On the Analysis of the Fisher Information of a Perturbed Linear Model after Random Compression

Jiang Zhu, Lin Han, Rick S. Blum and Zhiwei Xu

Abstract

The impact of random compression on the Fisher information matrix (FIM) and the Cramér-Rao bound (CRB) is studied when estimating unknown complex parameters in the perturbed linear model. A random compression matrix is considered whose elements are i.i.d standard complex normal random variables. The FIM averaged over compression is equal to a scalar of the FIM before compression plus an additional term. The upper and lower bounds of the CRB averaged over the random compression matrix are also given. Finally, numerical results are conducted to verify our theoretical results.

Index Terms

Perturbation, Cramér-Rao bound, Fisher information, random compression

I. Introduction

One of the classical problems in statistical signal processing is to estimate an unknown deterministic parameter vector $\boldsymbol{\theta}$ from its observations $\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{n}$, where $\mathbf{A}$ is a measurement matrix, and $\mathbf{n}$ is an additive white Gaussian noise. This model is important as it can be applied in various applications such as radar processing and digital communications [1, 2].

The maximum likelihood estimator for the above model leads naturally to a least squares (LS) solution. In addition, this maximum likelihood estimator is unbiased so its performance limits can be characterized by the corresponding Cramér-Rao bound (CRB) [1–3]. Recently, the effects of random compression on the Fisher information matrix (FIM) and the CRB have been studied [4–6]. In [4] and [5, 6], the impact of compression on Fisher information has been studied in depth. The authors in [4] study a model where a noiseless signal is compressed, whereas in [5, 6], the effects of a random compression matrix on a signal-plus-noise model have been studied. The effects of random compression on a direction of arrival (DOA) problem are studied in [7]. In addition, using random compression matrix for dimension reduction in Kalman Filtering has been studied in [8].

In practice, the linear model may not be completely known. One way is to model the measurement matrix $\mathbf{A}$ as a random matrix with known statistics, and the CRB is derived in [9]. In addition, several works have extended [9] to the sparse recovery [10–14] and quantization settings [15, 16]. In this paper, we derive the FIM and CRB both before and after random compression for a model with a perturbed sensing matrix.

II. CRB without Random Compression

The linear regression problem with perturbed sensing matrix can be described as (extended to complex observations)

$$\mathbf{y} = (\mathbf{A} + \mathbf{E}\mathbf{C})\boldsymbol{\theta} + \boldsymbol{\upsilon}, \tag{1}$$

where $\mathbf{E} \in \mathbb{C}^{n\times r}$ is a perturbation matrix with entries which are i.i.d complex Gaussian distributed random variables with zero mean and variance σ_e^2, $\mathbf{C} \in \mathbb{C}^{r\times p}$ is a known matrix, $\boldsymbol{\theta} \in \mathbb{C}^p$ is a deterministic unknown complex vector, and $\boldsymbol{\upsilon} \sim \mathcal{CN}(\mathbf{0}, \sigma_\upsilon^2\mathbf{I}_n)$ is an additive white Gaussian noise vector, where $\mathcal{CN}$ denotes complex Gaussian distribution. Note that $\mathbf{C}$ imposes some structure on the sensing matrix $\mathbf{A}$. For example, the choice $\mathbf{C} = [\mathbf{0}_{r\times q}, \mathbf{I}_{(p-q)}]$, where $r = p - q$, corresponds to the scenario in which the first q columns of $\mathbf{A}$ are error free while the remaining $p - q$ are noisy [17]. In the following text, σ_e^2 is viewed as the strength of the perturbation.

Given the observation vector $\mathbf{y}$ which determines the estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$, a widely used criterion to characterize the performance of the estimator is the mean square error (MSE), given by

$$\mathrm{mse}(\hat{\boldsymbol{\theta}}(\mathbf{y})) = \mathrm{E}_{\mathbf{y};\boldsymbol{\theta}}[(\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta})^{\mathrm{H}}(\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta})]. \tag{2}$$

Here, $\mathrm{E}_{\mathbf{y};\boldsymbol{\theta}}[\cdot]$ denotes the expectation taken with respect to the probability density function (pdf) $p(\mathbf{y};\boldsymbol{\theta})$ of the measurement vector $\mathbf{y}$ parameterized by $\boldsymbol{\theta}$, $(\cdot)^{\mathrm{H}}$ represent the conjugate transpose operator. A commonly employed approach seeks unbiased estimators satisfying $\mathrm{E}_{\mathbf{y};\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}(\mathbf{y})] = \boldsymbol{\theta}$ which minimize MSE. Fortunately, the performance of an unbiased estimator can be characterised by the CRB.

To derive the complex CRB, let $\underline{\boldsymbol{\theta}} \in \mathbb{C}^{2p}$ be $\underline{\boldsymbol{\theta}} = [\boldsymbol{\theta}^{\mathrm{T}}, \boldsymbol{\theta}^{\mathrm{H}}]^{\mathrm{T}}$ [18], where $(\cdot)^{\mathrm{T}}$ represent the transpose operator. Consequently, under usual regularity conditions, the covariance of any unbiased estimator $\hat{\underline{\boldsymbol{\theta}}}(\mathbf{y})$ satisfies [18]

$$\mathrm{E}_{\mathbf{y};\boldsymbol{\theta}}[(\hat{\underline{\boldsymbol{\theta}}}(\mathbf{y}) - \underline{\boldsymbol{\theta}})(\hat{\underline{\boldsymbol{\theta}}}(\mathbf{y}) - \underline{\boldsymbol{\theta}})^{\mathrm{H}}] \succeq \mathbf{J}^{-1}(\underline{\boldsymbol{\theta}}) \triangleq \mathbf{CRB}(\underline{\boldsymbol{\theta}}), \tag{3}$$

where $\mathbf{J}(\underline{\boldsymbol{\theta}}) \in \mathbb{C}^{2p\times 2p}$ is the FIM given by

$$\mathbf{J}(\underline{\boldsymbol{\theta}}) = \mathrm{E}_{\mathbf{y};\boldsymbol{\theta}}\left[\left(\frac{\partial \log f(\mathbf{y};\boldsymbol{\theta})}{\partial \underline{\boldsymbol{\theta}}}\right)^{*}\left(\frac{\partial \log f(\mathbf{y};\boldsymbol{\theta})}{\partial \underline{\boldsymbol{\theta}}}\right)^{\mathrm{T}}\right]. \tag{4}$$

$(\cdot)^*$ represent the conjugate operator, and CRB follows by taking the inverse of the FIM [2]. According to (2) and (3), we have $\mathrm{mse}(\hat{\boldsymbol{\theta}}(\mathbf{y})) = \mathrm{mse}(\hat{\underline{\boldsymbol{\theta}}}(\mathbf{y}))/2 \geq \mathrm{tr}[\mathbf{CRB}(\underline{\boldsymbol{\theta}})]/2$.

To evaluate the CRB for model (1), the following lemma is utilized.

Lemma 1 *[18] (General Complex Gaussian CRLB) Assume the observations* $\mathbf{y}$ *follow the complex Gaussian distribution* $\mathbf{y} \sim \mathcal{CN}(\mathbf{s}(\boldsymbol{\theta}), \boldsymbol{\Gamma}(\boldsymbol{\theta}))$*, then the* (k,l)*th element of the complex FIM* $\mathbf{J}(\underline{\boldsymbol{\theta}})$ *is*

$$\begin{aligned}[\mathbf{J}(\underline{\boldsymbol{\theta}})]_{kl} = &\left[\frac{\partial \mathbf{s}}{\partial \underline{\theta}_k}\right]^{\mathrm{H}}\boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta})\left[\frac{\partial \mathbf{s}}{\partial \underline{\theta}_l}\right] + \left[\frac{\partial \mathbf{s}}{\partial \underline{\theta}_l^*}\right]^{\mathrm{H}}\boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta})\left[\frac{\partial \mathbf{s}}{\partial \underline{\theta}_k^*}\right] \\ &+ \mathrm{tr}\left\{\left[\frac{\partial \boldsymbol{\Gamma}}{\partial \underline{\theta}_k^*}\right]\boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta})\left[\frac{\partial \boldsymbol{\Gamma}}{\partial \underline{\theta}_l}\right]\boldsymbol{\Gamma}^{-1}(\boldsymbol{\theta})\right\},\end{aligned} \tag{5}$$

for $1 \leq k,\ l \leq 2p$.

In the following, we use $\mathbf{J}$ and $\mathbf{J}^{-1}$ instead of $\mathbf{J}(\underline{\boldsymbol{\theta}})$ and $\mathbf{J}^{-1}(\underline{\boldsymbol{\theta}})$. By defining

$$\mathbf{w} = \mathbf{E}\mathbf{C}\boldsymbol{\theta} + \boldsymbol{\upsilon}, \tag{6}$$

model (1) simplifies to $\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{w}$. Meanwhile, since $\mathbf{E}$ is independent of $\boldsymbol{\upsilon}$ and both of them follow a complex normal distribution, it can be shown that $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2\mathbf{I}_n)$, where

$$\sigma_w^2 = \sigma_e^2\|\mathbf{C}\boldsymbol{\theta}\|_2^2 + \sigma_\upsilon^2. \tag{7}$$

To apply Lemma 1 to our model, $\mathbf{s}(\boldsymbol{\theta}) = \mathbf{A}\boldsymbol{\theta}$, $\boldsymbol{\Gamma}(\boldsymbol{\theta}) = \sigma_w^2\mathbf{I}_n = (\sigma_e^2\|\mathbf{C}\boldsymbol{\theta}\|_2^2 + \sigma_\upsilon^2)\mathbf{I}_n$. By defining $\boldsymbol{\theta}_c = \mathbf{C}^{\mathrm{H}}\mathbf{C}\boldsymbol{\theta}$ and using Lemma 1 and Table I, the FIM can be calculated as

$$\mathbf{J} = \begin{bmatrix} \frac{\mathbf{A}^{\mathrm{H}}\mathbf{A}}{\sigma_w^2} & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \frac{(\mathbf{A}^{\mathrm{H}}\mathbf{A})^*}{\sigma_w^2}\end{bmatrix} + \frac{n\sigma_e^4}{\sigma_w^4}\begin{bmatrix}\boldsymbol{\theta}_c \\ \boldsymbol{\theta}_c^*\end{bmatrix}\begin{bmatrix}\boldsymbol{\theta}_c \\ \boldsymbol{\theta}_c^*\end{bmatrix}^{\mathrm{H}}. \tag{8}$$

Applying the rank one update formula the CRB can be obtained as

$$\begin{aligned}\mathbf{J}^{-1} = &\begin{bmatrix} \sigma_w^2(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1} & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \sigma_w^2[(\mathbf{A}^{\mathrm{H}}\mathbf{A})^*]^{-1}\end{bmatrix} - \frac{n\sigma_e^4\sigma_w^2}{d_0}\times \\ &\begin{bmatrix}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\boldsymbol{\theta}_c \\ [(\mathbf{A}^{\mathrm{H}}\mathbf{A})^*]^{-1}\boldsymbol{\theta}_c^*\end{bmatrix}\begin{bmatrix}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\boldsymbol{\theta}_c \\ [(\mathbf{A}^{\mathrm{H}}\mathbf{A})^*]^{-1}\boldsymbol{\theta}_c^*\end{bmatrix}^{\mathrm{H}},\end{aligned} \tag{9}$$

where

$$d_0 = \sigma_w^2 + n\sigma_e^4\left[\boldsymbol{\theta}_c^{\mathrm{H}}\left(\mathbf{A}^{\mathrm{H}}\mathbf{A}\right)^{-1}\boldsymbol{\theta}_c + \boldsymbol{\theta}_c^{\mathrm{T}}[(\mathbf{A}^{\mathrm{H}}\mathbf{A})^*]^{-1}\boldsymbol{\theta}_c^*\right]. \tag{10}$$

TABLE I
THE PARTIAL DERIVATIVES OF $\mathbf{s}(\boldsymbol{\theta})$ AND $\boldsymbol{\Gamma}(\boldsymbol{\theta})$ WITH RESPECT TO $\underline{\boldsymbol{\theta}}$. .

$k\&l\backslash\partial(\cdot)/\partial(\cdot)$	$\frac{\partial \mathbf{s}}{\partial \underline{\theta}_k}$	$\frac{\partial \mathbf{s}}{\partial \underline{\theta}_l}$	$\frac{\partial \mathbf{s}}{\partial \underline{\theta}_l^*}$	$\frac{\partial \mathbf{s}}{\partial \underline{\theta}_k^*}$	$\frac{\partial \boldsymbol{\Gamma}}{\partial \underline{\theta}_k^*}$	$\frac{\partial \boldsymbol{\Gamma}}{\partial \underline{\theta}_l}$
$1 \leq k \leq p, 1 \leq l \leq p$	$\mathbf{A}\mathbf{e}_k$	$\mathbf{A}\mathbf{e}_l$	$\mathbf{0}_n$	$\mathbf{0}_n$	$\sigma_e^2\mathbf{e}_k^{\mathrm{H}}(\mathbf{C}^{\mathrm{H}}\mathbf{C})\theta$	$\sigma_e^2\theta^{\mathrm{H}}(\mathbf{C}^{\mathrm{H}}\mathbf{C})\mathbf{e}_l$
$1 \leq k \leq p, p+1 \leq l \leq 2p$	$\mathbf{A}\mathbf{e}_k$	$\mathbf{0}_n$	$\mathbf{A}\mathbf{e}_{l-p}$	$\mathbf{0}_n$	$\sigma_e^2\mathbf{e}_k^{\mathrm{H}}(\mathbf{C}^{\mathrm{H}}\mathbf{C})\theta$	$\sigma_e^2\mathbf{e}_{l-p}^{\mathrm{H}}(\mathbf{C}^{\mathrm{H}}\mathbf{C})\theta$
$p+1 \leq k \leq 2p, 1 \leq l \leq p$	$\mathbf{0}_n$	$\mathbf{A}\mathbf{e}_l$	$\mathbf{0}_n$	$\mathbf{A}\mathbf{e}_{k-p}$	$\sigma_e^2\theta^{\mathrm{H}}(\mathbf{C}^{\mathrm{H}}\mathbf{C})\mathbf{e}_{k-p}$	$\sigma_e^2\theta^{\mathrm{H}}(\mathbf{C}^{\mathrm{H}}\mathbf{C})\mathbf{e}_l$
$p+1 \leq k \leq 2p, p+1 \leq l \leq 2p$	$\mathbf{0}_n$	$\mathbf{0}_n$	$\mathbf{A}\mathbf{e}_{l-p}$	$\mathbf{A}\mathbf{e}_{k-p}$	$\sigma_e^2\theta^{\mathrm{H}}(\mathbf{C}^{\mathrm{H}}\mathbf{C})\mathbf{e}_{k-p}$	$\sigma_e^2\mathbf{e}_{l-p}^{\mathrm{H}}(\mathbf{C}^{\mathrm{H}}\mathbf{C})\theta$

III. EFFECTS OF RANDOM COMPRESSION

In this section, the effects of random compression on the FIM and CRB are analyzed. In particular, the relationships between the FIMs and CRBs before and after compression are obtained.

For the original measurement model (1), let $\mathbf{z}$ denote the compressed observations and $\mathbf{z} = \boldsymbol{\Phi}\mathbf{y}$, where $\boldsymbol{\Phi} \in \mathbb{C}^{m\times n}$ represents a random compression matrix whose elements are i.i.d. standard complex normal random variables. It is assumed that $p \leq m \leq n-p$, which is representative in almost all interesting compression scenarios. Straightforward calculation shows that $\mathbf{z} \sim \mathcal{CN}(\boldsymbol{\Phi}\mathbf{A}\boldsymbol{\theta}, \sigma_w^2\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}})$. Now we can utilize Lemma 1 to evaluate the CRB in the compressed observations scenario. In this setting, $\mathbf{s}(\boldsymbol{\theta}) = \boldsymbol{\Phi}\mathbf{A}\boldsymbol{\theta}$ and $\boldsymbol{\Gamma}(\boldsymbol{\theta}) = \sigma_w^2\boldsymbol{\Phi}\boldsymbol{\Phi}^H$.

Consequently, according to Lemma 1, one obtains

$$\widetilde{\mathbf{J}} = \begin{bmatrix} \frac{\mathbf{A}^{\mathrm{H}}\mathbf{P}_{\mathbf{\Phi}^{\mathrm{H}}}\mathbf{A}}{\sigma_w^2} & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \frac{(\mathbf{A}^{\mathrm{H}}\mathbf{P}_{\mathbf{\Phi}^{\mathrm{H}}}\mathbf{A})^*}{\sigma_w^2} \end{bmatrix} + \frac{n\sigma_e^4}{\sigma_w^4}\begin{bmatrix} \boldsymbol{\theta}_c \\ \boldsymbol{\theta}_c^* \end{bmatrix}\begin{bmatrix} \boldsymbol{\theta}_c \\ \boldsymbol{\theta}_c^* \end{bmatrix}^{\mathrm{H}}, \tag{11}$$

where $\widetilde{\mathbf{J}}$ denotes the FIM under compressed observations, and $\mathbf{P}_{\mathbf{\Phi}^{\mathrm{H}}} = \mathbf{\Phi}^{\mathrm{H}}(\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{H}})^{-1}\mathbf{\Phi}$ is the orthogonal projection onto the row span of $\mathbf{\Phi}$. By using the rank one formula, the CRB under compressed observations is (12).

$$\widetilde{\mathbf{J}}^{-1} = \begin{bmatrix} \sigma_w^2(\mathbf{A}^{\mathrm{H}}\mathbf{P}_{\mathbf{\Phi}^{\mathrm{H}}}\mathbf{A})^{-1} & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \sigma_w^2[(\mathbf{A}^{\mathrm{H}}\mathbf{P}_{\mathbf{\Phi}^{\mathrm{H}}}\mathbf{A})^*]^{-1} \end{bmatrix} - \frac{n\sigma_w^2\sigma_e^4\begin{bmatrix} (\mathbf{A}^{\mathrm{H}}\mathbf{P}_{\mathbf{\Phi}^{\mathrm{H}}}\mathbf{A})^{-1}\boldsymbol{\theta}_c \\ ((\mathbf{A}^{\mathrm{H}}\mathbf{P}_{\mathbf{\Phi}^{\mathrm{H}}}\mathbf{A})^*)^{-1}\boldsymbol{\theta}_c^* \end{bmatrix}\begin{bmatrix} (\mathbf{A}^{\mathrm{H}}\mathbf{P}_{\mathbf{\Phi}^{\mathrm{H}}}\mathbf{A})^{-1}\boldsymbol{\theta}_c \\ ((\mathbf{A}^{\mathrm{H}}\mathbf{P}_{\mathbf{\Phi}^{\mathrm{H}}}\mathbf{A})^*)^{-1}\boldsymbol{\theta}_c^* \end{bmatrix}^{\mathrm{H}}}{\sigma_w^2 + n\sigma_e^4\left[\boldsymbol{\theta}_c^{\mathrm{H}}\left(\mathbf{A}^{\mathrm{H}}\mathbf{P}_{\mathbf{\Phi}^{\mathrm{H}}}\mathbf{A}\right)^{-1}\boldsymbol{\theta}_c + \boldsymbol{\theta}_c^{\mathrm{T}}[(\mathbf{A}^{\mathrm{H}}\mathbf{P}_{\mathbf{\Phi}^{\mathrm{H}}}\mathbf{A})^*]^{-1}\boldsymbol{\theta}_c^*\right]} \tag{12}$$

In the following, we will derive the relationships between the before and after compression FIM (CRB). Before presenting the results, the following lemma is introduced.

Lemma 2 *[5]With $\boldsymbol{\theta}$ being a real vector, let $\mathbf{J}_0 = \mathbf{A}^{\mathrm{H}}\mathbf{A}/\sigma_w^2$ denote the FIM and let $\widetilde{\mathbf{J}}_0 = \mathbf{A}^{\mathrm{H}}\mathbf{P}_{\mathbf{\Phi}^{\mathrm{H}}}\mathbf{A}/\sigma_w^2$ denote the FIM under compressed observations in the absence of perturbation. In addition, the corresponding CRBs are $\mathbf{J}_0^{-1} = \sigma_w^2(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}$ and $\widetilde{\mathbf{J}}_0^{-1} = \sigma_w^2(\mathbf{A}^{\mathrm{H}}\mathbf{P}_{\mathbf{\Phi}^{\mathrm{H}}}\mathbf{A})^{-1}$, respectively. It can be shown that*

$$\mathrm{E}_{\mathbf{\Phi}}[\widetilde{\mathbf{J}}_0] = (m/n)\mathbf{J}_0, \tag{13a}$$

$$\mathrm{E}_{\mathbf{\Phi}}[\widetilde{\mathbf{J}}_0^{-1}] = (n-p)/(m-p)\mathbf{J}_0^{-1}. \tag{13b}$$

Using (13a), the expectation of $\widetilde{\mathbf{J}}$ (11) is

$$\mathrm{E}_{\mathbf{\Phi}}[\widetilde{\mathbf{J}}] = \begin{bmatrix} \frac{m\mathbf{A}^{\mathrm{H}}\mathbf{A}}{n\sigma_w^2} & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \frac{m(\mathbf{A}^{\mathrm{H}}\mathbf{A})^*}{n\sigma_w^2} \end{bmatrix} + \frac{n\sigma_e^4}{\sigma_w^4}\begin{bmatrix} \boldsymbol{\theta}_c \\ \boldsymbol{\theta}_c^* \end{bmatrix}\begin{bmatrix} \boldsymbol{\theta}_c \\ \boldsymbol{\theta}_c^* \end{bmatrix}^{\mathrm{H}}. \tag{14}$$

Substituting (8) in (14), we obtain

$$\mathrm{E}_{\mathbf{\Phi}}[\widetilde{\mathbf{J}}] = \frac{m}{n}\mathbf{J} + \left(1 - \frac{m}{n}\right)\frac{n\sigma_e^4}{\sigma_w^4}\begin{bmatrix} \boldsymbol{\theta}_c \\ \boldsymbol{\theta}_c^* \end{bmatrix}\begin{bmatrix} \boldsymbol{\theta}_c \\ \boldsymbol{\theta}_c^* \end{bmatrix}^{\mathrm{H}}. \tag{15}$$

Thus after perturbation, the FIM averaged over the compression, is equal to the FIM before compression plus an additional term independent of $\mathbf{A}$. In addition, it can be shown that $\frac{(\widetilde{\mathbf{J}}_0)_{ii}}{(\mathbf{J}_0)_{ii}} = \frac{\mathbf{e}_i^{\mathrm{H}}\widetilde{\mathbf{J}}_0\mathbf{e}_i}{\mathbf{e}_i^{\mathrm{H}}\mathbf{J}_0\mathbf{e}_i} \sim B^I(m, n-m)$ [19], where $B^I(m, n-m)$ denotes the Type I complex multivariate beta distribution. Thus $\mathrm{var}[(\widetilde{\mathbf{J}}_0)_{ii}] = \frac{m(n-m)}{n^2(n+1)}((\mathbf{J}_0)_{ii})^2$. In the presence of perturbation, one has $\mathrm{var}[(\widetilde{\mathbf{J}})_{ii}] = \frac{m(n-m)}{n^2(n+1)}((\mathbf{J}_0)_{ii})^2$. Given a fixed compression ratio m/n, as m tends to infinity the variance of $(\widetilde{\mathbf{J}})_{ii}$ tends to zero, which means that in the high dimensional setting, random compression results in a stable Fisher information loss.

Now, we begin to analyze the relationship between the CRB before and after compression. From (12) and (13b), it is difficult to obtain the expectation of $\widetilde{\mathbf{J}}^{-1}$ over $\mathbf{\Phi}$. As a result, we provide upper and lower bounds. For the upper bound, the following result is presented.

Lemma 3 $\mathbf{f}(\mathbf{z}) = \frac{b\mathbf{z}\mathbf{z}^{\mathrm{H}}}{a+b\underline{\boldsymbol{\theta}}_c^{\mathrm{H}}\mathbf{z}}$ *is matrix convex with respect to $\mathbf{z}$, where $\underline{\boldsymbol{\theta}}_c^{\mathrm{H}}\mathbf{z} \geq 0$, $a > 0$ and $b > 0$.*

The proof follows by

$$(1-\alpha)\mathbf{f}(\mathbf{z}_1) + \alpha\mathbf{f}(\mathbf{z}_2) - \mathbf{f}[(1-\alpha)\mathbf{z}_1 + \alpha\mathbf{z}_2] \succeq \mathbf{0}. \tag{16}$$

Because $f(\mathbf{z})$ is matrix convex, Jensen's inequality implies that

$$\mathrm{E}_{\mathbf{z}}[\mathbf{f}(\mathbf{z})] \succeq \mathbf{f}(\mathrm{E}_{\mathbf{z}}[\mathbf{z}]). \tag{17}$$

The second term of (12) can be expressed as $\sigma_w^2\mathbf{f}(\mathbf{z})$, where $a = \sigma_w^2$, $b = n\sigma_e^4$ and

$$\mathbf{z} = \begin{bmatrix} (\mathbf{A}^{\mathrm{H}}\mathbf{P}_{\mathbf{\Phi}^{\mathrm{H}}}\mathbf{A})^{-1}\boldsymbol{\theta}_c \\ [(\mathbf{A}^{\mathrm{H}}\mathbf{P}_{\mathbf{\Phi}^{\mathrm{H}}}\mathbf{A})^*]^{-1}\boldsymbol{\theta}_c^* \end{bmatrix}.$$

Note that

$$\mathrm{E}_{\mathbf{z}}[\mathbf{z}] = \frac{n-p}{m-p}\begin{bmatrix} (\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\boldsymbol{\theta}_c \\ [(\mathbf{A}^{\mathrm{H}}\mathbf{A})^*]^{-1}\boldsymbol{\theta}_c^* \end{bmatrix}$$

is calculated according to (13b). Taking the expectation of $\widetilde{\mathbf{J}}^{-1}$ and substituting (13b) in (12), we obtain

$$\begin{aligned}\mathrm{E}_{\mathbf{\Phi}}\left[\widetilde{\mathbf{J}}^{-1}\right] = \frac{n-p}{m-p}\left[\begin{array}{cc}\sigma_w^2(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1} & \mathbf{0}_{p\times p}\\ \mathbf{0}_{p\times p} & \sigma_w^2[(\mathbf{A}^{\mathrm{H}}\mathbf{A})^*]^{-1}\end{array}\right]\\ -\sigma_w^2\mathrm{E}_{\mathbf{z}}[f(\mathbf{z})].\end{aligned}$$

With (17), along with fact that $\mathbf{f}(\mathbf{z})$ is also matrix convex, one has (18), where c_u is

$$\begin{aligned}\mathrm{E}_{\mathbf{\Phi}}\left[\widetilde{\mathbf{J}}^{-1}\right] &\preceq \frac{n-p}{m-p}\left[\begin{array}{cc}\sigma_w^2(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1} & \mathbf{0}_{p\times p}\\ \mathbf{0}_{p\times p} & \sigma_w^2[(\mathbf{A}^{\mathrm{H}}\mathbf{A})^*]^{-1}\end{array}\right] - \sigma_w^2\mathbf{f}\left(\mathrm{E}_{\mathbf{z}}[\mathbf{z}]\right)\\ &= \frac{n-p}{m-p}\mathbf{J}^{-1} - c_u\left[\begin{array}{c}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\boldsymbol{\theta}_c\\ ((\mathbf{A}^{\mathrm{H}}\mathbf{A})^*)^{-1}\boldsymbol{\theta}_c^*\end{array}\right]\left[\begin{array}{c}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\boldsymbol{\theta}_c\\ ((\mathbf{A}^{\mathrm{H}}\mathbf{A})^*)^{-1}\boldsymbol{\theta}_c^*\end{array}\right]^{\mathrm{H}} \triangleq \widetilde{\mathbf{J}}_u^{-1}\end{aligned} \tag{18}$$

$$c_u = \frac{n(n-m)(n-p)\sigma_w^4\sigma_e^4}{(m-p)^2 d_0\left[\sigma_w^2 + (n-p)/(m-p)(d_0-\sigma_w^2)\right]} \geq 0, \tag{19}$$

and d_0 is (10).

Now we provide the lower bound of $\mathrm{E}_{\mathbf{\Phi}}[\widetilde{\mathbf{J}}^{-1}]$. According to [20], the function $\mathbf{X}^{-1}$ is matrix convex on $\mathbf{X}\succ\mathbf{0}$. Consequently, we have (20)

$$\begin{aligned}\mathrm{E}_{\mathbf{\Phi}}[\widetilde{\mathbf{J}}^{-1}] &\succeq (\mathrm{E}_{\mathbf{\Phi}}[\widetilde{\mathbf{J}}])^{-1} = \left[\left[\begin{array}{cc}\frac{m\mathbf{A}^{\mathrm{H}}\mathbf{A}}{n\sigma_w^2} & \mathbf{0}_{p\times p}\\ \mathbf{0}_{p\times p} & \frac{m(\mathbf{A}^{\mathrm{H}}\mathbf{A})^*}{n\sigma_w^2}\end{array}\right] + \frac{n\sigma_e^4}{\sigma_w^4}\left[\begin{array}{c}\boldsymbol{\theta}_c\\ \boldsymbol{\theta}_c^*\end{array}\right]\left[\begin{array}{c}\boldsymbol{\theta}_c\\ \boldsymbol{\theta}_c^*\end{array}\right]^{\mathrm{H}}\right]^{-1}\\ &= \frac{n}{m}\mathbf{J}^{-1} - c_l\left[\begin{array}{c}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\boldsymbol{\theta}_c\\ \left((\mathbf{A}^{\mathrm{H}}\mathbf{A})^*\right)^{-1}\boldsymbol{\theta}_c^*\end{array}\right]\left[\begin{array}{c}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\boldsymbol{\theta}_c\\ \left((\mathbf{A}^{\mathrm{H}}\mathbf{A})^*\right)^{-1}\boldsymbol{\theta}_c^*\end{array}\right]^{\mathrm{H}} \triangleq \widetilde{\mathbf{J}}_l^{-1},\end{aligned} \tag{20}$$

where c_l is

$$c_l = \frac{(n-m)n^2\sigma_e^4\sigma_w^4}{md_0\left[m\sigma_w^2 + n(d_0-\sigma_w^2)\right]} \geq 0. \tag{21}$$

In the case of $m=n$, both the lower and upper bounds are tight and equal to $\mathbf{J}^{-1}$. This makes sense because the observations are not compressed in this setting.

We introduce the concept of asymptotic relative efficiency (ARE) [3]. In our setting, the ARE is defined as the ratio of the trace of the CRBs obtained by uncompressed observations to that of compressed observations, which is

$$\gamma \triangleq \mathrm{tr}(\mathbf{J}^{-1})/\mathrm{tr}[\mathrm{E}_{\mathbf{\Phi}}(\widetilde{\mathbf{J}}^{-1})]. \tag{22}$$

Meanwhile, we introduce the upper and lower bounds of γ, which are $\gamma_l \triangleq \mathrm{tr}(\mathbf{J}^{-1})/\mathrm{tr}(\widetilde{\mathbf{J}}_u^{-1})$ and $\gamma_u \triangleq \mathrm{tr}(\mathbf{J}^{-1})/\mathrm{tr}(\widetilde{\mathbf{J}}_l^{-1})$, where $\widetilde{\mathbf{J}}_u^{-1}$ and $\widetilde{\mathbf{J}}_l^{-1}$ are given by (18) and (20). Obviously $0\leq\gamma\leq 1$. Now we analyze two extreme cases corresponding to $\sigma_v^2 \gg \sigma_e^2$ and $\sigma_v^2 \ll \sigma_e^2$.

Consider the situation where $\sigma_v^2 \gg \sigma_e^2$ such that any term containing σ_e^2 is ignored. As a result, c_u (19), c_l (21) is proportional to σ_e^4. Because σ_e^2 is small compared to σ_v^2, $\widetilde{\mathbf{J}}_u^{-1}$ and $\widetilde{\mathbf{J}}_l^{-1}$ are approximated by $\left((n-p)/(m-p)\right)\mathbf{J}^{-1}$ and $(n/m)\,\mathbf{J}^{-1}$, respectively, and $\gamma_u \approx m/n$, $\gamma_l \approx (m-p)/(n-p)$.

For the other case, i.e., the variance of additive noise is much smaller than the strength of perturbation, i.e., $\sigma_v^2 \ll \sigma_e^2$, it can be shown that c_u and c_l are almost independent of σ_e^2, while $\mathrm{tr}(\mathbf{J}^{-1})$ is approximately proportional to σ_e^2. As a consequence, the terms involving c_u and c_l can be ignored, and $\gamma_u \approx m/n$, $\gamma_l \approx (m-p)/(n-p)$.

To summarize, in the cases of $\sigma_v^2 \gg \sigma_e^2$ and $\sigma_v^2 \ll \sigma_e^2$, the ARE γ can be simply bounded by m/n and $(m-p)/(n-p)$.

IV. Numerical Results

In this section, numerical results are conducted to evaluate the effect of random compression on signal estimation. Let $\mathbf{a}_i$ denote the ith column of $\mathbf{A}$ and choose $\mathbf{a}_i = [1, e^{\mathrm{j}\omega_i}, e^{\mathrm{j}2\omega_i}, \cdots, e^{\mathrm{j}(n-1)\omega_i}]^{\mathrm{T}}$. Let $\boldsymbol{\theta} = [A_1e^{\mathrm{j}\phi_1}, A_2e^{\mathrm{j}\phi_2}, \cdots, A_pe^{\mathrm{j}\phi_p}]^{\mathrm{T}}$ denote the unknown complex parameter vector. The random compression matrix $\mathbf{\Phi}\in\mathbb{C}^{m\times n}$ has elements that are i.i.d. standard complex normal random variables. The number of Monte Carlo trials is 5000.

In Fig. 1, we plot the concentration ellipses for both uncompressed and realizations of random compressed observations. Because $\mathbf{J}^{-1}$ provides a lower bound on the error covariance matrix $\mathbf{\Delta} = \mathrm{E}[\underline{\mathbf{e}}\ \underline{\mathbf{e}}^{\mathrm{H}}]$ for unbiased errors $\underline{\mathbf{e}} = \hat{\underline{\boldsymbol{\theta}}} - \underline{\boldsymbol{\theta}}$, the concentration ellipse satisfies $\underline{\mathbf{e}}^{\mathrm{H}}\mathbf{\Delta}^{-1}\underline{\mathbf{e}} \leq \underline{\mathbf{e}}^{\mathrm{H}}\mathbf{J}\underline{\mathbf{e}}$ for all $\underline{\mathbf{e}}\in\mathbb{C}^{2p}$. In these numerical results, we set $n=20$, $p=1$, $m=10$, $\omega_1 = 6\pi/n$, $C = 1$, $(\sigma_v^2,\sigma_e^2) = (0.1, 0.1)$. The ellipses $\underline{\mathbf{e}}^{\mathrm{H}}\tilde{\mathbf{J}}\underline{\mathbf{e}} = J_{11}$ and $\underline{\mathbf{e}}^{\mathrm{H}}\mathbf{J}\underline{\mathbf{e}} = J_{11}$ are plotted with respect to the real part e_r and imaginary part e_i, as shown in Fig. 1. Here $\mathbf{J}$ and $\tilde{\mathbf{J}}$ are given by (8) and (11), respectively. It can be seen that compression inflates the ellipse, i.e., all the realizations of ellipses after compression completely contain the ellipse before

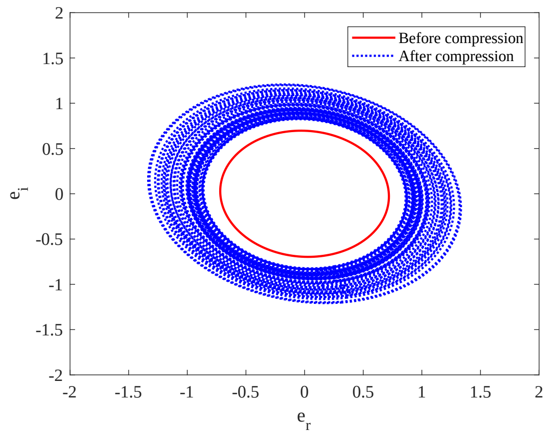

Fig. 1. Concentration ellipses for the FIM before and after compression. The red curve and blue curves are the locus of all points $\underline{\mathbf{e}}$ satisfying $\underline{\mathbf{e}}^{\mathrm{H}}\mathbf{J}\underline{\mathbf{e}} = J_{11}$ and $\underline{\mathbf{e}}^{\mathrm{H}}\widetilde{\mathbf{J}}\underline{\mathbf{e}} = J_{11}$ for 50 realizations of the Fisher information matrix after compression, respectively. The true value of $\boldsymbol{\theta}$ is $\boldsymbol{\theta} = 0.5 + 1\mathrm{j}$.

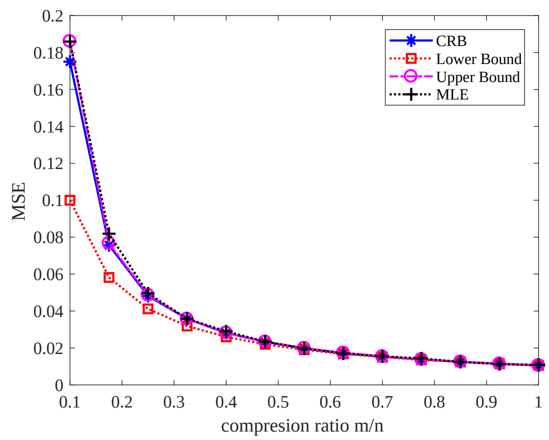

Fig. 2. The upper and lower bounds of $\mathrm{tr}[\mathrm{E}_{\mathbf{\Phi}}(\widetilde{\mathbf{J}}^{-1})]$ versus the compression ratio m/n. The true value of $\boldsymbol{\theta}$ is $\boldsymbol{\theta} = [-0.56-0.40\mathrm{j}; 0.34-0.00\mathrm{j}; -0.72-0.47\mathrm{j}; 0.89-0.76\mathrm{j}]^{\mathrm{T}}$.

compression, and the orientation of the concentration ellipse is nearly aligned with that of uncompressed ellipse. This shows that compression results in loss of Fisher information.

Next, we provide the comparison between the empirical performances of maximum likelihood estimator (MLE) and the analytical lower and upper bounds of $\mathrm{E}_{\mathbf{\Phi}}[\widetilde{\mathbf{J}}^{-1}]$. In this numerical result, we set $n = 80$, $p = 4$, $r = 2$, $(\sigma_v^2, \sigma_e^2) = (0.1, 0.1)$ and $\mathbf{C} = [\mathbf{0}_{2\times 2}, \mathbf{I}_2]$, m is an integer from 8 to 80, $\omega_1 = 6\pi/n$, $\omega_2 = 10\pi/n$, $\omega_3 = 14\pi/n$, $\omega_4 = 18\pi/n$. For the nonconvex ML estimation problem, numerical algorithms are not guaranteed to converge to the global optimum [17]. As a consequence, we use the least squares solution as an initial point for the gradient algorithm. Owing to $\boldsymbol{\theta}$ being a complex vector, we calculate the gradient of the objective function with respect to $\boldsymbol{\theta}^*$ [21], and use backtracking line search methods to choose the step size [20]. The results are shown in Fig. 2. It can be seen that as the compression rate gets closer to 1, the traces of both bounds approach to the theoretical value of $\mathrm{tr}[\mathrm{E}_{\mathbf{\Phi}}(\widetilde{\mathbf{J}}^{-1})]$, which is consistent with the conclusion that both bounds are equal to $\mathbf{J}^{-1}$ in the case of $m = n$. In addition, the MSE performance of ML estimator approaches to CRB as the compression ratio m/n increases.

Furthermore, we provide the comparison between the ARE and the MSE performance of MLE. The results are shown in Fig. 3. The parameters are set as follows: $n = 64$, $m = 32$, $p = 8$, $\sigma_v^2 = 10^{-3}$, $\mathbf{C} = [\mathbf{0}_{2\times 2}, \mathbf{I}_2]$, $\omega_i = 2\pi(2i-1)/n$, $i = 1, \cdots, p$. Numerical results are presented in Fig. 3. It can be seen that in the cases of $\sigma_e^2 \ll \sigma_v^2 = 10^{-3}$ and $\sigma_e^2 \gg \sigma_v^2 = 10^{-3}$, $(m-p)/(n-p)$ and m/n are very good approximations of γ_l and γ_u. Meanwhile, $(m-p)/(n-p)$ approximates γ well in the above two extreme scenarios, which is consistent with the results obtained in [5] in the absence of perturbation. For the MLE, its performance is consistent with γ given $\sigma_e^2 \leq 10^{-2}$. As σ_e^2 increases, the MSE performance of MLE deviates from γ.

V. Conclusions

In this paper, we have studied the impact of random compression on the FIM and the CRB for estimating unknown parameters in the perturbed linear model after random compression. The analytical closed-form expression to describe the relationship of FIM before and after compression is derived. The lower and upper bounds of the expected CRB after compression are the

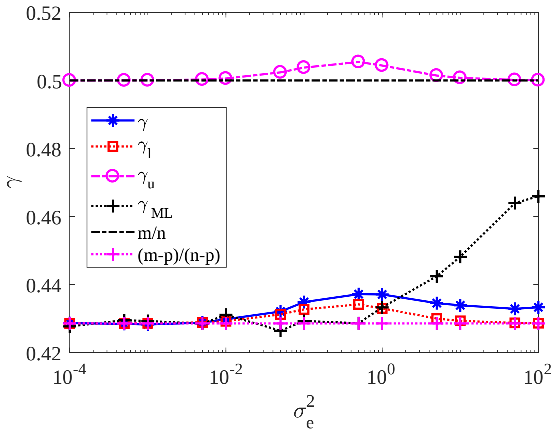

Fig. 3. The ARE γ versus the strength of perturbation σ_e^2. The true value of $\boldsymbol{\theta}$ is $\boldsymbol{\theta} = \mathbf{1} + \mathbf{0.5}$j.

scaled version of original CRB minus an additional rank one matrix dependent on the underlying parameter. When the strength of perturbation is much larger or smaller than the variance of additive noise, the relationship between CRB and compressed CRB averaged over compression matrix is almost independent of the underlying parameter.

References

[1] H. V. Poor, "An introduction to signal detection and estimation," Springer-Verlag New York, 1994.

[2] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, 1993, Englewood Cliffs, NJ: Prentice Hall.

[3] E. L. Lehmann, *Elements of Large-Sample Theory,* Springer-Verlag New York, inc., 1999.

[4] D. Ramasamy, S. Venkateswaran and U. Madhow, "Compressive parameter estimation in AWGN," *IEEE Trans. Signal Process.*, vol. 62, no. 8, pp. 2012-2027, 2014.

[5] 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," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6423-6428, 2015.

[6] J. K. Nielsen, M. G. Christensen and S. H. Jensen, "On compressed sensing and the estimation of continuous parameters from noisy observations," *ICASSP*, pp. 3609-3612, 2012.

[7] M. Shaghaghi and S. A. Vorobyov, "Cramér-Rao bound for sparse signals fitting the low-rank model with small number of parameters," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1497-1501, 2015.

[8] D. Berberidis and G. B. Giannakis, "Data sketching for large-scale Kalman filtering," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3688-3701, 2017.

[9] A. Wiesel, Y. C. Eldar and A. Yeredor, "Linear regression with Gaussian model uncertainty: Algorithms and bounds," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2194-2205, 2008.

[10] Y. Chi, L. Scharf, A. Pezeshki and A. Calderbank, "Sensitivility to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol 59, no. 4, pp. 2182-2195, 2011.

[11] M. A. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol 4, no. 2, pp. 342-349, 2010.

[12] L. Chen and Y. Gu, "Oracle-order recovery performance of greedy pursuits with replacement against general perturbations," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4625-4636, 2013.

[13] J. Ding, L. Chen and Y. Gu, "Perturbation analysis of orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 398-410, 2013.

[14] Y. Tang, L. Chen and Y. Gu, "On the performance bound of sparse estimation with sensing matrix perturbation," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4372-4386, Sep. 2013.

[15] J. Zhu, X. Wang, X. Lin and Y. Gu, "Maximum likelihood estimation from sign measurements with sensing matrix perturbation," *IEEE Trans. Signal Process.*, vol. 62, pp. 3741-3753, 2014.

[16] J. Zhu, X. Lin, R. S. Blum and Y. Gu, "Parameter estimation from quantized observations in multiplicative noise environments," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 4037-4050, 2015.

[17] A. Beck and Y. C. Eldar, "Structured total maximum likelihood: an alternative to structured total least squares," *SIAM J. Matrix Anal.*, vol. 31, no. 5, pp. 2623-2649, 2010.

[18] V. Nagesha and S. M. Kay, "Cramer-Rao lower bounds for complex parameters," available at www.ele.uri.edu/faculty/kay/New%20web/downloadable%20files/Nageha_complex%20CRLB.pdf.

[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.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[21] A. Hjrungnes, *Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications*, Cambridge University Press, 2011.