ANALYSIS OF FISHER INFORMATION AND THE CRAMER-RAO BOUND FOR NONLINEAR PARAMETER ESTIMATION AFTER COMPRESSED SENSING

Pooria Pakrooh,[1] *Louis L. Scharf,*[2] *Ali Pezeshki,*[3] *and Yuejie Chi*[4]

[1] ECE Department, Colorado State University, Fort Collins, CO.
[2] Departments of Mathematics and Statistics, Colorado State University, Fort Collins, CO.
[3] Departments of ECE and Mathematics, Colorado State University, Fort Collins, CO.
[4] Departmnets of ECE and Biomedical Informatics, The Ohio State University, Columbus, OH.

ABSTRACT

In this paper, we analyze the impact of compressed sensing with random matrices on Fisher information and the CRB for estimating unknown parameters in the mean value function of a multivariate normal distribution. We consider the class of random compression matrices that satisfy a version of the Johnson-Lindenstrauss lemma, and we derive analytical lower and upper bounds on the CRB for estimating parameters from randomly compressed data. These bounds quantify the potential loss in CRB as a function of Fisher information of the non-compressed data. In our numerical examples, we consider a direction of arrival estimation problem and compare the actual loss in CRB with our bounds.

Index Terms— Cramer-Rao bound, compressed sensing, Fisher information, Johnson-Lindenstrauss Lemma, parameter estimation

1. 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. The 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 which exploits sparse representations and sparse recovery for inversion.

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 reconstructed parameter vector.

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 compressed sensing on the Fisher information matrix and the Cramer-Rao bound (CRB). In our analysis, we consider compression matrices which satisfy an extension of the Johnson-Lindenstrauss lemma (cf. [8]) for $p-$dimensional subspaces, where p is the dimension of the underlying parameter vector.

Our prior work on compressed sensing 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 of the measurements in a multivariate Normal model. In this paper we derive analytical lower and upper bounds on the CRB for that case. These bounds quantify the potential loss in CRB and provide loose guidelines for selecting the compression ratio to manage the loss in CRB. Our results are simulated for the special example of DOA estimation in which the unknown parameters are nonlinearly embedded in the mean of a Gaussian distribution. The results are plotted to give the upper and lower bounds on the Mean-Squared Error (MSE) of any unbiased estimator after compressed sensing.

Other studies on the effect of compressed sensing on the CRB and the Fisher information matrix include [9]–[11]. Babadi *et al.* [9] proposed a so-called "Joint Typicality Estimator" to show the existence of an estimator which asymptotically achieves the CRB of sparse parameter estimation for random Gaussian compression matrices. Niazadeh *el al.* [10] generalize the results of [9] to a class of random compression matrices which satisfy the concentration of measures inequality. 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 2.

2. PROBLEM STATEMENT

Let $\mathbf{y} \in \mathbb{R}^n$ be a real random vector whose probability density function $f(\mathbf{y}; \boldsymbol{\theta})$ is parametrized by an unknown but deterministic parameter vector $\boldsymbol{\theta} \in \mathbb{R}^p$. The derivative of the

This work is supported by NSF under grant CCF-1018472.

log-likelihood function with respect to $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_p]$ is called the Fisher score, and the covariance matrix of the Fisher score is the Fisher information matrix which we denote by $\mathbf{J}(\boldsymbol{\theta})$:

$$\mathbf{J}(\boldsymbol{\theta}) = E[(\frac{\partial \log f(\mathbf{y};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})(\frac{\partial \log f(\mathbf{y};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^T]. \quad (1)$$

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

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

where $\mathbf{A} \succeq \mathbf{B}$ for matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ means $\mathbf{a}^T\mathbf{A}\mathbf{a} \geq \mathbf{a}^T\mathbf{B}\mathbf{a}$ for all $\mathbf{a} \in \mathbb{R}^n$. The i^{th} diagonal element of $\mathbf{J}^{-1}(\boldsymbol{\theta})$ is the Cramer-Rao bound 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., [12]).

When $f(\mathbf{y};\boldsymbol{\theta})$ is a multivariate normal density $\mathcal{N}(\mathbf{x}(\boldsymbol{\theta}), \mathbf{R})$ with unknown mean vector $\mathbf{x}(\boldsymbol{\theta})$ parametrized by $\boldsymbol{\theta}$, and known covariance $\mathbf{R} = \sigma^2\mathbf{I}$, the Fisher information matrix is the Grammian

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

The i^{th} column $\mathbf{g}_i$ of $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_p]$ 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 the i^{th} parameter θ_i. The CRB for estimating θ_i is given by

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

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 $\mathbf{G}_i$ [13]. 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 CRB, which is discussed in detail in [13].

If $\mathbf{y}$ is compressed by a compression matrix $\boldsymbol{\Phi}$ to produce $\hat{\mathbf{y}} = \boldsymbol{\Phi}\mathbf{y}$, then the probability density function of the compressed data $\hat{\mathbf{y}}$ is $\mathcal{N}[\boldsymbol{\Phi}\mathbf{x}(\theta), \sigma^2\boldsymbol{\Phi}\boldsymbol{\Phi}^T]$ and the Fisher information matrix $\hat{\mathbf{J}}(\boldsymbol{\theta})$ and the CRB $(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{ii}$ for estimating θ_i from $\hat{\mathbf{y}}$ are given by

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

and

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

where $\hat{\mathbf{g}}_i = \mathbf{P}_{\boldsymbol{\Phi}^T}\mathbf{g}_i$, $\hat{\mathbf{G}}_i = \mathbf{P}_{\boldsymbol{\Phi}^T}\mathbf{G}_i$, and $\mathbf{P}_{\boldsymbol{\Phi}^T} = \boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{-1}\boldsymbol{\Phi}$ is the orthogonal projection onto row span of $\boldsymbol{\Phi}$ [13].

Our aim is to bound the loss in CRB due to compression. In Section 4, we investigate this problem for the case where the compressed sensing matrix $\boldsymbol{\Phi}$ satisfies a subspace version of the Johnson-Lindenstrauss (JL) Lemma, as will be discussed in Section 3.

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 $\mathcal{N}[\boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta}), \sigma^2\mathbf{I}]$, in contrast to ours which is distributed as $\mathcal{N}[\boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta}), \sigma^2\boldsymbol{\Phi}\boldsymbol{\Phi}^T]$. In a signal-plus-noise scheme our model corresponds to compressing the noisy signal $\mathbf{x}(\boldsymbol{\theta}) + \mathbf{n}$, $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, to produce $\boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta}) + \boldsymbol{\Phi}\mathbf{n}$. In contrast, their model corresponds to compressing a noiseless signal $\mathbf{x}(\boldsymbol{\theta})$ to produce $\boldsymbol{\Phi}\mathbf{x}(\theta) + \mathbf{w}$, where $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ 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.

3. SUBSPACE JOHNSON-LINDENSTRAUSS LEMMA

Definition 1: For any $\epsilon \in (0, 1)$, a random linear transformation $\boldsymbol{\Phi} : \mathbb{R}^n \to \mathbb{R}^m$ is said to satisfy an $\epsilon-$JL type Lemma over a set of vectors $\mathcal{Q} \subset \mathbb{R}^n$ with probability at least $1 - \delta$ if

$$Pr\left(\forall \mathbf{q} \in \mathcal{Q} : (1-\epsilon)\|\mathbf{q}\|_2^2 \leq \|\boldsymbol{\Phi}\mathbf{q}\|_2^2 \leq (1+\epsilon)\|\mathbf{q}\|_2^2\right) \geq 1 - \delta. \quad (8)$$

For random matrices that satisfy $\mathbb{E}(\|\boldsymbol{\Phi}\mathbf{r}\|_2^2) = \|\mathbf{r}\|_2^2$ for any $\mathbf{r} \in \mathbb{R}^n$, [14] uses the union bound to show that

$$\delta \leq 2|\mathcal{Q}|e^{-mc_0(\epsilon)} \quad (9)$$

where $c_0(\cdot)$ is a function that depends on the distribution from which the entries Φ_{ij} of $\boldsymbol{\Phi}$ are drawn. When Φ_{ij}'s are i.i.d $\mathcal{N}(0, 1/m)$, then $c_0(\epsilon) = \epsilon^2/4 - \epsilon^3/6$.

Now for our bounding purposes in Section 4, we generalize the $\epsilon-$JL type Lemma to the case of affine p-dimensional subspaces in Lemma 1.

Lemma 1: Let $\boldsymbol{\Phi}: \mathbb{R}^n \to \mathbb{R}^m$ $(m < n)$, and $\epsilon \in (0, 1)$. Then $\boldsymbol{\Phi}$ satisfies the $\epsilon-$JL type Lemma over any arbitrary p-dimensional subspace $\langle \mathbf{V} \rangle$ of $\mathbb{R}^n$ with probability at least $1 - \delta$, provided that it satisfies the $\epsilon'-$JL type Lemma over any set $\mathcal{Q} \subset \mathbb{R}^n$ of $\lceil (2\sqrt{p}/\epsilon')^p \rceil$ vectors with probability at least $1 - \delta$, where ϵ' satisfies $(\frac{3\epsilon'}{1-\epsilon'})^2 + 2(\frac{3\epsilon'}{1-\epsilon'}) = \epsilon$.

Remark 2: The statement of Lemma 1 and its proof (omitted for lack of spaces) are inspired by [15]. However, our proof is different and the above way of stating the result makes it more readily applicable for our analysis in Section 4.

4. CRAMER-RAO BOUND ON PARAMETER ESTIMATION AFTER COMPRESSED SENSING

To bound the CRB $(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{ii}$ after compressed sensing, we use the following Lemma.

Lemma 2: Let $\mathbf{X}_{m\times n}$ and $\mathbf{Y}_{m\times p}$ be two arbitrary matrices. Let $\mathbf{Z} = [\mathbf{X} \quad \mathbf{Y}]$. Then,

$$\lambda_{max}(\mathbf{Z}^T\mathbf{Z}) \geq \lambda_{max}(\mathbf{X}^T\mathbf{P}_{\mathbf{Y}}^{\perp}\mathbf{X}) \tag{10}$$

$$\lambda_{min}(\mathbf{Z}^T\mathbf{Z}) \leq \lambda_{min}(\mathbf{X}^T\mathbf{P}_{\mathbf{Y}}^{\perp}\mathbf{X}) \tag{11}$$

where $\lambda_{max}(\cdot)$ and $\lambda_{min}(\cdot)$ take the largest and smallest eigenvalues of a matrix respectively.

Proof: Proof follows from eigenvalue analysis of the inverse of the block Grammian matrix $\mathbf{Z}^T\mathbf{Z}$.

Now, let $\mathbf{Z} = [\hat{\mathbf{g}}_i \quad \hat{\mathbf{G}}_i]$. Then, from Lemma 2 it follows that

$$\lambda_{min}(\mathbf{G}^T\mathbf{P}_{\boldsymbol{\Phi}^T}\mathbf{G}) \leq \hat{\mathbf{g}}_i^T(\mathbf{I}-\mathbf{P}_{\hat{\mathbf{G}}_\mathbf{i}})\hat{\mathbf{g}}_i \leq \lambda_{max}(\mathbf{G}^T\mathbf{P}_{\boldsymbol{\Phi}^T}\mathbf{G}). \tag{12}$$

Since $\mathbf{G}^T(\mathbf{I}-\mathbf{P}_{\boldsymbol{\Phi}^T})\mathbf{G}$ is a positive semidefinite matrix, we have $\mathbf{G}^T\mathbf{G} \succeq \mathbf{G}^T\mathbf{P}_{\boldsymbol{\Phi}^T}\mathbf{G}$. Consequently,

$$\lambda_{max}(\mathbf{G}^T\mathbf{G}) \geq \lambda_{max}(\mathbf{G}^T\mathbf{P}_{\boldsymbol{\Phi}^T}\mathbf{G}). \tag{13}$$

From the upper bounds in (12) and (13) we have $\hat{\mathbf{g}}_i^T(\mathbf{I} - \mathbf{P}_{\hat{\mathbf{G}}_\mathbf{i}})\hat{\mathbf{g}}_i \leq \lambda_{max}(\mathbf{G}^T\mathbf{P}_{\boldsymbol{\Phi}^T}\mathbf{G})$.

The lower bound in (12) is bounded as

$$\begin{aligned}\lambda_{min}(\mathbf{G}^T\mathbf{P}_{\boldsymbol{\Phi}^T}\mathbf{G}) &= \min_{\|\mathbf{a}\|_2=1} \mathbf{a}^T\mathbf{G}^T\boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{-1}\boldsymbol{\Phi}\mathbf{G}\mathbf{a} \\ &\geq \lambda_{min}((\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{-1})\|\boldsymbol{\Phi}\mathbf{G}\mathbf{a}\|_2^2 \\ &\geq C\|\boldsymbol{\Phi}\mathbf{G}\mathbf{a}\|_2^2 \end{aligned} \tag{14}$$

with probabbility $1-\delta'$. Here $1-\delta'$ is the probability that $\lambda_{min}((\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{-1})$ is larger than C. This probability can be calculated based on the distribution of the maximum eigenvalue of the Grammian $\boldsymbol{\Phi}\boldsymbol{\Phi}^T$. For example, if $\boldsymbol{\Phi}_{m\times n}$ has i.i.d. entries distributed as $\mathcal{N}(0, 1/m)$, then $\boldsymbol{\Phi}\boldsymbol{\Phi}^T$ has Wishart distribution and δ' is determined based on the results of [16] on the distribution of the largest eigenvalue of a Wishart matrix and the value of C.

If $\boldsymbol{\Phi}$ satisfies the $\epsilon-$JL type Lemma for any p-dimensional subspace (including $\langle\mathbf{G}\rangle$) with probability at least $1-\delta$, then, we have

$$\begin{aligned}\|\boldsymbol{\Phi}\mathbf{G}\mathbf{a}\|_2^2 &\geq (1-\epsilon)\|\mathbf{G}\mathbf{a}\|_2^2 \\ &\geq (1-\epsilon)\lambda_{min}(\mathbf{G}^T\mathbf{G})\end{aligned} \tag{15}$$

with probability at least $1-\delta$. This guarantee probability can be lower bounded using (9) for δ and by replacing ϵ in (9) with ϵ' from Lemma 1. Combining (14) and (15), we have

$$\lambda_{min}(\mathbf{G}^T\mathbf{P}_{\boldsymbol{\Phi}^T}\mathbf{G}) \geq C(1-\epsilon)\lambda_{min}(\mathbf{G}^T\mathbf{G}) \tag{16}$$

with probability at least $1-\delta-\delta'$.

Recalling that $\mathbf{J}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\mathbf{G}^T\mathbf{G}$, $\hat{\mathbf{J}}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\mathbf{G}^T\mathbf{P}_{\boldsymbol{\Phi}^T}\mathbf{G}$ and using (12) and (16), we can bound the CRB after compressed sensing $(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{ii}$ as

$$\lambda_{min}(\mathbf{J}^{-1}(\theta)) \leq (\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{ii} \leq \frac{\lambda_{max}(\mathbf{J}^{-1}(\boldsymbol{\theta}))}{C(1-\epsilon)} \tag{17}$$

where the upper bound is valid with probability at least $1-\delta-\delta'$. Also, because $\mathbf{G}^T\mathbf{G} \succeq \mathbf{G}^T\mathbf{P}_{\boldsymbol{\Phi}^T}\mathbf{G}$ we always have an increase in CRB after compression. That is, $\mathbf{J}(\boldsymbol{\theta}) \succeq \hat{\mathbf{J}}(\boldsymbol{\theta})$, which implies

$$(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{ii} \geq (\mathbf{J}^{-1}(\boldsymbol{\theta}))_{ii} \geq \lambda_{min}(\mathbf{J}^{-1}(\boldsymbol{\theta})). \tag{18}$$

In summary, we have

$$(\mathbf{J}^{-1}(\boldsymbol{\theta}))_{ii} \leq (\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{ii} \leq \frac{\lambda_{max}(\mathbf{J}^{-1}(\boldsymbol{\theta}))}{C(1-\epsilon)} \tag{19}$$

where the upper bound is valid with probability at least $1-\delta-\delta'$.

Remark 3: We can also derive bounds on $\mathrm{tr}\,(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))$ and $\det(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))$ using (19) and Hadamard's inequality:

$$\mathrm{tr}(\mathbf{J}^{-1}(\boldsymbol{\theta})) \leq \mathrm{tr}(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta})) \leq \frac{p\lambda_{max}(\mathbf{J}^{-1}(\boldsymbol{\theta}))}{C(1-\epsilon)} \tag{20}$$

$$\det(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta})) \leq \prod_{i=1}^{p}(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{ii} \leq (\frac{\lambda_{max}(\mathbf{J}^{-1}(\boldsymbol{\theta}))}{C(1-\epsilon)})^p. \tag{21}$$

Again, these upper bounds are valid with probability at least $1-\delta-\delta'$. The trace of $\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta})$ is the CRB for estimating the parameter vector $\boldsymbol{\theta}$ from $\hat{\mathbf{y}}$ and the $\det(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))$ is the volume of the concentration ellipse in that estimation.

Remark 4: We can generalize our results to the case where $\mathbf{y}$ has complex Gaussian distribution $\mathcal{CN}(\mathbf{x}(\boldsymbol{\theta}), \sigma^2\mathbf{I})$. To do so, we define $\mathbf{G}_r = real(\frac{\partial\mathbf{x}(\theta)}{\partial\boldsymbol{\theta}})$ and $\mathbf{G}_i = imag(\frac{\partial\mathbf{x}(\theta)}{\partial\boldsymbol{\theta}})$. Writing the Fisher information matrix as in (1) and after some calculations we get $\mathbf{J}(\boldsymbol{\theta}) = \frac{2(\mathbf{G}_r^T\mathbf{G}_r+\mathbf{G}_i^T\mathbf{G}_i)}{\sigma^2} = \frac{1}{\sigma^2}\mathbf{G}^T\mathbf{G}$, where $\mathbf{G}^T = \sqrt{2}[\mathbf{G}_r^T \quad \mathbf{G}_i^T]$ is the complex matrix of sensitivity vectors. Therefore, we can do our analysis for $\mathbf{G}$ with its columns as the newly defined sensitivity vectors and derive similar bounds as in (19).

5. NUMERICAL RESULTS

As a special example, we consider the effect of compression on DOA estimation with 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 $\mathbf{x}(\boldsymbol{\theta}) = \mathbf{x}(\boldsymbol{\theta}_1) + \mathbf{x}(\boldsymbol{\theta}_2)$, where

$$\mathbf{x}(\theta_i) = A_i e^{j\phi_i}[1 \quad e^{j\theta_i} \quad e^{j2\theta_i} \quad \cdots \quad e^{j(N-1)\theta_i}]^T. \tag{22}$$

Here A_i and ϕ_i are the amplitude and phase of the i^{th} source, which we assume to know. 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 θ_2. The CRB on the estimation of θ_1 is calculated for different values of θ_2. For our simulations, we use Gaussian compression matrices $\mathbf{\Phi}_{m\times n}$ whose elements are i.i.d. $\mathcal{N}(0, 1/m)$. Fig. 1 shows the results for $m = 32$, $n = 64$ and Fig. 2 shows the results for $m = 3000$, $n = 8192$. The red curves are the actual CRBs $(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{11}$ for several (10 in Fig. 1 and 5 in Fig. 2) independent realizations of random $\mathbf{\Phi}$, the blue curves represent the before compression CRBs which lower bound the CRBs $(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{11}$. The black curves represent the upper bounds on $(\hat{\mathbf{J}}^{-1}(\boldsymbol{\theta}))_{11}$. In computing these bounds we have set $C = 0.17$ in Fig.1 and $C = 0.14$ in Fig.2 to get $\delta' = 0.05$ in each case. The values of ϵ for the upper bounds are 0.66 (solid black) and 0.33 (dashed black). These bounds quantify the loss in the CRB after compressed sensing.

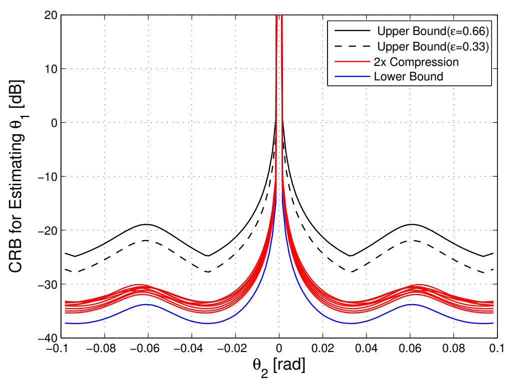

Fig. 1. Bounds on the CRB for $-2\pi/n \leq \theta_2 \leq 2\pi/n$ (Rayleigh limit), $m = 32, n = 64$

The upper bounds (black curves) for the CRB are valid with probability at least $1 - \delta - \delta'$, where $1 - \delta$ is the lower bound on the probablity with which the compression matrix $\mathbf{\Phi}$ is ϵ-JL over the set of $(p = 2)$-dimensional subpaces of $\mathbb{R}^n$ and $\delta' = Pr(\lambda_{min}((\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}) < C)$. For a Gaussian compression matrix, $\mathbf{\Phi}\mathbf{\Phi}^T$ is Wishart and δ' can be easily calculated using the result of [16]. A lower bound on the ϵ-JL probability $1 - \delta$ (or alternatively an upper bound on δ) can be calcuated using the arguments in [14] or [17]. Fig. 3 shows the plots of upper bounds for δ versus the number of measurements m for $\epsilon = 0.66$ (red curve) and $\epsilon = 0.33$ (green curve) for $n = 8192$. For any choice of m, Fig. 3 can be used to determine a lower bound on the confidence level $1 - \delta - \delta'$ for the upper bound on the after compression CRB for each value of ϵ. Alternatively, we can plot the curves versus ϵ for fixed values of m. 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 ϵ) with probability at least $1 - \delta - \delta'$. We note that, as with many probabilistic performance guarantees for compressed sensing, these probability bounds are too conservative, and sometimes they result in even trivial bounds, unless m and n are very large.

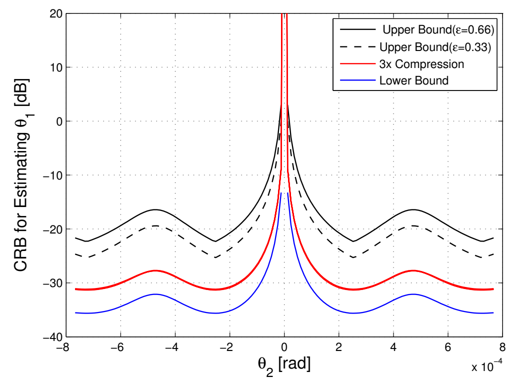

Fig. 2. Bounds on the CRB for $-2\pi/n \leq \theta_2 \leq 2\pi/n$ (Rayleigh limit), $m = 3000, n = 8192$

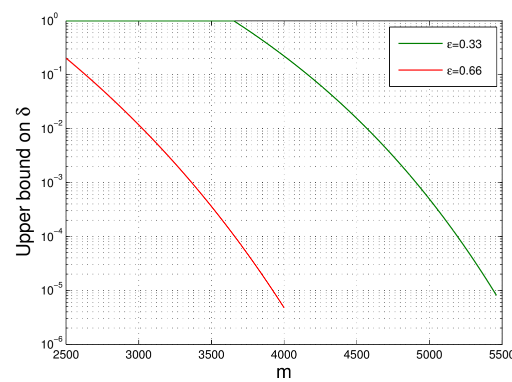

Fig. 3. The plot of an upper bound on δ versus the number of measurements m for $n = 8192$ and $\epsilon = 0.66$ (red) and $\epsilon = 0.33$ (green)

6. 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. The class of random compressors under study preserve the norm of vectors in any $p-$dimensional subspace up to a small multiplicative factor. The CRB before compressed sensing lower bounds the CRB after compression. An upper bound was derived for the CRB after compressed sensing.

7. REFERENCES

[1] E. J. Candes, "Compressive sampling," in *Proc.Int. Congress Math.*, 2006, vol. 3, pp. 1433–1452.

[2] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 –1306, 2006.

[3] R.G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118 –121, Jul. 2007.

[4] Y. Chi, L.L. Scharf, A. Pezeshki, and A.R. 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, Systs., Computs., Pacific Grove, CA,*, Nov. 2011, pp. 744 –748.

[7] L. L. Scharf, E. K. P. Chong, A. Pezeshki, and J. R. Luo, "Compressive sensing and sparse inversion in signal processing: Cautionary notes," in *Proc. 7th Workshop on Defence Applications of Signal Processing (DASP), Coolum, Queensland, Australia, Jul. 10-14, 2011*.

[8] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," in *Conf. Modern Analysis and Probability.*, 1984, pp. 189–206.

[9] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramer-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 Cramer-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, Systs., Computs., Pacific Grove, CA, Nov. 4-7, 2012, to appear.*

[12] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory (v. 1)*, Prentice-Hall, 1st edition, Apr. 1993.

[13] L. L. Scharf and L. T. McWhorter, "Geometry of the Cramer-Rao bound," *Signal Process.*, vol. 31, no. 3, pp. 301–311, Apr. 1993.

[14] R. Baraniuk, M. Davenport, R. Devore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx*, vol. 28, no. 3, pp. 253–263, Dec. 2008.

[15] T. Sarlos, "Improved approximation algorithms for large matrices via random projections," in *Proc. 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS), Berkeley, California*, Oct. 2006, pp. 143–152.

[16] R. W. Butler and R. L. Paige, "Exact distributional computations for Roy's statistic and the largest eigenvalue of a Wishart distribution," in *Proceedings of Statistics and Computing*, 2011, pp. 147–157.

[17] Dasgupta S. and Gupta A., "A simple proof of the restricted isometry property for random matrices," *Random Struc. Algorithms*, vol. 22, no. 1, pp. 60–65, 2003.