

Analysis of Fisher Information and the Cramér-Rao Bound for Nonlinear Parameter Estimation after Random Compression

Pooria Pakrooh, *Student Member, IEEE*, Ali Pezeshki, *Member, IEEE*, Louis L. Scharf, *Life Fellow, IEEE*, Douglas Cochran, *Senior Member, IEEE*, and Stephen D. Howard, *Member, IEEE*

Abstract—In this paper, we analyze the impact of compression with complex random matrices on Fisher information and the Cramér-Rao Bound (CRB) for estimating unknown parameters in the mean value function of a complex multivariate normal distribution. We consider the class of random compression matrices whose distribution is right-unitarily invariant. The compression matrix whose elements are i.i.d. standard complex normal random variables is one such matrix. We show that for all such compression matrices, the Fisher information matrix has a complex matrix beta distribution. We also derive the distribution of CRB. These distributions can be used to quantify the loss in CRB as a function of the Fisher information of the non-compressed data. In our numerical examples, we consider a direction of arrival estimation problem and discuss the use of these distributions as guidelines for choosing compression ratios based on the resulting loss in CRB.

Index Terms—Cramér-Rao bound, compressed sensing, Fisher information, matrix beta distribution, parameter estimation, random compression

I. INTRODUCTION

Inversion of a measurement for its underlying modes is an important topic which has applications in communications, radar/sonar signal processing and optical imaging. Classical methods for inversion are based on maximum likelihood, variations on linear prediction, subspace filtering, etc. Compressed sensing [1]–[3] is a relatively new theory that exploits compression (often with random matrices), sparse representations, and sparse recovery for inversion.

A preliminary version of parts of this paper was presented at the 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'13), Vancouver, BC, Canada, May 26-31, 2013.

This work is supported in part by NSF under grants CCF-1018472 and CCF-1422658.

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

P. Pakrooh is with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523, USA (e-mail: Pooria.Pakrooh@colostate.edu).

A. Pezeshki is with the Department of Electrical and Computer Engineering and the Department of Mathematics, Colorado State University, Fort Collins, CO 80523, USA (e-mail: Ali.Pezeshki@colostate.edu).

L. L. Scharf is with the Department of Mathematics and the Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA (e-mail: Louis.Scharf@colostate.edu).

D. Cochran is with the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ, 85287, USA (e-mail: cochran@asu.edu).

S. D. Howard is with the Defence Science and Technology Organisation, Edinburgh, SA 5111, Australia (e-mail: sdhoward@unimelb.edu.au).

In our previous work [4]–[7], the sensitivity of sparse inversion algorithms to basis mismatch and frame mismatch were studied. Our results show that mismatch between the actual basis in which a signal has a sparse representation and the basis (or frame) which is used for sparsity in a sparse reconstruction algorithm, e.g., basis pursuit, has performance consequences on the estimated parameter vector. These consequences have been identified and analyzed in detail, both analytically and numerically, in [4]–[7].

This paper addresses another fundamental question: *How much information is retained (or lost) in compressed noisy measurements for nonlinear parameter estimation?* To answer this question, we analyze the effect of random compression on the Fisher information matrix and the Cramér-Rao bound (CRB). We derive the distribution of the Fisher information matrix and the CRB for the class of random matrices whose distributions are invariant under right-unitary transformations. These include i.i.d. draws of spherically invariant matrix rows, including, for example, i.i.d. draws of standard complex normal matrix elements.

Our prior work on random compression and the Fisher information matrix [6], [7] contain numerical results that characterize the increase in CRB after random compression for the case where the parameters nonlinearly modulate the mean in a multivariate normal measurement model. Also, in [8] we derived analytical lower and upper bounds on the CRB for the same problem, and used our bounds to quantify the potential loss in CRB.

Other studies on the effect of random compression on the CRB and the Fisher information matrix include [9]–[11]. Babadi *et al.* [9] proposed a “Joint Typicality Estimator” to show the existence of an estimator that asymptotically achieves the CRB of sparse parameter estimation for random Gaussian compression matrices. Niazadeh *et al.* [10] generalize the results of [9] to a class of random compression matrices which satisfy the concentration of measures inequality. Nielsen *et al.* [12] derive the mean of the Fisher information matrix for the same class of random compression matrices that we are considering. Ramasamy *et al.* [11] derive bounds on the Fisher information matrix, but not for the model we are considering. We will clarify the distinction between our work and [11] after establishing our notation in Section II.

II. PROBLEM STATEMENT

Let $\mathbf{y} \in \mathbb{C}^n$ be a complex random vector whose probability density function $f(\mathbf{y}; \boldsymbol{\theta})$ is parameterized by an unknown but deterministic parameter vector $\boldsymbol{\theta} \in \mathbb{R}^p$. The derivative of the log-likelihood function with respect to $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_p]$ is called the Fisher score, and the covariance matrix of the Fisher score is the Fisher information matrix which we denote by $\mathbf{J}(\boldsymbol{\theta})$:

$$\mathbf{J}(\boldsymbol{\theta}) = E\left[\left(\frac{\partial \log f(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \log f(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^H\right] \in \mathbb{C}^{p \times p}. \quad (1)$$

The inverse $\mathbf{J}^{-1}(\boldsymbol{\theta})$ of the Fisher information matrix lower bounds the error covariance matrix for any unbiased estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$ of $\boldsymbol{\theta}$, that is

$$E[(\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta})^H] \succeq \mathbf{J}^{-1}(\boldsymbol{\theta}), \quad (2)$$

where $\mathbf{A} \succeq \mathbf{B}$ for matrices $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n \times n}$ means $\mathbf{a}^H \mathbf{A} \mathbf{a} \geq \mathbf{a}^H \mathbf{B} \mathbf{a}$ for all $\mathbf{a} \in \mathbb{C}^n$. We assume that $\mathbf{J}(\boldsymbol{\theta})$ is non-singular. The i^{th} diagonal element of $\mathbf{J}^{-1}(\boldsymbol{\theta})$ is the CRB for estimating θ_i and it gives a lower bound on the MSE of any unbiased estimator of θ_i from \mathbf{y} (see, e.g., [13]).

For $\mathbf{y} \in \mathbb{C}^n$ a proper complex normal random vector distributed as $\mathcal{CN}(\mathbf{x}(\boldsymbol{\theta}), \mathbf{C})$ with unknown mean vector $\mathbf{x}(\boldsymbol{\theta})$ parameterized by $\boldsymbol{\theta}$, and known covariance $\mathbf{C} = \sigma^2 \mathbf{I}$, the Fisher information matrix is the Gramian

$$\mathbf{J}(\boldsymbol{\theta}) = \mathbf{G}^H \mathbf{C}^{-1} \mathbf{G} = \frac{1}{\sigma^2} \mathbf{G}^H \mathbf{G}. \quad (3)$$

The i^{th} column \mathbf{g}_i of $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_p]$ is the partial derivative $\mathbf{g}_i = \frac{\partial}{\partial \theta_i} \mathbf{x}(\boldsymbol{\theta})$, which characterizes the sensitivity of the mean vector $\mathbf{x}(\boldsymbol{\theta})$ to variation of the i^{th} parameter θ_i .

The CRB for estimating θ_i is given by

$$(\mathbf{J}^{-1}(\boldsymbol{\theta}))_{ii} = \sigma^2 (\mathbf{g}_i^H (\mathbf{I} - \mathbf{P}_{\mathbf{G}_i}) \mathbf{g}_i)^{-1}, \quad (4)$$

where \mathbf{G}_i consists of all columns of \mathbf{G} except \mathbf{g}_i , and $\mathbf{P}_{\mathbf{G}_i}$ is the orthogonal projection onto the column space of \mathbf{G}_i [14]. This CRB can also be written as

$$(\mathbf{J}^{-1}(\boldsymbol{\theta}))_{ii} = \frac{\sigma^2}{\|\mathbf{g}_i\|_2^2 \sin^2(\psi_i)}, \quad (5)$$

where ψ_i is the principal angle between subspaces $\langle \mathbf{g}_i \rangle$ and $\langle \mathbf{G}_i \rangle$. These representations illuminate the geometry of the CRB, which is discussed in detail in [14].

If \mathbf{y} is compressed by a compression matrix $\Phi \in \mathbb{C}^{m \times n}$ to produce $\hat{\mathbf{y}} = \Phi \mathbf{y}$, then the probability density function of the compressed data $\hat{\mathbf{y}}$ is $\mathcal{CN}[\Phi \mathbf{x}(\boldsymbol{\theta}), \sigma^2 \Phi \Phi^H]$. The Fisher information matrix $\hat{\mathbf{J}}(\boldsymbol{\theta})$ is given by

$$\hat{\mathbf{J}}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \hat{\mathbf{G}}^H \hat{\mathbf{G}}, \quad (6)$$

where $\hat{\mathbf{G}} = \mathbf{P}_{\Phi^H} \mathbf{G}$. The CRB for estimating the i^{th} parameter is

$$(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{ii} = \sigma^2 (\hat{\mathbf{g}}_i^H (\mathbf{I} - \mathbf{P}_{\hat{\mathbf{G}}_i}) \hat{\mathbf{g}}_i)^{-1}, \quad (7)$$

where $\hat{\mathbf{G}}_i = \mathbf{P}_{\Phi^H} \mathbf{G}_i$, and $\mathbf{P}_{\Phi^H} = \Phi^H (\Phi \Phi^H)^{-1} \Phi$ is the orthogonal projection onto the row span of Φ .

Our aim is to study the effect of random compression on the Fisher information matrix and the CRB. In section III we investigate this problem by deriving the distributions of the Fisher information matrix and the CRB for the case in which the elements of the compression matrix Φ are distributed as i.i.d. standard complex normal random variables. Then we demonstrate that the same analysis holds for a wider range of random compression matrices.

Remark 1: In parallel to our work, Ramasamy et al. [11] have also looked at the impact of compression on Fisher information. However, they have considered a different parameter model. Specifically, their compressed data has density $\mathcal{CN}[\Phi \mathbf{x}(\boldsymbol{\theta}), \sigma^2 \mathbf{I}]$, in contrast to ours which is distributed as $\mathcal{CN}[\Phi \mathbf{x}(\boldsymbol{\theta}), \sigma^2 \Phi \Phi^H]$. Our model is a signal-plus-noise model, wherein the noisy signal $\mathbf{x}(\boldsymbol{\theta}) + \mathbf{n}$, $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$, is compressed to produce $\Phi \mathbf{x}(\boldsymbol{\theta}) + \Phi \mathbf{n}$. In contrast, their model corresponds to compressing a noiseless signal $\mathbf{x}(\boldsymbol{\theta})$ to produce $\Phi \mathbf{x}(\boldsymbol{\theta}) + \mathbf{w}$, where $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$ represents post-compression noise. Note that the Fisher information, CRB and corresponding bounds of these two models are different, as in our model noise enters at the input of the compressor, whereas in [11] noise enters at the output of the compressor. This is an important distinction.

III. DISTRIBUTION OF THE FISHER INFORMATION MATRIX AFTER COMPRESSION

Let \mathbf{W} be the normalized Fisher information matrix after compression, defined as

$$\mathbf{W} = \mathbf{J}^{-1/2} \hat{\mathbf{J}} \mathbf{J}^{-H/2} \in \mathbb{C}^{p \times p}, \quad (8)$$

where \mathbf{J} and $\hat{\mathbf{J}}$ are the Fisher information matrix before compression and the Fisher information matrix after compression, defined in (3) and (6), respectively. Our aim is to derive the distribution of the random matrix \mathbf{W} for the case that the elements of the compression matrix Φ_{ij} are i.i.d. random variables distributed as $\mathcal{CN}(0, 1)$. We assume $n - p \geq m$, which is typical in almost all compression scenarios of interest.

Using (3) and (6), \mathbf{W} may be written as

$$\mathbf{W} = \mathbf{H}^H \mathbf{P}_{\Phi^H} \mathbf{H}, \quad (9)$$

where $\mathbf{H} = \mathbf{G}(\mathbf{G}^H \mathbf{G})^{-H/2}$ is a left-unitary matrix, i.e., $\mathbf{H}^H \mathbf{H} = \mathbf{I}_p$. Define $\tilde{\mathbf{H}} \in \mathbb{C}^{n \times (n-p)}$ such that $\Lambda = [\mathbf{H} | \tilde{\mathbf{H}}]$ is an orthonormal basis for \mathbb{C}^n , i.e., $\Lambda \Lambda^H = \Lambda^H \Lambda = \mathbf{I}_n$. Then we have:

$$\mathbf{W} = \left[\mathbf{I}_p \mid \mathbf{0} \right] \Lambda^H \mathbf{P}_{\Phi^H} \Lambda \begin{bmatrix} \mathbf{I}_p \\ \mathbf{0} \end{bmatrix}, \quad (10)$$

where

$$\begin{aligned} \Lambda^H \mathbf{P}_{\Phi^H} \Lambda &= \Lambda^H \Phi^H (\Phi \Phi^H)^{-1} \Phi \Lambda \\ &= \Lambda^H \Phi^H (\Phi \Lambda \Lambda^H \Phi^H)^{-1} \Phi \Lambda \\ &= \mathbf{P}_{\Lambda^H \Phi^H}. \end{aligned} \quad (11)$$

Because the distribution of Φ is right-unitarily invariant, the distribution of $\Lambda^H \mathbf{P}_{\Phi^H} \Lambda$ is the same as the distribution of \mathbf{P}_{Φ^H} . Therefore, the distribution of \mathbf{W} is the same as the

distribution of $\mathbf{V} = \Phi_1^H (\Phi \Phi^H)^{-1} \Phi_1$, where $\Phi = [\Phi_1 | \Phi_2]$, $\Phi_1 \in \mathbb{C}^{m \times p}$ and $\Phi_2 \in \mathbb{C}^{m \times (n-p)}$. Now, write \mathbf{V} as $\mathbf{V} = \mathbf{Y} \mathbf{Y}^H$, with $\mathbf{Y} = \Phi_1^H \mathbf{Z}^{-1/2}$ and $\mathbf{Z} = \Phi \Phi^H$. Since $\mathbf{Z} = \Phi_1 \Phi_1^H + \Phi_2 \Phi_2^H$, and $\Phi_2 \Phi_2^H$ is distributed as complex Wishart $\mathcal{W}_c(\mathbf{I}_m, m, n-p)$ for $n-p \geq m$, given Φ_1 the pdf of \mathbf{Z} is

$$f(\mathbf{Z} | \Phi_1) = c_1 e^{-\text{tr}(\mathbf{Z} - \Phi_1 \Phi_1^H)} |\mathbf{Z} - \Phi_1 \Phi_1^H|^{n-m-p}. \quad (12)$$

The pdf of Φ_1 can be written as $c_2 e^{-\text{tr}(\Phi_1 \Phi_1^H)}$. Therefore, the joint pdf of \mathbf{Z} and Φ_1 is

$$f(\mathbf{Z}, \Phi_1) = c_3 e^{-\text{tr}(\mathbf{Z})} |\mathbf{Z} - \Phi_1 \Phi_1^H|^{n-m-p} \quad (13)$$

where c_1, c_2 , and $c_3 = c_1 c_2$ are normalization factors. Since $\mathbf{Y} = \Phi_1^H \mathbf{Z}^{-1/2}$, from (13) the joint pdf of \mathbf{Y} and \mathbf{Z} is

$$\begin{aligned} f(\mathbf{Y}, \mathbf{Z}) &= c_3 e^{-\text{tr}(\mathbf{Z})} |\mathbf{Z} - \mathbf{Z}^{1/2} \mathbf{Y}^H \mathbf{Y} \mathbf{Z}^{1/2}|^{n-m-p} |\mathbf{Z}|^p \\ &= c_3 e^{-\text{tr}(\mathbf{Z})} |\mathbf{I}_m - \mathbf{Y}^H \mathbf{Y}|^{n-m-p} |\mathbf{Z}|^{n-m}. \end{aligned} \quad (14)$$

This shows that \mathbf{Y} and \mathbf{Z} are independent and the pdf of \mathbf{Y} is

$$\begin{aligned} f(\mathbf{Y}) &= c_4 |\mathbf{I}_m - \mathbf{Y}^H \mathbf{Y}|^{n-m-p} \\ &= c_4 |\mathbf{I}_p - \mathbf{Y} \mathbf{Y}^H|^{n-m-p}, \end{aligned} \quad (15)$$

where c_4 is a normalization factor.

Let $g(\mathbf{Y} \mathbf{Y}^H) = f(\mathbf{Y})$. To derive the distribution of $\mathbf{V} = \mathbf{Y} \mathbf{Y}^H$ we use the following theorem.

Theorem 1: [15] If the density of $\mathbf{Y} \in \mathbb{C}^{p \times m}$ is $g(\mathbf{Y} \mathbf{Y}^H)$, then the density of $\mathbf{V} = \mathbf{Y} \mathbf{Y}^H$ is

$$\frac{|\mathbf{V}|^{m-p} g(\mathbf{V}) \pi^{mp}}{\tilde{\Gamma}_p(m)}, \quad (16)$$

where $\tilde{\Gamma}_m(\cdot)$ is the complex multivariate Gamma function.

Using Theorem 1 and (15), the pdf of \mathbf{V} is

$$c_5 |\mathbf{V}|^{m-p} |\mathbf{I}_p - \mathbf{V}|^{n-m-p} \quad \text{for } \mathbf{0} \preceq \mathbf{V} \preceq \mathbf{I}_p, \quad (17)$$

which is the Type I complex multivariate beta distribution $\mathbb{C}B_p^I(m, n-m)$ for $c_5 = \frac{\tilde{\Gamma}_p(n)}{\tilde{\Gamma}_p(m) \tilde{\Gamma}_p(n-m)}$. Recall that the distribution of the normalized Fisher information matrix after compression $\mathbf{W} = \mathbf{J}^{-1/2} \hat{\mathbf{J}} \mathbf{J}^{-H/2}$ is identical to that of \mathbf{V} . Therefore, \mathbf{W} is also distributed as $\mathbb{C}B_p^I(m, n-m)$.

Remark 2: It is important to note that the distribution of $\mathbf{W} = \mathbf{J}^{-1/2} \hat{\mathbf{J}} \mathbf{J}^{-H/2}$ is *invariant* to \mathbf{J} , and it depends only on the parameters p, m and n , where $p \leq m \leq n-p$. In this sense, this result for the distribution of $\mathbf{J}^{-1/2} \hat{\mathbf{J}} \mathbf{J}^{-H/2}$ is universal, and reminiscent of the classical result of Reed, Mallat, and Brennan [16] for normalized SNR in adaptive filtering.

Lemma 1: [17] Assume $\mathbf{A} \in \mathbb{C}^{p \times p}$ is a positive definite Hermitian random matrix with a pdf $h(\mathbf{A})$. Then, the joint pdf of eigenvalues $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ of \mathbf{A} is

$$\frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)} \prod_{i < j} (\lambda_i - \lambda_j)^2 \int_{\mathcal{U}(p)} h(\mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^H) d\mathbf{U}, \quad (18)$$

where $d\mathbf{U}$ is the invariant Haar measure on the Unitary group $\mathcal{U}(p)$.

Using Lemma 1, we can derive the joint distribution of the eigenvalues of \mathbf{W} . Replacing the pdf of $\mathbf{W} \sim \mathbb{C}B_p^I(m, n-m)$ in (18), the joint pdf of the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_p$ of $\mathbf{W} = \mathbf{J}^{-1/2} \hat{\mathbf{J}} \mathbf{J}^{-H/2}$ is given by

$$\frac{\pi^{p(p-1)} \tilde{\Gamma}_p(n)}{\tilde{\Gamma}_p(p) \tilde{\Gamma}_p(m) \tilde{\Gamma}_p(n-m)} \prod_{i < j} (\lambda_i - \lambda_j)^2 \prod_{i=1}^p \lambda_i^{m-p} (1 - \lambda_i)^{n-m-p}. \quad (19)$$

Now, from (17) and using the transformation $\hat{\mathbf{J}} = \mathbf{J}^{1/2} \mathbf{W} \mathbf{J}^{H/2}$, the Fisher information matrix after compression $\hat{\mathbf{J}}$ is distributed as

$$c_5 |\mathbf{J}|^{p-n} |\hat{\mathbf{J}}|^{m-p} |\mathbf{J} - \hat{\mathbf{J}}|^{n-m-p} \quad \text{for } \mathbf{0} \preceq \hat{\mathbf{J}} \preceq \mathbf{J} \quad (20)$$

and the inverse of the Fisher information matrix after compression $\hat{\mathbf{K}} = \hat{\mathbf{J}}^{-1}$ is distributed as

$$c_5 |\mathbf{J}|^{p-n} |\hat{\mathbf{K}}|^{-n} |\mathbf{J} \hat{\mathbf{K}} - \mathbf{I}_p|^{n-m-p} \quad \text{for } \hat{\mathbf{K}} \succeq \mathbf{J}^{-1}. \quad (21)$$

Remark 3: For the class of random compression matrices that have density functions of the form $g(\Phi \Phi^H)$, that is, the distribution of Φ is right-unitarily invariant, $\Phi^H (\Phi \Phi^H)^{-1/2}$ is uniformly distributed on the Stiefel manifold $\mathcal{V}_m(\mathbb{C}^n)$ [18]. Therefore, the distribution of the normalized Fisher information matrix for this class of compression matrices is the same as the one given in (17).

Remark 4: Using the properties of a complex multivariate beta distribution [19], we have:

$$E[\hat{\mathbf{J}}] = \frac{m}{n} \mathbf{J}, \quad (22)$$

and

$$E[\hat{\mathbf{J}}^{-1}] = \frac{n-p}{m-p} \mathbf{J}^{-1}. \quad (23)$$

This shows that, on average, compression results in a factor $\frac{n-m}{n}$ loss in the Fisher information and a factor $\frac{n-p}{m-p}$ increase in the CRB \mathbf{J}^{-1} .

The distribution of the CRB after compression can be derived using the following Lemma.

Lemma 2: [19] Assume $\mathbf{X} \sim \mathbb{C}B_p^I(a_1, a_2)$. Let \mathbf{z} be a complex vector independent of \mathbf{X} . Then, $x = \frac{\mathbf{z}^H \mathbf{z}}{\mathbf{z}^H \mathbf{X}^{-1} \mathbf{z}}$ is distributed as $B_p^I(a_1 - p + 1, a_2)$, which is a Type I univariate beta distribution with the pdf

$$\frac{\Gamma_p(a_1 + a_2)}{\Gamma_p(a_1) \Gamma_p(a_2)} x^{a_1-1} (1-x)^{a_2-1} \quad \text{for } 0 < x < 1. \quad (24)$$

Now consider the CRB on an unbiased estimator of parameter θ_i , after compression, normalized by the CRB before compression:

$$\frac{(\hat{\mathbf{J}}^{-1})_{ii}}{(\mathbf{J}^{-1})_{ii}} = \frac{\mathbf{e}_i^H \hat{\mathbf{J}}^{-1} \mathbf{e}_i}{\mathbf{e}_i^H \mathbf{J}^{-1} \mathbf{e}_i} = \frac{\mathbf{z}^H \mathbf{W}^{-1} \mathbf{z}}{\mathbf{z}^H \mathbf{z}}, \quad (25)$$

where $\mathbf{z} = \mathbf{J}^{-1/2} \mathbf{e}_i$, and $\mathbf{e}_i \in \mathbb{C}^p$ is a standard unit vector with 1 as its i th element and zeros as its other elements. By

Lemma 1, the above ratio is distributed as the inverse of a univariate beta random variable $B^I(m-p+1, n-m)$.

Remark 5: From the distribution of the CRB after compression, we have:

$$E[(\hat{\mathbf{J}}^{-1})_{ii}] = \left(\frac{n-p}{m-p}\right)(\mathbf{J}^{-1})_{ii}, \quad (26)$$

$$\text{var}[(\hat{\mathbf{J}}^{-1})_{ii}] = \frac{(n-m)(n-p)}{(m-p-1)(m-p)^2}((\mathbf{J}^{-1})_{ii})^2. \quad (27)$$

Remark 6: We can also look at the effect of random compression on the Kullback Leibler (KL) divergence of two normal probability distributions, for the class of random compressors already discussed in Remark 3. The KL divergence between $\mathcal{CN}(\mathbf{x}(\boldsymbol{\theta}), \mathbf{C})$ and $\mathcal{CN}(\mathbf{x}(\boldsymbol{\theta}'), \mathbf{C})$ is:

$$D(\boldsymbol{\theta}, \boldsymbol{\theta}') = (\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))^H \mathbf{C}^{-1} (\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}')). \quad (28)$$

After compression with Φ we have

$$\hat{D}(\boldsymbol{\theta}, \boldsymbol{\theta}') = (\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))^H \Phi^H (\Phi \mathbf{C} \Phi^H)^{-1} \Phi (\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}')). \quad (29)$$

For the case $\mathbf{C} = \sigma^2 \mathbf{I}$, the normalized KL divergence is

$$\frac{\hat{D}(\boldsymbol{\theta}, \boldsymbol{\theta}')}{D(\boldsymbol{\theta}, \boldsymbol{\theta}')} = \frac{(\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))^H \mathbf{P}_{\Phi^H} (\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))}{(\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))^H (\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))}. \quad (30)$$

Therefore, the normalized KL divergence, for random compression matrices Φ whose distributions are invariant to right-unitary transformations, is distributed as $B^I(m, n-m)$.

IV. NUMERICAL RESULTS

As a special example, we consider the effect of compression on DOA estimation using a uniform line array with n elements. In our simulations, we consider two sources whose electrical angles θ_1 and θ_2 are unknown. The mean vector $\mathbf{x}(\boldsymbol{\theta})$ is $\mathbf{x}(\boldsymbol{\theta}) = \mathbf{x}(\theta_1) + \mathbf{x}(\theta_2)$, where

$$\mathbf{x}(\theta_i) = A_i e^{j\phi_i} [1 \quad e^{j\theta_i} \quad e^{j2\theta_i} \quad \dots \quad e^{j(n-1)\theta_i}]^T. \quad (31)$$

Here A_i and ϕ_i are the amplitude and phase of the i^{th} source, which we assume known. We set $\phi_1 = \phi_2 = 0$ and $A_1 = A_2 = 1$. We wish to estimate θ_1 , whose true value in this example is zero, in the presence of the interfering source at electrical angle $\theta_2 = \pi/n$ (half the Rayleigh limit of the array). For our simulations, we use Gaussian compression matrices $\Phi_{m \times n}$ whose elements are i.i.d. $\mathcal{CN}(0, 1/m)$. The Fisher information matrix and the CRB on the estimation of θ_1 are calculated for different realizations of Φ . Fig. 1 shows the CRB on the estimation of θ_1 before compression divided by its corresponding value after compression, i.e. $\frac{(\hat{\mathbf{J}}^{-1})_{11}}{(\mathbf{J}^{-1})_{11}}$ for $m = 64, n = 128$. A histogram of actual values of $\frac{(\hat{\mathbf{J}}^{-1})_{11}}{(\mathbf{J}^{-1})_{11}}$ for 10^5 independent realizations of random Φ is shown in blue. The red curve represents the pdf of a $B^I(m-p+1, n-m)$ distributed random variable for $p = 2$. This figure simply provides an illustration of the result (25).

Recall that the inverse Fisher Information matrix \mathbf{J}^{-1} lower bounds the error covariance matrix $\boldsymbol{\Sigma} = E[\mathbf{e}\mathbf{e}^H]$ for unbiased

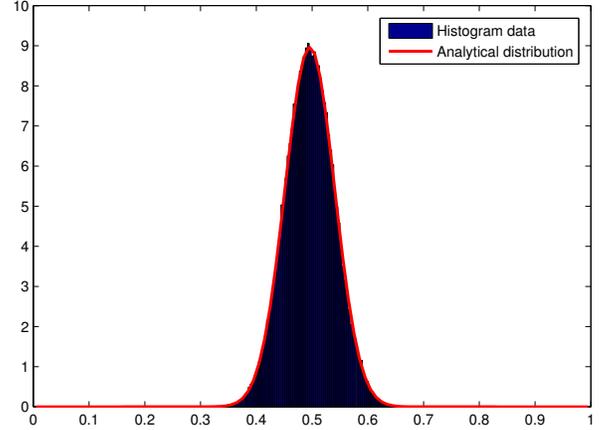


Fig. 1. Histogram data and analytical distributions for $\frac{(\hat{\mathbf{J}}^{-1})_{11}}{(\mathbf{J}^{-1})_{11}}$ using 10^5 realizations of i.i.d. Gaussian compression matrices with $n = 128$ and $m = 64$.

errors $\mathbf{e} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$. So the concentration ellipse $\mathbf{e}^H \boldsymbol{\Sigma}^{-1} \mathbf{e} \leq \mathbf{e}^H \mathbf{J} \mathbf{e}$ for all $\mathbf{e} \in \mathbb{C}^p$. The ellipses $\mathbf{e}^H \mathbf{J} \mathbf{e} = r^2$ and $\mathbf{e}^H \hat{\mathbf{J}} \mathbf{e} = r^2$, with $r^2 = \mathbf{J}_{11}$, are illustrated in Fig. 2, demonstrating the effect that compression inflates the concentration ellipse. The blue curve is the locus of all points $\mathbf{e} \in \mathbb{C}^p$, for which $\mathbf{e}^H \mathbf{J} \mathbf{e} = r^2$. The red curves are the loci of all points $\mathbf{e} \in \mathbb{C}^p$, for which $\mathbf{e}^H \hat{\mathbf{J}} \mathbf{e} = r^2$ for 50 realizations of the Fisher information matrix after compression. As can be seen, the concentration ellipse for the Fisher information matrix before compression has the smallest volume in comparison with all the realizations of the concentration ellipses after compression. Also, for each realization of the Gaussian compression, the orientation of the concentration ellipse is nearly aligned with that of the uncompressed ellipse.

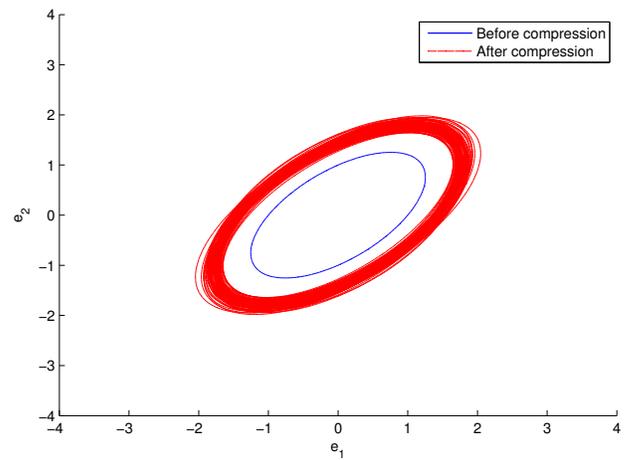


Fig. 2. Concentration ellipses for the Fisher information matrices before and after compression.

Figure 3 shows the compression ratio m/n needed so that the CRB after compression $(\hat{\mathbf{J}}^{-1})_{11}$ does not exceed κ times the CRB before compression $(\mathbf{J}^{-1})_{11}$, at two levels of confidence and for $n = 128$. These curves are plotted using the tail prob-

abilities of a univariate beta random variables. They can be used as guidelines for deriving a satisfactory compression ratio based on a tolerable level of loss in the CRB. Alternatively, we can plot the confidence level curves versus m for fixed values of κ . In that case, the plots may be useful to find a number of measurements that would guarantee that after compression CRB does not go above a desired bound (corresponding to a particular κ) with a certain level of confidence.

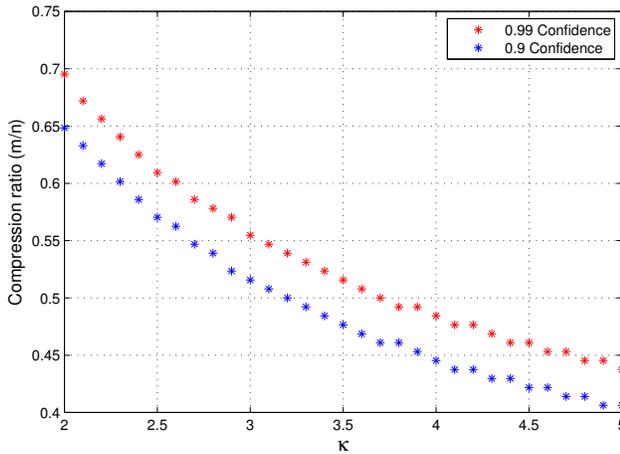


Fig. 3. Compression ratios needed so that $(\hat{\mathbf{J}}^{-1})_{11} < \kappa(\mathbf{J}^{-1})_{11}$ for different confidence levels.

V. CONCLUSION

In this paper, we have studied the effect of random compression of noisy measurements on the CRB for estimating parameters in a nonlinear model. We have considered the class of random compression matrices whose distributions are right-unitarily invariant. A random compression matrix with i.i.d. standard complex normal elements is one such compression matrix. The analytical distributions obtained in this paper can be used to quantify the amount of loss due to compression. Also, they can be used as guidelines for choosing a suitable compression ratio based on a tolerable loss in the CRB. Importantly, the distribution of the ratios of CRBs before and after compression depends only on the number of parameters and the number of measurements. The distribution is invariant to the underlying signal-plus-noise model, in the sense that it is invariant to the underlying (before compression) Fisher information matrix.

REFERENCES

- [1] E. J. Candès, “Compressive sampling,” in *Proc. Int. Congress Math.*, vol. 3, 2006, pp. 1433–1452.
- [2] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] R. Baraniuk, “Compressive sensing,” *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, Jul. 2007.
- [4] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank, “Sensitivity to basis mismatch in compressed sensing,” *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, May 2011.

- [5] Y. Chi, A. Pezeshki, L. L. Scharf, and R. Calderbank, “Sensitivity to basis mismatch in compressed sensing,” in *Proc. 2010 IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Dallas, TX, Mar. 2010, pp. 3930–3933.
- [6] L. L. Scharf, E. K. P. Chong, A. Pezeshki, and J. R. Luo, “Sensitivity considerations in compressed sensing,” in *Conf. Rec. 45th Annual Asilomar Conf. Signals, Systems, Computs., Pacific Grove, CA.*, Nov. 2011, pp. 744–748.
- [7] —, “Compressive sensing and sparse inversion in signal processing: Cautionary notes,” in *Proc. 7th Workshop on Defence Applications of Signal Processing (DASP)*, Coolumb, Queensland, Australia, Jul. 10–14, 2011.
- [8] P. Pakrooh, L. L. Scharf, A. Pezeshki, and Y. Chi, “Analysis of Fisher information and the Cramér-Rao bound for nonlinear parameter estimation after compressed sensing,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, BC, Canada, May 2013, pp. 6630–6634.
- [9] B. Babadi, N. Kalouptsidis, and V. Tarokh, “Asymptotic achievability of the Cramér-Rao bound for noisy compressive sampling,” *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1233–1236, Mar. 2009.
- [10] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, “On the achievability of Cramér-Rao bound in noisy compressed sensing,” *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 518–526, Jan. 2012.
- [11] D. Ramasamy, S. Venkateswaran, and U. Madhow, “Compressive estimation in AWGN: General observations and a case study,” in *Conf. Rec. 46th Annual Asilomar Conf. Signals, Systems, Computs., Pacific Grove, CA.*, Nov. 2012, pp. 953–957.
- [12] J. K. Nielsen, M. G. Christensen, and S. H. Jensen, “On compressed sensing and the estimation of continuous parameters from noisy observations,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 3609–3612.
- [13] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice-Hall, Apr. 1993.
- [14] L. L. Scharf and L. T. McWhorter, “Geometry of the Cramér-Rao bound,” *Signal Process.*, vol. 31, no. 3, pp. 301–311, Apr. 1993.
- [15] M. Srivastava, “On the complex Wishart distribution,” *The Annals of Mathematical Statistics*, pp. 313–315, 1965.
- [16] I. S. Reed, J. D. Mallett, and L. E. Brennan, “Rapid convergence rate in adaptive arrays,” *IEEE Transactions on Aerospace and Electronic Systems*, no. 6, pp. 853–863, 1974.
- [17] A. T. James, “Distributions of matrix variates and latent roots derived from normal samples,” *The Annals of Mathematical Statistics*, pp. 475–501, 1964.
- [18] Y. Chikuse, *Statistics on special manifolds*. New York: Springer, 2003.
- [19] A. K. Gupta, D. K. Nagar, and E. Bedoya, “Properties of the complex matrix variate Dirichlet distribution,” *Scientiae Mathematicae Japonicae*, vol. 66, no. 1, p. 53, 2007.



Pooria Pakrooh (S’13) received the B.Sc. degree in electrical engineering from Iran University of Science and Technology, Tehran, Iran, in 2008, and the M.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2011. He received his Ph.D. degree in electrical engineering from Colorado State University, Fort Collins, Colorado, in 2015. He is currently a postdoctoral research associate with the Department of Mathematics at Colorado State University. His current research interests include compressed sensing and statistical signal processing, and their applications in sensor and array signal processing.



Ali Pezeshki (S'95-M'05) received the B.Sc. and M.Sc. degrees in electrical engineering from University of Tehran, Tehran, Iran, in 1999 and 2001, respectively. He received his PhD degree in electrical engineering at Colorado State University in 2004. In 2005, he was a postdoctoral research associate with the Electrical and Computer Engineering Department at Colorado State University. From January 2006 to August 2008, he was a postdoctoral research associate with The Program in Applied and Computational Mathematics at Princeton University. In

August 2008, he joined the faculty of Colorado State University, where he is now an Associate Professor in the Department of Electrical and Computer Engineering, and the Department of Mathematics. His research interests are in statistical signal processing, coding theory, applied harmonic analysis, and bioimaging.



Douglas Cochran holds S.M. and Ph.D. degrees in applied mathematics from Harvard University and degrees in mathematics from MIT and the University of California, San Diego. Since 1989, he has been on the faculty of the School of Electrical, Computer, and Energy Engineering at Arizona State University (ASU) and is also affiliated with the School of Mathematical and Statistical Sciences. Between 2005 and 2008, he served as Assistant Dean for Research in the Ira A. Fulton Schools of Engineering at ASU. Between 2000 and 2005, Professor Cochran was Program Manager for Mathematics at the Defense Advanced Research Projects Agency (DARPA) and he held a similar position at the U.S. Air Force Office of Scientific Research between 2008 and 2010. Prior to joining the ASU faculty, he was Senior Scientist at BBN Systems and Technologies Inc., during which time he served as resident scientist at the DARPA Acoustic Research Center and the Naval Ocean Systems Center. He has been a visiting scientist at the Australian Defence Science and Technology Organisation and served as a consultant to several technology companies.

Professor Cochran was Technical Program co-Chair for the 2015 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015), and he was General Co-Chair of the ICASSP' 99 and Co-Chair of the 1997 U.S.-Australia Workshop on Defense Signal Processing. He has also served as Associate Editor for book series and journals, including the IEEE Transactions on Signal Processing. His research is in applied harmonic analysis and statistical signal processing.



Louis L. Scharf (S'67-M'69-SM'77-F'86-LF'07) received his Ph.D. from the University of Washington, Seattle. He is currently Research Professor of Mathematics and Emeritus Professor of Electrical and Computer Engineering, with a joint appointment in Statistics, at Colorado State University. He has served on the faculties of University of Rhode Island and University of Colorado, and held visiting positions at Ecole Supérieure d'électricité, Ecole Supérieure des Télécommunications, EURECOM, University of La Plata, Duke University, University of Wisconsin, University Tromsø, and University of Paderborn.

Professor Scharf's research interests are in statistical signal processing, as it applies to adaptive radar and sonar, electric power, cyber security, and wireless communication. He has authored three books: L. L. Scharf, "Statistical Signal Processing: Detection, Estimation, and Time Series Analysis," Addison-Wesley, 1991; L. L. Scharf, "A First Course in Electrical and Computer Engineering," Addison-Wesley, 1998; and P. J. Schreier and L. L. Scharf, "Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals," Cambridge University Press, 2010.

Professor Scharf was Technical Program Chair for the 1980 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Tutorial Chair for ICASSP 2001; and Technical Program Chair for the Asilomar Conference on Signals, Systems, and Computers, 2002. He is past Chair of the Fellow Committee for the IEEE Signal Processing Society. He has received numerous awards for his contributions to statistical signal processing, including an IEEE Distinguished Lectureship; an IEEE Third Millennium Medal; the Donald W. Tufts Award for Underwater Acoustic Signal Processing, the Technical Achievement Award from the IEEE Signal Processing Society (SPS); and the Society Award from the IEEE SPS. He is a Life Fellow of IEEE.



Stephen D. Howard (M'10) graduated from La Trobe University, Melbourne, Australia, in 1982. He received the M.S. and Ph.D. degrees in mathematics from La Trobe University in 1984 and 1990, respectively. He joined the Australian Defence Science and Technology Organisation (DSTO) in 1991, where he has been involved in research in the area electronic surveillance and radar systems. He has led the DSTO research effort into the development of algorithms in all areas of electronic surveillance, including radar pulse train deinterleaving, precision radar parameter estimation and tracking, estimation of radar intrapulse modulation, and advanced geolocation techniques. Since 2003, he has led the DSTO long-range research program in radar resource management and waveform design.