affected by the transformation $G(\cdot)$ in Eqn. 18.

Lemma 3: The 2-D pair $(\hat{m}_1, \hat{m}_2) = (\underline{y}_{1.2}^\dagger\mathbf{S}_{11.2}^{-1}\underline{y}_{1.2}, \underline{y}_2^\dagger\mathbf{S}_{22}^{-1}\underline{y}_2)$ is invariant to the transformation $G(\cdot)$ of Eqn. 18, for multi-rank signal subspaces.

Proof: To prove this, we identify the sum as

$$\hat{m}_0 = \hat{m}_1 + \hat{m}_2 = \underline{y}^\dagger\mathbf{S}^{-1}\underline{y}, \quad (30)$$

and show that $(\hat{m}_0, \hat{m}_2)$ is invariant. That $\hat{m}_1 = \hat{m}_0 - \hat{m}_2$ is invariant then follows trivially. We again refer to the pre and post multiplying matrices, as $\mathcal{A}$ and $\mathcal{U}$, as defined in Equation 19. The test vector and training data matrix are transformed as $\underline{y}' = \mathcal{A}\underline{y}$ and $\mathbf{X}' = \mathcal{A}\mathbf{X}\mathcal{U}$. Then m_0 is invariant by a straightforward calculation:

$$\begin{aligned} \hat{m}_0' &= \underline{y}'^{\dagger}\mathbf{S}'^{-1}\underline{y}' = \underline{y}^{\dagger}\mathcal{A}^{\dagger}\left(\mathcal{A}\mathbf{X}\mathcal{U}\mathcal{U}^{\dagger}\mathbf{X}^{\dagger}\mathcal{A}^{\dagger}\right)^{-1}\mathcal{A}\underline{y} \\ &= \underline{y}^{\dagger}\mathcal{A}^{\dagger}\mathcal{A}^{\dagger-1}\left(\mathbf{X}\mathbf{X}^{\dagger}\right)^{-1}\mathcal{A}^{-1}\mathcal{A}\underline{y} = \underline{y}^{\dagger}\left(\mathbf{X}\mathbf{X}^{\dagger}\right)^{-1}\underline{y} = \hat{m}_0. \end{aligned}$$

Similarly, for $\hat{m}_2$ we have $\underline{y}_2' = \mathbf{\Gamma}\underline{y}_2$ and $\mathbf{X}_2' = \mathbf{\Gamma}\mathbf{X}_2\mathbf{U}$. This gives

$$\begin{aligned} \hat{m}_2' &= \underline{y}_2'^{\dagger}\mathbf{S}_{22}'^{-1}\underline{y}_2' = \underline{y}_2^{\dagger}\mathbf{\Gamma}^{\dagger}\left(\mathbf{\Gamma}\mathbf{X}_2\mathbf{U}\mathbf{U}^{\dagger}\mathbf{X}_2^{\dagger}\mathbf{\Gamma}^{\dagger}\right)^{-1}\mathbf{\Gamma}\underline{y}_2 \\ &= \underline{y}_2^{\dagger}\mathbf{\Gamma}^{\dagger}\mathbf{\Gamma}^{\dagger-1}\left(\mathbf{X}_2\mathbf{X}_2^{\dagger}\right)^{-1}\mathbf{\Gamma}^{-1}\mathbf{\Gamma}\underline{y}_2 = \underline{y}_2^{\dagger}\left(\mathbf{X}_2\mathbf{X}_2^{\dagger}\right)^{-1}\underline{y}_2 = \hat{m}_2. \quad \blacksquare \end{aligned}$$

In the case of a dimension-1 signal subspace ($p = 1$), $\hat{m}_1$ and $\hat{m}_2$ in Eqn. 29 simplify to the 2-D maximal invariant identified by Bose and Steinhardt. Note that for a rank-1 signal, the steering vector in our rotated coordinate system is given by $\underline{\psi} = \underline{e}_1 = [1 \quad 0\cdots 0]^{\dagger}$. Then by using Eqn. 26, $\hat{m}_1$ in Eqn. 22 is given by Eqn. 29. Furthermore, Lemma 2 identifies the sum of $\hat{m}_1$ and $\hat{m}_2$ as $\underline{y}^{\dagger}\mathbf{S}^{-1}\underline{y}$, meaning that Eqn. 23 may now be obtained as follows:

$$\widehat{\kappa^2} = \frac{|\underline{\psi}^{\dagger}\mathbf{S}^{-1}\underline{y}|^2}{\underline{\psi}^{\dagger}\mathbf{S}^{-1}\underline{\psi}(1 + \underline{y}^{\dagger}\mathbf{S}^{-1}\underline{y})} = \frac{\hat{m}_1}{1 + \underline{y}^{\dagger}\mathbf{S}^{-1}\underline{y}} = \frac{\hat{m}_1}{1 + \hat{m}_1 + \hat{m}_2}. \tag{31}$$

Bose and Steinhardt showed that $(\hat{m}_1, \hat{m}_2)$ are not only an invariant of $G(\cdot)$, but a *maximal invariant*. This means that any two data matrices $[\underline{y}, \mathbf{X}]$ and $[\underline{y}', \mathbf{X}']$ that yield the same value for the pair $(\hat{m}_1, \hat{m}_2)$ are related by a linear transformation in the form of Eqn. 18. This result is extendable to the general multi-rank signal case, which we now state as Lemma 4.

Lemma 4: The pair $(\hat{m}_1, \hat{m}_2)$ of Eqn. 29 is a 2-D maximal invariant for the transformation of Eqn. 18, in the general case of multi-dimensional signal subspaces. *Proof:* This is shown in Appendix A.

C. ACE: a 1-D Maximal Invariant for an Inhomogeneous Detection Problem

The adaptive analogue to the hypothesis testing problem of Section II-B is one in which the test and training data are required to have the same covariance structure, but not the same *scaling*:

$$cov(\text{vec}([\underline{y}|\mathbf{X}])) = \left[\begin{array}{c|c} \sigma^2\mathbf{R} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{R}\otimes\mathbf{I}_K \end{array}\right]. \tag{32}$$

Accordingly, we can require invariance to the more general transformation:

$$G_\gamma([\underline{y}|\mathbf{X}]) = \begin{bmatrix} \mathbf{A} & \mathbf{B}^\dagger \\ \mathbf{0} & \mathbf{\Gamma} \end{bmatrix} \begin{bmatrix} \underline{y}_1 & \mathbf{X}_1 \\ \underline{y}_2 & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \gamma & \underline{0}^\dagger \\ \underline{0} & \mathbf{U} \end{bmatrix}. \tag{33}$$

This is identical to the transformation of Equation 18, except for the introduction of the scaling γ, whose only effect is to scale the test vector $\underline{y}$ by γ, and to accordingly scale its covariance by γ^2.

Again, the statistics $\hat{m}_1$ and $\hat{m}_2$ are no longer invariant to $G_\gamma(\cdot)$: $\hat{m}_1' = \gamma^2\hat{m}_1$ and $\hat{m}_2' = \gamma^2\hat{m}_2$. However, their ratio, $\hat{m}_3 = \frac{\hat{m}_1}{\hat{m}_2}$ *is* an invariant, and furthermore it is a maximal invariant to this transformation. The ratio $\hat{\beta} = \frac{\hat{m}_3}{1+\hat{m}_3}$ is the ACE statistic, and is also a maximal invariant. This brings us to our first key result.

Theorem 1: The ACE statistic $\hat{\beta} = \frac{\hat{m}_3}{1+\hat{m}_3}$ is a *maximal* invariant with respect to $G_\gamma(\cdot)$. *Proof:* Since $\hat{\beta}$ is a monotone function of $\hat{m}_3$ it suffices to show that $\hat{m}_3$ is a maximal invariant, which is proved in the same way that Proposition 1 was proved from Lemma 1 in Section II-B. ■

The adaptive statistics $\hat{\beta}$ and $\hat{m}_3$ are related as follows:

$$\frac{\hat{m}_1}{\hat{m}_2} = \hat{m}_3 = \hat{F} = \frac{\hat{\beta}}{1-\hat{\beta}}; \quad \hat{\beta} = \frac{\hat{F}}{1+\hat{F}} = \frac{\hat{m}_3}{1+\hat{m}_3} = \frac{\hat{m}_1}{\hat{m}_1+\hat{m}_2}. \tag{34}$$

Then in the cases of rank-1 and multi-rank signal, the maximal invariant ACE is given by

$$\begin{aligned} \hat{\beta} &= \frac{|\underline{\psi}^\dagger\mathbf{S}^{-1}\underline{y}|^2}{(\underline{\psi}^\dagger\mathbf{S}^{-1}\underline{\psi})(\underline{y}^\dagger\mathbf{S}^{-1}\underline{y})} \gtrless \eta \quad p=1; \\ \hat{\beta} &= \frac{\underline{y}^\dagger\mathbf{S}^{-1}\mathbf{H}(\mathbf{H}^\dagger\mathbf{S}^{-1}\mathbf{H})^{-1}\mathbf{H}^\dagger\mathbf{S}^{-1}\underline{y}}{\underline{y}^\dagger\mathbf{S}^{-1}\underline{y}} = \frac{\|\mathbf{P}_{\hat{\mathbf{G}}}\underline{\hat{z}}\|^2}{\|\underline{\hat{z}}\|^2} \gtrless \eta, \quad p>1, \end{aligned} \tag{35}$$

where $\hat{\mathbf{G}} = \mathbf{S}^{-\frac{1}{2}}\mathbf{H}$ and $\underline{\hat{z}} = \mathbf{S}^{-\frac{1}{2}}\underline{z}$.

IV. ACE has Monotone Likelihood Ratio

A. The Statistical Distribution of ACE

We consider the distribution in the general multi-rank case, where the signal subspace has dimension p to accomodate uncertain signal waveforms. We begin with a result in [22], [4], which is that the "F" version of ACE has a stochastic representation in terms of five statistically independent random variables,

$$\hat{F} = \frac{1}{h_1^2}\left|\underline{n}\frac{h_2}{g} + \underline{h_3}\right|^2, \tag{36}$$

where $\underline{n}$ and $\underline{h_3}$ are Gaussian distributed, and g^2, h_2^2 and h_1^2 are chi-squared distributed. More specifically, $\underline{n}$ and g^2 model the random realizations of the test-data vector $\underline{y}$, and are distributed as $\underline{n} \sim CN_p[\frac{\mu}{\sigma}(\mathbf{H}^\dagger\mathbf{R}^{-1}\mathbf{H})^{-\frac{1}{2}}\mathbf{H}^\dagger\mathbf{R}^{-1}\mathbf{H}\underline{\theta}, \mathbf{I}_p]$ and $g^2 \sim \chi^2_{N-p}[0]$. The random variables h_1^2, h_2^2 and $\underline{h}_3$ account for the random realizations of the training-data matrix $\mathbf{X}$, and are distributed as ${h_1}^2 \sim \chi^2_{K-N+1}[0]$, ${h_2}^2 \sim \chi^2_{K-N+p+1}[0]$, and $\underline{h}_3 \sim CN_p[\underline{0}, \mathbf{I}_p]$. In all derivations we assume the distributions are obtained from complex Gaussian random variables, wherein N degrees of freedom refers to a random variable generated from N proper complex normal random variables of unity variance, or equivalently from $2N$ real random variables of one-half variance.

The determination of the density function for $\hat{F}$ is simplified by conditioning on the beta-distributed random variable $b = \frac{h_2^2}{h_2^2+g^2} \sim \beta_{K-N+1+p,N-p}$ (an approach used by Kelly in his study of $\widehat{\kappa^2}$ [14]). Letting $L = K - N + 1$, the density of b is a beta density with $L+p$ and $N-p$ degrees of freedom:

$$f(b) = \frac{b^{L+p-1}(1-b)^{N-p-1}}{B(L+p, N-p)}, \quad 0 < b < 1. \tag{37}$$

The coefficient $B(L+p, N-p) = \frac{\Gamma(L+p)\Gamma(N-p)}{\Gamma(L+N)}$ is a beta function, where $\Gamma(\cdot)$ is a gamma function.

Then it can be shown [23], [24], [4] that conditioned on the beta-distributed random variable b, $\hat{F}$ in Equation 36 has a non-central F distribution:

$$\hat{F}|b \sim F_{p,L}[\alpha \cdot b] \cdot \frac{1}{1-b}, \tag{38}$$

where the parameter $\alpha = \mu^2\underline{\psi}^\dagger\mathbf{R}^{-1}\underline{\psi}$, $\underline{\psi} = \mathbf{H}\underline{\theta}$, is the non-centrality parameter. In the case of sensor array detection, it is the output signal-to-noise ratio (SNR) of the array. The arguments p and L are the numerator and denominator degrees of freedom, respectively.

It is well known that the non-central F density can be written as a mixture density, an infinite sum of central F densities, weighted by Poisson coefficents [11]:

$$f_{q,r}(x;\mu^2) = \sum_{m=0}^{\infty} P(m,\mu^2) f_{q+m,r}(x;0), \tag{39}$$

where the non-centrality parameter μ^2 appears in the Poisson coefficient $P(m,\mu^2) = \frac{(\mu^2)^m}{m!}e^{-\mu^2}$, and $f_{q+m,r}(x;0)$ is a central F density with $q+m$ numerator and r denominator degrees of freedom. Conditioned on the parameter b, the density of $\hat{F}$ is then given by

$$f(\hat{F}|b;\alpha) = \sum_{m=0}^{\infty} \frac{P(m,\alpha\cdot b)}{B(L,p+m)} \frac{(1-b)^{p+m}\hat{F}^{p+m-1}}{[1+(1-b)\hat{F}]^{L+p+m}}. \tag{40}$$

Making the change of variable $\hat{\beta} = \frac{\hat{F}}{1+\hat{F}}$, we have

$$f(\hat{\beta}|b) = \sum_{m=0}^{\infty} P(m,\alpha\cdot b)\frac{\hat{\beta}^{p+m-1}(1-\hat{\beta})^{L-1}}{B(p+m,L)} \frac{(1-b)^{p+m}}{[1-b\hat{\beta}]^{L+p+m}}. \tag{41}$$

Note that each term in this infinite series contains a beta density with $p+m$ and L degrees of freedom.

By integrating over the marginal density of b, the non-central density of ACE in its $\hat{\beta}$ form may be written as an integral:

$$\begin{aligned} f(\hat{\beta};\alpha) &= \int_0^1 f(\hat{\beta}|b;\alpha)f(b)\mathrm{d}b \\ &= \int_0^1 \sum_{m=0}^{\infty} P(m,\alpha\cdot b)\frac{\hat{\beta}^{p+m-1}(1-\hat{\beta})^{L-1}}{B(p+m,L)} \frac{(1-b)^{p+m}}{[1-b\hat{\beta}]^{L+p+m}} \frac{b^{L+p-1}(1-b)^{N-p-1}}{B(L+p,N-p)}\mathrm{d}b \quad (42) \\ &= \sum_{m=0}^{\infty} \frac{\alpha^m}{m!} \frac{1}{B(L+p,N-p)} \frac{\hat{\beta}^{p+m-1}(1-\hat{\beta})^{L-1}}{B(p+m,L)} \cdot \int_0^1 \frac{b^{L+p+m-1}(1-b)^{N+m-1}}{[1-b\hat{\beta}]^{L+p+m}} e^{-\alpha b}\mathrm{d}b. \quad (43) \end{aligned}$$

Under the H_0 hypothesis ($\alpha=0$), only the first term of the summation remains, and the density is given by

$$\begin{aligned} f(\hat{\beta};0) &= \int_0^1 f(\hat{\beta}|b;0)f(b)\mathrm{d}b \\ &= \int_0^1 \frac{\hat{\beta}^{p-1}(1-\hat{\beta})^{L-1}}{B(p,L)} \frac{(1-b)^{p}}{[1-b\hat{\beta}]^{L+p}} \frac{b^{L+p-1}(1-b)^{N-p-1}}{B(L+p,N-p)}\mathrm{d}b \quad (44) \\ &= \frac{1}{B(L+p,N-p)} \frac{\hat{\beta}^{p-1}(1-\hat{\beta})^{L-1}}{B(p,L)} \cdot \int_0^1 \frac{b^{L+p-1}(1-b)^{N-1}}{[1-b\hat{\beta}]^{L+p}}\mathrm{d}b. \quad (45) \end{aligned}$$

There is an interesting connection between the dependence of the distribution on the output signal-to-noise ratio α and invariance theory. Space constraints do not permit a full discussion of it here, but the transformation group $G_\gamma(\cdot)$ on the data matrix $[\underline{y}|\mathbf{X}]$ induces a transformation group $\overline{G}_\gamma(\cdot)$ that acts on the parameters $(\mu\underline{\psi},\mathbf{R},\sigma^2)$. The distribution of the statistic that is a maximal invariant with respect to $G_\gamma(\cdot)$ (i.e., $\hat{\beta}$) will then depend on the parameter which is a maximal invariant with respect to $\overline{G}_\gamma(\cdot)$ (this is Theorem 3 in Chapter 6 of [11]). In this case, $\alpha = \mu^2 \underline{\psi}^\dagger \mathbf{R}^{-1} \underline{\psi} = \mu^2 \underline{\theta}^\dagger \mathbf{H}^\dagger \mathbf{R}^{-1} \mathbf{H} \underline{\theta}$ is the maximal invariant with respect to the induced transformation $\overline{G}_\gamma(\cdot)$, and the distribution of $\hat{\beta}$ thus depends on α.

To summarize, Eqns. 42 and 44 are representations for the density of $\hat{\beta}$ for $\alpha > 0$ and for $\alpha = 0$. These can be used to obtain numerical approximations for the densities, as shown in Figure 1, for various SNRs. The corresponding log-likelihood ratios are shown in Figure 2. Plots such as these suggest that the conjecture that ACE has monotone likelihood ratio [18] might be true.

We wish to *prove* that this is true, for all possible values of N, p, K, $\alpha > 0$, and $0 < \hat{\beta} < 1$. In order to show this, our intent from this point forward is to show that the density of $\hat{\beta}$ possesses the property of "total positivity," to be defined below, and from this conclude that $\hat{\beta}$ has monotone likelihood ratio.

B. Total Positivity

Consider the likelihood ratio, which we denote by $\lambda(\hat{\beta}, \alpha)$:

$$\lambda(\hat{\beta}; \alpha) = \frac{f(\hat{\beta}; \alpha)}{f(\hat{\beta}; 0)}, \quad \alpha > 0. \tag{46}$$

A threshold test on the likelihood ratio can be replaced by a threshold test on the statistic $\hat{\beta}$ if the likelihood ratio is a monotone function of $\hat{\beta}$. This condition can be expressed as

$$\frac{f(\hat{\beta}_1; \alpha)}{f(\hat{\beta}_1; 0)} > \frac{f(\hat{\beta}_0; \alpha)}{f(\hat{\beta}_0; 0)}, \quad \text{for all} \quad 1 > \hat{\beta}_1 > \hat{\beta}_0 > 0 \quad \text{and} \quad \alpha > 0. \tag{47}$$

This is equivalent to positivity of the determinant of the matrix

$$\mathbf{F} = \begin{bmatrix} f(\hat{\beta}_0; 0) & f(\hat{\beta}_0; \alpha) \\ f(\hat{\beta}_1; 0) & f(\hat{\beta}_1; \alpha) \end{bmatrix}. \tag{48}$$

To establish the positivity of the determinant of $\mathbf{F}$, we will refer to a more general concept, termed "total positivity" by Karlin [20].

Definition 1 (Total Positivity, STP_n) A kernel function of two variables $f(x, \theta)$ is defined to be "totally positive" of order n (abbreviated TP_n) if given n^2 values of the function evaluated at the ordered points $x_0 \ldots x_{n-1}$ and $\theta_0 \ldots \theta_{n-1}$, the $n \times n$ matrix $\mathbf{F}$, whose (i, j) element is given by $\mathbf{F}_{i,j} = f(x_i, \theta_j)$, has a non-negative determinant:

$$\|\mathbf{F}\| \geq 0. \tag{49}$$

When the inequality is strict, this condition is called "strictly totally positive," and abbreviated STP_n.

So to show that the likelihood ratio is monotone, it is sufficient to interpret $f(\hat{\beta};\alpha)$ as a kernel function of two variables, and to show that it is strictly totally positive of order 2, or STP_2. We wish to show that

$$\|\mathbf{F}\| = \left\| \begin{matrix} f(\beta_0;\alpha_0) & f(\beta_0;\alpha_1) \\ f(\beta_1;\alpha_0) & f(\beta_1;\alpha_1) \end{matrix} \right\| > 0; \quad 1 > \beta_1 > \beta_0 > 0, \quad \alpha_1 > \alpha_0 = 0. \tag{50}$$

To show this we will make repeated use of a basic *composition* formula, originally due to Polya and Szego (see p.17 of [20], also [25], [26]). This formula can be stated generally in terms of n-dimensional integrals over order-n kernels. For simplicity, we will only state and prove it here for the order-2 case that is of interest to us.

Integral Composition. Suppose $h(\theta,\alpha)$ is the composition of two kernels g and f: $h(\theta,\alpha) = \int g(\theta,x) f(x,\alpha)\mathrm{d}x$. Then the corresponding determinant for h is given by an integral of determinants:

$$\|\mathbf{H}\| = \left\| \begin{matrix} h(\theta_0,\alpha_0) & h(\theta_0,\alpha_1) \\ h(\theta_1,\alpha_0) & h(\theta_1,\alpha_1) \end{matrix} \right\| = \iint\limits_{x<y} \left\| \begin{matrix} g(\theta_0,x) & g(\theta_0,y) \\ g(\theta_1,x) & g(\theta_1,y) \end{matrix} \right\| \left\| \begin{matrix} f(x,\alpha_0) & f(x,\alpha_1) \\ f(y,\alpha_0) & f(y,\alpha_1) \end{matrix} \right\| \mathrm{d}x\mathrm{d}y. \tag{51}$$

This can be shown by direct expansion of the determinants as follows:

$$\begin{aligned}
&\iint\limits_{x<y} \begin{Vmatrix} g(\theta_0, x) & g(\theta_0, y) \\ g(\theta_1, x) & g(\theta_1, y) \end{Vmatrix} \begin{Vmatrix} f(x, \alpha_0) & f(x, \alpha_1) \\ f(y, \alpha_0) & f(y, \alpha_1) \end{Vmatrix} \mathrm{d}x\mathrm{d}y \\
&= \iint\limits_{x<y} \left[g(\theta_0, x)g(\theta_1, y) - g(\theta_1, x)g(\theta_0, y)\right]\left[f(x, \alpha_0)f(y, \alpha_1) - f(x, \alpha_1)f(y, \alpha_0)\right] \mathrm{d}x\mathrm{d}y \\
&= \iint\limits_{x<y} \left[g(\theta_0, x)f(x, \alpha_0) \cdot g(\theta_1, y)f(y, \alpha_1) - g(\theta_0, y)f(y, \alpha_1) \cdot g(\theta_1, x)f(x, \alpha_0)\right] \mathrm{d}x\mathrm{d}y \\
&+ \iint\limits_{x<y} \left[g(\theta_0, y)f(y, \alpha_0) \cdot g(\theta_1, x)f(x, \alpha_1) - g(\theta_0, x)f(x, \alpha_1) \cdot g(\theta_1, y)f(y, \alpha_0)\right] \mathrm{d}x\mathrm{d}y \\
&= \iint \left[g(\theta_0, x)f(x, \alpha_0) \cdot g(\theta_1, y)f(y, \alpha_1) - g(\theta_0, y)f(y, \alpha_1) \cdot g(\theta_1, x)f(x, \alpha_0)\right] \mathrm{d}x\mathrm{d}y \\
&= \int g(\theta_0, x)f(x, \alpha_0)\mathrm{d}x \int g(\theta_1, y)f(y, \alpha_1)\mathrm{d}y - \int g(\theta_0, y)f(y, \alpha_1)\mathrm{d}y \int g(\theta_1, x)f(x, \alpha_0)\mathrm{d}x \\
&= h(\theta_0, \alpha_0)h(\theta_1, \alpha_1) - h(\theta_0, \alpha_1)h(\theta_1, \alpha_0). \qquad (52)
\end{aligned}$$

We will use the fact that the formula also holds if the composition is a sum rather than an integral.

Sum Composition. Suppose $h(\theta, \alpha) = \sum_l g(\theta, x_l)f(x_l, \alpha)$. Then we have,

$$\begin{Vmatrix} h(\theta_0, \alpha_0) & h(\theta_0, \alpha_1) \\ h(\theta_1, \alpha_0) & h(\theta_1, \alpha_1) \end{Vmatrix} = \sum_{l<m} \begin{Vmatrix} g(\theta_0, x_l) & g(\theta_0, x_m) \\ g(\theta_1, x_l) & g(\theta_1, x_m) \end{Vmatrix} \begin{Vmatrix} f(x_l, \alpha_0) & f(x_l, \alpha_1) \\ f(x_m, \alpha_0) & f(x_m, \alpha_1) \end{Vmatrix}. \qquad (53)$$

This is shown in same manner as the integral composition formula (note that the determinants are zero when $l = m$).

From these formulas it can easily be seen that the property of total positivity is preserved under such compositions. The following lemma makes this precise.

Lemma 5 (Preservation of Total Positivity) If $h(\theta, \alpha) = \int g(\theta, x)f(x, \alpha)\mathrm{d}x$, and f and g are both STP_2, then h is STP_2. *Proof:* This can be seen directly from Equation 51. The property of STP_2 for f and g insures strict positivity of the integrand, which insures strict positivity of the integral, and thus STP_2 for h. STP_2 is also preserved for sum compositions of Equation 53, provided the number of terms in the sum exceeds one (note again that the determinants are zero when $l = m$).

C. Total Positivity of ACE

Lemma 5 may be applied repeatedly to show that the density $f(\hat{\beta};\alpha)$ is STP_2 for positive α.

Lemma 6: The density $f(\hat{\beta};\alpha)$ is an STP_2 kernel with respect to $\hat{\beta}$ and α, for *positive* α. Equivalently, the likelihood ratio is monotone for positive SNRs:

$$\frac{f(\hat{\beta}_1;\alpha_1)}{f(\hat{\beta}_1;\alpha_0)} > \frac{f(\hat{\beta}_0;\alpha_1)}{f(\hat{\beta}_0;\alpha_0)}, \quad 1 > \beta_1 > \beta_0 > 0; \quad \alpha_1 > \alpha_0 > 0. \tag{54}$$

Proof: Referring to Eqn. 42, we make the change of variables $\xi = \alpha b$, and rewrite it as the integral composition of two kernels, $G(\hat{\beta},\xi;\alpha) = f(\hat{\beta}|b;\alpha)\big|_{b=\xi/\alpha}$ and $H(\xi,\alpha) = f(b)\big|_{b=\xi/\alpha}$:

$$f(\hat{\beta};\alpha) = \int_0^\infty G(\hat{\beta},\xi;\alpha)H(\xi,\alpha)\mathrm{d}\xi, \quad \alpha > 0 \tag{55}$$

The kernels are given by

$$G(\hat{\beta},\xi;\alpha) = \sum_{m=0}^{\infty} \frac{\hat{\beta}^{p+m-1}(1-\hat{\beta})^{L-1}}{B(p+m,L)} \frac{(1-\frac{\xi}{\alpha})^{p+m}}{(1-\frac{\xi}{\alpha}\hat{\beta})^{L+p+m}} P(m,\xi); \tag{56}$$

$$H(\xi,\alpha) = \frac{(\frac{\xi}{\alpha})^{L+p-1}(1-\frac{\xi}{\alpha})^{N-p-1}}{B(L+p,N-p)} \chi_{[0,\alpha)}(\xi), \tag{57}$$

where $P(m,\xi) = \frac{\xi^m}{m!}e^{-\xi}$ is a Poisson coefficient and $\chi_{[0,\alpha)}(\xi)$ is an indicator function (one on $[0,\alpha)$, zero elsewhere).

Furthermore, $G(\beta,\xi;\alpha)$ may be written as the sum composition of two kernels $Q(\hat{\beta},m;\xi,\alpha)$ and $P(m,\xi)$:

$$\begin{aligned} G(\hat{\beta},\xi;\alpha) &= \sum_{m=0}^{\infty} Q(\hat{\beta},m;\xi,\alpha)P(m,\xi); \\ Q(\hat{\beta},m;\xi,\alpha) &= \frac{\hat{\beta}^{p+m-1}(1-\hat{\beta})^{L-1}}{B(p+m,L)} \frac{(1-\frac{\xi}{\alpha})^{p+m}}{(1-\frac{\xi}{\alpha}\hat{\beta})^{L+p+m}}. \end{aligned} \tag{58}$$

The strategy for the remainder of the proof is as follows, with the details contained in Lemmas 7 through 10 in Appendix B: (1) Show by direct calculation that the kernels $Q(\hat{\beta},m;\xi,\alpha)$, $P(m,\xi)$, and $H(\xi,\alpha)$ are totally positive (STP_2) (Lemmas 7, 8, and 10). (2) Use the summation composition formula and Lemma 5 to show that $G(\hat{\beta},\xi;\alpha)$ is totally positive (Lemma 9). (3) Use the integral composition formula to show that $f(\hat{\beta};\alpha)$ is totally positive. ■

A continuity argument in Appendix C extends this result to the case $\alpha_0 = 0$. Then we have the following theorem, which is a key theoretical result of this paper:

Theorem 2: The likelihood ratio for ACE is monotone:

$$\frac{f(\hat{\beta}_1;\alpha_1)}{f(\hat{\beta}_1;0)} > \frac{f(\hat{\beta}_0;\alpha_1)}{f(\hat{\beta}_0;0)}, \quad 1 > \beta_1 > \beta_0 > 0; \quad \alpha_1 > 0. \tag{59}$$

Theorem 1 shows ACE to be a maximal invariant, and Theorem 2 shows its likelihood ratio to be monotone. Together they establish ACE as a UMP-invariant statistic for threshold detection. The following theorem makes this precise.

Theorem 3: A threshold test on $\hat{\beta}$ is UMP-invariant out of all tests that are invariant to the transformation group $G_\gamma(\cdot)$ of Eqn. 33. *Proof:* Because $\hat{\beta}$ is a maximal invariant, any test that is invariant to $G_\gamma(\cdot)$ depends on the data matrix only through $\hat{\beta}$. By the Neyman-Pearson lemma, a threshold test on the likelihood ratio of $\hat{\beta}$ is then Most Powerful (MP) out of all tests that are invariant to $G_\gamma(\cdot)$. That is, a Neyman-Pearson likelihood ratio test is a MP-invariant detector for a given value of output-SNR α. Because $\hat{\beta}$ has monotone likelihood ratio, the Karlin-Rubin Theorem [11] then makes a direct threshold test on $\hat{\beta}$ uniformly most powerful, that is, MP for all α. ■

V. Conclusions

We have extended the transformation group of Bose and Steinhardt to accommodate the problem for which the ACE is a GLRT, that is, where the covariance of the primary test vector can be scaled relative to that of the training data. With this extension, we have proved that ACE is a maximal invariant with respect to transformations that include such scaling.

With the additional property of total positivity, we have shown that ACE has monotone likelihood ratio, and thus that a threshold test on it is UMP-invariant. This provides a precise statement of the class of tests for which ACE has optimal detection performance. The structure of the transformation group gives additional insight into the scenario for which ACE can be expected to be advantageous, namely one in which there is a lack of scaling homogeneity between the test-data and training-data.

VI. Acknowledgments

This paper is based on work presented at sessions on "Invariance in Signal Processing" at the 2000 and 2002 Annual Asilomar Conference on Signals and Systems [18], [27]. We wish to thank

Michael Clark and Steven Smith for inviting us to present. We extend sincere appreciation to Douglas Turner for publishing support. This work was supported in part by a grant from NSERC Canada, Grant 262026, and by ONR USA, under Contract N00014-01-1-1019.

VII. Appendix

A. The Maximal Invariant for the Homogeneous Detection Problem

Proof of Lemma 4 We must show that if we are given two data matrices $[\underline{y}, \mathbf{X}]$ and $[\underline{y}', \mathbf{X}']$ such that $\hat{m}_1 = \hat{m}_1'$ and $\hat{m}_2 = \hat{m}_2'$, we can construct a transformation in the form of Eqn. 18 relating the two data matrices. The proof is a variation on the method presented by Raghavan, Pulsone, and McLaughlin in Appendix B of [8].

Introduce the vectors $\underline{z}_{1.2} = \mathbf{S}_{11.2}^{-\frac{1}{2}}\underline{y}_{1.2}$ and $\underline{z}_2 = \mathbf{S}_{22}^{-\frac{1}{2}}\underline{y}_2$, which are adaptively pre-whitened by the appropriate components of the covariance matrix. Then $\hat{m}_1 = \underline{z}_{1.2}^\dagger \underline{z}_{1.2}$ and $\hat{m}_2 = \underline{z}_2^\dagger \underline{z}_2$. Referring to the block-Cholesky factorization of $\mathbf{S}^{-1}$ in Eqn. 26, consider the block upper-triangular filtering matrix $\mathbf{W}$, and block diagonal pre-whitening matrix $\mathbf{Q}$ defined in Eqn. 27. Then we have

$$\begin{bmatrix} \underline{z}_{1.2} \\ \underline{z}_2 \end{bmatrix} = \mathbf{Q} \begin{bmatrix} \underline{y}_{1.2} \\ \underline{y}_2 \end{bmatrix} = \mathbf{QW} \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \end{bmatrix} = \mathbf{QW}\underline{y}. \tag{60}$$

The condition $\hat{m}_1' = \hat{m}_1$ implies that the vectors $\underline{z}_{1.2}'$ and $\underline{z}_{1.2}$ are equal in magnitude and related by a unitary transformation: $\underline{z}_{1.2}' = \mathbf{T}_{1.2}\underline{z}_{1.2}$. Similarly, $\underline{z}_2'$ and $\underline{z}_2$ must also be related by a unitary transformation: $\underline{z}_2' = \mathbf{T}_2\underline{z}_2$. Defining a new block-diagonal unitary matrix $\mathbf{T}$ formed from $\mathbf{T}_{1.2}$ and $\mathbf{T}_2$, we can express this as

$$\begin{bmatrix} \underline{z}_{1.2}' \\ \underline{z}_2' \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{1.2} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_2 \end{bmatrix} \begin{bmatrix} \underline{z}_{1.2} \\ \underline{z}_2 \end{bmatrix} = \mathbf{T} \begin{bmatrix} \underline{z}_{1.2} \\ \underline{z}_2 \end{bmatrix} \quad \text{or}$$

$$\mathbf{Q}'\mathbf{W}'\underline{y}' = \mathbf{TQW}\underline{y} \tag{61}$$

$$\rightarrow \underline{y}' = \mathbf{W}'^{-1}\mathbf{Q}'^{-1}\mathbf{TQW}\underline{y}. \tag{62}$$

Note that the matrices $\mathbf{W}'^{-1}$, $\mathbf{Q}'^{-1}$, $\mathbf{T}$, $\mathbf{Q}$, and $\mathbf{W}$ are all upper block triangular, and therefore so is their product. So $\underline{y}'$ and $\underline{y}$ are related by a block upper-triangular transformation in the form of the pre-multiplying matrix of Eqn. 19. We denote the matrix by $\mathcal{A}' = \mathbf{W}'^{-1}\mathbf{Q}'^{-1}\mathbf{TQW}$:

$$\underline{y}' = \mathbf{W}'^{-1}\mathbf{Q}'^{-1}\mathbf{TQW}\underline{y} = \mathcal{A}'\underline{y}. \tag{63}$$

Now consider the training-data matrices $\mathbf{X}$ and $\mathbf{X}'$, transformed the same as their associated test-data vectors in Eqn. 61:

$$\mathbf{Z}' = \mathbf{Q}'\mathbf{W}'\mathbf{X}'; \quad \mathbf{Z} = \mathbf{TQWX}. \tag{64}$$

From Eqns. 26 and 27, we have

$$\begin{aligned} \mathbf{S}^{-1} &= \mathbf{W}^{\dagger}\mathbf{Q}^{\dagger}\mathbf{QW}; \quad \rightarrow \mathbf{S} = \mathbf{W}^{-1}\mathbf{Q}^{-1}\mathbf{Q}^{\dagger -1}\mathbf{W}^{\dagger -1} \\ &\rightarrow \mathbf{QWSW}^{\dagger}\mathbf{Q}^{\dagger} = \mathbf{I}. \end{aligned} \tag{65}$$

So $\mathbf{QW}$ block diagonalizes the sample covariance $\mathbf{S}$. (This can be verified directly by multiplying the matrices out.) Up to this point, this proof has followed that of [8], except that in [8] $(\underline{y}, \mathbf{S})$ is transformed, rather than $(\underline{y}, \mathbf{X})$.

As a consequence we have that the rows of $\mathbf{Z}$ and $\mathbf{Z}'$ are orthonormal:

$$\begin{aligned} \mathbf{ZZ}^{\dagger} &= \mathbf{TQWXX}^{\dagger}\mathbf{W}^{\dagger}\mathbf{Q}^{\dagger}\mathbf{T}^{\dagger} = \mathbf{TQWSW}^{\dagger}\mathbf{Q}^{\dagger}\mathbf{T}^{\dagger} = \mathbf{TT}^{\dagger} = \mathbf{I} \\ \mathbf{Z}'\mathbf{Z}'^{\dagger} &= \mathbf{Q}'\mathbf{W}'\mathbf{X}'\mathbf{X}'^{\dagger}\mathbf{W}'^{\dagger}\mathbf{Q}'^{\dagger} = \mathbf{Q}'\mathbf{W}'\mathbf{S}'\mathbf{W}'^{\dagger}\mathbf{Q}'^{\dagger} = \mathbf{I}. \end{aligned} \tag{66}$$

Therefore we can extend $\mathbf{Z}$ and $\mathbf{Z}'$ to full unitary matrices $\mathbf{V}$ and $\mathbf{V}'$:

$$\mathbf{Z} = \left[\begin{array}{c|c} \mathbf{I} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \mathbf{Z} \\ \hline \mathbf{Z}_{\perp} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{I} & \mathbf{0} \end{array}\right] \mathbf{V}; \quad \mathbf{Z}' = \left[\begin{array}{c|c} \mathbf{I} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \mathbf{Z}' \\ \hline \mathbf{Z}'_{\perp} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{I} & \mathbf{0} \end{array}\right] \mathbf{V}'. \tag{67}$$

This gives $\mathbf{Z}'\mathbf{V}'^{\dagger} = \mathbf{ZV}^{\dagger} = [\mathbf{I}|\mathbf{0}]$, and thus

$$\mathbf{Z}' = \mathbf{ZV}^{\dagger}\mathbf{V}' = \mathbf{ZU}', \tag{68}$$

where $\mathbf{U}' = \mathbf{V}^{\dagger}\mathbf{V}'$ is also a unitary matrix. Now substituting in Eqn. 64 to write these matrices in terms of the original training-data matrices, we have

$$\begin{aligned} \mathbf{Q}'\mathbf{W}'\mathbf{X}' &= \mathbf{TQWXU}'; \\ &\rightarrow \mathbf{X}' = \mathbf{W}'^{-1}\mathbf{Q}'^{-1}\mathbf{TQWXU}' = \mathcal{A}'\mathbf{XU}'. \end{aligned} \tag{69}$$

Combining Eqns. 63 and 69, we have

$$\underline{y}' = \mathcal{A}'\underline{y}; \quad \mathbf{X}' = \mathcal{A}'\mathbf{X}\mathbf{U}'; \tag{70}$$

$$\rightarrow \left[\underline{y}'|\mathbf{X}'\right] = \mathcal{A}'\left[\underline{y}|\mathbf{X}\right]\begin{bmatrix} 1 & \underline{0}^{\dagger} \\ \underline{0} & \mathbf{U}' \end{bmatrix}, \tag{71}$$

where $\mathcal{A}'$ is block upper-triangular, and $\mathbf{U}'$ is unitary. Thus the Lemma is proved: given data matrices that produce the same value for the pair $(\hat{m}_1, \hat{m}_2)$, these data matrices are related by a linear transformation of the form of Eqn. 18. ■

B. Total Positivity for Positive SNRs, $\alpha > 0$

Lemma 7: The kernel $P(m, \xi)$ of Eqn. 56 is STP_2 with respect to m and ξ.

Proof: We want to show that

$$\frac{P(m_1, \xi_1)}{P(m_0, \xi_1)} > \frac{P(m_1, \xi_0)}{P(m_0, \xi_0)} \quad \text{for} \quad m_1 > m_0, \quad \xi_1 > \xi_0. \tag{72}$$

Let $\delta = m_1 - m_0 > 0$. Using the Pochammer notation to indicate a ratio of gamma functions, $(x)_\delta = \frac{\Gamma(x+\delta)}{\Gamma(x)}$, the ratio can be written as

$$\frac{P(m_1, \xi)}{P(m_0, \xi)} = \frac{\xi^\delta}{(m_0 - 1)_\delta}, \tag{73}$$

which shows the ratio to be monotone increasing in ξ for all $m_1 > m_0$. ■

Lemma 8: The kernel $Q(\hat{\beta}, m; \xi, \alpha)$ of Eqn. 58 is STP_2 with respect to $\hat{\beta}$ and m.

Proof: Equivalently, we show that the derivative of the ratio $Q(\hat{\beta}, m_1; \xi, \alpha)/Q(\hat{\beta}, m_0; \xi, \alpha)$ is positive. The ratio is given by

$$\frac{Q(\hat{\beta}, m_1; \xi, \alpha)}{Q(\hat{\beta}, m_0; \xi, \alpha)} = \left(\frac{\hat{\beta}\left(1 - \frac{\xi}{\alpha}\right)}{1 - \frac{\xi}{\alpha}\hat{\beta}}\right)^\delta \frac{(p + m_0 + L)_\delta}{(p + m_0)_\delta}, \tag{74}$$

and the derivative with respect to $\hat{\beta}$ is given by

$$\frac{\mathrm{d}}{\mathrm{d}\hat{\beta}}\left(\frac{Q(\hat{\beta}, m_1; \xi, \alpha)}{Q(\hat{\beta}, m_0; \xi, \alpha)}\right) = \frac{\delta\hat{\beta}^{\delta-1}\left(1 - \frac{\xi}{\alpha}\right)^\delta}{\left(1 - \frac{\xi}{\alpha}\hat{\beta}\right)^{\delta+1}} \frac{(p + m_0 + L)_\delta}{(p + m_0)_\delta} > 0, \tag{75}$$

$$0 < \xi < \alpha, \quad 0 < \hat{\beta} < 1. \quad \blacksquare \tag{76}$$

Lemma 9: The kernel $G(\hat{\beta}, \xi; \alpha) = \sum_{m=0}^{\infty} Q(\hat{\beta}, m; \xi, \alpha) P(m, \xi)$ of Eqn. 56 is STP_2 with respect to $\hat{\beta}$ and ξ. *Proof:* Both P and Q are STP_2 by Lemmas 7 and 8. Then the summation composition formula and Lemma 5 ensures that the composition G is STP_2. ■

Lemma 10: The kernel $H(\xi, \alpha)$ of Eqn. 57 is STP_2 with respect to ξ and α, for $\xi < \alpha$.

Proof: The determinant that defines total positivity, $H(\xi_0, \alpha_0)H(\xi_1, \alpha_1) - H(\xi_0, \alpha_1)H(\xi_1, \alpha_0)$, is zero for $\xi_1 > \xi_0 \geq \alpha_0$. It is non-negative for $\xi_1 \geq \alpha_0 > \xi_0$, being given by

$$H(\xi_0, \alpha_0)H(\xi_1, \alpha_1) \geq 0. \tag{77}$$

It is strictly positive for $\xi_0 < \xi_1 < \alpha_0$, which can be seen by considering the derivative of the ratio $H(\xi_1, \alpha)/H(\xi_0, \alpha)$ with respect to α. The ratio is given by

$$\frac{H(\xi_1, \alpha)}{H(\xi_0, \alpha)} = \left(\frac{\xi_1}{\xi_0}\right)^{L+p-1} \left(\frac{\alpha - \xi_1}{\alpha - \xi_0}\right)^{N-p-1}, \tag{78}$$

with a positive derivative given by

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\alpha}\left(\frac{H(\xi_1, \alpha)}{H(\xi_0, \alpha)}\right) &= \left(\frac{\xi_1}{\xi_0}\right)^{L+p-1} {}_{(N-p-1)}\frac{(\alpha - \xi_1)^{N-p-2}}{(\alpha - \xi_0)^{N-p}}(\xi_1 - \xi_0) > 0, \\ & \xi_0 < \xi_1 < \alpha. \quad ■ \end{aligned} \tag{79}$$

Proof of Lemma 6: The lemma states that the density $f(\hat{\beta}; \alpha) = \int_0^{\infty} G(\hat{\beta}, \xi; \alpha) H(\xi, \alpha) \mathrm{d}\xi$ is an STP_2 kernel with respect to $\hat{\beta}$ and α, for positive α. From Lemmas 9 and 10, we know that G is STP_2 and that H is STP_2 for $\xi < \alpha$. Then by the integral composition formula and Lemma 5, $f(\hat{\beta}; \alpha)$ is STP_2, with the additional observation that the integrand is non-negative, and strictly positive for $\xi < \alpha$. ■

C. Total Positivity for Non-Negative SNRs, $\alpha \geq 0$

In order to prove Theorem 2, we need to extend Lemma 6 to zero SNR, $\alpha_0 = 0$.

Proof of Theorem 2: An inequality for $\alpha_0 = 0$ can be established from the continuity of $f(\hat{\beta}; \alpha)$ with respect to α. We first show the soft inequality

$$\frac{f(\hat{\beta}_1; \alpha_1)}{f(\hat{\beta}_1; 0)} \geq \frac{f(\hat{\beta}_0; \alpha_1)}{f(\hat{\beta}_0; 0)}. \qquad (A) \tag{80}$$

We prove this by contradiction. Suppose Proposition A is not true ($\sim A$). This would imply

$$f(\hat{\beta}_0;0)f(\hat{\beta}_1;\alpha_1) - f(\hat{\beta}_0;\alpha_1)f(\hat{\beta}_1;0) = \nu < 0;$$

$$\rightarrow f(\hat{\beta}_0;\delta)f(\hat{\beta}_1;\alpha_1) - f(\hat{\beta}_0;\alpha_1)f(\hat{\beta}_1;\delta) + \epsilon = \nu, \quad \text{where} \tag{81}$$

$$\epsilon = \left[f(\hat{\beta}_0;0) - f(\hat{\beta}_0;\delta)\right] f(\hat{\beta}_1;\alpha_1) - \left[f(\hat{\beta}_1;0) - f(\hat{\beta}_1;\delta)\right] f(\hat{\beta}_0;\alpha_1). \tag{82}$$

Due to continuity of $f(\hat{\beta};\alpha)$, there exists a δ such that $|\epsilon| < |\nu|$, which would imply that

$$\frac{f(\hat{\beta}_1;\alpha_1)}{f(\hat{\beta}_1;\delta)} < \frac{f(\hat{\beta}_0;\alpha_1)}{f(\hat{\beta}_0;\delta)}. \qquad (B) \tag{83}$$

(So $\sim A \rightarrow B$, or equivalently $\sim B \rightarrow A$.) Proposition B would contradict Lemma 6, so A must be true.

To show that the inequality is strict, we again prove by contradiction, supposing that it would be possible to have equality:

$$\frac{f(\hat{\beta}_1;\alpha_1)}{f(\hat{\beta}_1;0)} = \frac{f(\hat{\beta}_0;\alpha_1)}{f(\hat{\beta}_0;0)}. \qquad (C) \tag{84}$$

From Lemma 6, we have

$$\frac{f(\hat{\beta}_1;\alpha_1/2)}{f(\hat{\beta}_1;\alpha_1)} < \frac{f(\hat{\beta}_0;\alpha_1/2)}{f(\hat{\beta}_0;\alpha_1)}. \tag{85}$$

Multiplying the left and right sides of Equation 84 by the associated sides of this inequality, we would have

$$\frac{f(\hat{\beta}_1;\alpha_1/2)}{f(\hat{\beta}_1;\alpha_1)}\frac{f(\hat{\beta}_1;\alpha_1)}{f(\hat{\beta}_1;0)} < \frac{f(\hat{\beta}_0;\alpha_1/2)}{f(\hat{\beta}_0;\alpha_1)}\frac{f(\hat{\beta}_0;\alpha_1)}{f(\hat{\beta}_0;0)}. \tag{86}$$

And after cancellation we would have

$$\frac{f(\hat{\beta}_1;\alpha_1/2)}{f(\hat{\beta}_1;0)} < \frac{f(\hat{\beta}_0;\alpha_1/2)}{f(\hat{\beta}_0;0)}, \tag{87}$$

which would contradict the soft inequality of Equation 80 ($C \rightarrow\sim A$, or equivalently $A \rightarrow\sim C$). So we cannot have equality, and the inequality of Equation 80 must in fact be strict, and the Theorem is proven. ■

References

[1] L. L. Scharf and L. Todd McWhorter, "Adaptive matched subspace detectors and adaptive coherence," in *Proceedings of the 30th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 1996.

[2] L. L. Scharf, *Statistical Signal Processing*, chapter 4, Addison-Wesley, 1991.

[3] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Processing*, vol. 42, no. 8, pp. 2146–2157, 1994.

[4] S. Kraut, L. L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Processing*, vol. 49, no. 1, pp. 1–16, 2001.

[5] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aero. and Electr. Systems*, vol. 10, no. 6, pp. 853–863, 1974.

[6] E. J. Kelly and K. M. Forsythe, *Adaptive Detection and Parameter Estimation for Multidimensional Signal Models, Technical Report 848*, M.I.T. Lincoln Laboratories, 1989.

[7] A. Steinhardt, "Adaptive multisensor detection and estimation," in *Adaptive Radar Detection and Estimation*, S. Haykin and A. Steinhardt, Eds. Wiley, 1992.

[8] R. S. Raghavan, N. Pulsone, and D. J. McLaughlin, "Performance of the GLRT for adaptive vector subspace detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 4, pp. 1473–1487, 1996.

[9] M. Rangaswamy, J. H. Michels, and B. Himed, "Statistical analysis of the nonhomogeneity detector for non-gaussian interference backgrounds," in *Proceedings of the IEEE Radar Conference*, April 2002, pp. 304–310.

[10] S. Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Trans. Signal Processing*, vol. 47, no. 9, pp. 2538–2541, Sept. 1999.

[11] E. L. Lehmann, *Testing Statistical Hypotheses*, John Wiley & Sons, 1959.

[12] T. S. Ferguson, *Mathematical Statistics: a Decision Theoretic Approach*, Academic Press, 1967.

[13] S. Bose and A. Steinhardt, "A maximal invariant framework for adaptive detection with structured and unstructured covariance matrices," *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2164–75, 1995.

[14] E. J. Kelly, "An adaptive detection algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 1, pp. 115–127, 1986.

[15] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. A. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 28, no. 1, pp. 208–216, 1992.

[16] Wai-Sheou Chen and Irving S. Reed, "A new CFAR detection test for radar," *Digital Signal Processing*, vol. 1, no. 4, pp. 198–214, 1991.

[17] S. Bose and A. O. Steinhardt, "Optimum array detector for a weak signal in unknown noise," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 3, pp. 911–922, 1996.

[18] S. Kraut and J. L. Krolik, "Application of maximal invariance to the ACE detection problem," in *Proceedings of the 34th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, October 2000.

[19] E. Conte, M. Lops, and G. Ricci, "Asymptotically optimum radar detection in compound-gaussian clutter," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 31, no. 2, pp. 617–625, 1995.

[20] S. Karlin, *Total Positivity*, Stanford University Press, 1968.

[21] S. Bose and A. O. Steinhardt, "Adaptive detection of uncertain rank one waveforms," *IEEE Transactions on Signal Processing*, vol. 44, no. 11, pp. 2801–2809, 1996.

[22] S. Kraut, L. T. McWhorter, and L. L. Scharf, "A canonical representation for distributions of adaptive matched subspace detectors," in *Proceedings of the 31st Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 1997.

[23] C. D. Richmond, "Performance of the adaptive sidelobe blanker detection algorithm in homogeneous environments," *IEEE Trans. on Signal Processing*, vol. 48, pp. 1235–1247, May 2000.

[24] C. D. Richmond, "Performance of a class of detection algorithms in nonhomogeneous environments," *IEEE Trans. on Signal Processing*, vol. 48, pp. 1248–1262, May 2000.

[25] S. Karlin and W. J. Studden, *Tchebycheff Systems: with Applications in Analysis and Statistics*, John Wiley & Sons, 1966.

[26] A. W. Marshal and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, 1979.

[27] S. Kraut, L. L. Scharf, and R. W. Butler, "ACE is UMP-invariant," in *36th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2002, Oral Presentation.

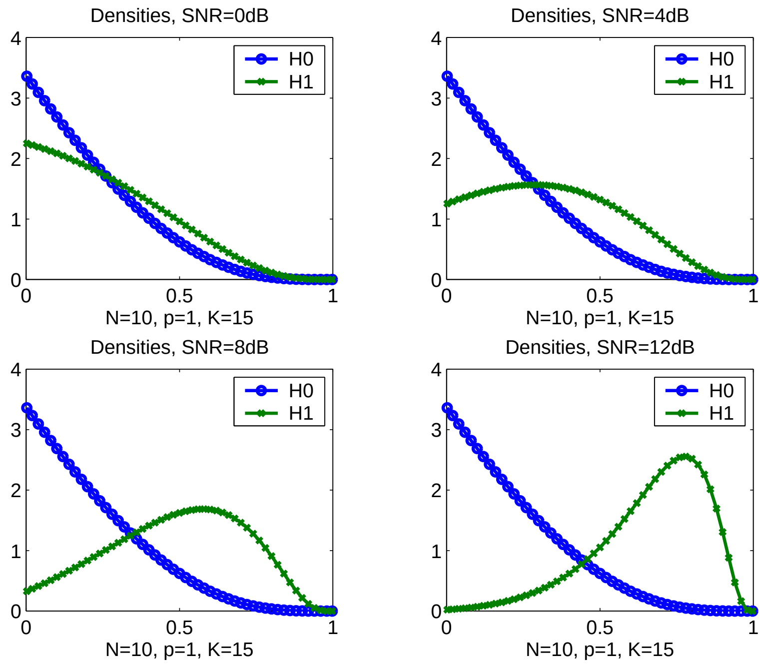

Fig. 1. Probability density functions of $\hat{\beta}$. Under H_1 the SNR's are 0, 4, 8 and 12 dB; data dimension is $N = 10$, signal dimension is $p = 1$, and training-data support is $K = 15$.

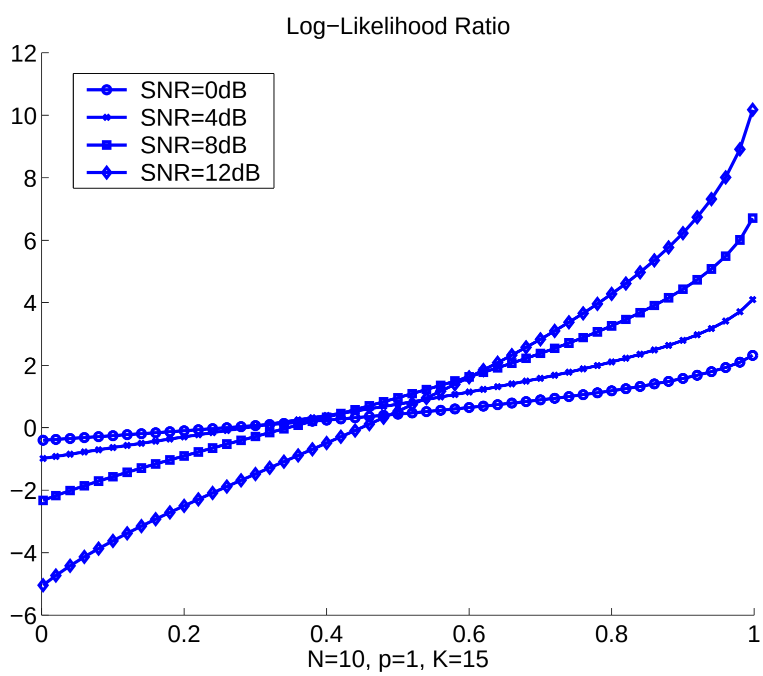

Fig. 2. Log-Likelihood ratios for $\hat{\beta}$, for the densities in Figure 1. Under H_1 the SNR's are 0, 4, 8 and 12 dB; $N = 10$, $p = 1$, and $K = 15$.