

# Constrained Hypothesis Testing and the Cramér-Rao Bound

Terrence J. Moore, Brian M. Sadler, *Fellow, IEEE*

**Abstract**— The classical Wald and Rao test statistics are asymptotically equivalent to the generalized likelihood ratio test statistics, while not requiring parameter estimation under both hypotheses, and so they provide lower complexity test statistics. In this paper we develop corresponding variations of the Wald and Rao test for nested hypothesis testing under parameter constraints. The resulting tests incorporate the constrained Cramér-Rao bound formulation from Stoica and Ng, and unify some asymptotic hypothesis testing results. Examples will illustrate key ideas and test performance.

**Index Terms**—Hypothesis testing, constrained Cramér-Rao bound, asymptotic analysis.

## I. INTRODUCTION

Gorman and Hero [1] developed a novel approach to the analysis of estimator performance when side information exists for the parameters. They showed that unbiased estimators for parameters that satisfy continuously differentiable constraints have a variance (or mean-square error) lower bounded by the constrained Cramér-Rao bound (CCRB), which depends only on the Fisher information matrix (FIM) for the original unconstrained model and the Jacobian of the constraint function. Stoica and Ng [2] extended this result<sup>1</sup> for the more general case when the FIM is singular by using the complement of the constraint's Jacobian, that, in general, corresponds to scenarios where the parameters in the original model are not identifiable [4].

Interestingly, these results were implicitly shown in the historical literature when considering the asymptotic normality of the constrained maximum likelihood estimate (CMLE) as well as for constrained hypothesis testing [5], [6], [7], [8]. This prior work typically developed asymptotic variance formulas for constrained parameters that in retrospect we recognize to contain the Gorman-Hero version of the CCRB. The asymptotic normality and convergence of the CMLE to the corresponding CCRB are shown in [9], generalizing classic maximum likelihood estimate (MLE) results, and incorporating the Stoica and Ng CCRB. In this paper, we generalize the Rao and Wald hypothesis tests under constraints, incorporating the CCRB. These provide simpler alternatives to the generalized likelihood ratio test (GLRT), and are asymptotically equivalent.

## II. PRELIMINARIES

We observe samples  $x_1, x_2, \dots, x_n$  that are i.i.d. with pdf  $p(x; \theta)$ , where  $\theta$  is a vector of unknown parameters in

B. M. Sadler and T. J. Moore are with the Army Research Laboratory, Adelphi, MD 20783 USA (e-mail: brian.sadler;terrence.moore@us.army.mil).

<sup>1</sup>Technically, the result of Stoica and Ng was derived for constrained estimators as opposed to constrained parameters. However, either assumption results in the same CCRB as was shown in [3].

the open set  $\Theta \subset \mathbb{R}^m$ . Let  $y_n = (x_1, x_2, \dots, x_n)$  be the collection of these samples so that we have the likelihood

$$l(y_n; \theta) = \prod_{i=1}^n p(x_i; \theta). \quad (1)$$

The Fisher information matrix (FIM) for  $p(x; \theta)$  is given by

$$I(\theta) = E_{\theta}\{s^T(\theta)s(\theta)\} \quad (2)$$

with the expectation taken with respect to the pdf, i.e.,  $E_{\theta}(\cdot) = \int_{\Omega} (\cdot)p(x; \theta)dx$ , and the Fisher score is given by

$$s(\theta) = \frac{\partial \log p(x; \theta)}{\partial \theta}. \quad (3)$$

The FIM and score depend on the model for  $x$ , but this notation is omitted for conciseness. The variance of any unbiased estimator of the parameters is lower bounded by the Cramér-Rao bound (CRB), defined as

$$\text{CRB}(\theta) = I^{-1}(\theta). \quad (4)$$

Note that for the likelihood function (1), we have  $I_{y_n}(\theta) = nI(\theta)$  and  $\text{CRB}_{y_n}(\theta) = \frac{1}{n}\text{CRB}(\theta)$  since  $s_{y_n}(\theta) = \frac{\partial l(y_n; \theta)}{\partial \theta}$ . These results can be found in many texts on statistical inference, e.g., [10].

### A. The Constrained Cramér-Rao Bound

It has been shown in [2], [3] that under parametric equality constraints

$$f(\theta) = \mathbf{0} \quad (5)$$

where  $f: \Theta \rightarrow \mathbb{R}^k$  is a nonredundant continuously differentiable function, then the variance of any unbiased estimator of  $\theta$  in the model for  $x$  subject to the constraints is lower bounded by the constrained Cramér-Rao bound (CCRB), defined as

$$\text{CCRB}(\theta) = U(\theta) (U^T(\theta)I(\theta)U(\theta))^{-1} U^T(\theta), \quad (6)$$

where the columns of  $U(\theta)$  form a linearly independent complement of the Jacobian of the constraints  $F(\theta) = \frac{\partial f(\theta)}{\partial \theta^T}$ , i.e.,

$$F(\theta)U(\theta) = \mathbf{0} \quad \text{and} \quad \text{rank}(U(\theta)) = m - \text{rank}(F(\theta)). \quad (7)$$

Similarly, for the likelihood function in (1), the CCRB is  $nU(\theta) (U^T(\theta)I(\theta)U(\theta))^{-1} U^T(\theta)$ . We shall define the constraint set to be  $\Theta_f = \{\theta \in \Theta : f(\theta) = \mathbf{0}\}$ .

## B. Hypothesis Testing

Consider the testing problem

$$H_0 : \mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}, \quad (8)$$

where  $\mathbf{h} : \Theta \rightarrow \mathbb{R}^r$  is a consistent, nonredundant, continuously differentiable function, and denote the Jacobian of this null hypothesis as

$$\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial \mathbf{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}. \quad (9)$$

A common statistic considered for this problem is the generalized likelihood ratio test (GLRT)

$$\lambda_n = \frac{l(\mathbf{y}_n; \hat{\boldsymbol{\theta}}_h)}{l(\mathbf{y}_n; \hat{\boldsymbol{\theta}})}, \quad (10)$$

where  $\hat{\boldsymbol{\theta}}_h$  is the maximum likelihood estimate (MLE) under the null hypothesis (i.e., subject to the constraint  $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$ ) and  $\hat{\boldsymbol{\theta}}$  is the maximum likelihood (ML) solution under the alternative hypothesis  $H_1 : \mathbf{h}(\boldsymbol{\theta}) \neq \mathbf{0}$ . Asymptotically, as  $n$  approaches  $\infty$ ,  $-2 \log \lambda_n$  is  $\chi_r^2$  in distribution (see [10] for more details).

The Wald and Rao (score) tests are asymptotically equivalent to the GLRT, yet these test statistics only require a single ML solution either under  $H_0 : \boldsymbol{\theta} \in \Theta$  or  $H_1 : \boldsymbol{\theta} \in \Theta_h$ , respectively, as opposed to requiring both. The Wald test is given by

$$\omega_n = n \mathbf{h}^T(\hat{\boldsymbol{\theta}}) [\mathbf{H}(\hat{\boldsymbol{\theta}}) \text{CRB}(\hat{\boldsymbol{\theta}}) \mathbf{H}^T(\hat{\boldsymbol{\theta}})]^{-1} \mathbf{h}(\hat{\boldsymbol{\theta}}), \quad (11)$$

and only requires the solution of the MLE  $\hat{\boldsymbol{\theta}}$ . Alternatively, the Rao score test is given by

$$\rho_n = \mathbf{s}^T(\hat{\boldsymbol{\theta}}_h) \text{CRB}(\hat{\boldsymbol{\theta}}_h) \mathbf{s}(\hat{\boldsymbol{\theta}}_h) / n, \quad (12)$$

and only requires the solution of the CMLE  $\hat{\boldsymbol{\theta}}_h$ , i.e., the MLE under  $\boldsymbol{\theta} \in \Theta_h$ . Again, further details of these results can be found, e.g., in [10].

## III. TESTING UNDER CONSTRAINTS

Now consider the scenario where a testing problem has a constrained alternative, i.e., under  $H_1$  the parameters are known to satisfy the constraint  $\mathbf{f}(\boldsymbol{\theta}) = \mathbf{0}$  as described in (5). It is assumed, additionally, that  $\mathbf{h}$  is consistent and nonredundant with respect to  $\mathbf{f}$ , i.e.,  $\Theta_h \cap \Theta_f = \emptyset$  and

$$\text{rank} \begin{pmatrix} \mathbf{H}(\boldsymbol{\theta}) \\ \mathbf{F}(\boldsymbol{\theta}) \end{pmatrix} = r + k \leq m, \quad (13)$$

respectively. This hypothesis testing problem can also be stated as

$$H_0 : \boldsymbol{\theta} \in \Theta_h \text{ vs. } H_1 : \boldsymbol{\theta} \in \Theta_f, \quad (14)$$

where now  $\Theta_h = \{\boldsymbol{\theta} \in \Theta_f : \mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}\}$ .<sup>2</sup>

The conditions for the implicit constraint  $\mathbf{f}$  defined in section II-A imply the existence of an explicit constraint, at least locally. Hence, instead of requiring the parameters to be zeros of a particular function  $\mathbf{f}$ , we can equivalently require

<sup>2</sup>In general, the alternative hypothesis in (14) is stated as  $H_1 : \boldsymbol{\theta} \in \Theta_f \setminus \Theta_h$ . However, by the definition of  $\mathbf{h}$  and provided  $r \geq 1$ , then  $\Theta_h$  is a set of (Lebesgue) measure zero in  $\Theta_f$  and therefore the hypotheses is equivalent.

the parameters to be solutions of a function  $\mathbf{g}$  of a reduced set of parameters. This equivalence follows from the Implicit Function Theorem (IFT), see, e.g., [3] and [11, theorems 5-1 and 5-2]. The difficulty of this step is finding an explicit reparameterization  $\mathbf{g}$ , however, it shall be shown that our results do not require this function directly. Instead, the results rely on the Jacobian of  $\mathbf{g}$ , defined as

$$\mathbf{G}(\boldsymbol{\xi}) = \frac{\partial \mathbf{g}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}^T}. \quad (15)$$

It has been shown in [3] that the Jacobian for any reparameterization under the conditions for the implicit constraint in section II-A corresponds to a linearly independent complement of  $\mathbf{F}(\boldsymbol{\theta})$  satisfying (7). Mathematically, the IFT states that there exist open sets  $\mathbb{O} \subset \Theta_f$  and  $\mathbb{P} \subset \mathbb{R}^{m-k}$  with  $\boldsymbol{\theta} \in \mathbb{O}$  and there exists a diffeomorphism  $\mathbf{g} : \mathbb{P} \rightarrow \mathbb{O}$  such that  $\boldsymbol{\theta} = \mathbf{g}(\boldsymbol{\xi})$  for some  $\boldsymbol{\xi} \in \mathbb{P}$ . This reparameterization  $\mathbf{g}$  depends on  $\boldsymbol{\theta}$  (since the diffeomorphism only holds locally about  $\boldsymbol{\theta}$ ) but this dependence is omitted here for the sake of conciseness.

Hence, for any  $\boldsymbol{\theta} \in \Theta_f$ , an asymptotically equivalent testing problem to (14) can be written in terms of the parameter as an explicit constraint, e.g.,

$$H_0 : \mathbf{h}(\mathbf{g}(\boldsymbol{\xi})) = \mathbf{0} \text{ vs. } H_1 : \mathbf{h}(\mathbf{g}(\boldsymbol{\xi})) \neq \mathbf{0}, \quad (16)$$

for which the Wald and Rao statistics are well-known (with respect to  $\boldsymbol{\xi}$ ).

### A. The general likelihood-ratio test statistic

The likelihood ratio statistic for (14) is simply

$$\lambda_n = \frac{l(\mathbf{y}_n; \hat{\boldsymbol{\theta}}_h)}{l(\mathbf{y}_n; \hat{\boldsymbol{\theta}}_f)}, \quad (17)$$

where  $\hat{\boldsymbol{\theta}}_h$  and  $\hat{\boldsymbol{\theta}}_f$  are the maximum likelihood solutions in  $\Theta_h$  and  $\Theta_f$ , respectively. This test is somewhat common and so will not be addressed further here.

### B. The generalized Wald test statistic

The Wald test statistic for (16) is given by

$$n \mathbf{h}^T(\mathbf{g}(\hat{\boldsymbol{\xi}})) [(\mathbf{H} \circ \mathbf{G})(\hat{\boldsymbol{\xi}}) \tilde{\mathbf{I}}^{-1}(\hat{\boldsymbol{\xi}}) (\mathbf{H} \circ \mathbf{G})^T(\hat{\boldsymbol{\xi}})]^{-1} \mathbf{h}(\mathbf{g}(\hat{\boldsymbol{\xi}})), \quad (18)$$

where  $(\mathbf{H} \circ \mathbf{G})(\hat{\boldsymbol{\xi}}) = \mathbf{H}(\mathbf{g}(\hat{\boldsymbol{\xi}})) \mathbf{G}(\hat{\boldsymbol{\xi}})$  and the Fisher information  $\tilde{\mathbf{I}}(\boldsymbol{\theta})$  is derived from the score  $\tilde{\mathbf{s}}(\boldsymbol{\theta})$  which is itself derived from the alternative likelihood  $\tilde{p}(\mathbf{x}; \boldsymbol{\xi}) = p(\mathbf{x}; \mathbf{g}(\boldsymbol{\xi}))$ . This expression is not particularly useful, as it requires knowledge of an explicit formulation of the constraint. To transform the formula into an expression relevant to the original parameter space, note from [3] that  $\mathbf{G}(\boldsymbol{\xi}) \tilde{\mathbf{I}}^{-1}(\boldsymbol{\xi}) \mathbf{G}(\boldsymbol{\xi}) = \text{CCRB}(\boldsymbol{\theta})$  and  $\mathbf{g}(\hat{\boldsymbol{\xi}}) = \hat{\boldsymbol{\theta}}_f$  (for large  $n$  certainly). Substitution into (18) produces

$$\omega_n = n \mathbf{h}^T(\hat{\boldsymbol{\theta}}_f) [\mathbf{H}(\hat{\boldsymbol{\theta}}_f) \text{CCRB}(\hat{\boldsymbol{\theta}}_f) \mathbf{H}^T(\hat{\boldsymbol{\theta}}_f)]^{-1} \mathbf{h}(\hat{\boldsymbol{\theta}}_f). \quad (19)$$

This statistic incorporates the CCRB formula and the CMLE (subject to  $\mathbf{f}$ ) into the usual Wald statistic. This is a generalization of the classical result presented by Aitchison [7, ‘ $\lambda_{21}(\boldsymbol{\theta})$ ’ on p. 240], which uses the Gorman-Hero variant of the CCRB

[1] that assumes a nonsingular FIM.<sup>3</sup> For the statistic in (19) to exist, it is necessary that  $\mathbf{H}(\hat{\boldsymbol{\theta}}_f)\text{CCRB}(\hat{\boldsymbol{\theta}}_f)\mathbf{H}^T(\hat{\boldsymbol{\theta}}_f)$  is nonsingular. However, this is the usual requirement for valid testing, i.e., that the function to be tested is identifiable.<sup>4</sup>

Moreover, the statistic in (19) preserves the asymptotic characteristics of the usual Wald statistic. This can be assumed from the derivation or shown directly: Using a first order Taylor series approximation of  $\mathbf{h}$  about  $\boldsymbol{\theta}$  evaluated at  $\hat{\boldsymbol{\theta}}$ , and the fact that  $\sqrt{n}(\hat{\boldsymbol{\theta}}_f - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \text{CCRB}(\boldsymbol{\theta}))$  (which is shown, for example, in [9] and [5]), then  $\mathbf{h}(\hat{\boldsymbol{\theta}}_f) - \mathbf{h}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_f - \boldsymbol{\theta}) + o(1)$  and  $\sqrt{n}(\mathbf{h}(\hat{\boldsymbol{\theta}}_f) - \mathbf{h}(\boldsymbol{\theta}))$  goes to  $\mathcal{N}(\mathbf{0}, \mathbf{H}(\boldsymbol{\theta})\text{CCRB}(\boldsymbol{\theta})\mathbf{H}^T(\boldsymbol{\theta}))$  in distribution as  $n \rightarrow \infty$ . Thus,  $\sqrt{n}(\mathbf{H}(\boldsymbol{\theta})\text{CCRB}(\boldsymbol{\theta})\mathbf{H}^T(\boldsymbol{\theta}))^{-1/2}(\mathbf{h}(\hat{\boldsymbol{\theta}}_f) - \mathbf{h}(\boldsymbol{\theta}))$  is asymptotically an  $r$ -dimensional standard normal distribution, and  $\omega_n \rightarrow \chi_r^2$  in distribution.

### C. The generalized Rao test statistic

The Rao test statistic for the problem in (16) is given by

$$\tilde{s}^T(\hat{\boldsymbol{\xi}}_{h \circ g})\tilde{\text{CRB}}(\hat{\boldsymbol{\xi}}_{h \circ g})\tilde{s}(\hat{\boldsymbol{\xi}}_{h \circ g})/n, \quad (20)$$

where the score is taken with respect to the  $\boldsymbol{\xi}$  parameter, in particular the score is  $\tilde{s}(\boldsymbol{\xi}) = \mathbf{G}(\boldsymbol{\xi})\mathbf{s}(\boldsymbol{\theta})$ , and it is evaluated at the CMLE with respect to  $\mathbf{h}$ . Note for sufficiently large  $n$ , the invariance property of the MLE holds, i.e.,  $\hat{\boldsymbol{\theta}}_{f, \mathbf{h}} = \mathbf{g}(\hat{\boldsymbol{\xi}}_{\mathbf{h}})$ , which justifies this reparameterization approach for the CMLE (with respect to  $\mathbf{f}$  and  $\mathbf{h}$ ). Hence, substituting into (20) the formula for the score in terms of  $\boldsymbol{\theta}$  and noting, again from [3], that  $\mathbf{G}^T(\boldsymbol{\xi})\tilde{\text{CRB}}(\boldsymbol{\xi})\mathbf{G}(\boldsymbol{\xi}) = \text{CCRB}(\boldsymbol{\theta})$  produces

$$\rho_n = \mathbf{s}^T(\hat{\boldsymbol{\theta}}_{h, \mathbf{f}})\text{CCRB}(\hat{\boldsymbol{\theta}}_{h, \mathbf{f}})\mathbf{s}(\hat{\boldsymbol{\theta}}_{h, \mathbf{f}})/n. \quad (21)$$

As in the general Wald test statistic in (19), this statistic incorporates the CCRB formula and the CMLE (subject to both  $\mathbf{f}$  and  $\mathbf{h}$ ) into the usual Rao score statistic.<sup>5</sup>

## IV. EXAMPLES

**Example 1: Detection of a subspace signal (linear model) subject to linear constraints in white Gaussian noise of known level.** Assume the parameters in the model  $\mathbf{x} = \mathbf{D}\boldsymbol{\theta} + \mathbf{n}$  satisfy the linear constraint  $\mathbf{F}\boldsymbol{\theta} + \mathbf{v} = \mathbf{0}$ , where  $\mathbf{D} \in \mathbb{R}^{d \times m}$  is a known observation matrix,  $\mathbf{n}$  is a white Gaussian vector with known covariance matrix  $\sigma^2\mathbf{I}_{d \times d}$ , and  $\mathbf{F}$  has full row rank  $k$ . For this model, the score is  $\mathbf{s}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\mathbf{D}^T(\mathbf{x} - \mathbf{D}\boldsymbol{\theta})$  and therefore the FIM is  $\mathbf{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\mathbf{D}^T\mathbf{D}$ . Consider the testing problem:

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \quad (23)$$

<sup>3</sup>Aitchison also discusses the more general scenario when the FIM is singular in [7, section 3.9]. Equivalently, this formula can be found in [5].

<sup>4</sup>Conditions for which identifiability holds under a constraint can be found in [4].

<sup>5</sup>There exists a corresponding Lagrange multiplier variant (originally developed by Silvey [6]) of this generalized Rao statistic given by

$$\hat{\boldsymbol{\lambda}}_{h, \mathbf{f}}^T \mathbf{H}(\hat{\boldsymbol{\theta}}_{h, \mathbf{f}})\text{CCRB}(\hat{\boldsymbol{\theta}}_{h, \mathbf{f}})\mathbf{H}^T(\hat{\boldsymbol{\theta}}_{h, \mathbf{f}})\hat{\boldsymbol{\lambda}}_{h, \mathbf{f}}/n, \quad (22)$$

where the Lagrange multiplier (LM) estimates are based on the first order conditions relating to the constraints  $\mathbf{h}$ . This result is consistent with a variant of this LM test [6, if one takes the Schur complement of equation (6.5)] that uses an alternative variant of the CCRB. The equivalence of the formula was not shown in Silvey, but can be shown using a general matrix identity in Khatri [12].

under the linear constraint. The formulation of the null hypothesis does not satisfy (13) and the matrix in (19) will be singular, and hence the pseudo-inverse may be considered, i.e.,

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \left[ \mathbf{U} (\mathbf{U}^T \mathbf{I}(\hat{\boldsymbol{\theta}}) \mathbf{U})^{-1} \mathbf{U}^T \right]^\dagger (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \quad (24)$$

since  $\mathbf{H}(\boldsymbol{\theta}) = \mathbf{I}_{m \times m}$ . The pseudo-inverse formula for a Gramian matrix is  $(\mathbf{A}\mathbf{A}^T)^\dagger = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-2}\mathbf{A}^T$  if  $\mathbf{A}$  is full column rank [13]. The resulting general Wald test is

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \mathbf{U} \mathbf{U}^T \mathbf{I}(\hat{\boldsymbol{\theta}}) \mathbf{U} \mathbf{U}^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \quad (25)$$

using  $\mathbf{A} = \mathbf{U} (\mathbf{U}^T \mathbf{I}(\hat{\boldsymbol{\theta}}) \mathbf{U})^{-1/2}$  and where  $\mathbf{U}$  satisfies (7). This is equivalent to the reparameterized approach: Given  $\boldsymbol{\theta} = \mathbf{g}(\boldsymbol{\xi}) = \mathbf{U}\boldsymbol{\xi} - \mathbf{F}^T(\mathbf{F}\mathbf{F}^T)^{-1}\mathbf{v}$  then the classical Wald test formula in (11) for this problem is  $(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0)^T \tilde{\mathbf{I}}(\hat{\boldsymbol{\xi}})(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0)$  where, as shown in [3],  $\tilde{\mathbf{I}}(\boldsymbol{\xi}) = \mathbf{U}^T \mathbf{I}(\boldsymbol{\theta}) \mathbf{U}$  and  $\mathbf{U}^T \boldsymbol{\theta} = \boldsymbol{\xi}$ . Note that even though the Jacobian of  $\mathbf{h}(\boldsymbol{\theta}) = \boldsymbol{\theta} - \boldsymbol{\theta}_0$  has full rank  $m$ , the statistic is  $\chi_{m-k}^2$  where  $k$  is the rank of the Jacobian of  $\mathbf{f}(\boldsymbol{\theta}) = \mathbf{F}\boldsymbol{\theta} + \mathbf{v}$  by the assumption of nonredundancy.

We can show this Wald test is equivalent to the corresponding Rao test for (23). Substituting the FIM and the MLE  $\hat{\boldsymbol{\theta}} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{x}$  in , then we have

$$(\mathbf{x} - \mathbf{D}\boldsymbol{\theta}_0)^T \mathbf{D} (\mathbf{D}^T \mathbf{D})^{-1/2} \times \left[ (\mathbf{D}^T \mathbf{D})^{1/2} \mathbf{U} (\mathbf{U}^T \mathbf{D}^T \mathbf{D} \mathbf{U})^{-1} \mathbf{U}^T (\mathbf{D}^T \mathbf{D})^{1/2} \right]^\dagger \times (\mathbf{D}^T \mathbf{D})^{-1/2} \mathbf{D}^T (\mathbf{x} - \mathbf{D}\boldsymbol{\theta}_0), \quad (26)$$

where we have pulled  $(\mathbf{D}^T \mathbf{D})^{-1/2}$  into the pseudo-inverse with the CCRB as this makes the resulting matrix inside idempotent and its own pseudo-inverse. Hence, we are left with

$$(\mathbf{x} - \mathbf{D}\boldsymbol{\theta}_0)^T \mathbf{D} \mathbf{U} (\mathbf{U}^T \mathbf{D}^T \mathbf{D} \mathbf{U})^{-1} \mathbf{U}^T \mathbf{D}^T (\mathbf{x} - \mathbf{D}\boldsymbol{\theta}_0), \quad (27)$$

which using the score and CCRB formula is

$$\mathbf{s}^T(\boldsymbol{\theta}_0)\text{CCRB}(\boldsymbol{\theta}_0)\mathbf{s}(\boldsymbol{\theta}_0), \quad (28)$$

which is exactly the Rao test statistic for (23). This is an expected result since the Wald and Rao agree for a linear test on a linear model and a linear constraint on a linear model is simply another linear model.

**Example 2: Detection of a signal over an unknown convolutive channel subject to a normed channel constraint and real-valued signaling.** We consider a single-input multiple-output finite-impulse response model, which is typically framed within a blind channel identification context. This particular model is given as

$$\mathbf{x} = \mathbf{K}\mathbf{s} + \mathbf{n} \quad (29)$$

where  $\mathbf{K}$  is a 2-sensor, 1-source convolutive channel matrix of order 3 generated from the example in [2],  $\mathbf{s} \in \mathbb{C}^{50 \times 1}$  includes the data symbols using binary phase shift keying symbols generated from a Binomial(1,.5) distribution, and the noise is white Gaussian with known variance  $\sigma^2\mathbf{I}_{47 \times 47}$ . We assume a norm constraint on the channel elements, i.e.,  $\|\text{vec}(\mathbf{K})\|^2 = 1$ , and additionally assume the signal elements are real-valued, i.e.,  $\text{imag}(\mathbf{s}) = \mathbf{0}$ . The MLE estimate for the Wald test is the

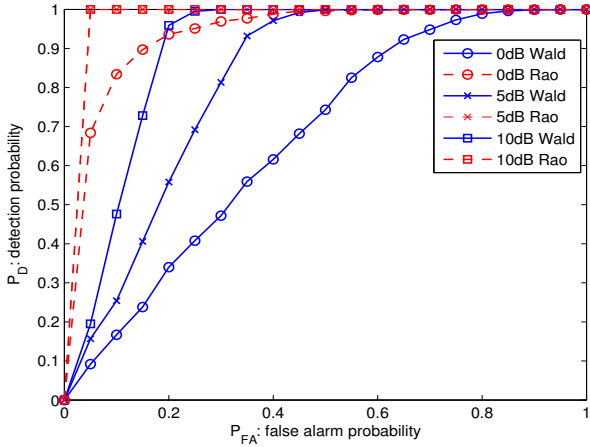


Fig. 1. ROC curves for detection of a signal in a blind channel identification context in Example 2 over 1000 trials.

subspace method in [14] with a smoothing factor of eight. We wish to test

$$H_0 : s = \mathbf{0} \text{ vs. } H_1 : s \neq \mathbf{0} \quad (30)$$

under the channel-norm and real-valued signal constraint. Since the test is redundant with respect to the constraints, either the test can be reformulated as  $\text{real}(s) = \mathbf{0}$  or the pseudoinverse can be used in (19). A more interesting complication is that under  $H_0$  there exists no meaningful estimate of the channel elements, which are unobservable under  $H_0$ . Hence any estimate for the channel is a CMLE, so we use the subspace method. Figure 1 displays simulation results for this example conducted under a range of signal-to-noise ratios (SNR) for both the Wald and Rao tests. For this particular scenario the Rao statistic outperforms the Wald statistic, and as the SNR grows the Rao exhibits perfect classification. It is important to note the scenario is of a single model sample or observation ( $n = 1$ ) so a  $\chi^2$  assumption would be optimistic.

**Example 3: Detection of unit modulus signaling over an unknown channel for multiple users subject to training or pilot symbols.** This model is a 2-channel, 2-source multiple-input multiple-output (MIMO) generalization of (29), i.e.,  $\mathbf{x} = \mathbf{H}^{(1)}\mathbf{s}^{(1)} + \mathbf{H}^{(2)}\mathbf{s}^{(2)} + \mathbf{n}$  except with zero order channels. We use a mixing model with 3-ray multipath subchannels expressed as a weighted sum of steering vectors with angle-of-arrivals and corresponding amplitudes being  $\{-1, 0, 4\}$  and  $\{\sqrt{0.2}\angle(-\frac{\pi}{6}), \sqrt{0.5}, \sqrt{0.15}\angle(-\frac{\pi}{5})\}$  for source  $\mathbf{s}^{(1)}$  (at SNR 20dB) and  $\{0, 5, 11\}$  and  $\{\sqrt{0.15}\angle(-\frac{\pi}{5}), \sqrt{0.6}, \sqrt{0.25}\angle(\frac{\pi}{3})\}$  for source  $\mathbf{s}^{(2)}$  (at SNR 15dB). The SNRs are accomplished by scaling with respect to the channel. Given an observation of 30 time snapshots for each source, we assume knowledge of the first two data elements (i.e., training) per source and test the remaining data elements for unit (constant) modulus signaling, i.e.,

$$H_0 : |s_j^{(i)}|^2 = 1 \quad \forall i, n \text{ vs. } H_1 : |s_j^{(i)}|^2 \neq 1 \text{ for some } i, n. \quad (31)$$

For the simulation, under  $H_1$ , we set the  $\mathbf{s}^{(1)}$  elements to have unit modulus and the  $\mathbf{s}^{(2)}$  elements to have modulus  $\frac{1}{4}$ . An initial estimate of  $\mathbf{H}$  is obtained via least-squares using the training symbols, the initialization of unknown elements of  $\mathbf{s}$

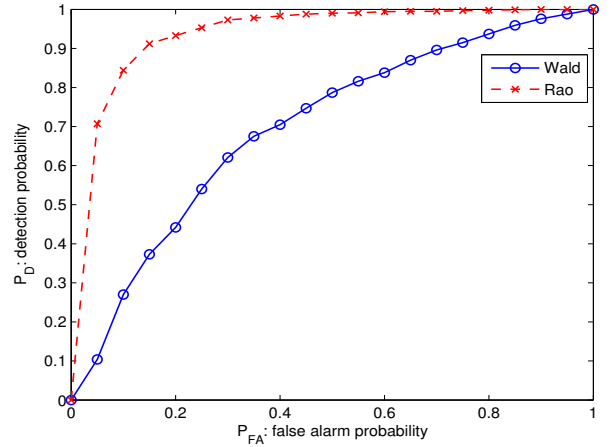


Fig. 2. ROC curves for detection of unit modulus signaling in the MIMO mixing model of Example 3 over 1000 trials.

are from  $\hat{\mathbf{H}}$  to find the best fit for the MLE. Under  $H_0$ , we use ZF-ACMA [15] and use the training to find the best fit for the CMLE. Figure 2 displays the simulation results. This is a difficult test as under  $H_1$  both sources each have constant modulus, but the Rao statistic still well detects the moduli are different.

## REFERENCES

- [1] J. D. Gorman, A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Transactions on Information Theory*, vol. 26, no. 6, pp. 1285-1301, November 1990.
- [2] P. Stoica, B. C. Ng, "On the Cramér-Rao bound under parametric constraints," *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 177-179, July 1998.
- [3] T. J. Moore, R. J. Kozick, B. M. Sadler, "The constrained Cramér-Rao bound from the perspective of fitting a model," *IEEE Signal Processing Letters*, vol. 14, no. 8, pp. 564-567, August 2007.
- [4] T. J. Rothenberg, "Identification in parametric models," *Econometrica*, vol. 39, no. 3, pp. 577-591, May 1971.
- [5] M. Crowder, "On constrained maximum likelihood estimation with non-i.i.d. observations," *Annals of the Institute of Statistical Mathematics*, vol. 36, no. A, pp. 239-249, 1984.
- [6] S. D. Silvey, "The Lagrange multiplier test," *Annals of Mathematical Statistics*, vol. 30, pp. 389-407, 1959.
- [7] J. Aitchison, "Large sample restricted parametric tests," *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 24, no. 1, pp. 234-250, 1962.
- [8] C. Gourieroux, A. Monfort, *Statistics and Econometric Models: Volume 1, General Concepts, Estimation, Prediction and Algorithms*, Cambridge University Press, 1995.
- [9] T. J. Moore, B. M. Sadler, R. J. Kozick, "Maximum-likelihood estimation, the Cramér-Rao bound, and the method of scoring with parameter constraints," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 895-908, March 2008.
- [10] J. Shao, *Mathematical Statistics*. New York, NY: Springer-Verlag, 2003.
- [11] M. Spivak, *Calculus on Manifolds*. Reading, MA: Addison Wesley, 1965.
- [12] C. G. Khatri, "A note on a MANOVA model applied to problems in growth curve," *Annals of the Institute of Statistical Mathematics*, vol. 18, pp. 75-86, 1966.
- [13] A. Ben-Israel, T. N. E. Greville, *Generalized Inverses: Theory and Applications*, New York: Springer-Verlag, 2003.
- [14] E. Moulines, P. Duhamel, J. F. Cardoso, S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 516-525, February 1995.
- [15] A. van der Veen, "Asymptotic properties of the algebraic constant modulus algorithm," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1796-1807, August 2001.