Cramer-Rao Lower Bounds for Complex Parameters*

Venkatesh Nagesha† and Steven Kay†
Dept. of Electrical Engineering
University of Rhode Island
Kingston, RI 02881

Abstract

A tutorial introduction to Cramer-Rao lower bounds for complex-valued parameters is presented. The vector parameterizing the probability density function of the data is assumed to contain some complex-valued and some real-valued parameters. The traditional approach in such problems is to form a real-valued parameter vector by using the real and imaginary parts of the complex-valued parameters. The resulting algebra is often tedious and clumsy. We present a *direct* approach to this problem which, we believe, leads to elegant algebraic manipulations. Application to a simple but common signal processing problem is included to illustrate the calculations.

1 Introduction

The Cramer-Rao lower bound (CRLB) serves as a fundamental tool in estimation theory by providing a means to study the limits of estimator performance [5, 10]. It may be used (i) to produce the minimum variance unbiased (MVU) estimator, (ii) for comparing performances of different estimators, (iii) to describe the asymptotic performance of maximum likelihood estimators (MLE), (iv) to analyze the effect of different model parameters on estimation

*EDICS Category 5.2

†This work was supported by the Air Force Office of Scientific Research under contract AFOSR-91-0024

accuracy, (v) as an instrument in designing signal processing systems and algorithms. In addition, as opposed to other lower bounds such as the Bhattacharya or Barankin bounds [10], the CRLB is usually simpler to compute. In many cases, closed form expressions for the CRLB can be derived, from which many useful interpretations can be made.

In elementary derivations of the CRLB, one assumes that the parameter vector $\underline{\boldsymbol{\theta}} \in \mathbb{R}^p$ and that the parameterized probability density function (PDF) $p(\mathbf{x}\,;\,\underline{\boldsymbol{\theta}})$ obeys certain "regularity" conditions. However, in many signal processing applications, the PDF is parameterized by some complex-valued and real-valued parameters, or $\underline{\boldsymbol{\theta}} \in \mathbb{C}^{p_1} \times \mathbb{R}^{p_2}$. The customary approach is to form a real-valued vector of parameters by concatenating the real and imaginary parts and then applying the usual theory. Although straightforward, from an algebraic point of view, this is very tedious and cumbersome. It is better to work directly with complex quantities and exploit the associated algebra. The resulting expressions are considerably more intuitive and have the *same form* as corresponding real-valued cases. Such an approach is commonplace in other similar situations, viz. minimization of Hermitian forms, certain complex Gaussian PDFs for complex random vectors, Fourier series and transforms, etc. [2, 5, 6]. Although extensions of most principles of statistical inference for complex data and/or parameters have been attempted before [5, 6], the concept of CRLB has not be adequately addressed. Some attempts, but incorrect, can be found in [11]. (The notion of CRLB has been generalized along other lines, for example, real separable Banach spaces [4], closed subsets in $\mathbb{R}^p$ [3].) This paper represents a step in this direction. We emphasize that *no new theory* is presented, only an algebra to *efficiently manipulate* complex-valued quantities is explored.

An examination of a simple derivation of the CRLB (for example, see [5, Chap. 3]) reveals that such an extension requires (i) some "regularity" conditions on the PDF $p(\mathbf{x}\,;\,\underline{\boldsymbol{\theta}})$, ($ii$) some sort of Cauchy-Schwartz inequality, and (iii) a notion of derivative with respect to (w.r.t.) the unknown parameters. While the first two requirements are met fairly easily, the third one is a little complicated. This is because the PDF $p(\mathbf{x}\,;\,\underline{\boldsymbol{\theta}})$, being a real-valued function, is *not* an analytic function of the complex-valued parameters (in the sense that it does not satisfy the Cauchy-Riemann conditions). Fortunately, we can define a limiting operator that *behaves like* a derivative [1] and this is quite adequate for our purpose.

The paper begins with an introduction to defining a "derivative" of non-analytic functions of a complex variable such as real-valued functions. The purpose is to only motivate the plausibility of such a definition, other details in this regard can be found in [1]. The next section presents a theorem that states the Cramer-Rao inequality and the existence of efficient estimators for complex parameters. Proof of the theorem can be found in Appendix A. The final section illustrates the use of the theorems in computing the CRLB for a simple parameter estimation problem that arises in many signal processing applications. It represents a significant generalization of the example considered in [9, 11]. An expression for the elements of the Fisher information matrix (FIM) in the general complex Gaussian case with information in the mean and covariance is included. A word about the notation used is in order. We will use superscripts T to denote transpose, $*$ to denote complex conjugation, H to denote Hermitian or complex conjugate-transpose, and subscripts r and i to indicate real and imaginary parts, respectively. For a matrix $\mathbf{T}$, $[\mathrm{T}]_{k,l}$ and $[\mathrm{T}]^{k,l}$ will denote the $k\,l$-th element of $\mathbf{T}$ and T^{-1}, respectively.

NOTATIONS AND TERMINOLOGY USED:

$\underline{\boldsymbol{\theta}} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix}$: parameter vector to be estimated. Contains p_1 complex-valued $(\boldsymbol{\theta}_1)$ and p_2 real-valued $(\boldsymbol{\theta}_2)$ variables.

$\boldsymbol{\theta}^{(r)} = \begin{bmatrix} \boldsymbol{\theta}_{1r} \\ \boldsymbol{\theta}_{1i} \\ \boldsymbol{\theta}_2 \end{bmatrix}$: real-valued parameter vector formed from $\underline{\boldsymbol{\theta}}$ by concatenating real and imaginary parts of $\boldsymbol{\theta}_1$. Used in conventional approach to CRLBs.

$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_1^* \\ \boldsymbol{\theta}_2 \end{bmatrix}$: complex-valued parameter vector formed from $\underline{\boldsymbol{\theta}}$. Used in proposed approach to CRLBs.

$\mathrm{I}(\boldsymbol{\theta}^{(r)})$: real Fisher information matrix, a $(2p_1 + p_2) \times (2p_1 + p_2)$ real symmetric matrix

$\mathrm{I}(\boldsymbol{\theta})$: complex Fisher information matrix, a $(2p_1 + p_2) \times (2p_1 + p_2)$ complex Hermitian matrix

$\mathrm{I}^{-1}(\boldsymbol{\theta}^{(r)})$: real CRLB matrix, a $(2p_1 + p_2) \times (2p_1 + p_2)$ real symmetric matrix

$\mathrm{I}^{-1}(\boldsymbol{\theta})$: complex CRLB matrix, a $(2p_1 + p_2) \times (2p_1 + p_2)$ complex Hermitian matrix

2 Differentiation with respect to Complex Variables

To begin with, consider an analytic function $h(z)$ of a complex variable $z = x + \jmath y$. Write $h(z) = f_r(x, y) + \jmath f_i(x, y)$, where f_r and f_i are the real and imaginary parts of $h(z)$. Note that the real-valued functions f_r and f_i are "special" in that they are differentiable w.r.t. x and y (to all orders) and their partial derivatives satisfy the so-called Cauchy-Riemann conditions. Since $h(z)$ is analytic,

$$\frac{dh}{dz} = \lim_{\triangle z \to 0} \frac{h(z + \triangle z) - h(z)}{\triangle z} \tag{1}$$

where $\triangle z \to 0$ along any contour in the z-plane. In particular, letting $\triangle z \to 0$ along the real and imaginary axes we have

$$\frac{dh}{dz} = \frac{\partial f_r(x, y)}{\partial x} + \jmath \frac{\partial f_i(x, y)}{\partial x} \tag{2a}$$

and

$$\begin{aligned} \frac{dh}{dz} &= \frac{\partial f_r(x, y)}{\partial (\jmath y)} + \jmath \frac{\partial f_i(x, y)}{\partial (\jmath y)} \\ &= -\jmath \left[\frac{\partial f_r(x, y)}{\partial y} + \jmath \frac{\partial f_i(x, y)}{\partial y} \right] . \end{aligned} \tag{2b}$$

Combining equations $(2a)$ and $(2b)$ we obtain

$$\begin{aligned} \frac{dh}{dz} &= \frac{1}{2} \left[\frac{\partial f_r}{\partial x} + \jmath \frac{\partial f_i}{\partial x} \right] - \frac{\jmath}{2} \left[\frac{\partial f_r}{\partial y} + \jmath \frac{\partial f_i}{\partial y} \right] \\ &= \frac{1}{2} \left[\frac{\partial f(x, y)}{\partial x} - \jmath \frac{\partial f(x, y)}{\partial y} \right] , \end{aligned} \tag{2c}$$

where $f(x, y) = f_r(x, y) + \jmath f_i(x, y)$. By a similar reasoning, for $\tilde{h}(z^*)$, an analytic function of z^*, we obtain

$$\frac{d\tilde{h}}{dz^*} = \frac{1}{2} \left[\frac{\partial \tilde{f}(x, y)}{\partial x} + \jmath \frac{\partial \tilde{f}(x, y)}{\partial y} \right] , \tag{2d}$$

where $\tilde{h}(z^*) = \tilde{f}(x, y) = \tilde{f}_r(x, y) + \jmath \tilde{f}_i(x, y)$. Similar expressions for analytic functions of several complex variables are possible. For instance, let $h(z_1, z_2)$ be an analytic function of two complex variables, $z_1 = x_1 + \jmath y_1$, $z_2 = x_2 + \jmath y_2$. Let $h(z_1, z_2) = f(x_1, x_2, y_1, y_2)$, then

$$\frac{\partial h}{\partial z_k} = \frac{1}{2} \left[\frac{\partial f}{\partial x_k} - \jmath \frac{\partial f}{\partial y_k} \right] , \tag{3a}$$

for $k = 1, 2$. Likewise, for $\tilde{h}(z_1^*, z_2^*) = \tilde{f}(x_1, x_2, y_1, y_2)$, an analytic function of z_1^*, z_2^*, we obtain

$$\frac{\partial \tilde{h}}{\partial z_k^*} = \frac{1}{2}\left[\frac{\partial \tilde{f}}{\partial x_k} + \jmath \frac{\partial \tilde{f}}{\partial y_k}\right] , \tag{3b}$$

for $k = 1, 2$, and so on.

A natural question to ask is whether equations (2*c*) and (2*d*) make any sense when the functions are *not* analytic. Obviously the function must be such that the partial derivatives on the right hand side (RHS) of these equations exist, i.e. the limit in (1) must exist *at least* along the real and imaginary axes in the z-plane (and these need not be related). This is far less restrictive than analyticity and is satisfied by a large class of functions of a complex variable. Hence, for such functions, we may be able to use the RHS of equations (2*c*), (2*d*) as *definitions* of an *operator* on non-analytic functions. We emphasize that for non-analytic functions this is, strictly speaking, not a derivative. By sheer abuse of notation, we continue to use the symbol and nomenclature of (partial) derivative to represent these operators (as is common in the literature). For some applications, these operators are adequate, one instance is when the function $h(z)$ is real-valued [1].

Consider now a real-valued function $g'(z)$ of a complex variable $z = x + \jmath y$. Clearly g' is not analytic as it does not satisfy the Cauchy-Riemann conditions. Furthermore, since g' is real-valued, it *must also* depend *explicitly* on z^*, or we can write $g'(z) = g(z, z^*) = f(x, y)$, where f is a real-valued function of two (real) variables. We assume f is differentiable w.r.t. x and y. We now assume $g(z, z^*)$ is a function of *two* complex variables such that, for any complex numbers a and b, the functions $g_a(z) = g(z, a)$ and $g_b(z^*) = g(b, z^*)$ are analytic functions of z and z^*, respectively.

We then note that operating the RHS of equations (2*c*) and (2*d*) on functions z^* and z, respectively, yields zero. This seems to indicate that when definitions (2*c*) and (2*d*) are used, z and z^* behave like "independent" complex variables. With this in mind, we now think of $g(z, z^*)$ as being an analytic function of two independent complex variables z and z^*. Upon making this identification, we define partial derivatives of $g(z, z^*)$ w.r.t. z and z^* by

$$\frac{\partial g(z, z^*)}{\partial z} = \frac{1}{2}\left[\frac{\partial f}{\partial x} - \jmath \frac{\partial f}{\partial y}\right] , \tag{4a}$$

$$\frac{\partial g(z, z^*)}{\partial z^*} = \frac{1}{2}\left[\frac{\partial f}{\partial x} + \jmath \frac{\partial f}{\partial y}\right] . \tag{4b}$$

Note that all the theory associated with differentiating analytic functions of two variables can now be applied because the functions g_a, g_b are analytic functions of z and z^*, respectively (hence RHS of equations (4a) and (4b) are the true partials of g_a and g_b evaluated at $a = z^*$ and $b = z$, respectively).

Finally, as a further generalization, let $g'(\mathbf{z}) = g(\mathbf{z}, \mathbf{z}^*) = f(\mathbf{x}, \mathbf{y})$ be a function of several complex variables, $\mathbf{z} = [z_1\, z_2\, \cdots\, z_p]^T$ with $z_k = x_k + \jmath y_k$, and so on. We define the following partial derivatives

$$\frac{\partial g(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}} = \tfrac{1}{2}\begin{bmatrix} \frac{\partial f}{\partial x_1} - \jmath\frac{\partial f}{\partial y_1} \\ \frac{\partial f}{\partial x_2} - \jmath\frac{\partial f}{\partial y_2} \\ \vdots \\ \frac{\partial f}{\partial x_p} - \jmath\frac{\partial f}{\partial y_p} \end{bmatrix}, \quad \frac{\partial g(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^*} = \tfrac{1}{2}\begin{bmatrix} \frac{\partial f}{\partial x_1} + \jmath\frac{\partial f}{\partial y_1} \\ \frac{\partial f}{\partial x_2} + \jmath\frac{\partial f}{\partial y_2} \\ \vdots \\ \frac{\partial f}{\partial x_p} + \jmath\frac{\partial f}{\partial y_p} \end{bmatrix}. \tag{5}$$

We emphasize that (5) also serves as a definition of derivative of (analytic and non-analytic) complex-valued functions f. For example, for a complex-valued function $f = f_r + \jmath f_i$, we have $\frac{\partial f}{\partial x_k} = \frac{\partial f_r}{\partial x_k} + \jmath\frac{\partial f_i}{\partial x_k}$, and so on. More details and examples can be found in [1].

3 Cramer-Rao lower bounds

Let $\mathbf{x}$ denote the observed data with a PDF $p(\mathbf{x}\,;\,\underline{\boldsymbol{\theta}})$ parameterized by a vector of unknown parameters $\underline{\boldsymbol{\theta}} \in \mathbb{C}^{p_1} \times \mathbb{R}^{p_2}$. We will denote

$$\underline{\boldsymbol{\theta}} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix} \in \mathbb{C}^{p_1} \times \mathbb{R}^{p_2}, \quad \boldsymbol{\theta}^{(r)} = \begin{bmatrix} \boldsymbol{\theta}_{1r} \\ \boldsymbol{\theta}_{1i} \\ \boldsymbol{\theta}_2 \end{bmatrix} \in \mathbb{R}^{2p_1+p_2}, \quad \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_1^* \\ \boldsymbol{\theta}_2 \end{bmatrix} \in \mathbb{C}^{2p_1} \times \mathbb{R}^{p_2}$$

where $\boldsymbol{\theta}_1 \in \mathbb{C}^{p_1}$ and $\boldsymbol{\theta}_2 \in \mathbb{R}^{p_2}$, and $p = 2p_1 + p_2$. Let $\underline{\boldsymbol{\alpha}} = \underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}) \in \mathbb{C}^{q_1} \times \mathbb{R}^{q_2}$ be a vector function of $\underline{\boldsymbol{\theta}}$ whose estimate $\underline{\hat{\boldsymbol{\alpha}}}$ is of interest. Again, we will denote

$$\underline{\boldsymbol{\alpha}} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{bmatrix} \in \mathbb{C}^{q_1} \times \mathbb{R}^{q_2}, \quad \boldsymbol{\alpha}^{(r)} = \begin{bmatrix} \boldsymbol{\alpha}_{1r} \\ \boldsymbol{\alpha}_{1i} \\ \boldsymbol{\alpha}_2 \end{bmatrix} \in \mathbb{R}^{2q_1+q_2}, \quad \boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_1^* \\ \boldsymbol{\alpha}_2 \end{bmatrix} \in \mathbb{C}^{2q_1} \times \mathbb{R}^{q_2}$$

where $\boldsymbol{\alpha}_1 \in \mathbb{C}^{q_1}$ and $\boldsymbol{\alpha}_2 \in \mathbb{R}^{q_2}$, and $q = 2q_1 + q_2$. Corresponding estimates are denoted by $\underline{\hat{\boldsymbol{\alpha}}}$, $\hat{\boldsymbol{\alpha}}^{(r)}$, and $\hat{\boldsymbol{\alpha}}$. Also the k-th component of a vector $\boldsymbol{\beta}$ will be denoted by $\beta[k]$.

Definition 1 *(Complex Gradient) For any real/complex scalar function* $h(\underline{\boldsymbol{\theta}})$

$$\frac{\partial h}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial h}{\partial \boldsymbol{\theta}_1} \\ \frac{\partial h}{\partial \boldsymbol{\theta}_1^*} \\ \frac{\partial h}{\partial \boldsymbol{\theta}_2} \end{bmatrix}, \quad where \ \frac{\partial h}{\partial \boldsymbol{\theta}_2} = \begin{bmatrix} \frac{\partial h}{\partial \boldsymbol{\theta}_2[1]} \\ \frac{\partial h}{\partial \boldsymbol{\theta}_2[2]} \\ \vdots \\ \frac{\partial h}{\partial \boldsymbol{\theta}_2[p_2]} \end{bmatrix}$$

and the complex derivatives $\frac{\partial h}{\partial \boldsymbol{\theta}_1}, \frac{\partial h}{\partial \boldsymbol{\theta}_1^*}$ *are as defined in equation (5).*

Definition 2 *(Complex Jacobian) For any real/complex s–dimensional vector function* $\mathrm{h}(\underline{\boldsymbol{\theta}})$

$$\frac{\partial \mathrm{h}}{\partial \boldsymbol{\theta}} = \left[\frac{\partial \mathrm{h}}{\partial \theta_1[1]} \ \frac{\partial \mathrm{h}}{\partial \theta_1[2]} \cdots \frac{\partial \mathrm{h}}{\partial \theta_1[p_1]} \ \frac{\partial \mathrm{h}}{\partial \theta_1^*[1]} \ \frac{\partial \mathrm{h}}{\partial \theta_1^*[2]} \cdots \frac{\partial \mathrm{h}}{\partial \theta_1^*[p_1]} \ \frac{\partial \mathrm{h}}{\partial \theta_2[1]} \ \frac{\partial \mathrm{h}}{\partial \theta_2[2]} \cdots \frac{\partial \mathrm{h}}{\partial \theta_2[p_2]}\right]$$

where

$$\frac{\partial \mathrm{h}}{\partial \theta_1[l]} = \frac{1}{2}\begin{bmatrix} \frac{\partial h[1]}{\partial \theta_{1r}[l]} - \jmath\frac{\partial h[1]}{\partial \theta_{1i}[l]} \\ \frac{\partial h[2]}{\partial \theta_{1r}[l]} - \jmath\frac{\partial h[2]}{\partial \theta_{1i}[l]} \\ \vdots \\ \frac{\partial h[s]}{\partial \theta_{1r}[l]} - \jmath\frac{\partial h[s]}{\partial \theta_{1i}[l]} \end{bmatrix}, \quad \frac{\partial \mathrm{h}}{\partial \theta_1^*[l]} = \frac{1}{2}\begin{bmatrix} \frac{\partial h[1]}{\partial \theta_{1r}[l]} + \jmath\frac{\partial h[1]}{\partial \theta_{1i}[l]} \\ \frac{\partial h[2]}{\partial \theta_{1r}[l]} + \jmath\frac{\partial h[2]}{\partial \theta_{1i}[l]} \\ \vdots \\ \frac{\partial h[s]}{\partial \theta_{1r}[l]} + \jmath\frac{\partial h[s]}{\partial \theta_{1i}[l]} \end{bmatrix}, \quad \frac{\partial \mathrm{h}}{\partial \theta_2[k]} = \begin{bmatrix} \frac{\partial h[1]}{\partial \theta_2[k]} \\ \frac{\partial h[2]}{\partial \theta_2[k]} \\ \vdots \\ \frac{\partial h[s]}{\partial \theta_2[k]} \end{bmatrix},$$

for $1 \leq l \leq p_1$, $1 \leq k \leq p_2$.

Here we assume that the real and imaginary parts of these functions are all differentiable w.r.t. the elements of $\boldsymbol{\theta}^{(r)}$. These definitions simply mean that the partials w.r.t. real variables are defined as usual and the partials w.r.t. complex variables are interpreted as in section 2.

Theorem 1 *Under usual regularity conditions, the covariance matrix of any unbiased estimator* $\hat{\boldsymbol{\alpha}}$ *is bounded from below by*

$$\mathbf{C}_{\hat{\alpha}} = \mathcal{E}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^H \geq \left(\frac{\partial \mathrm{g}}{\partial \boldsymbol{\theta}}\right) \mathbf{I}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial \mathrm{g}}{\partial \boldsymbol{\theta}}\right)^H, \tag{6}$$

where the complex Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ *is given by*

$$\mathbf{I}(\boldsymbol{\theta}) = \mathcal{E}\left(\left[\frac{\partial \ln p}{\partial \boldsymbol{\theta}}\right]^* \left[\frac{\partial \ln p}{\partial \boldsymbol{\theta}}\right]^T\right) = \mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\theta}^*}\frac{\partial \ln p}{\partial \boldsymbol{\theta}^*}^H\right) , \tag{7}$$

or, for $1 \leq k,\ l \leq p$,

$$[\mathbf{I}(\boldsymbol{\theta})]_{k,l} = -\mathcal{E}\left(\frac{\partial^2 \ln p}{\partial \theta^*[k]\,\partial \theta[l]}\right) . \tag{8}$$

The derivatives are all evaluated at the true values of the parameters. In particular, for $\mathrm{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}$, *we have* $\mathbf{C}_{\hat{\theta}} \geq \mathbf{I}^{-1}(\boldsymbol{\theta})$. *We call* $\mathbf{I}^{-1}(\boldsymbol{\theta})$ *the complex CRLB of* $\boldsymbol{\theta}$.

When $\dfrac{\partial \ln p(\mathrm{x}\,;\,\underline{\theta})}{\partial \boldsymbol{\theta}^*} = \mathbf{I}(\boldsymbol{\theta})\,[\mathrm{t}(\mathrm{x}) - \boldsymbol{\theta}]$ *then the estimator* $\hat{\boldsymbol{\theta}} = \mathrm{t}(\mathrm{x})$ *is unbiased, attains the complex CRLB, and is said to be the efficient estimator.*

Proof. See Appendix *A*. ◇

As seen in the proof of the theorem, only assumptions needed to derive a CRLB for the real-valued parameter vector $\boldsymbol{\theta}^{(r)}$ are used, i.e., no "new" assumptions are required. Also the presence of both $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_1^*$ in the vector $\boldsymbol{\theta}$ seems to indicate that $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_1^*$ are *to be treated* as being functionally independent parameters. (This is akin to definitions of derivatives in section 2.) It also means that there is some redundancy in the complex Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$, this is explored later.

As expected the results are very similar to the corresponding analogues for the real-valued case. The curious thing however is that the derivative of the log PDF w.r.t. $\boldsymbol{\theta}^*$, $\dfrac{\partial \ln p}{\partial \boldsymbol{\theta}^*}$, seems to take on the role played by $\dfrac{\partial \ln p}{\partial \boldsymbol{\theta}}$ in the real-valued case. This seems to be the case in other situations too, viz. minimizing Hermitian forms [1] where a complex gradient is defined to be the derivative w.r.t. $\boldsymbol{\theta}^*$.

Connections to real CRLB: Since the complex gradient $\dfrac{\partial \ln p}{\partial \boldsymbol{\theta}^*}$ and the real gradient $\dfrac{\partial \ln p}{\partial \boldsymbol{\theta}^{(r)}}$ are related as

$$\frac{\partial \ln p}{\partial \boldsymbol{\theta}^*} = \begin{bmatrix} \dfrac{\partial \ln p}{\partial \boldsymbol{\theta}_1{}^*} \\ \dfrac{\partial \ln p}{\partial \boldsymbol{\theta}_1} \\ \dfrac{\partial \ln p}{\partial \boldsymbol{\theta}_2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\mathbf{I} & \frac{j}{2}\mathbf{I} & \mathbf{0} \\ \frac{1}{2}\mathbf{I} & -\frac{j}{2}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \dfrac{\partial \ln p}{\partial \mathrm{Re}\{\boldsymbol{\theta}_1\}} \\ \dfrac{\partial \ln p}{\partial \mathrm{Im}\{\boldsymbol{\theta}_1\}} \\ \dfrac{\partial \ln p}{\partial \boldsymbol{\theta}_2} \end{bmatrix} = \mathbf{T}\frac{\partial \ln p}{\partial \boldsymbol{\theta}^{(r)}} , \tag{9}$$

the complex Fisher information matrix, $\mathbf{I}(\boldsymbol{\theta})$, and the real Fisher information matrix, $\mathbf{I}(\boldsymbol{\theta}^{(r)})$, can be related as

$$\mathbf{I}(\boldsymbol{\theta}) = \mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\theta}^*}\frac{\partial \ln p}{\partial \boldsymbol{\theta}^*}^H\right) = \mathbf{T}\ \mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\theta}^{(r)}}\frac{\partial \ln p}{\partial \boldsymbol{\theta}^{(r)}}^T\right)\ \mathbf{T}^H = \mathbf{T}\ \mathbf{I}(\boldsymbol{\theta}^{(r)})\ \mathbf{T}^H \ . \tag{10}$$

We then obtain

$$\mathrm{CRLB}\,(\boldsymbol{\theta}) = \mathbf{I}^{-1}(\boldsymbol{\theta}) = \mathbf{T}^{H-1}\ \mathbf{I}^{-1}(\boldsymbol{\theta}^{(r)})\ \mathbf{T}^{-1} = \mathbf{T}^{H-1}\ \mathrm{CRLB}\left(\boldsymbol{\theta}^{(r)}\right)\ \mathbf{T}^{-1} \tag{11}$$

and

$$\mathrm{CRLB}\left(\boldsymbol{\theta}^{(r)}\right) = \mathbf{T}^H\ \mathrm{CRLB}\,(\boldsymbol{\theta})\ \mathbf{T}\ , \tag{12}$$

where, from (9),

$$\mathbf{T}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{I} & \mathbf{0} \\ -\jmath\mathbf{I} & \jmath\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} . \tag{13}$$

This relates the complex CRLB derived from Theorem 1, $\mathrm{CRLB}\,(\boldsymbol{\theta})$, with the real CRLB, $\mathrm{CRLB}\left(\boldsymbol{\theta}^{(r)}\right)$, derived from the usual theory for real-valued parameter vectors.

Finally, we apply Theorem 1 for a parameter transformation $\underline{\mathrm{g}}(\boldsymbol{\theta}) = \boldsymbol{\theta}^{(r)}$. We compute the complex gradient,

$$\frac{\partial \mathrm{g}}{\partial \boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}}\begin{bmatrix} (\boldsymbol{\theta}_1 + \boldsymbol{\theta}_1^*)/2 \\ (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_1^*)/2\jmath \\ \boldsymbol{\theta}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\mathbf{I} & \frac{1}{2}\mathbf{I} & \mathbf{0} \\ -\frac{\jmath}{2}\mathbf{I} & \frac{\jmath}{2}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} = \mathbf{T}^H ,$$

so that from Theorem 1 we have

$$\mathrm{CRLB}\left(\mathrm{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}^{(r)}\right) = \left(\frac{\partial \mathrm{g}}{\partial \boldsymbol{\theta}}\right)\mathrm{CRLB}\,(\boldsymbol{\theta})\left(\frac{\partial \mathrm{g}}{\partial \boldsymbol{\theta}}\right)^H = \mathbf{T}^H\mathrm{CRLB}\,(\boldsymbol{\theta})\,\mathbf{T}\ .$$

This means that the CRLBs derived from Theorem 1 and by application of the usual theory for real-valued parameter vectors always lead to the same results. Also, using (9) and (10) and $\boldsymbol{\theta}^{(r)} = \mathbf{T}^H\boldsymbol{\theta}$ in the efficiency condition $\frac{\partial \ln p\,(\mathrm{x}\,;\,\underline{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}^*} = \mathbf{I}(\boldsymbol{\theta})\,[\mathrm{t}(\mathrm{x}) - \boldsymbol{\theta}]$, we obtain $\frac{\partial \ln p\,(\mathrm{x}\,;\,\underline{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}^{(r)}} = \mathbf{I}(\boldsymbol{\theta}^{(r)})\left[\mathbf{T}^H\mathrm{t}(\mathrm{x}) - \boldsymbol{\theta}^{(r)}\right]$, which is the efficiency condition is the usual real-valued theory. In other words, the two approaches are theoretically equivalent, Theorem 1 only provides a *simpler* algebraic approach to CRLB computations. ◇

Remark. Since $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^{(r)}$ are related by a linear transformation, one may conjecture that Theorem 1 can be proved by direct application of the CRLB theory for real-valued parameters [5]. This is *not* so because only real-valued transformations are permitted there. Furthermore, since $\underline{\mathbf{g}}(\underline{\boldsymbol{\theta}}) : \mathbb{C}^{p_1} \times \mathbb{R}^{p_2} \to \mathbb{C}^{q_1} \times \mathbb{R}^{q_2}$, from Theorem 1 we can *now* calculate CRLB's for complex-valued functions of real-valued parameters (e.g., set $p_1 = 0$, $q_2 = 0$). This can be considered to be yet another generalization.

The only disturbing thing about Theorem 1 is that it is not obvious that the bottom right $p_2 \times p_2$ block in $\mathbf{I}^{-1}(\boldsymbol{\theta})$ will be real-valued (being the lower bound on the covariance of $\boldsymbol{\theta}_2$, a real-valued parameter vector) as one might expect. This is indeed true as the following lemma states.

Lemma 1 *(Structure of complex FIM and complex CRLB) Partition*

$$\mathbf{I}(\boldsymbol{\theta}) = \mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\theta}^*}\frac{\partial \ln p}{\partial \boldsymbol{\theta}^*}^H\right) = \mathcal{E}\begin{bmatrix} \frac{\partial \ln p}{\partial \boldsymbol{\theta}_1^*} \\ \frac{\partial \ln p}{\partial \boldsymbol{\theta}_1} \\ \frac{\partial \ln p}{\partial \boldsymbol{\theta}_2} \end{bmatrix} \begin{bmatrix} \frac{\partial \ln p}{\partial \boldsymbol{\theta}_1^*}^H & \frac{\partial \ln p}{\partial \boldsymbol{\theta}_1}^H & \frac{\partial \ln p}{\partial \boldsymbol{\theta}_2}^T \end{bmatrix}$$

$$= \left[\begin{array}{cc|c} \mathbf{A} & \mathbf{B}^* & \mathbf{P}^H \\ \mathbf{B} & \mathbf{A}^* & \mathbf{P}^T \\ \hline \mathbf{P} & \mathbf{P}^* & \mathbf{Q} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{I}_{11} & \mathbf{I}_{12} \\ \hline \mathbf{I}_{21} & \mathbf{I}_{22} \end{array}\right] = \begin{bmatrix} 2p_1 \times 2p_1 & 2p_1 \times p_2 \\ p_2 \times 2p_1 & p_2 \times p_2 \end{bmatrix}$$

where $\dfrac{\partial \ln p}{\partial \boldsymbol{\theta}_1^*} = \dfrac{\partial \ln p}{\partial \boldsymbol{\theta}_1}^*$ *was used, and*

$$\mathbf{I}_{11} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^* \\ \mathbf{B} & \mathbf{A}^* \end{bmatrix}, \quad \mathbf{I}_{21} = [\mathbf{P} \;\; \mathbf{P}^*], \quad \mathbf{I}_{12} = \begin{bmatrix} \mathbf{P}^H \\ \mathbf{P}^T \end{bmatrix} = \mathbf{I}_{21}^H, \quad \mathbf{I}_{22} = \mathbf{Q} = \mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\theta}_2}\frac{\partial \ln p}{\partial \boldsymbol{\theta}_2}^T\right),$$

$$\mathbf{A} = \mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\theta}_1^*}\frac{\partial \ln p}{\partial \boldsymbol{\theta}_1^*}^H\right), \quad \mathbf{B} = \mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\theta}_1}\frac{\partial \ln p}{\partial \boldsymbol{\theta}_1}^T\right), \quad \mathbf{P} = \mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\theta}_2}\frac{\partial \ln p}{\partial \boldsymbol{\theta}_1^*}^H\right).$$

Note that $\mathbf{I}_{11}$ *is complex-valued, Hermitian, and* $\mathbf{I}_{22}$ *is real-valued, symmetric.*

Then

$$CRLB(\boldsymbol{\theta}_2) = \left[\mathbf{I}_{22} - 2\,Re\left\{\mathbf{P}\mathbf{C}\mathbf{P}^H + \mathbf{P}^*\mathbf{D}\mathbf{P}^H\right\}\right]^{-1} \tag{14}$$

where $\mathbf{C} = \left(\mathbf{A} - \mathbf{B}^*\mathbf{A}^{*-1}\mathbf{B}\right)^{-1}$, *and* $\mathbf{D} = -\mathbf{A}^{*-1}\mathbf{B}\mathbf{C} = -\mathbf{C}^*\mathbf{B}\mathbf{A}^{-1}$. *Note that* $CRLB(\boldsymbol{\theta}_2)$ *is real-valued, symmetric as required.*

Proof. By the partioned matrix inversion lemma, $\text{CRLB}(\boldsymbol{\theta}_2) = \left[\mathbf{I}_{22} - \mathbf{I}_{21}\mathbf{I}_{11}^{-1}\mathbf{I}_{12}\right]^{-1}$. As shown in lemma *B*.4 in Appendix *B*, $\mathbf{I}_{11}^{-1} = \begin{bmatrix} \mathbf{C} & \mathbf{D}^* \\ \mathbf{D} & \mathbf{C}^* \end{bmatrix}$, where $\mathbf{C} = \left(\mathbf{A} - \mathbf{B}^*\mathbf{A}^{*-1}\mathbf{B}\right)^{-1}$, and $\mathbf{D} = -\mathbf{A}^{*-1}\mathbf{B}\mathbf{C} = -\mathbf{C}^*\mathbf{B}\mathbf{A}^{-1}$. Computing

$$\begin{aligned}
\mathbf{I}_{21}\mathbf{I}_{11}^{-1}\mathbf{I}_{12} &= [\mathbf{P}\ \mathbf{P}^*]\begin{bmatrix} \mathbf{C} & \mathbf{D}^* \\ \mathbf{D} & \mathbf{C}^* \end{bmatrix}\begin{bmatrix} \mathbf{P}^H \\ \mathbf{P}^T \end{bmatrix} = [\mathbf{P}\ \mathbf{P}^*]\begin{bmatrix} \mathbf{C}\mathbf{P}^H + \mathbf{D}^*\mathbf{P}^T \\ \mathbf{D}\mathbf{P}^H + \mathbf{C}^*\mathbf{P}^T \end{bmatrix} \\
&= \mathbf{P}\mathbf{C}\mathbf{P}^H + \mathbf{P}\mathbf{D}^*\mathbf{P}^T + \mathbf{P}^*\mathbf{D}\mathbf{P}^H + \mathbf{P}^*\mathbf{C}^*\mathbf{P}^T = 2\,\text{Re}\left\{\mathbf{P}\mathbf{C}\mathbf{P}^H + \mathbf{P}^*\mathbf{D}\mathbf{P}^H\right\} ,
\end{aligned}$$

from which (14) follows. Similarly, one can show that

$$\begin{aligned}
\text{CRLB}(\boldsymbol{\theta}_1) &= \mathbf{C} + \left(\mathbf{C}\mathbf{P}^H + \mathbf{D}^*\mathbf{P}^T\right)\ \text{CRLB}(\boldsymbol{\theta}_2)\ \left(\mathbf{P}\mathbf{C}^H + \mathbf{P}^*\mathbf{D}^T\right) \\
&= [\text{CRLB}(\boldsymbol{\theta}_1^*)]^* \ .
\end{aligned} \tag{15}$$

Expression (14) is particularly useful in signal processing applications because the complex-valued parameter $\boldsymbol{\theta}_1$ is usually a nuisance parameter (a signal amplitude) and $\boldsymbol{\theta}_2$ contains the structural information that is of greater importance. For example, we cite the common problem of sinusoidal parameter estimation [5] wherein $\boldsymbol{\theta}_1$ is the sinusoidal amplitude and $\boldsymbol{\theta}_2$ is the sinusoidal frequency. ◇

Note that the complex Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ is Hermitian so the lower bound on the complex covariance matrix in (6) is also Hermitian. In particular, their diagonal elements are real-valued. For example, applying the inequality (6) for the [1, 1] element, we obtain $\mathcal{E}\left(|\hat{\theta}_1[1] - \theta_1[1]|^2\right) \geq \mathbf{I}^{1,1}(\boldsymbol{\theta})$. That is to say, we have a bound for the *sum* of the variances of the real and imaginary parts of $\hat{\theta}_1[1]$. If a bound on the variance of only the real part of $\hat{\theta}_1[1]$ is desired, we use $g(\boldsymbol{\theta}) = (\theta_1[1] + \theta_1^*[1])/2$ in Theorem 1 and obtain

$$\begin{aligned}
\mathcal{E}\left(|\hat{\theta}_{1r}[1] - \theta_{1r}[1]|^2\right) &\geq \frac{1}{4}\left[\mathbf{I}^{1,1}(\boldsymbol{\theta}) + \mathbf{I}^{1,p_1+1}(\boldsymbol{\theta}) + \mathbf{I}^{p_1+1,1}(\boldsymbol{\theta}) + \mathbf{I}^{p_1+1,p_1+1}(\boldsymbol{\theta})\right] \\
&= \frac{1}{2}\left[\mathbf{I}^{1,1}(\boldsymbol{\theta}) + \text{Re}\left\{\mathbf{I}^{1,p_1+1}(\boldsymbol{\theta})\right\}\right] \ .
\end{aligned} \tag{16}$$

Note that, from (15), $\mathbf{I}^{1,1}(\boldsymbol{\theta}) = \mathbf{I}^{p_1+1,p_1+1}(\boldsymbol{\theta})$. Similarly, the variance of the imaginary part of $\hat{\theta}_1[1]$ can be shown to be bounded by $\frac{1}{2}\left[\mathbf{I}^{1,1}(\boldsymbol{\theta}) - \text{Re}\left\{\mathbf{I}^{1,p_1+1}(\boldsymbol{\theta})\right\}\right]$. In general, the bounds on the variances of the real and imaginary parts will be unequal (depending on the value of $\text{Re}\left\{\mathbf{I}^{1,p_1+1}(\boldsymbol{\theta})\right\}$). Also note that the sum of the bounds on the individual variances of the real and imaginary parts of $\theta_1[1]$ is exactly the bound on (total) variance of $\theta_1[1]$ that is given by $\mathbf{I}^{1,1}(\boldsymbol{\theta})$.

Example (DC level in complex white Gaussian noise): Consider a simple example of observing a constant signal A in additive noise. We shall assume A is complex-valued and is the unknown parameter of interest. The samples of noise are assumed to be uncorrelated, complex Gaussian [5] unit variance, random variables. The PDF of the data vector $\mathbf{x} = [x[0]\, x[1] \ldots x[N-1]]^T$ is then given by

$$p(\mathbf{x}\,;\,A) = \frac{1}{\pi^N} \exp\left\{-\sum_{n=0}^{N-1} |x[n] - A|^2\right\} ,$$

so that

$$\ln p(\mathbf{x}\,;\,A) = -N\ln\pi - \sum_{n=0}^{N-1} |x[n] - A|^2 = -N\ln\pi - (\mathbf{x} - A\mathbf{1})^H(\mathbf{x} - A\mathbf{1}) ,$$

where $\mathbf{1}$ denotes the length-N vector of ones. Using the complex gradient formulas in lemma $B.6$ Appendix B, we have

$$\frac{\partial \ln p}{\partial A} = (\mathbf{x} - A\mathbf{1})^H\mathbf{1} = \mathbf{1}^H(\mathbf{x} - A\mathbf{1})^*$$

$$\frac{\partial \ln p}{\partial A^*} = \mathbf{1}^H(\mathbf{x} - A\mathbf{1})$$

and using (7) with $p_1 = 1$, $p_2 = 0$, $p = 2p_1 + p_2 = 2$, we have

$$\begin{aligned}
\mathcal{E}\begin{bmatrix} \frac{\partial \ln p}{\partial A^*} \\ \frac{\partial \ln p}{\partial A} \end{bmatrix}\begin{bmatrix} \frac{\partial \ln p}{\partial A^*} \\ \frac{\partial \ln p}{\partial A} \end{bmatrix}^H &= \mathcal{E}\begin{bmatrix} \mathbf{1}^H(\mathbf{x} - A\mathbf{1}) \\ \mathbf{1}^H(\mathbf{x} - A\mathbf{1})^* \end{bmatrix}\begin{bmatrix} (\mathbf{x} - A\mathbf{1})^H\mathbf{1} & (\mathbf{x} - A\mathbf{1})^T\mathbf{1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{1}^H\mathcal{E}(\mathbf{x} - A\mathbf{1})(\mathbf{x} - A\mathbf{1})^H\mathbf{1} & \mathbf{1}^H\mathcal{E}(\mathbf{x} - A\mathbf{1})(\mathbf{x} - A\mathbf{1})^T\mathbf{1} \\ \mathbf{1}^H\mathcal{E}(\mathbf{x} - A\mathbf{1})^*(\mathbf{x} - A\mathbf{1})^H\mathbf{1} & \mathbf{1}^H\mathcal{E}(\mathbf{x} - A\mathbf{1})^*(\mathbf{x} - A\mathbf{1})^T\mathbf{1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{1}^H\mathbf{I}\mathbf{1} & \mathbf{1}^H\mathbf{0}\mathbf{1} \\ \mathbf{1}^H\mathbf{0}\mathbf{1} & \mathbf{1}^H\mathbf{I}\mathbf{1} \end{bmatrix} = \begin{bmatrix} N & 0 \\ 0 & N \end{bmatrix}
\end{aligned}$$

where properties $\mathcal{E}(\mathbf{w}\mathbf{w}^H) = \mathrm{cov}(\mathbf{w})$ and $\mathbf{E}(\mathbf{w}\mathbf{w}^T) = \mathbf{0}$ for complex Gaussian random vectors $\mathbf{w}$ have been used [5]. From Theorem 1 it then follows that $\mathrm{var}(\hat{A}) \geq 1/N$. Also, using the transformation formula in Theorem 1 as in (16), we obtain $\mathrm{var}(\mathrm{Re}\{\hat{A}\}) \geq 1/2N$ and $\mathrm{var}(\mathrm{Im}\{\hat{A}\}) \geq 1/2N$.

To illustrate the conventional approach to this problem, we compute

$$\frac{\partial \ln p}{\partial \mathrm{Re}\{A\}} = (\mathbf{x} - A\mathbf{1})^H \mathbf{1} + \mathbf{1}^H(\mathbf{x} - A\mathbf{1}) = \mathbf{1}^T\, 2\,\mathrm{Re}\{\mathbf{x} - A\mathbf{1}\}$$

$$\frac{\partial \ln p}{\partial \mathrm{Im}\{A\}} = (\mathbf{x} - A\mathbf{1})^H \jmath\mathbf{1} - \jmath\mathbf{1}^H(\mathbf{x} - A\mathbf{1}) = \jmath\mathbf{1}^H\left[(\mathbf{x} - A\mathbf{1})^* - (\mathbf{x} - A\mathbf{1})\right] = \mathbf{1}^T\, 2\,\mathrm{Im}\{\mathbf{x} - A\mathbf{1}\}$$

so that

$$\begin{aligned}
\mathcal{E}\begin{bmatrix} \dfrac{\partial \ln p}{\partial \mathrm{Re}\{A\}} \\ \dfrac{\partial \ln p}{\partial \mathrm{Im}\{A\}} \end{bmatrix}\begin{bmatrix} \dfrac{\partial \ln p}{\partial \mathrm{Re}\{A\}} \\ \dfrac{\partial \ln p}{\partial \mathrm{Im}\{A\}} \end{bmatrix}^T &= \mathcal{E}\begin{bmatrix} 2\mathbf{1}^T\mathrm{Re}\{\mathbf{x} - A\mathbf{1}\} \\ 2\mathbf{1}^T\mathrm{Im}\{\mathbf{x} - A\mathbf{1}\} \end{bmatrix}\begin{bmatrix} 2\mathrm{Re}^T\{\mathbf{x} - A\mathbf{1}\}\mathbf{1} & 2\mathrm{Im}^T\{\mathbf{x} - A\mathbf{1}\}\mathbf{1} \end{bmatrix} \\
&= 4\begin{bmatrix} \mathbf{1}^T\mathcal{E}\mathrm{Re}\{\mathbf{x} - A\mathbf{1}\}\mathrm{Re}^T\{\mathbf{x} - A\mathbf{1}\}\mathbf{1} & \mathbf{1}^T\mathcal{E}\mathrm{Re}\{\mathbf{x} - A\mathbf{1}\}\mathrm{Im}^T\{\mathbf{x} - A\mathbf{1}\}\mathbf{1} \\ \mathbf{1}^T\mathcal{E}\mathrm{Im}\{\mathbf{x} - A\mathbf{1}\}\mathrm{Re}^T\{\mathbf{x} - A\mathbf{1}\}\mathbf{1} & \mathbf{1}^T\mathcal{E}\mathrm{Im}\{\mathbf{x} - A\mathbf{1}\}\mathrm{Im}^T\{\mathbf{x} - A\mathbf{1}\}\mathbf{1} \end{bmatrix} \\
&= 2\begin{bmatrix} \mathbf{1}^T\mathbf{I}\mathbf{1} & \mathbf{1}^T\mathbf{0}\mathbf{1} \\ \mathbf{1}^T\mathbf{0}\mathbf{1} & \mathbf{1}^T\mathbf{I}\mathbf{1} \end{bmatrix} = \begin{bmatrix} 2N & 0 \\ 0 & 2N \end{bmatrix}
\end{aligned}$$

since for complex white Gaussian random variables [5] the real and imaginary parts are uncorrelated, Gaussian random variables with identical covariances. It then follows that the CRLBs for $\mathrm{Re}\{A\}$ and $\mathrm{Im}\{A\}$ are $1/2N$ each. Although the algebraic derivations of the two approaches seem to be equally simple, this is because the signal vector ($\mathbf{1}$) is real-valued in this example. The reader is urged to attempt more general problems (see section 4) to appreciate the algebraic simplicity involved in the proposed approach. ◇

Finally one wonders whether $\mathcal{E}\left(\frac{\partial \ln p}{\partial \underline{\boldsymbol{\theta}}}\frac{\partial \ln p}{\partial \underline{\boldsymbol{\theta}}}^H\right)$ can be used as a candidate Fisher information matrix. (Recall that $\underline{\boldsymbol{\theta}}$ contains only $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ while $\boldsymbol{\theta}$ contains $\boldsymbol{\theta}_1^*$ in addition.) An inspection of the proof of Theorem 1 indicates that this is possible but a CRLB so defined has some inconsistencies. First, lemma 1 does *not* hold, or the bottom right $p_2 \times p_2$ block in the inverse of $\mathcal{E}\left(\frac{\partial \ln p}{\partial \underline{\boldsymbol{\theta}}}\frac{\partial \ln p}{\partial \underline{\boldsymbol{\theta}}}^H\right)$, which would be a lower bound for the covariance of $\boldsymbol{\theta}_2$, is not guaranteed to be real-valued. Secondly, the bounds derived via the usual real-valued parameter vector approach (by forming $\boldsymbol{\theta}^{(r)}$) would, in general, be different from that provided by the inverse of $\mathcal{E}\left(\frac{\partial \ln p}{\partial \underline{\boldsymbol{\theta}}}\frac{\partial \ln p}{\partial \underline{\boldsymbol{\theta}}}^H\right)$. Finally, the inverse of $\mathcal{E}\left(\frac{\partial \ln p}{\partial \underline{\boldsymbol{\theta}}}\frac{\partial \ln p}{\partial \underline{\boldsymbol{\theta}}}^H\right)$ may not be the greatest lower bound, i.e., it may be too conservative and not attainable. These can be demonstrated by examining some simple examples.

Remark. Since $\frac{\partial \ln p}{\partial \boldsymbol{\theta}}$ and $\frac{\partial \ln p}{\partial \boldsymbol{\theta}^{(r)}}$ are related by a linear transformation, Yau and Bresler [11] propose to compute the CRLB($\boldsymbol{\theta}^{(r)}$) by inverting $\mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\theta}}\frac{\partial \ln p}{\partial \boldsymbol{\theta}}^T\right)$ and pre- and post-multiplying by the transformation matrix. While this strategy may be viable, they call the inverse of $\mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\theta}}\frac{\partial \ln p}{\partial \boldsymbol{\theta}}^T\right)$ a CRLB for $\boldsymbol{\theta}$, which is quite incorrect as Theorem 1 demonstrates. Note that this matrix is complex symmetric but a covariance matrix is Hermitian. Furthermore, in many applications, the main diagonal-blocks of $\mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\theta}}\frac{\partial \ln p}{\partial \boldsymbol{\theta}}^T\right)$ will be zero (in the DC level example, this matrix was $\begin{bmatrix} 0 & N \\ N & 0 \end{bmatrix}$). This also makes the inversion a little tedious.

4 Example

We now illustrate the use of these theorems in two generic problems that frequently arise in signal processing applications. The algebra, we believe, is considerably simpler than a straightforward use [9] of the real-valued theory. The CRLB expressions derived in this section have exactly the same form for an equivalent problem with real data and parameters. Therefore these results should, with obvious modifications, find application in corresponding real-valued problems [8].

Example 1. Consider a quasi-linear model described by

$$\mathbf{x} = \mathbf{H}(\boldsymbol{\beta})\,\boldsymbol{\theta}_1 + \mathbf{w} \tag{17}$$

where $\boldsymbol{\theta}_1$, $\boldsymbol{\beta}$ are unknown parameters. We observe $\mathbf{x}$, a noisy version of the signal $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}_1$ that lies in a parameterized subspace spanned by the p columns of $\mathbf{H}$. The p elements of $\boldsymbol{\theta}_1$ are the complex amplitudes of the component signals in $\mathbf{s}$. The noise vector $\mathbf{w}$ is assumed to be zero-mean, white complex Gaussian distributed with unknown variance σ^2. We assume $\boldsymbol{\beta}$ is real-valued for purposes of illustration. This is a common problem in signal processing [5] and in other areas [8] as well. Examples include parameter estimation for damped/undamped sinusoids in white noise [5], resolution of overlapping echos [11], etc. and their multi-dimensional extensions. Hence $\boldsymbol{\theta}_2 = \left[\boldsymbol{\beta}^T \sigma^2\right]^T$, $\underline{\theta} = \left[\boldsymbol{\theta}_1^T\, \boldsymbol{\beta}^T\, \sigma^2\right]^T$ and $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_1^T\, \boldsymbol{\theta}_1^H\, \boldsymbol{\beta}^T\, \sigma^2\right]^T$. The log PDF of the observed data is given by [2, 5]

$$\ln p(\mathbf{x}\,;\,\underline{\theta}) = -N \ln \pi\sigma^2 - \frac{(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}_1)^H(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}_1)}{\sigma^2}\ ,$$

so that, from lemma $B.6$ in Appendix B, we have

$$\begin{aligned}
\frac{\partial \ln p}{\partial \boldsymbol{\theta}_1} &= \mathbf{H}^T\,(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}_1)^* \big/ \sigma^2 \\
\frac{\partial \ln p}{\partial \boldsymbol{\theta}_1^*} &= \mathbf{H}^H\,(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}_1) \big/ \sigma^2 \\
\frac{\partial \ln p}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}\left[\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}^H (\mathbf{x}-\mathbf{s}) + \frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}^T (\mathbf{x}-\mathbf{s})^*\right] \\
\frac{\partial \ln p}{\partial \sigma^2} &= -\frac{N}{\sigma^2} + \frac{(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}_1)^H(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}_1)}{\sigma^4}\ .
\end{aligned}$$

Here the Jacobian $\dfrac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}$ is given by

$$\begin{aligned}
\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}} &= \frac{\partial(\mathbf{H}\boldsymbol{\theta}_1)}{\partial \boldsymbol{\beta}} = \left[\frac{\partial \mathbf{H}\boldsymbol{\theta}_1}{\partial \beta[1]}\ \frac{\partial \mathbf{H}\boldsymbol{\theta}_1}{\partial \beta[2]} \cdots \frac{\partial \mathbf{H}\boldsymbol{\theta}_1}{\partial \beta[p_2]}\right] = \left[\frac{\partial \mathbf{H}}{\partial \beta[1]}\boldsymbol{\theta}_1\ \frac{\partial \mathbf{H}}{\partial \beta[2]}\boldsymbol{\theta}_1 \cdots \frac{\partial \mathbf{H}}{\partial \beta[p_2]}\boldsymbol{\theta}_1\right] \\
&= \left[\frac{\partial \mathbf{H}}{\partial \beta[1]}\ \frac{\partial \mathbf{H}}{\partial \beta[2]} \cdots \frac{\partial \mathbf{H}}{\partial \beta[p_2]}\right]\begin{bmatrix} \boldsymbol{\theta}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\theta}_1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\theta}_1 \end{bmatrix} = \frac{\partial \mathbf{H}}{\partial \boldsymbol{\beta}}\left(\mathbf{I}_{p_2} \otimes \boldsymbol{\theta}_1\right)
\end{aligned}$$

where $\mathbf{I}_{p_2}$ denotes the identity matrix of size p_2 by p_2, and $\otimes$ denotes the Kronecker product. Recall that differentiating a matrix (vector) w.r.t. a real-valued scalar means replacing each element in the matrix (vector) by the corresponding derivative w.r.t. the scalar.

It is easily checked that $\mathcal{E}\left(\dfrac{\partial \ln p}{\partial \boldsymbol{\theta}}\right) = \mathbf{0}$, or the regularity conditions hold. Differentiating w.r.t. σ^2 and taking expectations, it is easy to check that

$$\mathcal{E}\left(\frac{\partial^2 \ln p}{\partial \sigma^2\, \partial \boldsymbol{\theta}_1}\right) = \mathbf{0}\ ,\ \mathcal{E}\left(\frac{\partial^2 \ln p}{\partial \sigma^2\, \partial \boldsymbol{\theta}_1^*}\right) = \mathbf{0}\ ,\ \mathcal{E}\left(\frac{\partial^2 \ln p}{\partial \sigma^2\, \partial \boldsymbol{\beta}}\right) = \mathbf{0}\ ,$$

and

$$-\mathcal{E}\left(\frac{\partial^2 \ln p}{\partial {\sigma^2}^2}\right) = \mathcal{E}\left(-\frac{N}{\sigma^4} + \frac{2\,(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}_1)^H(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}_1)}{\sigma^6}\right) = \frac{N}{\sigma^4}\ .$$

Since the odd order moments of $\mathbf{x} - \mathbf{H}\boldsymbol{\theta}_1 = \mathbf{w}$ are zero, and $\mathcal{E}(\mathbf{w}\mathbf{w}^T) = \mathbf{0}$ (follows from properties of complex Gaussian random vectors [5]), we obtain

$$\mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\theta}_1^*}\frac{\partial \ln p}{\partial \boldsymbol{\theta}_1}^T\right) = \mathbf{H}^H\mathbf{H}/\sigma^2\,,\ \mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\theta}_1}\frac{\partial \ln p}{\partial \boldsymbol{\theta}_1}^T\right) = \mathbf{0}\,,\ \mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\beta}}\frac{\partial \ln p}{\partial \boldsymbol{\theta}_1}^T\right) = \frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}^H \mathbf{H}/\sigma^2\,,$$

$$\mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\beta}}\frac{\partial \ln p}{\partial \boldsymbol{\beta}}^T\right) = -\mathcal{E}\left(\frac{\partial^2 \ln p}{\partial \boldsymbol{\beta}\,\partial \boldsymbol{\beta}}\right) = \frac{1}{\sigma^2}\left[\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}^H\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}} + \frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}^T\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}^*\right]\,.$$

Using symmetry as in lemma 1, we form

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\left[\begin{array}{cc|cc} \mathbf{H}^H\mathbf{H} & \mathbf{0} & \mathbf{H}^H\dfrac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}^T\mathbf{H}^* & \mathbf{H}^T\dfrac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}^* & \mathbf{0} \\ \hline \dfrac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}^H\mathbf{H} & \dfrac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}^T\mathbf{H}^* & 2\mathrm{Re}\left\{\dfrac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}^H\dfrac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}\right\} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dfrac{N}{\sigma^2} \end{array}\right].$$

It is easy to see that $\mathrm{CRLB}(\sigma^2) = \sigma^4/N$. By the block diagonal nature of $\mathbf{I}(\boldsymbol{\theta})$, the complex CRLB of $\boldsymbol{\theta}_1$ and $\boldsymbol{\beta}$ is found by simply inverting the complex FIM formed from $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_1^*$ and $\boldsymbol{\beta}$. In other words, CRLB $(\boldsymbol{\theta}_1)$ and CRLB $(\boldsymbol{\beta})$ may be found by thinking of only $\boldsymbol{\theta}_1$, $\boldsymbol{\beta}$ as the unknown parameters. Then using lemma 1 we obtain

$$\begin{aligned}\mathrm{CRLB}(\boldsymbol{\beta}) &= \left[\frac{2}{\sigma^2}\mathrm{Re}\left\{\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}^H\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}\right\} - 2\mathrm{Re}\left\{\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}^H\mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}\right\}/\sigma^2\right]^{-1} \\ &= \left[\frac{2}{\sigma^2}\mathrm{Re}\left\{\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}^H\mathbf{P}_H^\perp\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}\right\}\right]^{-1} = \frac{\sigma^2}{2}\left[\mathrm{Re}\left\{(\mathbf{I}_{p_2}\otimes\boldsymbol{\theta}_1)^H\frac{\partial \mathbf{H}}{\partial \boldsymbol{\beta}}^H\mathbf{P}_H^\perp\frac{\partial \mathbf{H}}{\partial \boldsymbol{\beta}}(\mathbf{I}_{p_2}\otimes\boldsymbol{\theta}_1)\right\}\right]^{-1}\end{aligned} \tag{18}$$

where $\mathbf{P}_H^\perp = \mathbf{I} - \mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H$ and

$$\mathrm{CRLB}(\boldsymbol{\theta}_1) = \sigma^2\left(\mathbf{H}^H\mathbf{H}\right)^{-1} + \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}\,\mathrm{CRLB}(\boldsymbol{\beta})\,\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}^H\mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\,. \tag{19}$$

Equation (18) is of particular interest in many signal processing applications because $\boldsymbol{\beta}$ contains important physical information about the underlying signal (e.g. range, doppler, frequency, arrival angle, etc.). We consider a few special cases which are of some interest.

1. $\boldsymbol{\beta}$ and σ^2 are known. Then

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_1^* \end{bmatrix} \;,\; \mathbf{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\begin{bmatrix} \mathbf{H}^H\mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}^T\mathbf{H}^* \end{bmatrix} \;,\; \frac{\partial \ln p}{\partial \boldsymbol{\theta}^*} = \begin{bmatrix} \mathbf{H}^H\mathbf{x} - \mathbf{H}^H\mathbf{H}\boldsymbol{\theta}_1 \\ \mathbf{H}^T\mathbf{x}^* - \mathbf{H}^T\mathbf{H}^*\boldsymbol{\theta}_1{}^* \end{bmatrix} .$$

Hence we have

$$\frac{\partial \ln p}{\partial \boldsymbol{\theta}^*} = \mathbf{I}(\boldsymbol{\theta}) \begin{bmatrix} \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{x} - \boldsymbol{\theta}_1 \\ \left(\mathbf{H}^T\mathbf{H}^*\right)^{-1}\mathbf{H}^T\mathbf{x}^* - \boldsymbol{\theta}_1{}^* \end{bmatrix} ,$$

so that by Theorem 1, $\hat{\boldsymbol{\theta}}_1 = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{x}$ is the efficient estimator.

2. Let $p_1 = p_2$ and let the k-th column of $\mathbf{H}$, denoted by $\mathbf{h}_k$, depend only on $\beta[k]$. An example is sinusoidal parameter estimation [5]. Then

$$\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}} = \left[\frac{\partial \mathbf{h}_1}{\partial \beta[1]}\,\frac{\partial \mathbf{h}_2}{\partial \beta[2]}\cdots\frac{\partial \mathbf{h}_{p_1}}{\partial \beta[p_1]}\right]\begin{bmatrix} \theta_1[1] & 0 & \cdots & 0 \\ 0 & \theta_1[2] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \theta_1[p_1] \end{bmatrix} = \mathbf{G}\,\boldsymbol{\Theta}_1 \;,$$

and hence (18) becomes $\mathrm{CRLB}\,(\boldsymbol{\beta}) = \frac{\sigma^2}{2}\left[\mathrm{Re}\left\{\boldsymbol{\Theta}_1{}^H\mathbf{G}^H\mathbf{P}_H^{\perp}\mathbf{G}\boldsymbol{\Theta}_1\right\}\right]^{-1}$. Next, write $\theta_1[k] = |\theta_1[k]|\,e^{\jmath\phi_1[k]}$ and denote $|\boldsymbol{\Theta}_1| = \mathrm{diag}\left\{|\theta_1[1]|, \cdots, |\theta_1[p_1]|\right\}$, $\boldsymbol{\Phi}_1 = \mathrm{diag}\left\{e^{\jmath\phi_1[1]}, \cdots, e^{\jmath\phi_1[p_1]}\right\}$, so that $\boldsymbol{\Theta}_1 = |\boldsymbol{\Theta}_1|\,\boldsymbol{\Phi}_1$. Using this we obtain

$$\mathrm{CRLB}\,(\boldsymbol{\beta}) = \frac{\sigma^2}{2}|\boldsymbol{\Theta}_1|^{-1}\left[\mathrm{Re}\left\{\boldsymbol{\Phi}_1{}^H\mathbf{G}^H\mathbf{P}_H^{\perp}\mathbf{G}\,\boldsymbol{\Phi}_1\right\}\right]^{-1}|\boldsymbol{\Theta}_1|^{-1}$$

and

$$\mathrm{CRLB}(\beta[k]) = \frac{\sigma^2}{2|\theta_1[k]|^2}\left[\mathrm{Re}\left\{\boldsymbol{\Phi}_1{}^H\mathbf{G}^H\mathbf{P}_H^{\perp}\mathbf{G}\,\boldsymbol{\Phi}_1\right\}\right]^{k,k}$$

which brings out the explicit dependence on the signal to noise ratio of the k-th signal component, $|\theta_1[k]|^2/\sigma^2$. To obtain $\mathrm{CRLB}\,(\boldsymbol{\theta}_1)$, we use $\frac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}} = \mathbf{G}\,\boldsymbol{\Theta}_1$ and above expression for $\mathrm{CRLB}\,(\boldsymbol{\beta})$ in (19). Hence we have

$$\begin{aligned}\mathrm{CRLB}\,(\boldsymbol{\theta}_1) &= \sigma^2\Big[\left(\mathbf{H}^H\mathbf{H}\right)^{-1} + \tfrac{1}{2}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{G}\,\boldsymbol{\Theta}_1\,|\boldsymbol{\Theta}_1|^{-1} \\ &\qquad \left[\mathrm{Re}\left\{\boldsymbol{\Phi}_1{}^H\mathbf{G}^H\mathbf{P}_H^{\perp}\mathbf{G}\,\boldsymbol{\Phi}_1\right\}\right]^{-1}|\boldsymbol{\Theta}_1|^{-1}\,\boldsymbol{\Theta}_1{}^H\mathbf{G}\mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\Big] \\ &= \sigma^2\Big[\left(\mathbf{H}^H\mathbf{H}\right)^{-1} + \tfrac{1}{2}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{G}\,\boldsymbol{\Phi}_1\left[\mathrm{Re}\left\{\boldsymbol{\Phi}_1{}^H\mathbf{G}^H\mathbf{P}_H^{\perp}\mathbf{G}\,\boldsymbol{\Phi}_1\right\}\right]^{-1}\boldsymbol{\Phi}_1{}^H\mathbf{G}\mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\Big] .\end{aligned}$$

Note that the bound on the complex amplitudes, CRLB $(\boldsymbol{\theta}_1)$, depends on the true value of the amplitudes only via the phase differences $(\phi_i - \phi_j)$, it does not depend on the magnitudes. If $\boldsymbol{\beta}$ were to be known the bound would have been just $\sigma^2\left(\mathbf{H}^H\mathbf{H}\right)^{-1}$, i.e. the second term in above expression represents the increase in the bound because $\boldsymbol{\beta}$ is also unknown. For other interpretations of these expressions, see [7, 11]. This example can be generalized to include models wherein the columns $\mathbf{h}_k$ are parameterized by vector parameters $\boldsymbol{\beta}_k$, also see [11].

3. (Multiple snapshot models):

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{L-1} \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} \mathbf{H}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H}_{L-1} \end{bmatrix}, \quad \boldsymbol{\theta}_1 = \begin{bmatrix} \boldsymbol{\theta}_{1,0} \\ \boldsymbol{\theta}_{1,1} \\ \vdots \\ \boldsymbol{\theta}_{1,L-1} \end{bmatrix}.$$

Usually all the $\mathbf{H}_k$'s are equal and are parameterized by a vector $\boldsymbol{\beta}$ common to all blocks [9, 11]. Further details can be worked out by simplifying $\dfrac{\partial \mathbf{s}}{\partial \boldsymbol{\beta}}$.

Example 2. (General Complex Gaussian CRLB) We now further generalize the previous example by considering the data to be simply complex Gaussian with a mean $\mathbf{s}$ and covariance matrix $\mathbf{C}$ jointly parameterized by a vector $\underline{\boldsymbol{\theta}}$. As before, $\underline{\boldsymbol{\theta}}$ may consist of some real-valued and some complex-valued parameters. Then the k, l-th element of the complex Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ can be computed as

$$[\mathbf{I}(\boldsymbol{\theta})]_{k,l} = \frac{\partial \mathbf{s}}{\partial \theta[k]}^H \mathbf{C}^{-1} \frac{\partial \mathbf{s}}{\partial \theta[l]} + \frac{\partial \mathbf{s}}{\partial \theta^*[l]}^H \mathbf{C}^{-1} \frac{\partial \mathbf{s}}{\partial \theta^*[k]} + \mathrm{tr}\left\{ \frac{\partial \mathbf{C}}{\partial \theta^*[k]} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta[l]} \mathbf{C}^{-1} \right\}, \tag{20}$$

for $1 \leq k,\, l \leq p$. This result is proved as lemma B.5 in Appendix B. In fact this general result can be used in Example 1 but our goal there was to illustrate the derivations from first principles.

It is believed that most of the other expressions for the CRLBs for real-valued parameters (for example, asymptotic CRLB for complex wide sense stationary Gaussian processes) can be extended along these lines as well.

5 Conclusions

An introduction to Cramer-Rao lower bounds for complex-valued parameters was presented. The vector parameterizing the PDF of the data is assumed to contain some complex-valued and some real-valued parameters. We present a *direct* approach to this problem by working with the complex-valued quantities *per se*. This, we believe, leads to elegant algebraic manipulations unlike customary approaches which lead to somewhat clumsy algebra. Explicit expressions for the CRLB in a simple but common signal processing problem are given.

References

[1] D. H. Brandwood, "A Complex Gradient Operator and its application in adaptive array theory," *IEE Proc. Pts. F and H,* pp. 11–16, Feb. 1983.

[2] N. R. Goodman, "Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution," *Ann. Math. Stat.,* pp. 152–177, 1963.

[3] J. D. Gorman and A. O. Hero, "Lower Bounds for Parametric Estimation with Constraints," *IEEE Trans. Info. Theory,* pp. 1285–1301, Nov. 1990.

[4] U. Grenander, *Abstract Inference.* New York, NY: John Wiley, 1981.

[5] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* . Englewood Cliffs, NJ: Prentice-Hall, to appear.

[6] K. S. Miller, *Complex Stochastic Processes.* Reading, MA: Addison-Wesley, 1974.

[7] D. C. Rife and R. R. Boorstyn, "Multiple Tone Parameter Estimation from discrete-time Observations," *Bell Syst. Tech. J.,* pp. 1389–1410, Nov. 1976.

[8] G. A. F. Seber and C. J. Wild, *Nonlinear Regression.* New York, NY: John Wiley, 1989.

[9] P. Stoica and A. Nehorai, "MUSIC, Maximum Likelihood, and the Cramer-Rao Bound," *IEEE Trans. Acoust., Speech, Signal Process.,* pp. 720–741, May 1989.

[10] H. L. Van Trees, *Detection, Estimation and Modulation Theory*, vol. 1. New York, NY: John Wiley, 1971.

[11] S. F. Yau and Y. Bresler, "A Compact Cramer-Rao Bound Expression for Parametric Estimation of Superimposed Signals," *IEEE Trans. Signal Process.*, pp. 1226–1230, May 1992.

Appendix A

Proof of Theorem 1

We begin by checking that

$$\mathcal{E}\left[\frac{\partial \ln p}{\partial \boldsymbol{\theta}}\right] = \mathcal{E}\begin{bmatrix} \dfrac{\partial \ln p}{\partial \boldsymbol{\theta}_1} \\ \dfrac{\partial \ln p}{\partial \boldsymbol{\theta}_1^*} \\ \dfrac{\partial \ln p}{\partial \boldsymbol{\theta}_2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\mathbf{I} & -\frac{j}{2}\mathbf{I} & \mathbf{0} \\ \frac{1}{2}\mathbf{I} & \frac{j}{2}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \mathcal{E}\left[\frac{\partial \ln p}{\partial \boldsymbol{\theta}^{(r)}}\right] = \mathbf{0} \ , \qquad (\mathrm{A}-1)$$

from (B-2). Since $\hat{\boldsymbol{\alpha}}^{(r)}$ is an unbiased estimate of $\boldsymbol{\alpha}^{(r)}$, it follows that $\hat{\boldsymbol{\alpha}}$ is an unbiased estimate of $\boldsymbol{\alpha}$. From lemma $B.2$ in Appendix B and definition 2, it follows that

$$\frac{\partial \mathrm{g}}{\partial \boldsymbol{\theta}} = \left[\frac{\partial \mathrm{g}}{\partial \theta[1]}\ \frac{\partial \mathrm{g}}{\partial \theta[2]} \cdots \frac{\partial \mathrm{g}}{\partial \theta[p]}\right] = \int_{\mathbf{X}} \hat{\boldsymbol{\alpha}}\, \frac{\partial \ln p}{\partial \boldsymbol{\theta}}^T\, p(\mathrm{x}\,;\,\boldsymbol{\theta})\ \mathrm{dx}\ . \qquad (\mathrm{A}-2)$$

Now

$$\begin{aligned} \int_{\mathbf{X}} (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})\, \frac{\partial \ln p}{\partial \boldsymbol{\theta}}^T\, p(\mathrm{x}\,;\,\boldsymbol{\theta})\ \mathrm{dx} &= \int_{\mathbf{X}} \hat{\boldsymbol{\alpha}}\, \frac{\partial \ln p}{\partial \boldsymbol{\theta}}^T\, p(\mathrm{x}\,;\,\boldsymbol{\theta})\ \mathrm{dx} - \int_{\mathbf{X}} \boldsymbol{\alpha}\, \frac{\partial \ln p}{\partial \boldsymbol{\theta}}^T\, p(\mathrm{x}\,;\,\boldsymbol{\theta})\ \mathrm{dx} \\ &= \int_{\mathbf{X}} \hat{\boldsymbol{\alpha}}\, \frac{\partial \ln p}{\partial \boldsymbol{\theta}}^T\, p(\mathrm{x}\,;\,\boldsymbol{\theta})\ \mathrm{dx} - \boldsymbol{\alpha}\, \mathcal{E}^T\left[\frac{\partial \ln p}{\partial \boldsymbol{\theta}}\right] \ = \ \frac{\partial \mathrm{g}}{\partial \boldsymbol{\theta}}\ , \end{aligned}$$

from (A-1) and (A-2). For any $\mathbf{a} \in \mathbb{C}^q$, $\mathbf{b} \in \mathbb{C}^p$, we then obtain

$$\int_{\mathbf{X}} \mathrm{a}^H (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})\, \frac{\partial \ln p}{\partial \boldsymbol{\theta}}^T\, \mathrm{b}\, p(\mathrm{x}\,;\,\boldsymbol{\theta})\ \mathrm{dx} = \mathrm{a}^H \frac{\partial \mathrm{g}}{\partial \boldsymbol{\theta}} \mathrm{b}\ ,$$

so that using the Cauchy-Schwartz inequality (lemma $B.3$ in Appendix B) we obtain

$$\begin{aligned} \left|\mathrm{a}^H \frac{\partial \mathrm{g}}{\partial \boldsymbol{\theta}} \mathrm{b}\right|^2 &\le \mathrm{a}^H \left[\int_{\mathbf{X}} (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^H p(\mathrm{x}\,;\,\boldsymbol{\theta})\ \mathrm{dx}\right] \mathrm{a} \cdot \mathrm{b}^H \left[\int_{\mathbf{X}} \frac{\partial \ln p}{\partial \boldsymbol{\theta}}^* \frac{\partial \ln p}{\partial \boldsymbol{\theta}}^T p(\mathrm{x}\,;\,\boldsymbol{\theta})\ \mathrm{dx}\right] \mathrm{b} \\ &= \mathrm{a}^H \mathbf{C}_{\hat{\alpha}}\, \mathrm{a} \cdot \mathrm{b}^H \mathbf{I}(\boldsymbol{\theta})\, \mathrm{b}\ , \qquad (\text{A-3}) \end{aligned}$$

where we define the $p \times p$ Hermitian matrix

$$\mathbf{I}(\boldsymbol{\theta}) = \mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\theta}}^* \frac{\partial \ln p}{\partial \boldsymbol{\theta}}^T\right) = \mathcal{E}\left(\frac{\partial \ln p}{\partial \boldsymbol{\theta}^*} \frac{\partial \ln p}{\partial \boldsymbol{\theta}^*}^H\right)\ , \qquad (\mathrm{A}-4)$$

as the complex Fisher information matrix. In particular, choosing $\mathbf{b} = \mathbf{I}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}\right)^H \mathbf{a}$ in (A-3) we obtain

$$\left| \mathbf{a}^H \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}^H \mathbf{a} \right|^2 \leq \left(\mathbf{a}^H \mathbf{C}_{\hat{\alpha}} \mathbf{a} \right) \left(\mathbf{a}^H \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}^H \mathbf{a} \right) .$$

Since $\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}^H$ is Hermitian and at least positive semidefinite, cancelling terms on both sides we obtain

$$\mathbf{a}^H \left[\mathbf{C}_{\hat{\alpha}} - \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}^H \right] \mathbf{a} \geq 0 \ , \qquad \text{(A} - 5)$$

for any $\mathbf{a} \in \mathbb{C}^q$. This means that the difference matrix in (A-5) is positive semidefinite or

$$\left[\mathbf{C}_{\hat{\alpha}} - \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}^H \right] \geq \mathbf{0} \ .$$

In particular

$$[\mathbf{C}_{\hat{\alpha}}]_{k,k} = \mathcal{E} \left| \hat{\alpha}[k] - \alpha[k] \right|^2 = \operatorname{var}(\hat{\alpha}[k]) \geq \left[\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}^H \right]_{k,k} .$$

Note that $\mathbf{I}(\boldsymbol{\theta})$ defined in (A-4) is only guaranteed to be positive semidefinite, we assume that it is positive definite.

Next, we relate the elements of $\mathbf{I}(\boldsymbol{\theta})$ to the second derivatives of the log PDF, this would complete the proof for the first part of Theorem 1. For $1 \leq k \leq p_1$, $1 \leq l \leq p_1$, consider

$$\begin{aligned} \frac{\partial^2 \ln p}{\partial \theta^*[k] \, \partial \theta[l]} &= \frac{\partial}{\partial \theta^*[k]} \left(\frac{\partial \ln p}{\partial \theta[l]} \right) = \frac{\partial}{\partial \theta^*[k]} \left[\frac{1}{2} \frac{\partial \ln p}{\partial \theta_r[l]} - \frac{\jmath}{2} \frac{\partial \ln p}{\partial \theta_i[l]} \right] \\ &= \left(\frac{1}{2} \frac{\partial}{\partial \theta_r[k]} + \frac{\jmath}{2} \frac{\partial}{\partial \theta_i[k]} \right) \left[\frac{1}{2} \frac{\partial \ln p}{\partial \theta_r[l]} - \frac{\jmath}{2} \frac{\partial \ln p}{\partial \theta_i[l]} \right] \\ &= \frac{1}{4} \left[\frac{\partial^2 \ln p}{\partial \theta_r[k] \partial \theta_r[l]} + \frac{\partial^2 \ln p}{\partial \theta_i[k] \partial \theta_i[l]} \right] + \frac{\jmath}{4} \left[\frac{\partial^2 \ln p}{\partial \theta_i[k] \partial \theta_r[l]} - \frac{\partial^2 \ln p}{\partial \theta_r[k] \partial \theta_i[l]} \right] . \end{aligned}$$

Upon taking expectations and using (B-4) we have

$$\begin{aligned} \mathcal{E} \left(\frac{\partial^2 \ln p}{\partial \theta^*[k] \, \partial \theta[l]} \right) &= \mathcal{E} \ \frac{1}{4} \left[\frac{\partial \ln p}{\partial \theta_r[k]} \frac{\partial \ln p}{\partial \theta_r[l]} + \frac{\partial \ln p}{\partial \theta_i[k]} \frac{\partial \ln p}{\partial \theta_i[l]} \right] + \frac{\jmath}{4} \left[\frac{\partial \ln p}{\partial \theta_i[k]} \frac{\partial \ln p}{\partial \theta_r[l]} - \frac{\partial \ln p}{\partial \theta_r[k]} \frac{\partial \ln p}{\partial \theta_i[l]} \right] \\ &= \mathcal{E} \ \left[\frac{1}{2} \frac{\partial \ln p}{\partial \theta_r[k]} + \frac{\jmath}{2} \frac{\partial \ln p}{\partial \theta_i[k]} \right] \cdot \left[\frac{1}{2} \frac{\partial \ln p}{\partial \theta_r[l]} - \frac{\jmath}{2} \frac{\partial \ln p}{\partial \theta_i[l]} \right] \\ &= \mathcal{E} \ \left(\frac{\partial \ln p}{\partial \theta^*[k]} \frac{\partial \ln p}{\partial \theta[l]} \right) = [\mathbf{I}(\boldsymbol{\theta})]_{k,l} \ . \end{aligned}$$

Other cases of k and l are handled similarly and this completes the proof for the first part of Theorem 1.

To prove the second part, we note that the Cauchy-Schwartz inequality holds exactly if and only if $g(\mathbf{x}) = ch^*(\mathbf{x})$ for some constant c. In (A-3) where the inequality was applied, this condition means that equality is attained if and only if

$$\left[\mathbf{a}^H(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})\right]^* = c(\boldsymbol{\theta})\frac{\partial \ln p}{\partial \boldsymbol{\theta}}^T \mathbf{b} = c(\boldsymbol{\theta})\frac{\partial \ln p}{\partial \boldsymbol{\theta}}^T \mathbf{I}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}^H \mathbf{a}$$

or

$$(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^H \mathbf{a} = c(\boldsymbol{\theta})\frac{\partial \ln p}{\partial \boldsymbol{\theta}^*}^H \mathbf{I}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}^H \mathbf{a}$$

holds for all $\mathbf{a} \in \mathbb{C}^p$. Hence we need

$$c^*(\boldsymbol{\theta})\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}\mathbf{I}^{-1}(\boldsymbol{\theta})\frac{\partial \ln p}{\partial \boldsymbol{\theta}^*} = \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \ . \qquad (\mathrm{A}-6)$$

Let $\mathbf{T} = \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}\mathbf{I}^{-1}(\boldsymbol{\theta})$, a q by p matrix, so that (A-6) can be written as

$$c^*(\boldsymbol{\theta})\sum_{l=1}^{p}[\mathbf{T}]_{k,l}\frac{\partial \ln p}{\partial \theta^*[l]} = \hat{\alpha}[k] - \alpha[k]$$

for $1 \le k \le q$. Differentiating both sides w.r.t. $\theta[n]$ for $1 \le n \le p$ and taking expected values w.r.t. $\mathbf{x}$ we obtain (using the chain rule and regularity condition)

$$c^*(\boldsymbol{\theta})\sum_{l=1}^{p}[\mathbf{T}]_{k,l}[\mathbf{I}(\boldsymbol{\theta})]_{l,n} = \left[\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}\right]_{k,n} .$$

But the left hand side of above equation simplifies to

$$c^*(\boldsymbol{\theta})[\mathbf{T}\mathbf{I}(\boldsymbol{\theta})]_{k,n} = c^*(\boldsymbol{\theta})\left[\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}\mathbf{I}^{-1}(\boldsymbol{\theta})\mathbf{I}(\boldsymbol{\theta})\right]_{k,n} = c^*(\boldsymbol{\theta})\left[\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}\right]_{k,n} ,$$

so that we need $c(\boldsymbol{\theta}) = 1$. Therefore equality holds in Theorem 1 if and only if

$$\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}\mathbf{I}^{-1}(\boldsymbol{\theta})\frac{\partial \ln p}{\partial \boldsymbol{\theta}^*} = \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \ . \qquad (\mathrm{A}-7)$$

Taking expectations in (A-7), and using the regularity condition in (A-1), we conclude that $\hat{\boldsymbol{\alpha}}$ is unbiased. Although (A-7) in itself is seldom of much use, it simplifies when $\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}$. In such a case, we have

$$\frac{\partial \ln p}{\partial \boldsymbol{\theta}^*} = \mathbf{I}(\boldsymbol{\theta})\left[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right] , \qquad (\mathrm{A}-8)$$

which helps identify the efficient estimator as stated in Theorem 1. Finally, it is easy to establish that the real-valued estimator $\hat{\boldsymbol{\theta}}^{(r)}$ formed from $\hat{\boldsymbol{\theta}}$ is unbiased and attains the real CRLB given by $\mathbf{I}^{-1}(\boldsymbol{\theta}^{(r)})$, i.e. efficient in the usual sense [5, 10].

Appendix B

Lemma B.1 *The usual approach to CRLB for the unknown parameters* $\underline{\theta}$ *(or* $\underline{\alpha} = \underline{\mathrm{g}}(\underline{\theta})$*) is to consider a vector of real-valued parameters,*

$$\theta^{(r)} = \begin{bmatrix} \theta_{1r} \\ \theta_{1i} \\ \theta_2 \end{bmatrix}, \quad \alpha^{(r)} = \begin{bmatrix} \alpha_{1r} \\ \alpha_{1i} \\ \alpha_2 \end{bmatrix}, \quad \hat{\alpha}^{(r)} = \begin{bmatrix} \hat{\alpha}_{1r} \\ \hat{\alpha}_{1i} \\ \hat{\alpha}_2 \end{bmatrix},$$

and use the available theory. The following conditions are then assumed in a simple derivation of the CRLB theorem [5].

Unbiased estimates. $\mathcal{E}\left(\hat{\alpha}^{(r)}\right) = \alpha^{(r)}$, or

$$\mathcal{E}\begin{bmatrix} \hat{\alpha}_{1r} \\ \hat{\alpha}_{1i} \\ \hat{\alpha}_2 \end{bmatrix} = \begin{bmatrix} \alpha_{1r} \\ \alpha_{1i} \\ \alpha_2 \end{bmatrix} . \tag{B-1}$$

Regularity conditions. $\mathcal{E}\left(\frac{\partial \ln p}{\partial \theta^{(r)}}\right) = \mathbf{0}$, or

$$\mathcal{E}\begin{bmatrix} \frac{\partial \ln p}{\partial \theta_{1r}} \\ \frac{\partial \ln p}{\partial \theta_{1i}} \\ \frac{\partial \ln p}{\partial \theta_2} \end{bmatrix} = \mathbf{0} . \tag{B-2}$$

This basically states that the order of differentiation w.r.t. $\theta^{(r)}$ and integration over x can be interchanged, more elaborately (B-2) means that

$$\frac{\partial \alpha^{(r)}[k]}{\partial \theta^{(r)}[l]} = \int_{\mathbf{x}} \hat{\alpha}^{(r)}[k] \, \frac{\partial p\left(\mathrm{x}\,;\,\theta^{(r)}\right)}{\partial \theta^{(r)}[l]} \mathrm{dx} . \tag{B-3}$$

for $1 \le k \le q$, $1 \le l \le p$. Next, for $1 \le k \le q$, $1 \le l \le p$, we also have

$$\mathcal{E}\left(\frac{\partial \ln p}{\partial \theta^{(r)}[k]} \frac{\partial \ln p}{\partial \theta^{(r)}[l]}\right) = -\mathcal{E}\left(\frac{\partial^2 \ln p}{\partial \theta^{(r)}[k] \, \partial \theta^{(r)}[l]}\right) . \tag{B-4}$$

Lemma B.2 *Recall that* $\mathrm{g}(\boldsymbol{\theta}) = \boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_1^* \\ \boldsymbol{\alpha}_2 \end{bmatrix}$, $\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_1^* \\ \boldsymbol{\theta}_2 \end{bmatrix}$, $p = 2p_1 + p_2$, $q = 2q_1 + q_2$. *Then*

$$\frac{\partial g[k]}{\partial \theta[l]} = \int_{\mathbf{X}} \hat{\alpha}[k] \, \frac{\partial \ln p}{\partial \theta[l]} \, p(\mathrm{x}\,;\,\boldsymbol{\theta}) \; d\mathrm{x} \; . \qquad (\mathrm{B}-5)$$

for $1 \le k \le q$, $1 \le l \le p$. *If* $\theta[l]$ *is real-valued, the derivative has the usual meaning, if* $\theta[l]$ *is complex-valued, the derivative is as defined in section* 2.

Proof. Consider the case $1 \le k \le q_1$, $1 \le l \le p_1$. Then

$$\begin{aligned} \frac{\partial g[k]}{\partial \theta[l]} &= \frac{\partial}{\partial \theta[l]} \, \mathcal{E}\left(\hat{\alpha}[k]\right) = \frac{\partial}{\partial \theta[l]} \int_{\mathbf{X}} \hat{\alpha}[k] \, p(\mathrm{x}\,;\,\boldsymbol{\theta}) \; \mathrm{dx} \\ &= \frac{1}{2}\left(\frac{\partial}{\partial \theta_r[l]} - \jmath \frac{\partial}{\partial \theta_i[l]}\right) \int_{\mathbf{X}} \left(\hat{\alpha}_r[k] + \jmath \hat{\alpha}_i[k]\right) \, p(\mathrm{x}\,;\,\boldsymbol{\theta}) \; \mathrm{dx} \\ &= \int_{\mathbf{X}} \left(\hat{\alpha}_r[k] + \jmath \hat{\alpha}_i[k]\right) \frac{1}{2}\left(\frac{\partial p}{\partial \theta_r[l]} - \jmath \frac{\partial p}{\partial \theta_i[l]}\right) \mathrm{dx} \\ &= \int_{\mathbf{X}} \hat{\alpha}[k] \frac{\partial p}{\partial \theta[l]} \, \mathrm{dx} = \int_{\mathbf{X}} \hat{\alpha}[k] \frac{\partial \ln p}{\partial \theta[l]} \, p(\mathrm{x}\,;\,\boldsymbol{\theta}) \; \mathrm{dx} \; . \end{aligned}$$

Step 1 followed from equation (B-1), steps 2, 3, and 5 by definitions, and step 4 follows by repeated application of (B-3). Other cases of k and l are handled similarly. This proves the lemma.

Lemma B.3 *(Cauchy-Schwartz inequality). Let* $g(\mathrm{x})$, $h(\mathrm{x})$ *be arbitrary complex-valued functions of* $\mathrm{x} \in \mathbb{R}^N$ *or* $\mathbb{C}^N$, *and let* $w(\mathrm{x})$ *be a non-negative real-valued function. Then*

$$\left|\int_{\mathbf{X}} g(\mathrm{x}) \, h(\mathrm{x}) \, w(\mathrm{x}) \, d\mathrm{x}\right|^2 \le \left[\int_{\mathbf{X}} |g(\mathrm{x})|^2 \, w(\mathrm{x}) \, d\mathrm{x}\right] \left[\int_{\mathbf{X}} |h(\mathrm{x})|^2 \, w(\mathrm{x}) \, d\mathrm{x}\right] \; , \qquad (\mathrm{B}-6)$$

with strict equality if and only if $g(\mathrm{x}) = c\, h^*(\mathrm{x})$ *for all* x, *for some complex constant* c.

Proof. easy to check.

Lemma B.4 *Let* $\begin{bmatrix} \mathbf{A} & \mathbf{B}^* \\ \mathbf{B} & \mathbf{A}^* \end{bmatrix}$ *be a nonsingular Hermitian matrix where* $\mathbf{A}$ *is a* p_1 *by* p_1 *nonsingular Hermitian matrix and* $\mathbf{B}$ *is a* p_1 *by* p_1 *complex symmetric matrix. Then*

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^* \\ \mathbf{B} & \mathbf{A}^* \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C} & \mathbf{D}^* \\ \mathbf{D} & \mathbf{C}^* \end{bmatrix} , \qquad (\mathrm{B}-7)$$

where $\mathbf{C} = \left(\mathbf{A} - \mathbf{B}^*\mathbf{A}^{*-1}\mathbf{B}\right)^{-1}$, *and* $\mathbf{D} = -\,\mathbf{A}^{*-1}\mathbf{B}\mathbf{C} = -\,\mathbf{C}^*\mathbf{B}\mathbf{A}^{-1}$.

Proof. Using the partitioned matrix inversion lemma we have

$$\begin{aligned}\begin{bmatrix} \mathbf{A} & \mathbf{B}^* \\ \mathbf{B} & \mathbf{A}^* \end{bmatrix}^{-1} &= \begin{bmatrix} \left(\mathbf{A}-\mathbf{B}^*\mathbf{A}^{*-1}\mathbf{B}\right)^{-1} & -\left(\mathbf{A}-\mathbf{B}^*\mathbf{A}^{*-1}\mathbf{B}\right)^{-1}\mathbf{B}^*\mathbf{A}^{*-1} \\ -\left(\mathbf{A}^*-\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^*\right)^{-1}\mathbf{B}\mathbf{A}^{-1} & \left(\mathbf{A}^*-\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^*\right)^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{C} & -\mathbf{C}\mathbf{B}^*\mathbf{A}^{*-1} \\ -\mathbf{C}^*\mathbf{B}\mathbf{A}^{-1} & \mathbf{C}^* \end{bmatrix} = \begin{bmatrix} \mathbf{C} & -\mathbf{A}^{-1}\mathbf{B}^*\mathbf{C}^* \\ -\mathbf{A}^{*-1}\mathbf{B}\mathbf{C} & \mathbf{C}^* \end{bmatrix},\end{aligned}$$

because the matrix is Hermitian. This proves the lemma.

Lemma B.5 *Let* $\mathrm{x} \sim \mathcal{CN}(\mathrm{s}(\underline{\theta}), \mathrm{C}(\underline{\theta}))$ *i.e., the observed data is complex Gaussian distributed with information in the mean and covariance. The elements of the complex Fisher information matrix are given by*

$$[\mathbf{I}(\boldsymbol{\theta})]_{k,l} = \frac{\partial \mathrm{s}}{\partial\theta[k]}^H \mathrm{C}^{-1}\frac{\partial \mathrm{s}}{\partial\theta[l]} + \frac{\partial \mathrm{s}}{\partial\theta^*[l]}^H \mathrm{C}^{-1}\frac{\partial \mathrm{s}}{\partial\theta^*[k]} + tr\left\{\frac{\partial \mathrm{C}}{\partial\theta^*[k]}\mathrm{C}^{-1}\frac{\partial \mathrm{C}}{\partial\theta[l]}\mathrm{C}^{-1}\right\} . \qquad (\mathrm{B}-8)$$

Proof. It is well known that the real Fisher information matrix $\mathbf{I}(\boldsymbol{\theta}^{(r)})$ can be computed by [5, 6]

$$\left[\mathbf{I}(\boldsymbol{\theta}^{(r)})\right]_{k,l} = 2\,\mathrm{Re}\left\{\frac{\partial \mathrm{s}}{\partial\theta^{(r)}[k]}^H \mathrm{C}^{-1}\frac{\partial \mathrm{s}}{\partial\theta^{(r)}[l]}\right\} + \mathrm{tr}\left\{\frac{\partial \mathrm{C}}{\partial\theta^{(r)}[k]}\mathrm{C}^{-1}\frac{\partial \mathrm{C}}{\partial\theta^{(r)}[l]}\mathrm{C}^{-1}\right\} \qquad (\mathrm{B}-9)$$

for $1 \le k,\ l \le p$. Consider first the case of for $1 \le k \le p_1$, $1 \le l \le p_1$ in (B-8). Then, by definition, we have

$$\begin{aligned}[\mathbf{I}(\boldsymbol{\theta})]_{k,l} &= \frac{1}{4}\left[\left[\mathbf{I}(\boldsymbol{\theta}^{(r)})\right]_{k,l} + \left[\mathbf{I}(\boldsymbol{\theta}^{(r)})\right]_{k+p_1,l+p_1}\right] + \frac{\jmath}{4}\left[\left[\mathbf{I}(\boldsymbol{\theta}^{(r)})\right]_{k+p_1,l} - \left[\mathbf{I}(\boldsymbol{\theta}^{(r)})\right]_{k,l+p_1}\right] \\ &= \frac{1}{2}\left[\mathrm{Re}\left\{\frac{\partial \mathrm{s}}{\partial\theta_r[k]}^H \mathrm{C}^{-1}\frac{\partial \mathrm{s}}{\partial\theta_r[l]}\right\} + \mathrm{Re}\left\{\frac{\partial \mathrm{s}}{\partial\theta_i[k]}^H \mathrm{C}^{-1}\frac{\partial \mathrm{s}}{\partial\theta_i[l]}\right\}\right] \\ &\quad + \frac{\jmath}{2}\left[\mathrm{Re}\left\{\frac{\partial \mathrm{s}}{\partial\theta_i[k]}^H \mathrm{C}^{-1}\frac{\partial \mathrm{s}}{\partial\theta_r[l]}\right\} - \mathrm{Re}\left\{\frac{\partial \mathrm{s}}{\partial\theta_r[k]}^H \mathrm{C}^{-1}\frac{\partial \mathrm{s}}{\partial\theta_i[l]}\right\}\right] \\ &\quad + \frac{1}{4}\mathrm{tr}\left\{\frac{\partial \mathrm{C}}{\partial\theta_r[k]}\mathrm{C}^{-1}\frac{\partial \mathrm{C}}{\partial\theta_r[l]}\mathrm{C}^{-1} + \frac{\partial \mathrm{C}}{\partial\theta_i[k]}\mathrm{C}^{-1}\frac{\partial \mathrm{C}}{\partial\theta_i[l]}\mathrm{C}^{-1}\right\} \\ &\quad + \frac{\jmath}{4}\mathrm{tr}\left\{\frac{\partial \mathrm{C}}{\partial\theta_i[k]}\mathrm{C}^{-1}\frac{\partial \mathrm{C}}{\partial\theta_r[l]}\mathrm{C}^{-1} - \frac{\partial \mathrm{C}}{\partial\theta_r[k]}\mathrm{C}^{-1}\frac{\partial \mathrm{C}}{\partial\theta_i[l]}\mathrm{C}^{-1}\right\} .\end{aligned}$$

Then using the fact that $2\mathrm{Re}\{\mathrm{a}^H\mathbf{T}\mathrm{b}\} = \mathrm{a}^H\mathbf{T}\mathrm{b} + \mathrm{b}^H\mathbf{T}\mathrm{a}$ we have

$$[\mathbf{I}(\boldsymbol{\theta})]_{k,l} = \frac{1}{4}\left[\frac{\partial \mathrm{s}}{\partial\theta_r[k]}^H \mathrm{C}^{-1}\frac{\partial \mathrm{s}}{\partial\theta_r[l]} + \frac{\partial \mathrm{s}}{\partial\theta_r[l]}^H \mathrm{C}^{-1}\frac{\partial \mathrm{s}}{\partial\theta_r[k]} + \frac{\partial \mathrm{s}}{\partial\theta_i[k]}^H \mathrm{C}^{-1}\frac{\partial \mathrm{s}}{\partial\theta_i[l]} + \frac{\partial \mathrm{s}}{\partial\theta_i[l]}^H \mathrm{C}^{-1}\frac{\partial \mathrm{s}}{\partial\theta_i[k]}\right]$$

$$
\begin{aligned}
& + \frac{\jmath}{4}\left[\frac{\partial \mathbf{s}}{\partial\theta_i[k]}^H \mathbf{C}^{-1}\frac{\partial \mathbf{s}}{\partial\theta_r[l]} + \frac{\partial \mathbf{s}}{\partial\theta_r[l]}^H \mathbf{C}^{-1}\frac{\partial \mathbf{s}}{\partial\theta_i[k]} - \frac{\partial \mathbf{s}}{\partial\theta_r[k]}^H \mathbf{C}^{-1}\frac{\partial \mathbf{s}}{\partial\theta_i[l]} - \frac{\partial \mathbf{s}}{\partial\theta_i[l]}^H \mathbf{C}^{-1}\frac{\partial \mathbf{s}}{\partial\theta_r[k]}\right] \\
& + \frac{1}{2}\mathrm{tr}\left\{\frac{\partial \mathbf{C}}{\partial\theta_r[k]}\mathbf{C}^{-1}\left(\frac{1}{2}\frac{\partial \mathbf{C}}{\partial\theta_r[l]} - \frac{\jmath}{2}\frac{\partial \mathbf{C}}{\partial\theta_i[l]}\right)\mathbf{C}^{-1}\right\} \\
& + \frac{\jmath}{2}\mathrm{tr}\left\{\frac{\partial \mathbf{C}}{\partial\theta_i[k]}\mathbf{C}^{-1}\left(\frac{1}{2}\frac{\partial \mathbf{C}}{\partial\theta_r[l]} - \frac{\jmath}{2}\frac{\partial \mathbf{C}}{\partial\theta_i[l]}\right)\mathbf{C}^{-1}\right\} \\
= \; & \frac{1}{2}\frac{\partial \mathbf{s}}{\partial\theta_r[k]}^H \mathbf{C}^{-1}\left(\frac{1}{2}\frac{\partial \mathbf{s}}{\partial\theta_r[l]} - \frac{\jmath}{2}\frac{\partial \mathbf{s}}{\partial\theta_i[l]}\right) + \frac{\jmath}{2}\frac{\partial \mathbf{s}}{\partial\theta_i[k]}^H \mathbf{C}^{-1}\left(\frac{1}{2}\frac{\partial \mathbf{s}}{\partial\theta_r[l]} - \frac{\jmath}{2}\frac{\partial \mathbf{s}}{\partial\theta_i[l]}\right) \\
& + \frac{1}{2}\frac{\partial \mathbf{s}}{\partial\theta_r[l]}^H \mathbf{C}^{-1}\left(\frac{1}{2}\frac{\partial \mathbf{s}}{\partial\theta_r[k]} + \frac{\jmath}{2}\frac{\partial \mathbf{s}}{\partial\theta_i[k]}\right) - \frac{\jmath}{2}\frac{\partial \mathbf{s}}{\partial\theta_i[l]}^H \mathbf{C}^{-1}\left(\frac{1}{2}\frac{\partial \mathbf{s}}{\partial\theta_r[k]} + \frac{\jmath}{2}\frac{\partial \mathbf{s}}{\partial\theta_i[k]}\right) \\
& + \mathrm{tr}\left\{\left(\frac{1}{2}\frac{\partial \mathbf{C}}{\partial\theta_r[k]} + \frac{\jmath}{2}\frac{\partial \mathbf{C}}{\partial\theta_i[k]}\right)\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial\theta[l]}\mathbf{C}^{-1}\right\} \\
= \; & \left(\frac{1}{2}\frac{\partial \mathbf{s}}{\partial\theta_r[k]} - \frac{\jmath}{2}\frac{\partial \mathbf{s}}{\partial\theta_i[k]}\right)^H \mathbf{C}^{-1}\frac{\partial \mathbf{s}}{\partial\theta[l]} + \left(\frac{1}{2}\frac{\partial \mathbf{s}}{\partial\theta_r[l]} + \frac{\jmath}{2}\frac{\partial \mathbf{s}}{\partial\theta_i[l]}\right)^H \mathbf{C}^{-1}\frac{\partial \mathbf{s}}{\partial\theta^*[k]} \\
& + \mathrm{tr}\left\{\frac{\partial \mathbf{C}}{\partial\theta^*[k]}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial\theta[l]}\mathbf{C}^{-1}\right\} \\
= \; & \frac{\partial \mathbf{s}}{\partial\theta[k]}^H \mathbf{C}^{-1}\frac{\partial \mathbf{s}}{\partial\theta[l]} + \frac{\partial \mathbf{s}}{\partial\theta^*[l]}^H \mathbf{C}^{-1}\frac{\partial \mathbf{s}}{\partial\theta^*[k]} + \mathrm{tr}\left\{\frac{\partial \mathbf{C}}{\partial\theta^*[k]}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial\theta[l]}\mathbf{C}^{-1}\right\} .
\end{aligned}
$$

Note that $\frac{\partial \mathbf{C}}{\partial\theta[l]}$ will not be Hermitian even though $\mathbf{C}$ is Hermitian, hence the third term in above expression will be complex in general. (Note that sometimes it may be easier to compute the partials w.r.t. $\mathbf{C}^{-1}$ so that the third term may be replaced by $\mathrm{tr}\left\{\frac{\partial \mathbf{C}^{-1}}{\partial\theta^*[k]}\mathbf{C}\frac{\partial \mathbf{C}^{-1}}{\partial\theta[l]}\mathbf{C}\right\}$.)

The other cases of k and l can be handled similarly so that this completes the proof of the lemma.

Lemma B.6 *Let* $\mathbf{z}$ *and* $\mathbf{a}$ *be* $N \times 1$ *complex-valued vectors, and let* $\mathbf{W}$ *be* $N \times N$, *complex-valued Hermitian matrix. Then,*

$$
\begin{aligned}
\frac{\partial \mathbf{a}^H\mathbf{z}}{\partial \mathbf{z}} &= \mathbf{a}^* \quad , \quad \frac{\partial \mathbf{a}^H\mathbf{z}}{\partial \mathbf{z}^*} = \mathbf{0} \\
\frac{\partial \mathbf{z}^H\mathbf{a}}{\partial \mathbf{z}} &= \mathbf{0} \quad , \quad \frac{\partial \mathbf{z}^H\mathbf{a}}{\partial \mathbf{z}^*} = \mathbf{a} \\
\frac{\partial \mathbf{z}^H\mathbf{W}\mathbf{z}}{\partial \mathbf{z}} &= (\mathbf{W}\mathbf{z})^* \quad , \quad \frac{\partial \mathbf{z}^H\mathbf{W}\mathbf{z}}{\partial \mathbf{z}^*} = \mathbf{W}\mathbf{z}
\end{aligned}
$$

Proof. See [1].