The Cramer-Rao Lower Bound in the Phase Retrieval Problem

Radu Balan
Department of Mathematics and
Center for Scientific Computations and Mathematical Modeling
University of Maryland
College Park, MD 20742
Email: rvbalan@math.umd.edu , http://www.math.umd.edu/∼rvbalan/

***Abstract*—This paper presents an analysis of Cramer-Rao lower bounds (CRLB) in the phase retrieval problem. Previous papers derived Fisher Information Matrices for the phaseless reconstruction setup. Two estimation setups are presented. In the first setup the global phase of the unknown signal is determined by a correlation condition with a fixed reference signal. In the second setup an oracle provides the optimal global phase. The CRLB is derived for each of the two approaches. Surprisingly (or maybe not) they are different.**

I. Introduction

The *phase retrieval problem* (also known as the *phaseless reconstruction problem*) can simply be stated as the reconstruction of a signal from the magnitudes of a redundant representation (see [4]). Recently, there has been made progress on three problems: injectivity conditions, stability bounds, and reconstruction algorithms. In the next section we review existing results on the second problem, stability bounds, which is the focus of this paper. Recent papers computed the Fisher Information Matrix (FIM) for two noise mixing models: the additive white Gaussian noise model (AWGN), and a non-additive white Gaussian noise model (non-AWGN). This paper derives Cramer-Rao Lower Bounds (CRLB) for more general setups. As will become clear later, the CRLB is not just simply the inverse of FIM as is the case in the standard estimation theory ([10]). The difficulty comes from the non-identifiability of the general problem. To address the identifiability issue two estimation setups are proposed below. Associated to each of the two setups a CRLB is derived in section III.

Consider the case of the n-dimensional Hilbert space $H = \mathbb{C}^n$ as the signal space. Fix a frame (i.e. a spanning set in this finite dimensional case - see [9] for more information on frames) $\mathcal{F} = \{f_1, \ldots, f_m\}$ in H. For an unknown signal $x \in \mathbb{C}^n$ consider a measurement process $y = (y_k)_{1\leq k\leq r}$ where the distribution of y depends on the magnitudes of $\langle x, f_k\rangle$:

$$p(y; x) = f(s_1, \ldots, s_m, y) \quad , \quad s_k = |\langle x, f_k\rangle|, 1 \leq k \leq m \tag{1.1}$$

as in Figure 1. For instance this is the case for the measurements

$$y_k = |\langle x, f_k\rangle + \mu_k|^a + \nu_k \quad , \quad 1 \leq k \leq r = m, \tag{1.2}$$

where $a > 0$ is a fixed exponent (typically 1 or 2) and $(\mu_k)_{1\leq k\leq m}, (\nu_k)_{1\leq k\leq m}$ are realizations of independent noise

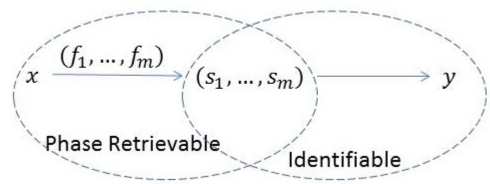

Fig. 1. Measurement process

processes with μ_k circular. In the absence of noise it is obvious that the signal x cannot be recovered from the set of parameters $(s_k = |\langle x, f_k\rangle|)_{1\leq k\leq m}$ since $e^{i\varphi}x$ will produce the same set of intermediary variables $(s_1, \ldots, s_m)$. The phase retrieval problem refers to estimating the original signal x up to a global phase factor. In order to formalize this concept, consider the equivalence relation between two vectors $x, y \in \mathbb{C}^n$: $x \sim y$ if there is a real t so that $x = e^{it}y$. Let $\widehat{\mathbb{C}^n} = \mathbb{C}^n / \sim$ denote the set of equivalence classes. A frame $\mathcal{F}$ is called *phase retrievable* if the nonlinear map

$$\alpha : \widehat{\mathbb{C}^n} \to \mathbb{R}^m \quad , \quad \alpha(x) = (|\langle x, f_k\rangle|)_{1\leq k\leq m} \tag{1.3}$$

is injective. Throughout this paper we assume the frame $\mathcal{F}$ is phase retrievable, unless otherwise said. Additionally we assume the parameters $(s_1, \ldots, s_m)$ are *identifiable* from the measurement y, meaning that if $f(s_1^1, \ldots, s_m^1, y) = f(s_1^2, \ldots, s_m^2, y)$ for all $y \in \mathbb{R}^r$ then $s_1^1 = s_1^2, \ldots, s_m^1 = s_m^2$.

The quotient space $\widehat{\mathbb{C}^n} = \mathbb{C}^n / \sim$ can be thought of as the quotient $\mathbb{C}^n / T^1$. As described in [6], [7] the quotient space admits two metric space structures which are topologically equivalent (generate the same open stes), but are not Lipschitz equivalent. In [2] we related the lower Lipschitz constant of the map α to the FIM of a non-AWGN, whereas in [6] we related the lower Lipschitz constant of the nonlinear map

$$\beta : \widehat{\mathbb{C}^n} \to \mathbb{R}^m \quad , \quad \beta(x) = (|\langle x, f_k\rangle|^2)_{1\leq k\leq m}$$

to the FIM of the AWGN model. We review these results in the next section. Now we present the two estimation setups considered in this paper.

A. Setup I: The Reference Signal Based Estimation

The first setup uses a reference signal to fix the global phase: Fix $z_0 \in \mathbb{C}^n$ a unit-norm vector, $\|z_0\| = 1$, and define the following set

$$V_{z_0} = \{x \in \mathbb{C}^n \ : \ \mathrm{imag}(\langle x, z_0\rangle) = 0 \ , \ \mathrm{real}(\langle x, z_0\rangle) > 0 \ \}. \quad (1.4)$$

For this setup we assume the unknown and to-be-estimated signal x is not orthogonal to z_0. In this case, from its equivalence class $\hat{x}$ we pick the representative that lies in V_{z_0}. Specifically we assume from the outset that the signal to-be-estimated x correlates positively with z_0. Note that this is a mild requirement since we can never find the global phase of the true signal x just from magnitude measurements $\alpha(x)$. The only loss of generality is due to the non-orthogonality assumption: $\langle x, z_0\rangle \neq 0$. In effect we exclude a linear subspace of $\mathbb{C}^n$ of complex dimension $n-1$. Under the assumption $x \in V_{z_0}$, when the frame is phase retrievable, the measurements (1.2) define an identifiable process. In this case an estimator is given by a map

$$o : \mathbb{R}^r \rightarrow E_{z_0} \quad (1.5)$$

where

$$E_{z_0} = \{x \in \mathbb{C}^n \ : \ \mathrm{imag}(\langle x, z_0\rangle) = 0\} = \mathrm{span}_{\mathbb{R}}(V_{z_0}). \quad (1.6)$$

Note E_{z_0} is a real linear space of real dimension $2n-1$. The estimator o is said *unbiased with respect to the first setup* if

$$\mathbb{E}[o(y); x] = x \quad , \quad \forall x \in V_{z_0}. \quad (1.7)$$

Theorem 3.2 in section III presents the Cramer-Rao Lower bound associated to this setup.

B. Setup II: The Oracle Based Signal Estimation

The second setup uses an oracle to provide the global phase. Specifically, the estimation procedure is performed in two steps: first a nonlinear function

$$o : \mathbb{R}^r \rightarrow \mathbb{C}^n$$

that estimates the equivalence class of the unknown signal x. Technically $o : \mathbb{R}^r \rightarrow \widehat{\mathbb{C}^n}$ but then for each $y \in \mathbb{R}^r$ choose differentiably a representative in the class $o(y)$. We overload the notation and use the same letter o to denote this latter differentiable estimator. Then an oracle computes the phase that minimizes the approximation error

$$\min_t \|x - e^{it} o(y)\|.$$

We choose the Euclidean norm in which case the optimal phase is given by $\langle x, o(y)\rangle / |\langle x, o(y)\rangle|$ (our scalar product is linear in the first term and anti-linear in the second term). Thus, the overall estimator has the form

$$\tilde{o} : \mathbb{R}^r \rightarrow \mathbb{C}^n \ , \ \tilde{o}(y; x) = \frac{\langle x, o(y)\rangle}{|\langle x, o(y)\rangle|} o(y). \quad (1.8)$$

The estimator $\tilde{o}$ is said *unbiased with respect to the second setup* (or simply, unbiased) if

$$\mathbb{E}\left[\frac{\langle x, o(y)\rangle}{|\langle x, o(y)\rangle|} o(y); x\right] = x \quad , \quad \forall x \in \mathbb{C}^n. \quad (1.9)$$

Notice the unbiasedness condition is slightly stronger than in (1.7) because it applies to all vectors $x \in \mathbb{C}^n$ without restriction. On the other hand it requires access to the unknown signal x in order to correct for the unknown global phase factor. Theorem 3.3 in section III presents the CRLB for this setup.

II. Existing Results

A. Notations

First we present the "realification" procedure as introduced in [1]. Our complex scalar product is given by $\langle a, b\rangle = a_1\overline{b_1} + \cdots + a_n\overline{b_n}$, where $a, b \in \mathbb{C}^n$. Throughout the paper the letter I denotes the identity matrix of appropriate size, whereas $\mathbb{I}$ denotes the Fisher Information Matrix. Consider the $\mathbb{R}$-linear map

$$\jmath : \mathbb{C}^n \rightarrow \mathbb{R}^{2n} \ , \ \jmath(x) = \begin{bmatrix} \mathrm{real}(x) \\ \mathrm{imag}(x) \end{bmatrix}.$$

We use Latin letters to denote C^n-vectors, and we use Greek letters to denote their $\mathbb{R}^{2n}$ correspondents. One exception: $y \in \mathbb{R}^m$ will always denote the real vector of m measurements. Let J denote the $2n \times 2n$ matrix

$$J = \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix},$$

where I denotes the $n \times n$ identity matrix. Note $\jmath(ix) = J\jmath(x)$ for every $x \in \mathbb{C}^n$. Let $\xi = \jmath(x)$, $\zeta_0 = \jmath(z_0$, and $\varphi_k = \jmath(f_k)$ for $1 \le k \le m$. For the reference signal z_0 we introduced sets V_{z_0} and E_{z_0}. Their counterparts in $\mathbb{R}^{2n}$ are denoted by $\mathcal{V}_{\zeta_0}$ and $\mathcal{E}_{\zeta_0}$:

$$\begin{aligned} \mathcal{V}_{\zeta_0} &= \{\eta \in \mathbb{R}^{2n} \ : \ \langle \eta, J\zeta_0\rangle = 0 \ , \ \langle \eta, \zeta_0\rangle > 0\} \\ \mathcal{E}_{\zeta_0} &= \{\eta \in \mathbb{R}^{2n} \ : \ \langle \eta, J\zeta_0\rangle = 0\} = \{J\zeta_0\}^{\perp}. \end{aligned} \quad (2.1)$$

Let $\Pi^{\perp}_{J\xi}$ and $\Pi^{\perp}_{J\zeta_0}$ denote the orthogonal projections onto the orthogonal complements of $J\xi$, and $J\zeta_0$, respectively:

$$\begin{aligned} \Pi^{\perp}_{J\xi} &= I - \frac{1}{\|\xi\|^2} J\xi\xi^T J^T \\ \Pi^{\perp}_{J\zeta_0} &= I - J\zeta_0\zeta_0^T J^T, \end{aligned}$$

where I denotes the $2n \times 2n$ identity matrix. To each frame vector f_k we associate the rank-2 matrix

$$\Phi_k = \varphi_k\varphi_k^T + J\varphi_k\varphi_k^T J^T.$$

A direct computation shows that $|\langle x, f_k\rangle| = \sqrt{\langle \Phi_k\xi, \xi\rangle}$. Let $\mathcal{R}(\xi)$ denote the following matrix

$$\mathcal{R}(\xi) = \sum_{k=1}^{m} \Phi_k\xi\xi^T\Phi_k.$$

For a square matrix M, we let $M^{\dagger}$ denote the Moore-Penrose pseudoinverse of M. For the estimator $o : \mathbb{R}^r \rightarrow E_{z_0}$ we let ω denote its realification

$$\omega : \mathbb{R}^r \rightarrow \mathcal{E}_{\zeta_0} \quad , \quad \omega(y) = \jmath(o(y)).$$

As shown in Theorem 3.2 below, the estimator ω is unbiased with respect to the first setup iff

$$\mathbb{E}[\omega(y); \xi] = \xi \quad , \quad \forall \xi \in \mathcal{V}_{\zeta_0}. \quad (2.2)$$

For the second setup, let $o : \mathbb{R}^r \to \mathbb{C}^n$ be the signal class estimator, and let $\omega = \jmath(o)$ denote its realification. Note that

$$\jmath(e^{it}o(y)) = cos(t)\omega(y) + sin(t)J\omega(y) =: U(t)\omega(y), \quad (2.3)$$

where $\{U(t) := cos(t)I + sin(t)J \; ; \; 0 \leq t < 2\pi\}$ is a 1-dimensional group of orthogonal matrices. As we show in Theorem 3.3 below, the unbiasedness condition (1.9) turns into:

$$\mathbb{E}\left[\frac{\langle \xi, \omega(y)\rangle}{\sqrt{(\langle \xi, \omega(y)\rangle)^2 + (\langle J\xi, \omega(y)\rangle)^2}}\omega(y) + \frac{\langle \xi, J\omega(y)\rangle}{\sqrt{(\langle \xi, \omega(y)\rangle)^2 + (\langle J\xi, \omega(y)\rangle)^2}}J\omega(y);\xi\right] = \xi \quad (2.4)$$

for every $\xi \in \mathbb{R}^{2n}$.

B. Existing FIM and CRLB Results

In this subsection we review existing expressions for the Fisher Information Matrix for two stochastic models and an existing Cramer-Rao Lower Bound.

The first model is the Additive White Gaussian Noise (AWGN) model

$$y_k = |\langle x, f_k\rangle|^2 + \nu_k \quad , \quad 1 \leq k \leq m, \quad (2.5)$$

where $(\nu_k)_{1\leq k\leq m}$ are independent and identically distributed (i.i.d.) realizations of a normal random variable of zero mean and variance σ^2. The second process is a non-Additive White Gaussian Noise (nonAWGN) model where the noise is added prior to taking the absolute value:

$$y_k = |\langle x, f_k\rangle + \mu_k|^2 \quad , \quad 1 \leq k \leq m, \quad (2.6)$$

where $(\mu_k)_{1\leq k\leq m}$ are i.i.d. realizations of a Gaussian complex process with zero mean and variance ρ^2. For either stochastic model we present the Fisher Information Matrix (FIM). The FIM is expressed in terms of the real vector $\xi = \jmath(x)$. Once the likelihood function $P(y;\xi)$ has been established, the FIM is computed by (see [10])

$$\mathbb{I}(\xi) = \mathbb{E}\left[(\nabla_\xi \log P(y;\xi))(\nabla_\xi \log P(y;\xi))^T\right]. \quad (2.7)$$

Following [8] and [1] for the AWGN model (2.5) we obtain:

$$\mathbb{I}^{\text{AWGN}}(\xi) = \frac{4}{\sigma^2}\mathcal{R}(\xi) = \frac{4}{\sigma^2}\sum_{k=1}^{m}\Phi_k\xi\xi^T\Phi_k. \quad (2.8)$$

In [2] the Fisher information matrix for the nonAWGN model (2.6) is shown to have the following form:

$$\begin{aligned}\mathbb{I}^{\text{nonAWGN}}(\xi) &= \frac{4}{\rho^4}\sum_{k=1}^{m}\left(G_1\left(\frac{\langle \Phi_k\xi,\xi\rangle}{\rho^2}\right) - 1\right)\Phi_k\xi\xi^*\Phi_k \\ &= \frac{4}{\rho^2}\sum_{k=1}^{m}G_2\left(\frac{\langle \Phi_k\xi,\xi\rangle}{\rho^2}\right)\frac{1}{\langle \Phi_k\xi,\xi\rangle}\Phi_k\xi\xi^*\Phi_k, \quad (2.9)\end{aligned}$$

where the two universal scalar functions $G_1, G_2 : \mathbb{R}^+ \to \mathbb{R}^+$ are given by

$$\begin{aligned}G_1(a) &= \frac{e^{-a}}{a}\int_0^\infty \frac{I_1^2(2\sqrt{at})}{I_0(2\sqrt{at})}te^{-t}dt \quad (2.10) \\ &= \frac{e^{-a}}{8a^3}\int_0^\infty \frac{I_1^2(t)}{I_0(t)}t^3e^{-\frac{t^2}{4a}}dt \\ G_2(a) &= a(G_1(a) - 1),\end{aligned}$$

where I_0 and I_1 are the modifed Bessel functions of the first kind and order 0 and 1, respectively. Both Fisher information matrices have the same null space spanned by $J\xi$.

For the first estimation setup based on a reference signal z_0, the paper [1] showed that the Cramer-Rao Lower Bound of an unbiased estimator $\omega : \mathbb{R}^m \to \mathcal{E}_{\zeta_0}$ for the AWGN model (2.5) is given by:

$$Cov[\omega] \geq \left(\Pi^\perp_{J\zeta_0}\mathbb{I}^{\text{AWGN}}(\xi)\Pi^\perp_{J\zeta_0}\right)^\dagger. \quad (2.11)$$

The same result was extended to the non-AWGN model (2.6) in [2]:

$$Cov[\omega] \geq \left(\Pi^\perp_{J\zeta_0}\mathbb{I}^{\text{nonAWGN}}(\xi)\Pi^\perp_{J\zeta_0}\right)^\dagger \quad (2.12)$$

where again ω is an unbiased estimator for the reference signal based estimation setup.

III. Main Results

In this section we present the new results of this paper. First we prove an analytic property of Fisher Information Matrix when the likelihood function satisfies the conditions of this paper.

Proposition 3.1: Assume the likelihood function $P(y;\xi) = p(y;x)$ of a measurement process y is a function of $(s_k = |\langle x, f_k\rangle| = \sqrt{\langle \Phi_k\xi,\xi\rangle})_{1\leq k\leq m}$ only as in equation (1.1), where $\mathcal{F} = \{f_k; 1 \leq k \leq m\}$ is a phase retrievable frame, and the parameters $(s_1,\ldots,s_m)$ are identifiable. Assume also the likelihood satisfies the usual regularity condition $\mathbb{E}[\nabla_\xi \log P(y;\xi)] = 0$ for all $\xi \in \mathbb{R}^{2n}$. Let $\mathbb{I}(\xi)$ denote the Fisher Information Matrix defined by (2.7). Then $\ker\mathbb{I}(\xi) = \text{span}_\mathbb{R}(J\xi)$ and thus $\text{rank}(\mathbb{I}(\xi)) = 2n-1$ for every $\xi \neq 0$. Conversely if $(s_1,\ldots,s_m)$ are identifiable from y and $\text{rank}(\mathbb{I}(\xi)) = 2n-1$ for every $\xi \neq 0$ then the frame is phase retrievable.

Next we extend the formulas (2.11) and (2.12) to any stochastic model where the likelihood depends on the magnitudes of frame coefficients. For this extension we use a different approach than the one used in [1]. Along the way we obtain a formally different expression that turns out to be a different factorization of the lower bound.

Theorem 3.2: Consider the reference signal based estimation setup. Let $z_0 \in \mathbb{C}^n$ and $\zeta_0 = \jmath(z_0)$ be the unit-norm reference signal and $\mathcal{V}_{\zeta_0}$ and $\mathcal{E}_{\zeta_0}$ as in (2.1). Assume the likelihood function $P(y;\xi) = p(y;x)$ satisifes the assumptions of Proposition 3.1. Let $\mathbb{I}(\xi)$ denote the Fisher Information Matrix defined by (2.7). Let $\omega : \mathbb{R}^r \to \mathcal{E}_{\zeta_0}$ be an unbiased estimator. Then the following hold true:

1) The equation (2.2) is satisfied for every $\xi \in \mathcal{V}_{\zeta_0}$;

2) For any $\xi \in \mathcal{V}_{\zeta_0}$ the covariance matrix is bounded below by any of the following two equal bounds:

$$Cov[\omega(y);\xi] \geq \left(\Pi^{\perp}_{J\zeta_0}\mathbb{I}(\xi)\Pi^{\perp}_{J\zeta_0}\right)^{\dagger} = L^T(\mathbb{I}(\xi))^{\dagger}L \quad (3.1)$$

where $L = I - \frac{1}{\langle\xi,\zeta_0\rangle}J\zeta_0\xi^T J^T$.

The third result refers to the oracle based estimation. We derive the appropriate CRLB for this case. Unfortunately the lower bound turns out to be dependent on the actual estimator.

Theorem 3.3: Consider the oracle based estimation setup. Assume the likelihood function $P(y;\xi) = p(y;x)$ satisfies the assumptions of Proposition 3.1. Let $\mathbb{I}(\xi)$ denote the Fisher Information Matrix defined by (2.7). Let $o : \mathbb{R}^r \to \mathbb{C}^n$ be an estimator of the equivalence class, and let $\tilde{o} : \mathbb{R}^r \to \mathbb{C}^n$ be given by (1.8). Assume $\tilde{o}$ is an unbiased estimator for x. Denote by $\omega = \jmath(o)$ and $\tilde{\omega} = \jmath(\tilde{o})$. Then the following hold true:

1) The equation (2.4) is satisfied for every $\xi \in \mathbb{R}^{2n}$;
2) For any $\xi \neq 0$ the covariance matrix is bounded below as follows

$$Cov[\tilde{\omega}(y);\xi] \geq (I-\Delta)(\mathbb{I}(\xi))^{\dagger}(I-\Delta) \quad (3.2)$$

where

$$\Delta = \mathbb{E}\left[\frac{(\langle\omega,J\xi\rangle)^2}{((\langle\omega,\xi\rangle)^2+(\langle\omega,J\xi\rangle)^2)^{3/2}}\omega\omega^T + \frac{\langle\omega,\xi\rangle\langle\omega,J\xi\rangle}{((\langle\omega,\xi\rangle)^2+(\langle\omega,J\xi\rangle)^2)^{3/2}}(J\omega\omega^T+\omega\omega^T J^T) + \frac{(\langle\omega,\xi\rangle)^2}{((\langle\omega,\xi\rangle)^2+(\langle\omega,J\xi\rangle)^2)^{3/2}}J\omega\omega^T J^T\right]. \quad (3.3)$$

3) The matrix Δ satisfies:

$$\Delta = \Delta^T \geq I - \Pi^{\perp}_{J\xi} \geq 0.$$

and $\Delta J\xi = J\xi$, $\Delta\xi = 0$.

IV. Proofs of the Main Results

A. Proof of Proposition 3.1

We denote by $p(y;x)$ the likelihood function parametrized by the unknown complex n-vector x, and we let $P(y;\xi)$ denote the same likelihood where $|\langle x, f_k\rangle|$ is replaced by $\sqrt{\langle\Phi_k\xi,\xi\rangle}$. A direct computation shows (2.3) as well as the commutation relation $U(t)\Phi_k = \Phi_k U(t)$ for every $t \in \mathbb{R}$ and $1 \leq k \leq m$. Thus $\langle\Phi_k U(t)\xi, U(t)\xi\rangle = \langle\Phi_k\xi,\xi\rangle$ which implies $P(y;U(t)\xi) = P(y;\xi)$ for all t. This invariance relation lifts to the Fisher Information Matrix (2.7):

$$\mathbb{I}(U(t)\xi) = U(t)\mathbb{I}(\xi)U(t)^T.$$

On the other hand let $q(t_1,\ldots,t_m,y) = \log f(\sqrt{t_1},\ldots,\sqrt{t_m},y)$ so that $\nabla_\xi \log P(y;\xi) = 2\sum_{k=1}^m \frac{\partial q}{\partial t_k}\Phi_k\xi$. Then

$$\mathbb{I}(\xi) = 4\sum_{k,j=1}^m \mathbb{E}[\frac{\partial q}{\partial t_k}\frac{\partial q}{\partial t_j}]\Phi_k\xi\xi^T\Phi_j$$

The identifiability condition implies the matrix $\widetilde{\mathbb{I}}(s) = (\mathbb{E}[\frac{\partial q}{\partial t_k}\frac{\partial q}{\partial t_j}])_{1\leq k,j\leq m}$ is strictly positive (hence invertible). Thus

$$\langle\mathbb{I}(\xi)v,v\rangle = 4\sum_{k,j=1}^m \widetilde{\mathbb{I}(s)}_{k,j}\langle\Phi_k\xi,v\rangle\langle\Phi_j\xi,v\rangle$$

This show that $v \in \ker\mathbb{I}(\xi)$ if and only if $\langle\Phi_k\xi,v\rangle = 0$ for all k. Thus $\ker(\mathbb{I}(\xi)) = \ker(\mathcal{R}(\xi))$. But Theorem 4 in [8], or Theorem 3.1 in [1] implies that $\mathcal{F}$ is a phase retrievable frame if and only if $\ker(\mathbb{I}(\xi)) = \mathrm{span}_{\mathbb{R}}(J\xi)$, for $\xi \neq 0$, which means $rank(\mathbb{I}(\xi)) = 2n-1$, for $\xi \neq 0$.

B. Proof of Theorem 3.2

The approach used to prove Theorem 4.3 in [1] can be used here to show that $(\Pi^{\perp}_{J\zeta_0}\mathbb{I}(\xi)\Pi^{\perp}_{J\zeta_0})^{\dagger}$ is a lower bound of the covariance matrix. However we prefer to use a different approach and obtain the entire (3.1).

First we justify the first claim by saying that (2.2) is just a rewriting of (1.7) using $\omega = \jmath(o)$.

As is customary in the standard CRLB derivation (see [10]) we start from the unbiasedness equation. However we first extend this equation from $\mathcal{V}_{\zeta_0}$ to the open set $\Omega_{\zeta_0} = \mathbb{R}^{2n}\setminus\{\xi \in \mathbb{R}^{2n} : \langle\xi,\zeta_0\rangle = \langle\xi,J\zeta_0\rangle = 0\}$. This is accomplished as follows. Let $\xi \in \Omega$. Then there is a unique $t = t_\xi \in [0,2\pi)$ so that $U(t)\xi \in \mathcal{V}_{\zeta_0}$. A direct computation shows:

$$U(t_\xi)\xi = \frac{\langle\xi,\zeta_0\rangle}{\sqrt{(\langle\xi,\zeta_0\rangle)^2+(\langle J\xi,\zeta_0\rangle)^2}}\xi + \frac{\langle J\xi,\zeta_0\rangle}{\sqrt{(\langle\xi,\zeta_0\rangle)^2+(\langle J\xi,\zeta_0\rangle)^2}}J\xi. \quad (4.1)$$

On the other hand $P(y;U(t_\xi)\xi) = P(y;\xi)$ due to invariance to a global phase factor. Hence (2.2) is equivalent to the following condition:

$$\mathbb{E}[\omega(y);\xi] = U(t_\xi)\xi \quad , \quad \forall\xi \in \Omega_{\zeta_0}$$

Next we take the gradient with respect to ξ on both sides and use the regularity condition to obtain (4.2).

Next we specialize (4.2) to the case $\xi \in \mathcal{V}_{\zeta_0}$, thus $\langle\xi,J\zeta_0\rangle = 0$ and $\langle\xi,\zeta_0\rangle > 0$. This simplifies (4.2) to:

$$\mathbb{E}[(\omega(y)-U(t_\xi)\xi)(\nabla_\xi \log P(y;\xi))^T] = I - \frac{1}{\langle\xi,\zeta_0\rangle}J\xi\zeta_0^T J^T = L^T.$$

Next take $\eta,\mu \in \mathbb{R}^{2n}$ and use the Cauchy-Schwartz inequality to obtain:

$$\langle Cov(\omega)\eta,\eta\rangle\langle\mathbb{I}(\xi)\mu,\mu\rangle \geq (\langle L^T\mu,\eta\rangle)^2.$$

Note next that $L^T J\xi = J\xi - J\xi = 0$. Hence $\ker(\mathbb{I}(\xi)) \subset \ker(L^T)$. Thus we get:

$$\langle Cov(\omega)\eta,\eta\rangle \geq \max_{\mu\in(\ker(\mathbb{I}(\xi)))^{\perp}}\frac{(\langle L^T\mu,\eta\rangle)^2}{\langle\mathbb{I}(\xi)\mu,\mu\rangle}.$$

Then using standard properties of the pseudoinverse we obtain:

$$Cov(\omega) \geq L^T(\mathbb{I}(\xi))^{\dagger}L$$

which is part of equation (3.1). The proof ends once we establish the equality $L^T(\mathbb{I}(\xi))^{\dagger}L = (\Pi^{\perp}_{J\zeta_0}\mathbb{I}(\xi)\Pi^{\perp}_{J\zeta_0})^{\dagger}$. This last equality is obtained once we noticed that $L^T\eta$ is the

$$\mathbb{E}[(\omega(y)-U(t_\xi)\xi)(\nabla_\xi \log P(y;\xi))^T] = \frac{(\langle \xi, J\zeta_0\rangle)^2}{((\langle \xi,\zeta_0\rangle)^2+(\langle \xi, J\zeta_0\rangle)^2)^{3/2}}\xi\zeta_0^T - \frac{\langle \xi,\zeta_0\rangle\langle \xi, J\zeta_0\rangle}{((\langle \xi,\zeta_0\rangle)^2+(\langle \xi, J\zeta_0\rangle)^2)^{3/2}}(\xi\zeta_0 J^T - J\xi\zeta_0)$$
$$- \frac{(\langle \xi,\zeta_0\rangle)^2}{((\langle \xi,\zeta_0\rangle)^2+(\langle \xi, J\zeta_0\rangle)^2)^{3/2}} J\xi\zeta_0^T J^T + \frac{\langle \xi,\zeta_0\rangle}{\sqrt{(\langle \xi,\zeta_0\rangle)^2+(\langle \xi, J\zeta_0\rangle)^2}} I - \frac{\langle \xi,J\zeta_0\rangle}{\sqrt{(\langle \xi,\zeta_0\rangle)^2+(\langle \xi, J\zeta_0\rangle)^2}} J \quad (4.2)$$

oblique projection of the vector η onto $\mathcal{E}_{\zeta_0}$ along the subspace $\ker(\mathbb{I}(\xi)) = \mathrm{span}_{\mathbb{R}}(J\xi)$. Thus any $\mu \in \mathbb{R}^{2n}$ can be written uniquely as $\mu = aJ\xi + L^T\mu$ with $a = \frac{\langle \mu, J\zeta_0\rangle}{\langle \xi,\zeta_0\rangle}$. Thus

$$\langle L^T(\mathbb{I}(\xi))^\dagger L\eta,\eta\rangle = \max_\mu \frac{(\langle L^T\mu,\eta\rangle)^2}{\langle \mathbb{I}(\xi)\mu,\mu\rangle} = \max_{\varepsilon\in\mathcal{E}_{\zeta_0}} \frac{(\langle \varepsilon,\eta\rangle)^2}{\langle \mathbb{I}(\xi)\varepsilon,\varepsilon\rangle} =$$
$$= \max_\varepsilon \frac{(\langle \Pi_{J\zeta_0}^\perp \varepsilon,\eta\rangle)^2}{\langle \Pi_{J\zeta_0}^\perp \mathbb{I}(\xi)\Pi_{J\zeta_0}^\perp\varepsilon,\varepsilon\rangle} = \langle (\Pi_{J\zeta_0}^\perp \mathbb{I}(\xi)\Pi_{J\zeta_0}^\perp)^\dagger\eta,\eta\rangle$$

which ends the proof.

C. Proof of Theorem 3.3

First we note that similar to equation (4.1), the argument of the expectation in (2.4) is exactly $\jmath(\tilde{\omega}(y))$. This proves the first claim, namely that equation (2.4) is the realification of (1.9).

For the second claim of the theorem we take the gradient of (2.4) with respect to ξ and use the regularity condition to get:

$$I = \Delta + \mathbb{E}[(\tilde{\omega}(y)-\xi)(\nabla_\xi \log P(y;\xi))^T]$$

with Δ given by (3.3). Then use (2.4) to obtain $\Delta J\xi = J\xi$. Therefore $\ker \mathbb{I}(\xi) = \mathrm{span}_{\mathbb{R}}(J\xi) \subset \ker(I-\Delta)$ and by arguments similar to ones used in Theorem 3.2 we conclude that

$$Cov(\tilde{\omega}) \geq (I-\Delta)(\mathbb{I}(\xi))^\dagger(I-\Delta).$$

It is obvious that $\Delta^T = \Delta$. The only remaining claims that need to be proved are $\Delta\xi = 0$ and $\Delta \geq I - \Pi_{J\xi}^\perp$. To show these, observe

$$\langle \Delta\eta,\eta\rangle = \mathbb{E}\left[\frac{(\langle \omega, J\xi\rangle\langle \omega,\eta\rangle + \langle \omega,\xi\rangle\langle J\omega,\eta\rangle)^2}{((\langle \omega,\xi\rangle)^2+(\langle \omega,J\xi\rangle)^2)^{3/2}}\right] \geq 0.$$

Thus $\langle \Delta\xi,\xi\rangle = 0$ and since $\Delta \geq 0$ it follows $\Delta\xi = 0$. Take now $\eta \in \mathbb{R}^{2n}$ and use $\Delta J\xi = J\xi$ to compute:

$$(\langle J\xi,\eta\rangle)^2 = \left(\mathbb{E}\left[\frac{\langle \omega, J\xi\rangle\langle \omega,\eta\rangle + \langle \omega,\xi\rangle\langle J\omega,\eta\rangle}{\sqrt{(\langle \xi,\omega\rangle)^2+(\langle J\xi,\omega\rangle)^2}}\right]\right)^2.$$

Use $(\mathbb{E}[XY]) \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$ to obtain

$$(\langle J\xi,\eta\rangle)^2 \leq \mathbb{E}\left[\frac{(\langle \omega, J\xi\rangle\langle \omega,\eta\rangle + \langle \omega,\xi\rangle\langle J\omega,\eta\rangle)^2}{((\langle \omega,\xi\rangle)^2+(\langle \omega,J\xi\rangle)^2)^{3/2}}\right] \times \mathbb{E}\left[\sqrt{(\langle \xi,\omega\rangle)^2+(\langle J\xi,\omega\rangle)^2}\right].$$

Using again (2.4) we obtain

$$\|\xi\|^2 = \mathbb{E}\left[\sqrt{(\langle \xi,\omega\rangle)^2+(\langle J\xi,\omega\rangle)^2}\right]$$

and therefore

$$\Delta \geq \frac{1}{\|\xi\|^2} J\xi\xi^T J^T = I - \Pi_{J\xi}^\perp.$$

This concludes the proof of Theorem 3.3.

V. Conclusion

In this paper we presented two Cramer-Rao Lower Bounds each corresponding to a specific estimation setup in the phase retrieval problem. The first setup assumes a unit-norm reference signal used to select the global phase factor of the unknown signal. The second setup uses an oracle that returns the phase factor that selects the unique representative in the estimated class that is closest to the unknown signal with respect to the Euclidean norm. For the first setup we derived two equal expressions of the lower bound that are independent of the unbiased estimator. For the second setup our lower bound depends on the unbiased estimator. Further research might obtain a different lower bound (possibly sub-optimal) that is independent of the estimator.

Acknowledgment

Financial support from NSF Grant DMS-1413249 and ARO Grant W911NF-16-1-0008 is gratefully acknowledged.

References

[1] R. Balan, *Reconstruction of Signals from Magnitudes of Redundant Representations: The Complex Case*, available online arXiv:1304.1839v1 [math.FA] 6 April 2013, to appear in Foundations of Computational Mathematics (2015).

[2] R. Balan, *The Fisher Information Matrix and the CRLB in a Non-AWGN Model for the Phase Retrieval Problem*, proceedings of SampTA 2015.

[3] R. Balan, *Frames and Phaseless Reconstruction*, available online arXiv: 1601.03425v1, to appear in the Proceedings of Symposia in Applied Mathematics, AMS 2016.

[4] R. Balan, P. Casazza, D. Edidin, *On signal reconstruction without phase*, Appl. Comput. Harmon. Anal. **20** (2006), 345–356.

[5] R. Balan, B. Bodmann, P. Casazza, D. Edidin, *Painless reconstruction from Magnitudes of Frame Coefficients*, J.Fourier Anal.Applic., **15** (4) (2009), 488–501.

[6] R. Balan and D. Zou, *On Lipschitz Inversion of Nonlinear Redundant Representations*, Contemporary Mathematics 650, "Trends in Harmonic Analysis and Its Applications", 15–22, 2015.

[7] R. Balan and D. Zou, *On Lipschitz Analysis and Lipschitz Synthesis for the Phase Retrieval Problem*, available online arXiv:1506.02092, to appear in Linear Algebra and Applications 2016.

[8] A.S. Bandeira, J. Cahill, D.G. Mixon, A.A. Nelson, *Saving phase: Injectivity and stability for phase retrieval*, available online arXiv:1302.4618v2. Appl. Comp. Harm. Anal. **37** (1) (2014), 106–125.

[9] P. Casazza, *The art of frame theory*, Taiwanese J. Math., (2) **4** (2000), 129–202.

[10] S. M. Kay, **Fundamentals of Statistical Signal Processing. I. Estimation Theory**, Prentice Hall PTR, 18th Printing, 2010.