Matched Subspace Detectors

Louis L. Scharf, *Fellow, IEEE,* and Benjamin Friedlander, *Fellow, IEEE*

***Abstract*—In this paper we formulate a general class of problems for detecting subspace signals in subspace interference and broadband noise. We derive the generalized likelihood ratio (GLR) for each problem in the class. We then establish the invariances for the GLR and argue that these are the natural invariances for the problem. In each case, the GLR is a maximal invariant statistic, and the distribution of the maximal invariant statistic is monotone. This means that the GLR test (GLRT) is the uniformly most powerful invariant detector. We illustrate the utility of this finding by solving a number of problems for detecting subspace signals in subspace interference and broadband noise. In each case we give the distribution for the detector and compute performance curves.**

I. Introduction

The matched filter, or more accurately the *matched signal detector*, is one of the basic building blocks of signal processing; however, in many applications the rank-1 matched signal detector is replaced by a multirank *matched subspace detector*. In fact, the matched subspace detector is really the general building block, and the matched signal detector is a special case. In sonar signal processing, the matched subspace detector is called a *matched field detector*.

In [1], one of the authors developed a theory of matched subspace detectors based on the construction of invariant statistics. In this paper we extend this work in two ways. First, we include structured interference in the measurement model, and second we use the principle of the generalized likelihood ratio test (GLRT) to derive matched subspace detectors. By studying the invariance classes for these GLRT's, we are able to establish that the GLRT's are invariant to a natural set of invariances and optimum within the class of detectors which share these invariances. This establishes once and for all the optimality of the GLRT for solving matched subspace detection problems and answers "no" to the question, "can the GLRT be improved upon?" This result holds for all finite sample sizes, thereby improving on the standard asymptotic theory of the GLRT.

Our program in this paper is to derive GLRT's for the class of problems studied in [1], [4]–[6] and generalize them to include subspace interferences. These problems involve unknown parameters in the mean and covariance of a multivariate normal (MVN) distribution. For each problem in the class, we establish invariances for the GLR and find that they are identical to the natural invariances for the problem. We show that a monotone function of the GLRT equals one of the uniformly most powerful invariant (UMP-invariant) tests derived in [1]. This means that the GLRT is itself UMP-invariant. In addition to tying up the theories of invariance and the GLRT, our results generalize and extend previous work on these problems published in [1]–[6].

We begin our development by establishing the invariances of the GLRT in the MVN problem. We then specialize our results for structured means in order to derive UMP-invariant GLRT detectors for matched subspace filtering in subspace interference. The GLRT produces an UMP-invariant detector, which is CFAR if the noise variance is unknown. As we shall find, the optimum detector may be interpreted as a null steering or interference rejecting processor followed by a matched subspace detector.

Manuscript received July 3, 1993; revised November 11, 1993. This work was supported by the Office of Naval Research, Mathematics Division, Statistics and Probability Branch, under Contracts N00014-89-J1070 and N00014-91-J1602, and by the National Science Foundation under Grant MIP-90-17221. The associate editor coordinating the review of this paper and approving it for publication was Prof. Kevin M. Buckley.

L. L. Scharf is with the Department of Electrical & Computer Engineering, University of Colorado, Boulder, Colorado 80309-0425 USA.

B. Friedlander is with the Department of Electrical & Computer Engineering, University of California-Davis, Davis, CA 95616 USA.

IEEE Log Number 9401912.

II. Detection Problems

The detection problems to be studied in this paper may be described as follows. We are given N samples from a real, scalar time series $\{y(n), n = 0, 1, \dots, N-1\}$ which are assembled into the N-dimensional measurement vector $\boldsymbol{y} = [y(0), y(1), \dots, y(N-1)]^{\mathrm{T}}$. Based on these data, we must decide between two possible hypotheses regarding how the data was generated. The null hypothesis H_0 says that the data consist of noise $\boldsymbol{v}$ only. The alternative hypothesis H_1 says that the data consist of a sum of signal $\mu\boldsymbol{x}$ and noise $\boldsymbol{v}$; that is,

$$\boldsymbol{y} = \mu\boldsymbol{x} + \boldsymbol{v} \tag{2.1}$$

where $\mu = 0$ under H_0 and $\mu > 0$ under H_1. This is the standard detection problem wherein the polarity of the signal $\boldsymbol{x}$ is assumed known. Near the end of Section V we replace $H_1 : \mu > 0$ with $H_1 : \mu \neq 0$ in order to model problems where polarity is unknown.

We shall assume that the signal $\boldsymbol{x}$ obeys the linear subspace model

$$\boldsymbol{x} = \boldsymbol{H\theta}, \qquad \boldsymbol{H} \in \mathbb{R}^{N\times p}, \quad \boldsymbol{\theta} \in \mathbb{R}^{p} \tag{2.2}$$

and the noise is MVN with mean $\boldsymbol{S\phi}$ and covariance $\boldsymbol{R} = \sigma^2\boldsymbol{R}_0$:

$$\boldsymbol{v} : \mathcal{N}[\boldsymbol{S\phi}, \sigma^2\boldsymbol{R}_0], \qquad \boldsymbol{S} \in \mathbb{R}^{N\times t}, \quad \boldsymbol{\phi} \in \mathbb{R}^{t}, \quad t < N-p$$
$$\boldsymbol{R}_0 > 0 \in \mathbb{R}^{N\times N}. \tag{2.3}$$

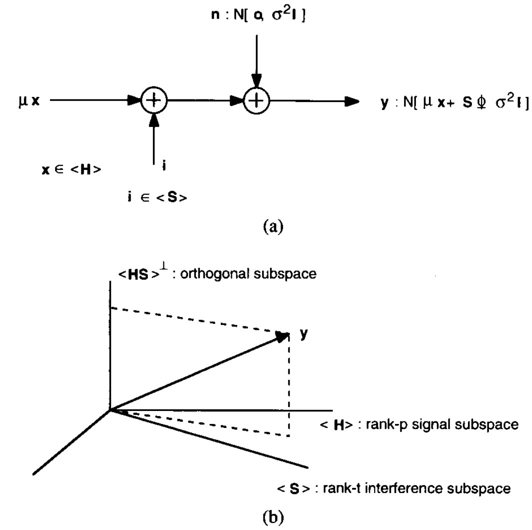

Fig. 1. Illustrating the detection problem: (a) Detecting a signal in interference plus additive noise; (b) resolving a measurement into signal plus interference plus noise.

We shall assume that $\boldsymbol{H}$, $\boldsymbol{S}$, and $\boldsymbol{R}_0$ are known, although we will offer some ways to relax this assumption in our conclusions. Without loss of generality, we assume that $\boldsymbol{R}_0 = \boldsymbol{I}$. Then the detection problem is a test of distributions:

$$H_0 : \boldsymbol{y} : \mathcal{N}[\boldsymbol{S\phi}, \sigma^2 \boldsymbol{I}] \text{ vs } H_1 : \boldsymbol{y} : \mathcal{N}[\mu \boldsymbol{H\theta} + \boldsymbol{S\phi}, \sigma^2 \boldsymbol{I}]. \quad (2.4)$$

The component $\mu\boldsymbol{H\theta}$ is an information-bearing signal that lies in the subspace $\langle \boldsymbol{H} \rangle$, and the component $\boldsymbol{S\phi}$ is an interference that lies in the subspace $\langle \boldsymbol{S} \rangle$. The noise $\boldsymbol{n} = \boldsymbol{y} - \mu\boldsymbol{H\theta} - \boldsymbol{S\phi} = \boldsymbol{v} - \boldsymbol{S\phi} : N[0, \sigma^2 \boldsymbol{I}]$ is additive white Gaussian noise. The subspaces $\langle \boldsymbol{H} \rangle$ and $\langle \boldsymbol{S} \rangle$ are *not* orthogonal (i.e. $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{S} \neq \boldsymbol{0}$), but they are linearly independent, meaning that there is no element of $\langle \boldsymbol{H} \rangle$ which can be written as a linear combination of vectors in $\langle \boldsymbol{S} \rangle$. Linear independence is much weaker than orthogonality. We assume that $\boldsymbol{H}$ and $\boldsymbol{S}$ are each full-rank matrices, meaning that $\langle \boldsymbol{H} \rangle$ and $\langle \boldsymbol{S} \rangle$ are each full-rank subspaces.

Fig. 1 illustrates the detection problem two ways: first, as a communication problem of detecting a signal when a channel adds background noise and structured interference and second, as an algebraic problem of determining which subspaces of $\mathbb{R}^N$ better model the measurement $\boldsymbol{y}$. In the second illustration, the problem is to determine whether $\boldsymbol{y}$ is more probably described by signal plus noise plus interference or by noise plus interference. The signal subspace $\langle \boldsymbol{H} \rangle$ and the interference subspace $\langle \boldsymbol{S} \rangle$, illustrated in Fig. 1, are generally of dimension greater than 1. When these subspaces are very close, then the resolution of hypotheses is difficult.

The probability density function for the MVN vector $\boldsymbol{y}$ is

$$f(\boldsymbol{y}; \boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{ -\frac{1}{2\sigma^2} \|\boldsymbol{n}\|_2^2 \right\} \quad (2.5)$$

where $\boldsymbol{y}$ is the variable of the function and $\boldsymbol{\beta} = (\mu\boldsymbol{\theta}, \boldsymbol{\phi})$ is the parameter of the density. The noise $\boldsymbol{n}$ is

$$\boldsymbol{n} = \boldsymbol{y} - \mu\boldsymbol{H\theta} - \boldsymbol{S\phi}. \quad (2.6)$$

The likelihood function for this MVN distribution is

$$l(\boldsymbol{\beta}, \sigma^2; \boldsymbol{y}) = (2\pi\sigma^2)^{-N/2} \exp\left\{ -\frac{1}{2\sigma^2} \|\boldsymbol{n}\|_2^2 \right\} \quad (2.7)$$

which is a function of $(\boldsymbol{\beta}, \sigma^2)$ with the data $\boldsymbol{y}$ playing the role of a parameter. For any two values $(\boldsymbol{\beta}_1, \sigma_1^2)$ and $(\boldsymbol{\beta}_0, \sigma_0^2)$, the *likelihood ratio* is defined to be

$$l(\boldsymbol{y}) = \frac{l(\boldsymbol{\beta}_1, \sigma_1^2; \boldsymbol{y})}{l(\boldsymbol{\beta}_0, \sigma_0^2; \boldsymbol{y})}. \quad (2.8)$$

We expect $l(\boldsymbol{y})$ to be greater than one whenever the parameters $(\boldsymbol{\beta}_1, \sigma_1^2)$ better model $\boldsymbol{y}$ than do the parameters $(\boldsymbol{\beta}_0, \sigma_0^2)$.

The detection problem outlined here applies to the detection of lines or modal signals in broadband noise and narrowband inteferences or to the detection of propagating fields (whether planar or not) in propagating interferences and broadband noise. Then the matrix $\boldsymbol{H}$ is a matrix of Vandermonde columns or of autoregressive impulse responses. The matrix $\boldsymbol{S}$ may be a Vandermonde matrix, one of the matrices illustrated in the following examples, or almost anything which characterizes structured noise.

Example 1—Detection in Unknown Bias: When there is an unknown bias added to a measurement, then we may say $\boldsymbol{y} = \mu\boldsymbol{x} + \boldsymbol{v}$ where $\boldsymbol{v} : \mathcal{N}[b\boldsymbol{1}, \sigma^2 \boldsymbol{I}]$. The bias is $b\boldsymbol{1} = (bb \cdots b)^{\mathrm{T}}$ with b unknown. In this case, the interference subspace $\langle \boldsymbol{S} \rangle$ is the rank-1 subspace $\langle \boldsymbol{1} \rangle$.

Example 2—Detection in Sinusoidal Interference: When there is a sinusoidal interference of known frequency but unknown amplitude and phase, $s_t = A\cos(\omega_0 t - \phi) = A\cos\phi\cos\omega_0 t + A\sin\phi\sin\omega_0 t$, then we may say $\boldsymbol{y} = \mu\boldsymbol{x} + \boldsymbol{v}$ where $\boldsymbol{v} : \mathcal{N}[\boldsymbol{S\phi}, \sigma^2 \boldsymbol{I}]$. The interference is

$$\boldsymbol{S\phi} = \begin{bmatrix} 1 & 0 \\ \cos\omega_0 & \sin\omega_0 \\ \vdots & \vdots \\ \cos(N-1)\omega_0 & \sin(N-1)\omega_0 \end{bmatrix} \begin{bmatrix} A\cos\phi \\ A\sin\phi \end{bmatrix}.$$

In this case the interference subspace $\langle \boldsymbol{S} \rangle$ is the rank-2 subspace above with cosine and sine columns.

III. Linear Algebraic Preliminaries

When we say that the signal $\boldsymbol{x}$ obeys the linear subspace model $\boldsymbol{x} = \boldsymbol{H\theta}$, we are saying that the vector $\boldsymbol{x} \in \mathbb{R}^N$ actually lies in a p-dimensional subspace of $\mathbb{R}^N$ which we denote $\langle \boldsymbol{H} \rangle$. The subspace $\langle \boldsymbol{H} \rangle$ is the range of the transformation $\boldsymbol{H}$. It is spanned by the columns of the matrix $\boldsymbol{H}$. These columns comprise a *basis* for the subspace, and the elements of $\boldsymbol{\theta} = (\theta_1\, \theta_2 \cdots \theta_p)^{\mathrm{T}}$ are the coordinates of $\boldsymbol{x}$ with respect to this basis. Similarly, the interference $\boldsymbol{S\phi}$ lies in a t-dimensional subspace of $\mathbb{R}^N$. This subspace, spanned by the columns of $\boldsymbol{S}$, is denoted $\langle \boldsymbol{S} \rangle$.

Together, the columns of the concatenated matrix $(\boldsymbol{HS})$ span the $(p+t)$-dimensional subspace $\langle \boldsymbol{HS} \rangle$. This subspace is illustrated in Fig. 2. The typical *orthogonal projection* of $\boldsymbol{y} \in \mathbb{R}^N$ onto $\langle \boldsymbol{HS} \rangle$ is denoted $\boldsymbol{P_{HS}y}$, where $\boldsymbol{P_{HS}}$ is the projection

$$\boldsymbol{P_{HS}} = [\boldsymbol{H}, \boldsymbol{S}]\left[[\boldsymbol{H}, \boldsymbol{S}]^{\mathrm{T}}[\boldsymbol{H}, \boldsymbol{S}]\right]^{-1}[\boldsymbol{H}, \boldsymbol{S}]^{\mathrm{T}}. \quad (3.1)$$

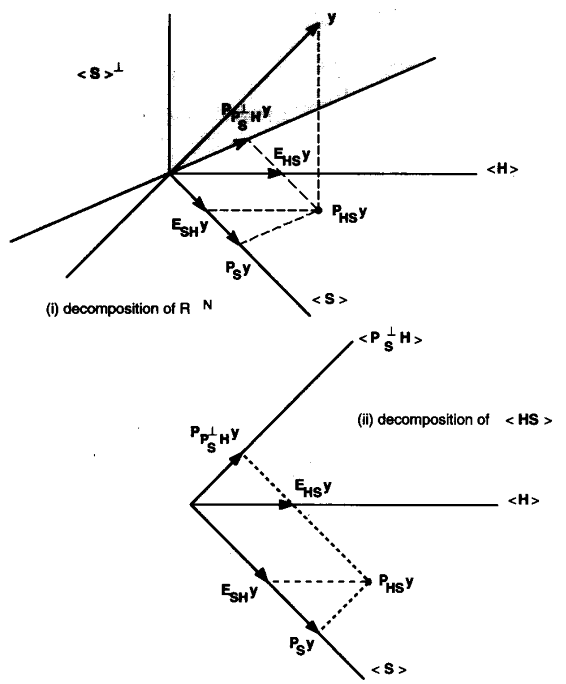

Fig. 2. Signal and interference subspaces, and various projections onto them.

As illustrated in the figure, this projection may be decomposed with respect to the subspaces $\langle \boldsymbol{H} \rangle$ and $\langle \boldsymbol{S} \rangle$ in two different ways:

$$\begin{aligned} \boldsymbol{P_{HS}} &= \boldsymbol{E_{HS}} + \boldsymbol{E_{SH}} \\ \boldsymbol{P_{HS}} &= \boldsymbol{P_S} + \boldsymbol{P_{P_S^\perp H}}. \end{aligned} \tag{3.2}$$

In the first of these decompositions, the orthogonal projection $\boldsymbol{P_{HS}}$ is resolved into *oblique* projections $\boldsymbol{E_{HS}}$ and $\boldsymbol{E_{SH}}$, where $\boldsymbol{E_{HS}}$ and $\boldsymbol{E_{SH}}$ have respective range spaces $\langle \boldsymbol{H} \rangle$ and $\langle \boldsymbol{S} \rangle$ and respective null spaces $\langle \boldsymbol{S} \rangle$ and $\langle \boldsymbol{H} \rangle$:

$$\begin{aligned} \boldsymbol{E_{HS}} &= \boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}} \boldsymbol{P_S^\perp} \boldsymbol{H})^{-1} \boldsymbol{H}^{\mathrm{T}} \boldsymbol{P_S^\perp} \\ \boldsymbol{E_{SH}} &= \boldsymbol{S}(\boldsymbol{S}^{\mathrm{T}} \boldsymbol{P_H^\perp} \boldsymbol{S})^{-1} \boldsymbol{S}^{\mathrm{T}} \boldsymbol{P_H^\perp} \\ \boldsymbol{E_{HS}} \boldsymbol{H} &= \boldsymbol{H}; \quad \boldsymbol{E_{HS}} \boldsymbol{S} = \boldsymbol{0} \\ \boldsymbol{E_{SH}} \boldsymbol{S} &= \boldsymbol{S}; \quad \boldsymbol{E_{SH}} \boldsymbol{H} = \boldsymbol{0}. \end{aligned} \tag{3.3}$$

That is, any vector $\boldsymbol{y} \in \langle \boldsymbol{HS} \rangle$ may be written as $\boldsymbol{y} = \boldsymbol{H\theta} + \boldsymbol{S\phi} = \boldsymbol{P_{HS} y} = (\boldsymbol{E_{HS}} + \boldsymbol{E_{SH}})\boldsymbol{y}$.

The second decomposition resolves $\boldsymbol{P_{HS}}$ into the orthogonal projections $\boldsymbol{P_S}$ and $\boldsymbol{P_{P_S^\perp H}}$. The subspace $\langle \boldsymbol{P_S^\perp H} \rangle$ is the subspace spanned by columns of the matrix $\boldsymbol{P_S^\perp H}$; the projector $\boldsymbol{P_S^\perp}$ projects onto the subspace $\langle \boldsymbol{S} \rangle^\perp$. Geometrically, the subspace $\langle \boldsymbol{P_S^\perp H} \rangle$ is the part of $\langle \boldsymbol{H} \rangle$ which is unaccounted for by the subspace $\langle \boldsymbol{S} \rangle$, when $\langle \boldsymbol{H} \rangle$ is resolved into $\langle \boldsymbol{S} \rangle \oplus \langle \boldsymbol{S} \rangle^\perp$. The ranges of $\boldsymbol{P_{P_S^\perp H}}$ and $\boldsymbol{P_S}$ are $\langle \boldsymbol{P_S^\perp H} \rangle$ and $\langle \boldsymbol{S} \rangle$, and the null spaces of $\boldsymbol{P_{P_S^\perp H}}$ and $\boldsymbol{P_S}$ are $\langle \boldsymbol{P_S^\perp H} \rangle^\perp$ and $\langle \boldsymbol{S} \rangle^\perp$:

$$\begin{aligned} \boldsymbol{P_{P_S^\perp H}} &= \boldsymbol{P_S^\perp H}(\boldsymbol{H}^{\mathrm{T}} \boldsymbol{P_S^\perp H})^{-1} \boldsymbol{H}^{\mathrm{T}} \boldsymbol{P_S^\perp} = \boldsymbol{P_S^\perp P_{P_S^\perp H} P_S^\perp} \\ \boldsymbol{P_S} &= \boldsymbol{S}(\boldsymbol{S}^{\mathrm{T}} \boldsymbol{S})^{-1} \boldsymbol{S}^{\mathrm{T}}. \end{aligned} \tag{3.4}$$

We will simplify our notation by defining the matrix $\boldsymbol{G} = \boldsymbol{P_S^\perp H}$, the subspace $\langle \boldsymbol{G} \rangle = \langle \boldsymbol{P_S^\perp H} \rangle$, and the projector $\boldsymbol{P_G} = \boldsymbol{P_{P_S^\perp H}}$. Then we may summarize our decomposition of $\boldsymbol{P_{HS}}$ as

$$\begin{aligned} \boldsymbol{P_{HS}} &= \boldsymbol{E_{HS}} + \boldsymbol{E_{SH}} \\ &= \boldsymbol{P_S} + \boldsymbol{P_S^\perp P_G P_S^\perp} \\ \boldsymbol{P_G} &\triangleq \boldsymbol{P_{P_S^\perp H}} = \boldsymbol{P_S^\perp E_{HS}}. \end{aligned} \tag{3.5}$$

The corresponding identity for $\boldsymbol{P_{HS}^\perp}$ is

$$\begin{aligned} \boldsymbol{P_{HS}^\perp} &= \boldsymbol{I} - \boldsymbol{P_{HS}} = \boldsymbol{I} - \boldsymbol{P_S} - \boldsymbol{P_S^\perp P_G P_S^\perp} \\ &= \boldsymbol{P_S^\perp P_G^\perp P_S^\perp} = \boldsymbol{P_S^\perp}(\boldsymbol{I} - \boldsymbol{E_{HS}})\boldsymbol{P_S^\perp}. \end{aligned} \tag{3.6}$$

Note that $\boldsymbol{P_S^\perp} - \boldsymbol{P_{HS}^\perp}$ is just the projection

$$\boldsymbol{P_S^\perp} - \boldsymbol{P_{HS}^\perp} = \boldsymbol{P_S^\perp P_G P_S^\perp} = \boldsymbol{P_S^\perp E_{HS} P_S^\perp}. \tag{3.7}$$

(3.6) and (3.7) are key identities. They will allow us to write quadratic forms in measurements $\boldsymbol{y}$ as follows:

$$\begin{aligned} \boldsymbol{y}^{\mathrm{T}}(\boldsymbol{P_S^\perp} - \boldsymbol{P_{HS}^\perp})\boldsymbol{y} &= \boldsymbol{y}^{\mathrm{T}} \boldsymbol{P_S^\perp P_G P_S^\perp y} = \boldsymbol{y}^{\mathrm{T}} \boldsymbol{P_S^\perp E_{HS} P_S^\perp y} \\ \boldsymbol{y}^{\mathrm{T}} \boldsymbol{P_{HS}^\perp y} &= \boldsymbol{y}^{\mathrm{T}} \boldsymbol{P_S^\perp P_G^\perp P_S^\perp y} = \boldsymbol{y}^{\mathrm{T}} \boldsymbol{P_S^\perp}(\boldsymbol{I} - \boldsymbol{E_{HS}})\boldsymbol{P_S^\perp y}. \end{aligned} \tag{3.8}$$

These quadratic forms are fundamental to our study of the GLRT.

Note: Whenever the subspace $\boldsymbol{H}$ is the rank-1 matrix $\boldsymbol{x}$, then $\boldsymbol{P_{HS}^\perp}$ is $\boldsymbol{P_{xS}^\perp} = \boldsymbol{P_S^\perp P_G^\perp P_S^\perp} = \boldsymbol{P_S^\perp}(\boldsymbol{I} - \boldsymbol{E_{HS}})\boldsymbol{P_S^\perp}$, where $\boldsymbol{P_G}$ is the rank-1 orthogonal projector $\boldsymbol{P_G} = \boldsymbol{P_{P_S^\perp x}} = \boldsymbol{P_S^\perp x}(\boldsymbol{x}^{\mathrm{T}} \boldsymbol{P_S^\perp x})^{-1} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{P_S^\perp}$ and $\boldsymbol{E_{HS}}$ is the rank-1 oblique projector $\boldsymbol{E_{HS}} = \boldsymbol{x}(\boldsymbol{x}^{\mathrm{T}} \boldsymbol{P_S^\perp x})^{-1} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{P_S^\perp}$. These are the projectors that arise in the study of known-form signal detection problems.

IV. The GLRT and Its Natural Invariances

The question we pose is this: "What can we say about the (generalized) likelihood ratio when unknown parameters are replaced by maximum likelihood estimates (MLE's) of them?" In other words, what kinds of invariances does the estimated likelihood ratio have, and how is it distributed? As we shall see, these questions underlie a systematic discussion of the GLRT, its invariances, and its optimality.

When the parameters (β_i, σ_i^2) are replaced by their MLE's$(\hat{\beta}_i, \hat{\sigma}_i^2)$, then the corresponding MLE of the likelihood ratio is called the *generalized likelihood ratio* (GLR):

$$\begin{aligned} \hat{l}(\boldsymbol{y}) &= \frac{l(\hat{\beta}_1, \hat{\sigma}_1^2; \boldsymbol{y})}{l(\hat{\beta}_0, \hat{\sigma}_0^2; \boldsymbol{y})} \\ &= \left(\frac{2\pi\hat{\sigma}_1^2}{2\pi\hat{\sigma}_0^2}\right)^{-N/2} \exp\left\{-\frac{1}{2\hat{\sigma}_1^2}\|\hat{\boldsymbol{n}}_1\|_2^2 + \frac{1}{2\hat{\sigma}_0^2}\|\hat{\boldsymbol{n}}_0\|_2^2\right\}. \end{aligned} \tag{4.1}$$

Note that $\hat{\boldsymbol{\beta}}_1 = (\hat{\mu}\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$, whereas $\hat{\boldsymbol{\beta}}_0 = (\mathbf{0}, \hat{\phi})$, with $\hat{\boldsymbol{\phi}}$ differing under H_1 and H_0. Thus, in this formula, $\hat{\boldsymbol{n}}_1$ and $\hat{\boldsymbol{n}}_0$ are the MLE's

$$\begin{aligned}\hat{\boldsymbol{n}}_1 &= \boldsymbol{y} - \hat{\mu}\boldsymbol{H}\hat{\boldsymbol{\theta}} - \boldsymbol{S}\hat{\boldsymbol{\phi}} \\ \hat{\boldsymbol{n}}_0 &= \boldsymbol{y} - \boldsymbol{S}\hat{\boldsymbol{\phi}}.\end{aligned} \tag{4.2}$$

We shall have more to say about these MLE's shortly. When σ^2 is known, then there is no need to estimate σ^2, and it will be convenient to replace the GLR by the *logarithmic GLR*

$$L_1(\boldsymbol{y}) = 2\ln \hat{l}(\boldsymbol{y}) = \frac{1}{\sigma^2}\left[||\hat{\boldsymbol{n}}_0||_2^2 - ||\hat{\boldsymbol{n}}_1||_2^2\right]. \tag{4.3}$$

When σ^2 is unknown, then $\hat{\sigma}_i^2 = \frac{1}{N}||\hat{\boldsymbol{n}}_i||_2^2$. Then it will be convenient to replace the GLR by the $(N/2)$-root GLR

$$L_2(\boldsymbol{y}) = \left(\hat{l}(\boldsymbol{y})\right)^{2/N} = \frac{||\hat{\boldsymbol{n}}_0||_2^2}{||\hat{\boldsymbol{n}}_1||_2^2}. \tag{4.4}$$

These two forms will play a key role in our studies of invariance. Although it is a slight abuse of terminology, we shall refer to $L_1(\boldsymbol{y})$ and $L_2(\boldsymbol{y})$ as GLRs.

The GLRT: The generalized likelihood ratio *test* (GLRT) is a natural extension of the Neyman-Pearson likelihood ratio test:

$$\begin{aligned}\phi(\boldsymbol{y}) &= \begin{cases} 1 \sim H_1, & L(\boldsymbol{y}) > \eta \\ 0 \sim H_0, & L(\boldsymbol{y}) \le \eta \end{cases} \\ P_{\mathrm{FA}} &= \sup E_{H_0}\phi(\boldsymbol{y}) = \sup P\left[L(\boldsymbol{y}) > \eta | H_0\right]\end{aligned} \tag{4.5}$$

where $L(\boldsymbol{y})$ is $L_1(\boldsymbol{y})$ or $L_2(\boldsymbol{y})$, depending on whether or not σ^2 is known. The function $\phi(\boldsymbol{y})$ selects H_1 when the GLR $L(\boldsymbol{y})$ exceeds a threshold η. The sup in (4.5) is the sup over all parameters $(\boldsymbol{\beta}, \sigma^2)$ under H_0. When the distribution of $L(\boldsymbol{y})$ is known under H_0, then the threshold η may be set to give a desired constant false alarm rate (CFAR) P_{FA}. The probability of detection is

$$P_{\mathrm{D}} = P\left[L(\boldsymbol{y}) > \eta | H_1\right]. \tag{4.6}$$

Invariance of the GLRT: We shall say that the GLRT is $\mathcal{T}$-invariant if the GLR $L(\boldsymbol{y})$ is invariant to transformations $T \in \mathcal{T}$:

$$L\left(T(\boldsymbol{y})\right) = L(\boldsymbol{y}); \qquad T \in \mathcal{T}. \tag{4.7}$$

By studying the invariance class $\mathcal{T}$, we gain geometrical insight into the mathematical structure of the GLR. Furthermore, we will be able to show that, of all detectors that are invariant-$\mathcal{T}$, the GLRT is the uniformly most powerful (UMP). This is the strongest statement of optimality that we could hope to make about a test of H_0 versus H_1, meaning that the GLRT cannot be improved upon by any detector which shares its invariances. We will argue that the invariances are so natural that no detector would be accepted which did not have them.

With these preliminaries established, we now undertake a study of four closely related problems, ranging from the detection of known-form signals in subspace interference and Gaussian noise to the detection of subspace signals in subspace interference and Gaussian noise of unknown level.

V. Known Signal in Subspace Interference and Noise of Known Level

The problem here is to test $H_0 : \boldsymbol{y} = \boldsymbol{S\phi} + \boldsymbol{n}$ versus $H_1 : \boldsymbol{y} = \mu\boldsymbol{x} + \boldsymbol{S\phi} + \boldsymbol{n}$ where $\mu > 0$. The signal $\boldsymbol{x}$ is known, the subspace interference $\boldsymbol{S\phi}$ lies in the rank-t subspace $\langle \boldsymbol{S} \rangle$, and the noise $\boldsymbol{n}$ is drawn from the normal distribution $\mathcal{N}[\mathbf{0}, \sigma^2\boldsymbol{I}]$ with σ^2 known. We may write the detection problem as a test of distributions:

$$H_0 : \boldsymbol{y} : \mathcal{N}[\boldsymbol{S\phi}, \sigma^2\boldsymbol{I}] \quad \text{versus} \quad H_1 : \boldsymbol{y} : \mathcal{N}[\mu\boldsymbol{x} + \boldsymbol{S\phi}, \sigma^2\boldsymbol{I}]. \tag{5.1}$$

We are testing the hypothesis that the mean of the distribution lies in the intersection of the subspace $\langle \boldsymbol{xS} \rangle$ and the positive half-line of the subspace $\langle \boldsymbol{x} \rangle$.

MLE's: The noise defined in (2.6) is

$$\boldsymbol{n} = \boldsymbol{y} - [\boldsymbol{x}, \boldsymbol{S}]\begin{bmatrix}\mu \\ \boldsymbol{\phi}\end{bmatrix}. \tag{5.2}$$

The MLE's for $\hat{\boldsymbol{n}}$ are, following (4.2)

$$\begin{aligned}\hat{\boldsymbol{n}}_1 &= \boldsymbol{y} - [\boldsymbol{x}, \boldsymbol{S}]\begin{bmatrix}\hat{\mu}_1 \\ \hat{\boldsymbol{\phi}}_1\end{bmatrix} \\ \hat{\boldsymbol{n}}_0 &= \boldsymbol{y} - \boldsymbol{S}\hat{\boldsymbol{\phi}}_0.\end{aligned} \tag{5.3}$$

The subscripts on the estimates remind us that the estimates are dependent upon the hypothesis.

If we proceed as if μ is unconstrained to be non-negative, then the MLE's for $\hat{\mu}$ and $\hat{\boldsymbol{\phi}}$ are obtained by writing $\boldsymbol{E_{xS}y}$ as $\boldsymbol{x}\hat{\mu}$ and $\boldsymbol{E_{Sx}y}$ as $\boldsymbol{S}\hat{\phi}$ [8]:

$$\begin{aligned}\hat{\mu} &= (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x})^{-1}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y} \\ \hat{\boldsymbol{\phi}}_1 &= (\boldsymbol{S}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{x}}^{\perp}\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{x}}^{\perp}\boldsymbol{y} \\ \hat{\boldsymbol{\phi}}_0 &= (\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{y}.\end{aligned} \tag{5.4}$$

The estimate $\hat{\mu}$ is distributed as $\mathcal{N}\left[\mu, \sigma^2(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x})^{-1}\right]$ under H_1, and the estimates $\hat{\boldsymbol{\phi}}_i$ are distributed as $\mathcal{N}\left[\boldsymbol{\phi}, \sigma^2(\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S})^{-1}\right]$ and $\mathcal{N}\left[\boldsymbol{\phi}, \sigma^2(\boldsymbol{S}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{x}}^{\perp}\boldsymbol{S})^{-1}\right]$ under their respective hypotheses. This makes them ML, MVUB, etc. (see Chapter 3 in [1]).

If we now enforce the constraint that $\hat{\mu}_1 \ge 0$, then the MLE's are

$$\begin{aligned}\hat{\mu}_1 &= \max(0, \hat{\mu}) \\ \hat{\boldsymbol{\phi}}_1 &= \begin{cases}\hat{\boldsymbol{\phi}}_0, & \hat{\mu} \le 0 \\ (\boldsymbol{S}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{x}}^{\perp}\boldsymbol{S})^{-1}\boldsymbol{S}^{T}\boldsymbol{P}_{\boldsymbol{x}}^{\perp}\boldsymbol{y}, & \hat{\mu} > 0\end{cases}.\end{aligned} \tag{5.5}$$

The corresponding MLE's for the noises $\boldsymbol{n}_1$ and $\boldsymbol{n}_0$ are

$$\begin{aligned}\hat{\boldsymbol{n}}_1 &= \begin{cases}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}, & \hat{\mu} \le 0 \\ \boldsymbol{P}_{\boldsymbol{xS}}^{\perp}\boldsymbol{y}, & \hat{\mu} > 0\end{cases} \\ \hat{\boldsymbol{n}}_0 &= \boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}.\end{aligned} \tag{5.6}$$

GLR: With these results for the estimated noises, we may write the logarithmic GLR as

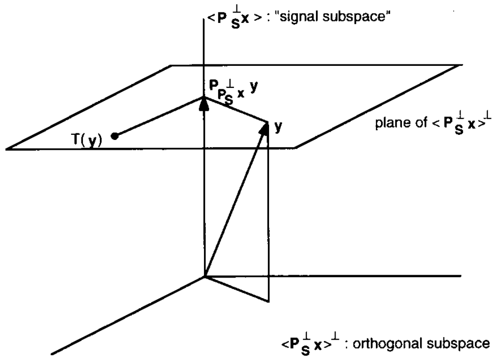

Fig. 3. Invariances of the GLRT for detecting known signal in subspace interference and noise of known level.

Invariances: The GLR $L_1(\boldsymbol{y})$ is invariant to transformations $T \in \mathcal{T}_1$, where $\mathcal{T}_1$ is the set of rotations and translations of $\boldsymbol{y}$ in the space $\langle \boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}\rangle^{\perp}$. However, since translations subsume rotations *and* translations, we say that the GLR is invariant to translations in the space $\langle \boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}\rangle^{\perp}$. This $(N-1)$-dimensional subspace is illustrated in Fig. 3. Why is $L_1(\boldsymbol{y})$ invariant-$\mathcal{T}_1$? Because $\hat{\mu}(T\boldsymbol{y}) = \hat{\mu}(\boldsymbol{y}+\underline{\nu}) = \hat{\mu}(\boldsymbol{y})$ whenever $\underline{\nu} \in \langle \boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}\rangle^{\perp}$ and $L_1(T\boldsymbol{y}) = L_1(\boldsymbol{y}+\underline{\nu}) = L_1(\boldsymbol{y})$. These invariances are illustrated in Fig. 3. Furthermore, if

$$L_1(\boldsymbol{y}) = \begin{cases} 0, & \hat{\mu} \le 0 \\ \frac{1}{\sigma^2}[\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y} - \boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{xS}}^{\perp}\boldsymbol{y}], & \hat{\mu} > 0 \end{cases}$$
$$= \begin{cases} 0, & \hat{\mu} \le 0 \\ \frac{1}{\sigma^2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}}\boldsymbol{y}, & \hat{\mu} > 0 \end{cases}$$
$$\hat{\mu} = (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x})^{-1}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}. \tag{5.7}$$

$L_1(\boldsymbol{y}_1) = L_1(\boldsymbol{y}_2)$, then there exists a transformation $T \in \mathcal{T}_1$ such that $\boldsymbol{y}_2 = T(\boldsymbol{y}_1)$. This makes the logarithmic GLR a maximal invariant statistic [7], meaning that every $\mathcal{T}_1$-invariant test of H_0 versus H_1 must be a function of $L_1(\boldsymbol{y})$.

The subspace $\langle \boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}\rangle$ is the space where the signal $\boldsymbol{x}$ lies after it has passed through the null steering operator $\boldsymbol{P}_{\boldsymbol{S}}^{\perp}$. Any component of the measurement $\boldsymbol{y}$ that lies in the subspace orthogonal to $\langle \boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}\rangle$ is—and should be—invisible to the matched subspace filter $\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}}$. Therefore, the invariances of $L_1(\boldsymbol{y})$ are, indeed, the natural invariances for this problem.

Optimality and Performance: The logarithmic GLR $L_1(\boldsymbol{y})$ is the unique invariant statistic for testing H_0 versus H_1 in the sense that *every* $\mathcal{T}_1$-invariant test of H_0 versus H_1 must be a function of it. The logarithmic GLR has a mixed discrete-continuous distribution because of the way it is defined with respect to $\hat{\mu}$. The distribution of the *matched filter* statistic $\hat{\mu}$ is

$$\hat{\mu} : \mathcal{N}[\mu, \sigma^2(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x})^{-1}]; \quad \begin{matrix} \mu = 0 \text{ under } H_0 \\ \mu > 0 \text{ under } H_1. \end{matrix} \tag{5.8}$$

The probability that $\hat{\mu} \le 0$ is therefore

$$\gamma(\mu) = P[\hat{\mu} \le 0] = P\left[\mathcal{N}\left[\frac{\mu}{\sigma}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x})^{1/2}, 1\right] \le 0\right]. \tag{5.9}$$

The notation $P\left[\mathcal{N}[m,\sigma^2] \le 0\right]$ denotes the probability that a normal random variable with mean m and variance σ^2 is less than or equal to zero. The random variable $\sigma^{-1}\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}}\boldsymbol{y}$ is distributed as follows:

$$\frac{1}{\sigma}\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}}\boldsymbol{y} : \mathcal{N}\left[\frac{\mu}{\sigma}\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}}\boldsymbol{x}, \boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}}\right]. \tag{5.10}$$

Therefore, $\sigma^{-2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}}\boldsymbol{y}$ is chi-squared distributed with one degree of freedom (the rank of $\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}}$) and noncentrality parameter λ^2 [1]:

$$\frac{1}{\sigma^2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}}\boldsymbol{y} : \chi_1^2(\lambda^2)$$
$$\lambda^2 = \frac{\mu^2}{\sigma^2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}}\boldsymbol{x} = \frac{\mu^2}{\sigma^2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}. \tag{5.11}$$

With these results, we see that the GLR $L_1(\boldsymbol{y})$ has a mixed distribution which we write as

$$z = L_1(\boldsymbol{y}) : \gamma(\mu)\delta(z) + (1-\gamma(\mu))\chi_1^2(\lambda^2). \tag{5.12}$$

That is, $z = L_1(\boldsymbol{y})$ has discrete probability mass of $\gamma(\mu)$ at $z = 0$ and continuous probability $(1-\gamma(\mu))\chi_1^2(\lambda^2)$ on the positive real axis. The distribution of $L_1(\boldsymbol{y})$ is monotone in the parameter $\mu \ge 0$, so by the Karlin-Rubin theorem [7] the GLRT

$$\phi(\boldsymbol{y}) = \begin{cases} 1 \sim H_1, & L_1(\boldsymbol{y}) > \eta \\ 0 \sim H_0, & L_1(\boldsymbol{y}) \le \eta \end{cases}$$
$$L_1(\boldsymbol{y}) = \begin{cases} 0, & \hat{\mu} \le 0 \\ \frac{1}{\sigma^2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}}\boldsymbol{y}, & \hat{\mu} > 0 \end{cases}$$
$$\hat{\mu} = (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x})^{-1}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y} \tag{5.13}$$

is uniformly most powerful (UMP) invariant for testing H_0 versus H_1. This is the strongest statement of optimality we could hope to make about a detector of H_0 versus H_1. The false alarm and detection probabilities are

$$P_{\mathrm{FA}} = (1-\gamma(0))(1 - P[\chi_1^2(0) \le \eta])$$
$$P_{\mathrm{D}} = (1-\gamma(\mu))(1 - P[\chi_1^2(\lambda^2) \le \eta])$$
$$\gamma(\mu) = P\left[\mathcal{N}\left[\frac{\mu}{\sigma}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x})^{1/2}, 1\right] \le 0\right]$$
$$\lambda^2 = \frac{\mu^2}{\sigma^2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}. \tag{5.14}$$

The noncentrality parameter λ^2 is the SNR in units of power.

We lend a different interpretation to $L_1(\boldsymbol{y})$ by noting that it is a monotone function of $\hat{\mu}$. This means that we may replace the logarithmic GLR $L_1(\boldsymbol{y})$ by the monotone function (there is no need to invent a new notation)

$$\begin{aligned} L_1(\boldsymbol{y}) &= \max\left[0, \left(\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}}{\sigma^2}\right)^{1/2}\hat{\mu}\right] \\ &= \max\left[0, \left(\frac{1}{\sigma^2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}}\right)^{1/2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}\right]. \end{aligned} \tag{5.15}$$

This is the familiar *matched filter*, censored to be nonnegative. The distribution of $(\sigma^2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x})^{-1/2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}$ is $\mathcal{N}[\lambda, 1]$. Therefore, the GLRT

$$\phi(\boldsymbol{y}) = \begin{cases} 1 \sim H_1, & L_1(\boldsymbol{y}) > \eta \\ 0 \sim H_0, & L_1(\boldsymbol{y}) \le \eta \end{cases} \tag{5.16}$$

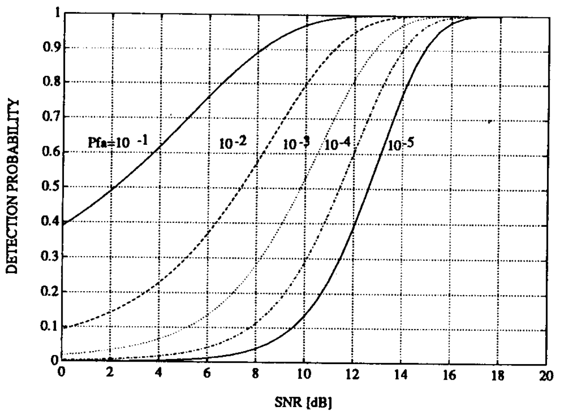

Fig. 4. ROC for known signal in subspace interference and noise of known level.

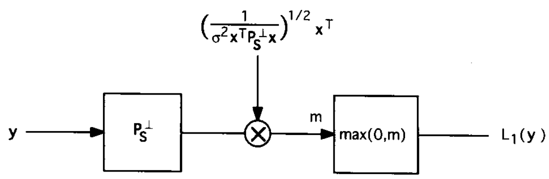

Fig. 5. Detector diagram.

is UMP-invariant with false-alarm and detection probabilities

$$\begin{aligned} P_{\mathrm{FA}} &= 1 - P[\mathcal{N}[0,1] \le \eta], \qquad \eta > 0 \\ P_{\mathrm{D}} &= 1 - P[\mathcal{N}[\lambda,1] \le \eta] \\ \lambda &= \frac{\mu}{\sigma}(\boldsymbol{x}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{x})^{1/2}. \end{aligned} \tag{5.17}$$

The parameter λ is the SNR in units of voltage; λ^2 is the signal energy after it has passed through the null-steering operator $\boldsymbol{P}_{\boldsymbol{S}}^{\perp}$.

The receiver operating characteristics (ROC's) for this detector are given in Fig. 4, and the detector diagram is given in Fig. 5. Note the interference rejecting filter $\boldsymbol{P}_{\boldsymbol{S}}^{\perp}$ followed by a matched filter.

Note: When the test H_0 versus H_1 is replaced by the two-sided test $H_1 : \mu \neq 0$ versus $H_1 : \mu = 0$, then the constraint that $\hat{\mu}_1 > 0$ is not enforced. The logarithmic GLR is then

$$L_1(\boldsymbol{y}) = \frac{1}{\sigma^2} \boldsymbol{y}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{x}} \boldsymbol{y}. \tag{5.18}$$

This statistic is invariant to $T \in \mathcal{T}_1$, and it has monotone likelihood ratio. Therefore, the test

$$\phi(\boldsymbol{y}) = \begin{cases} 1 \sim H_1, & L_1(\boldsymbol{y}) > \eta \\ 0 \sim H_0, & L_1(\boldsymbol{y}) \le \eta \end{cases} \tag{5.19}$$

is UMP-invariant. The distribution of $L_1(\boldsymbol{y})$ is chi-squared with one degree of freedom and nonconcentrality parameter λ^2:

$$\begin{aligned} L_1(\boldsymbol{y}) &: \chi_1^2(\lambda^2) \\ \lambda^2 &= \frac{\mu^2}{\sigma^2} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{x}. \end{aligned} \tag{5.20}$$

The false alarm and detection probabilities are

$$\begin{aligned} P_{\mathrm{FA}} &= 1 - P[\chi_1^2(0) \le \eta] \\ P_{\mathrm{D}} &= 1 - P[\chi_1^2(\lambda^2) \le \eta]. \end{aligned} \tag{5.21}$$

These results apply to the detection of rank-1 signals whose polarity can be changed by a reflection mechanism. As this problem is a special case of the more general problem to be treated in Section VII, we defer its more complete discussion until then. The results of this section generalize the results of [1], [5]–[6].

VI. Known Signal in Subspace Interference and Noise of Unknown Level

The problem here is to test

$$H_0 : \boldsymbol{y} : \mathcal{N}[\boldsymbol{S\phi}, \sigma^2 \boldsymbol{I}] \text{ versus } H_1 : \boldsymbol{y} : \mathcal{N}[\mu \boldsymbol{x} + \boldsymbol{S\phi}, \sigma^2 \boldsymbol{I}]; \quad \mu > 0, \quad \sigma^2 \text{unknown}. \tag{6.1}$$

For this problem, the MLE's $\hat{\boldsymbol{n}}_0$ and $\hat{\boldsymbol{n}}_1$ remain unchanged from Section V, but now the estimated variances are $\hat{\sigma}_1^2 = ||\hat{\boldsymbol{n}}_1||^2/N$ and $\hat{\sigma}_0^2 = ||\hat{\boldsymbol{n}}_0||^2/N$. (These results for $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are obtained by differentiating log-likelihood or by positing them and then using a variational argument.)

GLR: The $N/2$-root GLR is

$$L_2(\boldsymbol{y}) = \begin{cases} 1, & \hat{\mu} \le 0 \\ \dfrac{\boldsymbol{y}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{xS}}^{\perp} \boldsymbol{y}}, & \hat{\mu} > 0 \end{cases}$$
$$\hat{\mu} = (\boldsymbol{x}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{x})^{-1} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{y}. \tag{6.2}$$

It is actually more natural to reference $L_2(\boldsymbol{y})$ to unity, in which case the monotone function $L_2(\boldsymbol{y}) - 1$ may be written

$$L_2(\boldsymbol{y}) - 1 = \begin{cases} 0, & \hat{\mu} \le 0 \\ \dfrac{\boldsymbol{y}^{\mathrm{T}} (\boldsymbol{P}_{\boldsymbol{S}}^{\perp} - \boldsymbol{P}_{\boldsymbol{xS}}^{\perp}) \boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{xS}}^{\perp} \boldsymbol{y}}, & \hat{\mu} > 0. \end{cases} \tag{6.3}$$

We now call $L_2(\boldsymbol{y}) - 1$ simply $L_2(\boldsymbol{y})$ and use the identities of (3.6) and (3.7) to write this GLR as

$$L_2(\boldsymbol{y}) = \begin{cases} 0, & \hat{\mu} \le 0 \\ \dfrac{\boldsymbol{y}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{P}_{\boldsymbol{G}} \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{P}_{\boldsymbol{G}}^{\perp} \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{y}}, & \hat{\mu} > 0 \end{cases} \tag{6.4}$$

where $\boldsymbol{P}_{\boldsymbol{G}}$ is the projector $\boldsymbol{P}_{\boldsymbol{G}} = \boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{x}}$. The GLR may also be written as

$$L_2(\boldsymbol{y}) = \begin{cases} 0, & \hat{\mu} \le 0 \\ \dfrac{\boldsymbol{y}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{E}_{\boldsymbol{xS}} \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{S}}^{\perp} (\boldsymbol{I} - \boldsymbol{E}_{\boldsymbol{Sx}}) \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{y}}, & \hat{\mu} > 0 \end{cases} \tag{6.5}$$

where $\boldsymbol{E}_{\boldsymbol{xS}}$ is the oblique projection

$$\boldsymbol{E}_{\boldsymbol{xS}} = \boldsymbol{x}(\boldsymbol{x}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{x})^{-1} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{S}}^{\perp}. \tag{6.6}$$

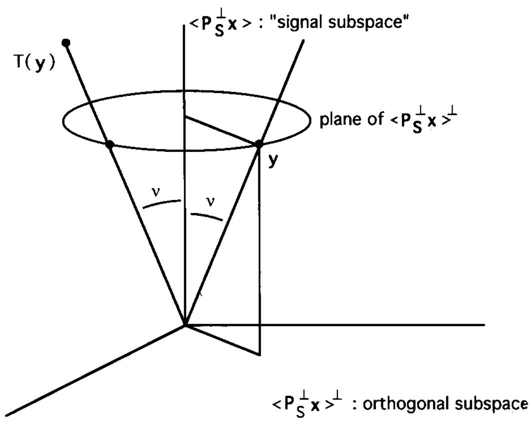

Fig. 6. Invariances for $L_2(\boldsymbol{y})$.

Invariances: The GLR $L_2(\boldsymbol{y})$ is invariant to transformations $T \in \mathcal{T}_2$, where $\mathcal{T}_2$ is the set of rotations in $\langle \boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}\rangle^{\perp}$ (rotations around $\langle \boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}\rangle$) and scalings $\gamma > 0$ illustrated in Fig. 6. That is, $L_2(T\boldsymbol{y}) = L_2(\boldsymbol{y})$. Furthermore, if $L_2(\boldsymbol{y}_2) = L_2(\boldsymbol{y}_1)$, then there exists a $T \in \mathcal{T}_2$ such that $\boldsymbol{y}_2 = T(\boldsymbol{y}_1)$. (Fig. 6 is geometrically convincing, but a rigorous proof requires an algebraic proof.) This makes the GLR $L_2(\boldsymbol{y})$ a maximal invariant statistic.

Again, the subspace $\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}$ is the subspace where the signal $\boldsymbol{x}$ lies after it has passed through the null-steering operator $\boldsymbol{P}_{\boldsymbol{S}}^{\perp}$. The ratio of energies that define $L_2(\boldsymbol{y})$ are—and ought to be—invariant to rotations around this subspace and to scalings, because scalings are what introduce unknown variances. Therefore, the invariances of $L_2(\boldsymbol{y})$ are, indeed, the natural invariances for this problem.

Optimality and Performance: The $(N/2)$-root GLR $L_2(\boldsymbol{y})$ is the unique invariant statistic for testing H_0 versus H_1. The distribution of the statistic $\hat{\mu}$ is the distribution given in (5.8), and the probability that $\hat{\mu} \le 0$ is given in equation (5.9). The statistics $\sigma^{-1}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}$ and $\sigma^{-1}\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}$ are distributed as follows:

$$\begin{aligned} \sigma^{-1}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y} &: N\left[\frac{\mu}{\sigma}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}, \boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\right] \\ \sigma^{-1}\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y} &: N[0, \boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}]. \end{aligned} \tag{6.7}$$

It is easy to see that these two statistics are independent. This means that the following quadratic forms are independent χ^2 random variables:

$$\begin{aligned} \frac{1}{\sigma^2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y} &: \chi_1^2(\lambda^2) \\ \frac{1}{\sigma^2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y} &: \chi_{N-t-1}^2(0). \end{aligned} \tag{6.8}$$

This means that the ratio is F-distributed with degrees of freedom $(1, N-t-1)$ [1]:

$$\frac{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}}{(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y})/(N-1)} : F_{1,N-t-1}(\lambda^2). \tag{6.9}$$

The noncentrality parameter λ^2 is defined in (5.20).

With these results we see that the GLR $L_2(\boldsymbol{y})$ has a mixed distribution, which we write as

$$z = L_2(\boldsymbol{y}) : \gamma(\mu)\delta(z) + \big(1-\gamma(\mu)\big)F_{1,N-t-1}(\lambda^2) \tag{6.10}$$

that is, $z = L_2(\boldsymbol{y})$ has discrete probability mass of $\gamma(\mu)$ at $z = 0$ and continuous probability $\big(1-\gamma(\mu)\big)F_{1,N-t-1}(\lambda^2)$ on the positive real axis. The distribution of $L_2(\boldsymbol{y})$ is monotone in $\mu \ge 0$, so the GLRT

$$\begin{aligned} \phi(\boldsymbol{y}) &= \begin{cases} 1 \sim H_1, & L_2(\boldsymbol{y}) > \eta \\ 0 \sim H_0, & L_2(\boldsymbol{y}) \le \eta \end{cases} \\ L_2(\boldsymbol{y}) &= \begin{cases} 0, & \hat{\mu} \le 0 \\ \dfrac{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}}, & \hat{\mu} > 0 \end{cases} \\ \hat{\mu} &= (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x})^{-1}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y} \end{aligned} \tag{6.11}$$

is UMP-invariant for testing H_0 versus H_1. The false alarm and detection probabilities are

$$\begin{aligned} P_{\mathrm{FA}} &= \big(1-\gamma(0)\big)\big(1 - P[F_{1,N-t-1}(0) \le \eta]\big) \\ P_{\mathrm{D}} &= \big(1-\gamma(0)\big)\big(1 - P[F_{1,N-t-1}(\lambda^2) \le \eta]\big) \\ \gamma(0) &= P\mathcal{N}[0,1] \le 0] \\ \lambda^2 &= \frac{\mu^2}{\sigma^2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}. \end{aligned} \tag{6.12}$$

We may lend a different interpretation to the GLR $L_2(\boldsymbol{y})$ by noting that it is a monotone function of $\hat{\nu}$; that is,

$$L_2(\boldsymbol{y}) = (N-t-1)^{1/2}\hat{\nu}^2 \tag{6.13}$$

$$\hat{\nu} = \frac{\left[(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x})^{1/2}/\sigma\right]\hat{\mu}}{\left[\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}/(N-t-1)\sigma^2\right]^{1/2}}. \tag{6.14}$$

Furthermore, $\hat{\nu} < 0$ iff $\hat{\mu} < 0$. This means that $L_2(\boldsymbol{y})$ may be replaced by

$$L_2(\boldsymbol{y}) = \max[0, \hat{\nu}]. \tag{6.15}$$

The statistic $\hat{\nu}$ is the ratio of a $\mathcal{N}\left[(\mu/\sigma)(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x})^{1/2}, 1\right]$ random variable and an independent, scaled square root of a $\chi_{N-t-1}^2(0)$ random variable. This makes $\hat{\nu}$ a t-distributed random variable with parameters $(1, N-t-1)$ and noncentrality parameter λ. The UMP-invariant detector may therefore be written

$$\phi(\boldsymbol{y}) = \begin{cases} 1 \sim H_1, & L_2(\boldsymbol{y}) > \eta \\ 0 \sim H_0, & L_2(\boldsymbol{y}) \le \eta \end{cases} \tag{6.16}$$

$$L_2(\boldsymbol{y}) = \max\left[0, \hat{\nu} = \frac{(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x})^{1/2}\hat{\mu}}{\left[\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}/(N-t-1)\right]^{1/2}}\right]. \tag{6.17}$$

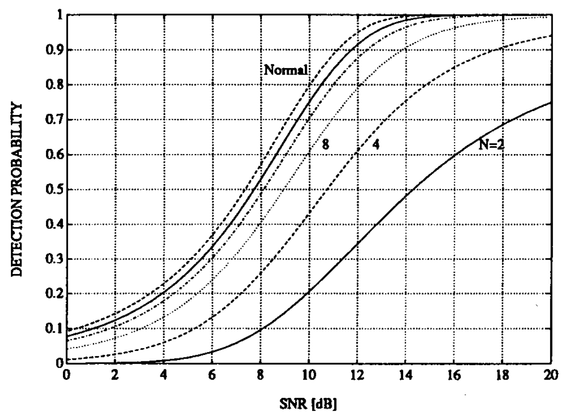

Fig. 7. ROC curves for known signal in subspace interference and noise of unknown level.

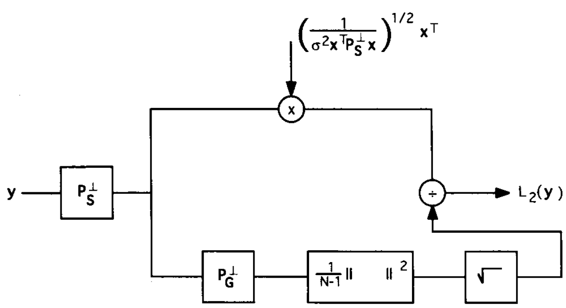

Fig. 8. Detector diagram.

The false alarm and detection probabilities are

$$P_{\rm FA} = 1 - P[t_{1,N-t-1}(0) \le \eta]$$
$$P_{\rm D} = 1 - P[t_{1,N-t-1}(\lambda) \le \eta]. \quad (6.18)$$

The ROC curves are given in Fig. 7, and the detector diagram is given in Fig. 8. In Fig. 7, the probability of false alarm is fixed at $P_{\rm FA} = 0.01$, and the sample size is varied from $N = 2$ to $N = 32$ in powers of 2. The normal ROC curve is plotted for reference. Note that the detector of Fig. 8 uses an interference rejecting filter followed by a matched filter in the top branch and a noise power estimator in the lower branch. In fact, $\boldsymbol{y}^{\rm T}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}/(N-1)$ is a maximum likelihood estimator of σ^2. The results of this section generalize results of [1]–[2], [5]–[6].

VII. Subspace Signal in Subspace Interference and Noise of Known Level

This problem is a generalization of the problem solved in Section V. The signal $\mu\boldsymbol{x}$ is replaced by the signal $\mu\boldsymbol{H\theta}$, where $\langle\boldsymbol{H}\rangle$ is a rank-p subspace. Now, as the elements of $\boldsymbol{\theta}$ may be positive or negative, we do not constrain $\mu\boldsymbol{H\theta}$ to lie in any particular orthant of $\langle\boldsymbol{H}\rangle$. Therefore, μ is absorbed into $\boldsymbol{\theta}$ and

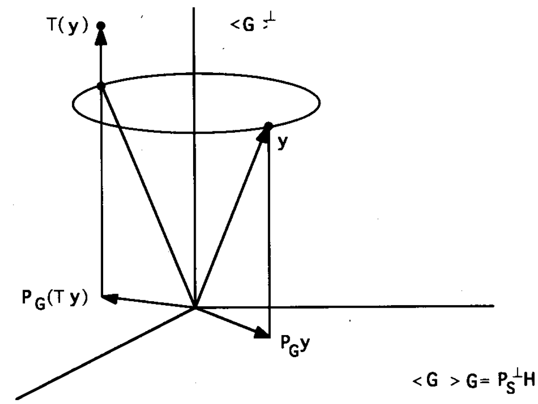

Fig. 9. Invariances for L_1.

the problem is to test the hypotheses

$$H_0 : \boldsymbol{y} : \mathcal{N}[\boldsymbol{S\phi}, \sigma^2\boldsymbol{I}]$$
$$\text{versus } \boldsymbol{y} : \mathcal{N}[\boldsymbol{H\theta} + \boldsymbol{S\phi}, \sigma^2\boldsymbol{I}], \quad \|\boldsymbol{\theta}\|_2^2 > 0. \quad (7.1)$$

For this problem, the MLE's $\hat{\boldsymbol{n}}_0$ and $\hat{\boldsymbol{n}}_1$ are

$$\hat{\boldsymbol{n}}_1 = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{HS}})\boldsymbol{y} = \boldsymbol{P}_{\boldsymbol{HS}}^{\perp}\boldsymbol{y}$$
$$\hat{\boldsymbol{n}}_0 = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{S}})\boldsymbol{y} = \boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}. \quad (7.2)$$

GLR: The GLR $L_1(\boldsymbol{y})$ is

$$\begin{aligned} L_1(\boldsymbol{y}) &= \frac{1}{\sigma^2}[\|\hat{\boldsymbol{n}}_0\|_2^2 - \|\hat{\boldsymbol{n}}_1\|_2^2] \\ &= \frac{1}{\sigma^2}\boldsymbol{y}^{\rm T}(\boldsymbol{P}_{\boldsymbol{S}}^{\perp} - \boldsymbol{P}_{\boldsymbol{HS}}^{\perp})\boldsymbol{y} = \frac{1}{\sigma^2}\boldsymbol{y}^{\rm T}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{y} \\ &= \frac{1}{\sigma^2}\boldsymbol{y}^{\rm T}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y} = \frac{1}{\sigma^2}\boldsymbol{y}^{\rm T}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{E}_{\boldsymbol{HS}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}. \end{aligned} \quad (7.3)$$

(We have used the identity of (3.7).) The identical quadratic forms $\boldsymbol{y}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}$ and $\boldsymbol{y}^{\rm T}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{y}$ are *generalized energy detectors*. Recall that $\boldsymbol{G} = \boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{H}$.

Invariances: The GLR $L_1(\boldsymbol{y})$ is invariant to transformations $T \in \mathcal{T}_3$ that rotate $\boldsymbol{y}$ within $\langle\boldsymbol{G}\rangle$ (or around $\langle\boldsymbol{G}\rangle^{\perp}$) and add a (bias) component in $\langle\boldsymbol{G}\rangle^{\perp}$. Furthermore, if $L_1(\boldsymbol{y}_1) = L_1(\boldsymbol{y}_2)$, then there exists a transformation $T \in \mathcal{T}_3$ such that $\boldsymbol{y}_2 = T(\boldsymbol{y}_1)$. This makes the logarithmic GLR a maximal invariant statistic, meaning that every $\mathcal{T}_3$-invariant test of H_0 versus H_1 must be a function of it.

The space $\langle\boldsymbol{G}\rangle$ is the space where $\boldsymbol{x} = \boldsymbol{H\theta}$ lies after it has passed through the null-steering operator $\boldsymbol{P}_{\boldsymbol{S}}^{\perp}$. As $\boldsymbol{\theta}$ is unknown and unconstrained, the signal to be detected can lie anywhere in $\langle\boldsymbol{G}\rangle$. No signal of constant energy in $\langle\boldsymbol{G}\rangle$ should be any more detectable than any other, so $L_1(\boldsymbol{y})$ should be invariant to rotations in $\langle\boldsymbol{G}\rangle$. The detector is—and should be—invariant to measurement components orthogonal to $\langle\boldsymbol{G}\rangle$. These natural invariances for this problem are illustrated in Fig. 9.

Optimality and Performance: The logarithmic GLR $L_1(\boldsymbol{y})$ is the unique invariant statistic for testing H_0 versus H_1. It is a quadratic form in the projection operator $\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}$. This quadratic form may be thought of as the norm-squared of the statistic $\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}$, which is distributed as

$$\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}:\begin{cases}\mathcal{N}[\boldsymbol{0},\sigma^2\boldsymbol{P}_{\boldsymbol{G}}], & H_0\\ \mathcal{N}[\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{H}\boldsymbol{\theta},\sigma^2\boldsymbol{P}_{\boldsymbol{G}}], & H_1.\end{cases} \tag{7.4}$$

Therefore, the quadratic form $L_1(\boldsymbol{y}) = \frac{1}{\sigma^2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}$ is χ^2 distributed [1]:

$$L_1(\boldsymbol{y}) = \begin{cases}\chi_p^2(0), & H_0\\ \chi_p^2(\lambda^2), & H_1\end{cases}$$
$$\lambda^2 = \frac{\mu^2}{\sigma^2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}. \tag{7.5}$$

This distribution is monotone in the noncentrality parameter λ^2, meaning that the test

$$\phi(\boldsymbol{y}) = \begin{cases}1\sim H_1, & L_1(\boldsymbol{y}) > \eta\\ 0\sim H_0, & L_2(\boldsymbol{y}) \le \eta\end{cases} \tag{7.6}$$

is UMP-invariant for testing H_0 versus H_1. Its false alarm and detection probabilities are

$$\begin{aligned}P_{\mathrm{FA}} &= 1 - P[\chi_p^2(0)\le\eta]\\ P_{\mathrm{D}} &= 1 - P[\chi_p^2(\lambda^2)\le\eta].\end{aligned} \tag{7.7}$$

The ROC curves for the GLRT are given in Fig. 10, and the detector diagram is given in Fig. 11. In Fig. 10, the probability of false alarm is fixed at $P_{\mathrm{FA}} = 0.01$, and the dimension of the subspace $\langle \boldsymbol{H}\rangle$ is varied from $p = 2$ to $p = 8$ in steps of 2. The normal ROC is plotted for reference. Note that the detector of Fig. 11 decomposes into a subspace filter for interference rejection, a subspace filter matched to the remaining signal, and an energy computation. These results generalize the results of [1], [3], [5]–[6].

VIII. Subspace Signal in Subspace Interference and Noise of Unknown Level

This problem is a generalization of the problems solved in Sections VI and VII. The signal $\mu\boldsymbol{x}$ is replaced by the signal $\mu\boldsymbol{H}\boldsymbol{\theta}$, and the noise level σ^2 is unknown. The parameter μ is absorbed into $\boldsymbol{\theta}$, and the problem is to test the hypotheses

$$\begin{aligned}H_0: \boldsymbol{y}: N[\boldsymbol{S}\boldsymbol{\phi},\sigma^2\boldsymbol{I}] \text{ versus } H_1: \boldsymbol{y}: N[\boldsymbol{H}\boldsymbol{\theta}+\boldsymbol{S}\boldsymbol{\phi},\sigma^2\boldsymbol{I}];\\ \|\boldsymbol{\theta}\|_2^2 > 0, \sigma^2 \text{ unknown}.\end{aligned} \tag{8.1}$$

For this problem, the MLE's are those of (7.2). The GLR

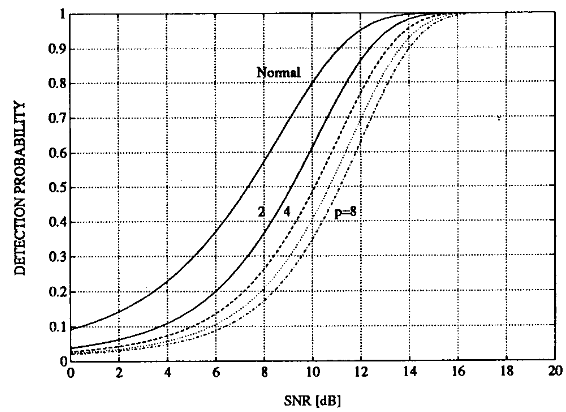

Fig. 10. ROC curves for detecting subspace signal in subspace interference and broadband noise of known variance.

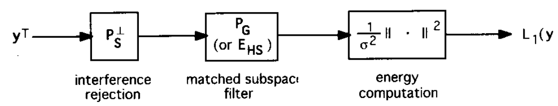

Fig. 11. Detector diagram.

$L_2(\boldsymbol{y})$ is therefore

$$\begin{aligned}L_2(\boldsymbol{y}) &= \frac{\|\hat{\boldsymbol{n}}_0\|_2^2}{\|\hat{\boldsymbol{n}}_1\|_2^2}\\ &= \frac{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{HS}}^{\perp}\boldsymbol{y}}\\ &= \frac{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}}.\end{aligned} \tag{8.2}$$

It is natural to reference $L_2(\boldsymbol{y})$ to unity, in which case the monotone function $L_2(\boldsymbol{y}) - 1$ is

$$\begin{aligned}L_2(\boldsymbol{y}) - 1 &= \frac{\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{P}_{\boldsymbol{S}}^{\perp} - \boldsymbol{P}_{\boldsymbol{HS}}^{\perp})\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{HS}}^{\perp}\boldsymbol{y}} = \frac{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}}\\ &= \frac{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{E}_{\boldsymbol{HS}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}(\boldsymbol{I} - \boldsymbol{E}_{\boldsymbol{HS}})\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}}.\end{aligned} \tag{8.3}$$

(We have used the identities of (3.6) and (3.7).) In what follows we shall call $L_2(\boldsymbol{y}) - 1$ simply $L_2(\boldsymbol{y})$.

Invariances: The GLR $L_2(\boldsymbol{y})$ is invariant to transformations $T \in \mathcal{T}_4$ that rotate in $\langle \boldsymbol{G}\rangle$ (or around $\langle \boldsymbol{G}\rangle^{\perp}$) and non-negatively scale $\boldsymbol{y}$, as illustrated in Fig. 12. They leave the angle ν invariant. Furthermore, if $L_2(\boldsymbol{y}_1) = L_2(\boldsymbol{y}_2)$, then $\boldsymbol{y}_2 = T(\boldsymbol{y}_1)$ for some $T \in \mathcal{T}_4$. This makes $L_2(\boldsymbol{y})$ a maximal invariant statistic.

Again, $\langle \boldsymbol{G}\rangle$ is the subspace where $\boldsymbol{x}$ lies after it has passed through the null-steering operator $\boldsymbol{P}_{\boldsymbol{S}}^{\perp}$. The detector should be invariant to rotations in this space for the reasons given in

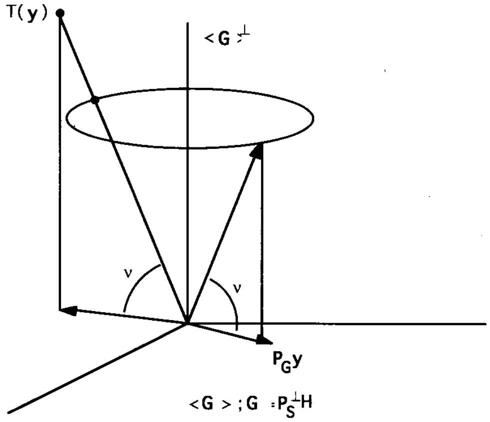

Fig. 12. Invariances for L_2.

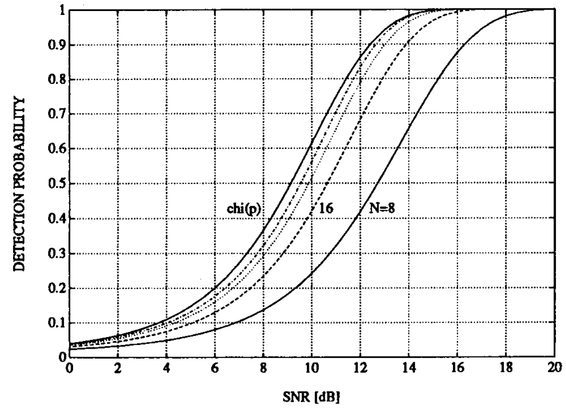

Fig. 13. ROC curves for detecting subspace signal in subspace interference and broadband noise of unknown level; $p = 2$ and N is variable.

Section VII, and it should be invariant to scalings that introduce unknown variances. These are the natural invariances for the problem.

Optimality and Performance: The $(N/2)$-root GLR is the unique invariant statistic for testing H_0 versus H_1. It is the ratio of quadratic forms in $\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}$ and $\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}$. Each of the quadratic forms may be thought of as a norm-squared of a statistic $\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}$ or $\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}$. These statistics are, respectively, distributed as

$$\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y} : \begin{cases} \mathcal{N}[\boldsymbol{0}, \sigma^2 \boldsymbol{P}_{\boldsymbol{G}}], & H_0 \\ \mathcal{N}[\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{H}\boldsymbol{\theta}, \sigma^2 \boldsymbol{P}_{\boldsymbol{G}}], & H_1 \end{cases}$$
$$\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y} : \mathcal{N}[\boldsymbol{0}, \sigma^2 \boldsymbol{P}_{\boldsymbol{G}}^{\perp}], \quad H_0 \text{ or } H_1. \tag{8.4}$$

Furthermore, the random vectors $\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{y}$ and $\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{y}$ are uncorrelated (and therefore independent in this multivariate normal case) by virtue of the fact that $\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}^{\perp} = \boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{G}}^{\perp} = \boldsymbol{0}$:

$$E\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\theta})\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}^{\perp} = \boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}^{\perp} = \boldsymbol{0}. \tag{8.5}$$

This means that the quadratic forms $\frac{1}{\sigma^2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}$ and $\frac{1}{\sigma^2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y}$ are *independent* χ^2 random variables:

$$\frac{1}{\sigma^2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y} : \chi_p^2(\lambda^2)$$
$$\frac{1}{\sigma^2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{P}_{\boldsymbol{G}}^{\perp}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{y} : \chi_{s-p}^2(0). \tag{8.6}$$

The parameter s is the dimension of $\langle \boldsymbol{S}\rangle^{\perp}$, namely $s = N - t > p$, the number of dimensions of $\mathbb{R}^N$ not occupied by $\langle \boldsymbol{S}\rangle$, and p is the dimension of $\langle \boldsymbol{G}\rangle = \langle \boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{H}\rangle$. The noncentrality parameter λ^2 is

$$\lambda^2 = \frac{\mu^2}{\sigma^2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{x}. \tag{8.7}$$

The GLR $\big((s-p)/p\big)L_2$ is distributed as

$$\frac{s-p}{p}L_2(\boldsymbol{y}) : \begin{cases} F_{p,s-p}(\lambda^2) & \text{under } H_1 \\ F_{p,s-p}(0) & \text{under } H_0 \end{cases} \tag{8.8}$$

where F denotes an F-distribution with parameters p and $s-p$. This distribution is monotone in $\lambda^2 \geq 0$, so the GLRT

$$\phi(\boldsymbol{y}) = \begin{cases} 1 \sim H_1, & L_2(\boldsymbol{y}) > \eta \\ 0 \sim H_0, & L_2(\boldsymbol{y}) \leq \eta \end{cases} \tag{8.9}$$

is UMP-invariant for testing H_0 versus H_1. Its false alarm and detection probabilities are

$$P_{\mathrm{FA}} = 1 - P\big[F_{p,s-p}(0) \leq \eta\big]$$
$$P_{\mathrm{D}} = 1 - P\big[F_{p,s-p}(\lambda^2) \leq \eta\big]. \tag{8.10}$$

The ROC curves for the GLRT are given in Figs. 13 and 14, and the detector diagram is given in Fig. 15. In Fig. 13, the probability of false alarm is fixed at $P_{\mathrm{FA}} = 0.01$, the dimension of the subspace $\langle \boldsymbol{H}\rangle$ is $p = 2$, and N is varied from $N = 8$ to $N = 64$ in powers of 2. The ROC for the χ^2 distributed matched subspace detector is plotted for reference. In Fig. 14, the probability of false alarm is fixed at $P_{\mathrm{FA}} = 0.01$, the number of measurements is fixed at $N = 16$, and the subspace dimension is varied from $p = 2$ to $p = 8$ in steps of 2. The normal ROC is plotted for reference. The detector of Fig. 15 decomposes into a subspace filter for interference rejection, a subspace filter matched to the remaining signal, and an energy computation, divided by the same operations with the matched subspace filter replaced by an orthogonal (or "noise") subspace filter. These results generalize the results of [1], [5]–[6].

In summary, the GLRT is UMP invariant for detecting subspace signals in subspace interferences and background noise whenever the noise is MVN. The conclusion holds whether or not the noise variance is known. When the interference is absent, then $\boldsymbol{P}_{\boldsymbol{S}}^{\perp}\boldsymbol{H} = \boldsymbol{H}$ and the GLRs are $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{H}}\boldsymbol{y}$ and $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{H}}\boldsymbol{y}/\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{H}}^{\perp}\boldsymbol{y}$, which are distributed as $\chi_p^2(\lambda^2)$ and $F_{p,N-p}(\lambda^2)$ as discussed in [1]. The parameter λ^2 is then $\lambda^2 = \frac{\mu^2}{\sigma^2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$.

IX. CONCLUSIONS

The generalized likelihood ratio test (GLRT) is a standard procedure for solving detection problems when (nuisance)

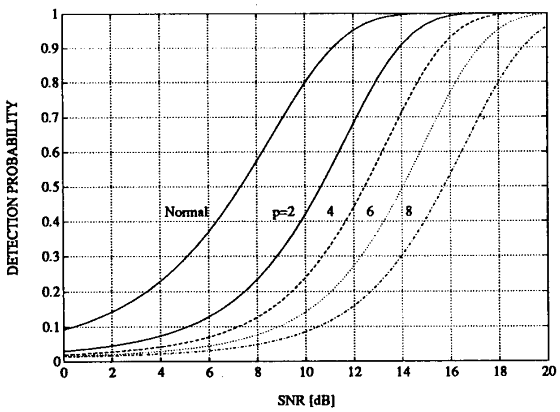

Fig. 14. ROC curves for detecting subspace signal in subspace interference and broadband noise of unknown level; p is variable and $N = 8$.

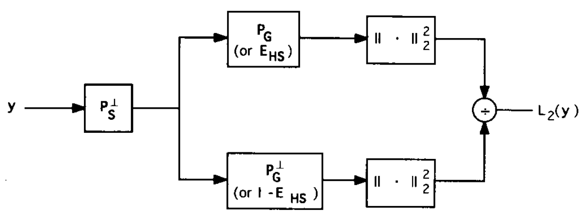

Fig. 15. Detector diagram.

parameters of the underlying distribution are unknown. Typically, nuisance parameters are things like bias, amplitude and phase of sinusoidal interference, noise variance, and so on. These parameters are of no intrinsic interest, but they defeat our efforts to state properties of optimality if we proceed along conventional lines.

The GLRT is easy to derive, and sometimes its distribution can be determined. In these cases, a detection threshold may be set to achieve a constant false alarm rate (CFAR). In spite of its tractability as a bootstrapping technique for solving detection problems, the GLRT has been difficult to characterize in terms of its optimality properties for the class of problems studied in this paper. In fact, it has not been clear whether or not the GLRT has any optimality properties at all for this class. So the question has remained, "can the GLRT be improved upon?"

In this paper we have constructed GLRT's for four detection problems which span a large subset of the practical detection problems encountered in time series analysis and multisensor array processing. For each class of problems we have derived the GLRT and established its invariances. Then we have drawn on the theory of invariance in hypothesis testing to establish that, within the class of invariant detectors which have the same invariances as the GLRT, the GLRT is uniformly most powerful (UMP) invariant. This is the strongest statement of optimality one could hope to make for a detector. For each class of problems, the invariances of the GLRT are just the invariances one would expect of a detector that claims to be optimum. The conclusion is that the GLRT cannot be improved upon for the classes of problems studied in this paper.

The geometrical interpretation of our results is this: Think of the plane $< S >^\perp$ in Fig. 2 (i) as a backplane onto which measurements $\boldsymbol{y}$ are projected to produce the interview-tree vector $P_S \perp \boldsymbol{y}$. This projection can be resolved into its two orthogonal components, $P_G P_S \perp \boldsymbol{y}$ and $P_G^\perp P_S \perp \boldsymbol{y}$. These two components are tested to see which of two competing hypotheses is in force. For detecting a deterministic rank one signal in interference and additive noise, when the noise level os known, the component $P_G P_S \perp \boldsymbol{y}$ is tested to see if it is positive and large enough. If the noise level is unknown, this component is tested to see if it is positive and large enough compared with the orthogonal component $P_G^\perp P_S \perp \boldsymbol{y}$. That is to say, the angle to the subspace $< G >$ must be less than $\pi/2$ radians and small enough. For detecting a subspace signal in interference and noise of known level, the energy in the component $P_G P_S \perp \boldsymbol{y}$ is tested to see if it is large enough. If the noise level is unknown, then the energy is tested to see if it is large enough compared with the energy in the orthogonal component $P_G^\perp P_S \perp \boldsymbol{y}$. That is to say, the angle to the subspace $< G >$ must be small enough. In summary, it is essentially the Pythagorean decomposition of $P_S \perp \boldsymbol{y}$ at the subspaces $< G >$ as $< G >^\perp$.

Our final remark is that each of the detectors studied in this paper may be realized as a generalized energy detector or as a ratio of generalized energy detectors. A typical detector first projects data onto a low-rank subspace where interference is removed. Such an operator is usually called a *null steering* or *interference rejecting filter*. Then the detector projects the data onto a low-rank subspace that is matched to the signal that remains in the data. This filter is usually called a *matched subspace filter* or *matched field filter*. The energy of the filter output is computed and compared with a threshold.

When projectors are replaced by time-invariant, frequency-selective digital filters, then the detectors look like band-selective filters followed by energy detectors. It is not hard to imagine the low-rank subspaces of $\mathbb{R}^N$ replaced by low-rank subspaces of ℓ_2 which are spanned by Fourier bases, wavelet bases, and the like. Then all of the formulas of this paper go through, with the formulas for $\boldsymbol{P_S^\perp}$, $\boldsymbol{P_G}$, $\boldsymbol{P_G^\perp}$, $\boldsymbol{E_{HS}}$, and $\boldsymbol{E_{SH}}$ replaced by their ℓ_2 analogs. This produces a theory of GLRT's that can be implemented in subbands of ℓ_2. The details of this extension will be reported in future work.

References

[1] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.

[2] L. L. Scharf and D. W. Lytle, "Signal detection in Gaussian noise of unknown level: An invariance application," *IEEE Trans Inform. Theory*, vol IT-17, no. 4, pp. 404–411, July 1971.

[3] S. M. Kay and L. L. Scharf, "Invariant detection of transient ARMA signals with unknown initial conditions," in *Proc. 1983 Int. Conf ASSP* (San Diego), 1983.

[4] M. J. Dunn, "Sufficiency and invariance principles applied to four detection problems," M.S. Thesis, Univ. of Colorado at Boulder, 1986.

[5] B. Friedlander and B. Porat, "Performance analysis of transient detectors based on linear data transforms," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 665-673, Mar. 1992.

[6] B. Porat and B. Friedlander, "Performance analysis of a class of transient detection algorithms—A unified framework," *IEEE Trans. Signal Processing*, vol. 40, no. 10, pp. 2536–2546, Oct. 1992.

[7] T. S. Ferguson, *Mathematical Statistics*. New York: Academic, 1967.

[8] R. T. Behrens and L. L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Trans. Signal Processing*, vol. 42, no. 6, pp. 1413–1424, June 1994.

Louis L. Scharf (S'67–M'69–SM'77–F'86) received the Ph.D. degree in electrical engineering from the University of Washington, Seattle, USA, in 1969.

From 1969 to 1971, he was a member of the Technical Staff at Honeywell's Marine Systems Center in Seattle. He was a faculty member at Colorado State University, Fort Collins, from 1971 to 1981, where he last served as Professor of Electrical Engineering and Statistics. From 1982 to 1985, he was Professor and Chairman of Electrical Engineering at the University of Rhode Island, Kingston. He is currently Professor of Electrical and Computer Engineering at the University of Colorado, Boulder, where he teaches and conducts research in signal processing. In 1974 he was a Visiting Associate Professor at Duke University, Durham, NC. In 1977 he was at the University of South Paris, Orsay, France, where he was a member of the Technical Staff in the CNRS Laboratoire des Signaux et Systemes, Gif-sur-Yvette. In 1981 he was a Visiting Professor at the Ecole Nationale Supérieure des Télécommunications, Paris. In 1992 he was a Visiting Professor at Institut Eurécom, Sophia-Antipolis, France. He has served as a consultant to Hewlett-Packard, Honeywell Inc., the Applied Physics Labs Seattle, the Research Triangle Institute, Green Mountain Geophysics, Ball Aerospace Corporation, and Siemens.

Dr. Scharf is a past member of the ASSP ADCOM, and he currently serves on the Editorial Boards of *Signal Processing* and the *Journal of Statistical Planning and Inference*. He is a past Associate Editor for the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. He was the Technical Program Chairman for the 1980 International Conference on ASSP. He is a member of Eta Kappu Nu.

Benjamin Friedlander (S'74–M'76–SM'82–F'87) received the B.Sc. and the M.Sc. degrees in electrical engineering from the Technion-Israel Institute of Technology in 1968 and 1972, respectively, and the Ph.D. degree in electrical engineering and the M.Sc. degree in statistics from Stanford University in 1976.

From 1976 to 1985 he was at Systems Control Technology, Inc., Palo Alto, CA. From November 1985 to July 1988, he was with Saxpy Computer Corporation, Sunnyvale, CA. Currently he is with the University of California at Davis and Signal Processing Technology, Ltd. in Palo Alto.

Dr. Friedlander is the recipient of the 1983 ASSP Senior Award, the 1985 Award for Best Paper of the Year from the European Association for Signal Processing (EURASIP), and the 1989 Technical Achievement Award of the Signal Processing Society.