ON COMPRESSED SENSING AND THE ESTIMATION OF CONTINUOUS PARAMETERS FROM NOISY OBSERVATIONS

Jesper Kjær Nielsen[†], *Mads Græsbøll Christensen*[‡], *Søren Holdt Jensen*[†]

[†]Aalborg University
Dept. of Electronic Systems
Niels Jernes Vej 12, DK-9220 Aalborg
{jkn,shj}@es.aau.dk

[‡]Aalborg University
Dept. of Architecture, Design & Media Technology
Niels Jernes Vej 14, DK-9220 Aalborg
mgc@create.aau.dk

ABSTRACT

Compressed sensing (CS) has in recent years become a very popular way of sampling sparse signals. This sparsity is measured with respect to some known dictionary consisting of a finite number of atoms. Most models for real world signals, however, are parametrised by continuous parameters corresponding to a dictionary with an infinite number of atoms. Examples of such parameters are the temporal and spatial frequency. In this paper, we analyse how CS affects the estimation performance of any unbiased estimator when we assume such infinite dictionaries. We base our analysis on the Cramer-Rao lower bound (CRLB) which is frequently used for benchmarking the estimation accuracy of unbiased estimators. For the popular sensing matrices such as the Gaussian sensing matrix, our analysis shows that compressed sensing on average degrades the estimation accuracy by at least the down-sample factor.

Index Terms— Compressed sensing, Cramer-Rao lower bound.

1. INTRODUCTION

For a wide range of applications such as compression, enhancement, identification, and separation, sparse decompositions have been a very useful tool. Mathematically speaking, a sparse decomposition of an N-dimensional complex vector $\boldsymbol{x} \in \mathbb{C}^N$ can be written as the linear model

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{s} + \boldsymbol{e} \qquad (1)$$

where $\boldsymbol{\Psi} \in \mathbb{C}^{N \times D}$ is referred to as the basis or dictionary, $\boldsymbol{s}$ is a D-dimensional S-sparse vector, and $\boldsymbol{e} \in \mathbb{C}^N$ is an error vector modelling noise and model inaccuracies. By S-sparse, we mean that $\boldsymbol{s}$ contains exactly S non-zero coefficients and $D - S$ zeros. Moreover, we say that $\boldsymbol{x}$ is an S-sparse or a compressible signal in the basis $\boldsymbol{\Psi}$ if $\boldsymbol{e} = 0$ or $\boldsymbol{e} \approx 0$, respectively. Traditionally, the non-zero coefficients of $\boldsymbol{s}$ are found by greedy or ℓ_1 optimisation algorithms such as matching pursuit [1] or basis pursuit [2]. In the usual cases where $N \gg S$, however, these methods may suffer from a large computational overhead as they have to work directly on $\boldsymbol{x}$. In compressed sensing (CS) [3], this overhead is decreased considerably by utilising the sparsity during the data acquisition. That is, instead of acquiring $\boldsymbol{x}$ by sampling at the Nyquist rate, we acquire $\boldsymbol{y} \in \mathbb{C}^M$ with $S < M \leq N$ by only collecting an amount of data close to the sparsity level S. Thus, CS may enable a faster computation of the parameters, data acquisition at a lower sample rate, and less demanding storage requirements. These properties are very important for most signal processing algorithms, and CS has therefore become very popular. Mathematically speaking, we model the relationship between $\boldsymbol{x}$ and $\boldsymbol{y}$ by $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}$ where $\boldsymbol{\Phi} \in \mathbb{C}^{M \times N}$ is referred to as the sensing matrix. If $\boldsymbol{x}$ is compressible, and $\boldsymbol{\Phi}$ is chosen appropriately, the vector $\boldsymbol{s}$ can be computed directly from $\boldsymbol{y}$ provided that the Restricted Isometry Property (RIP) holds [3, 4]. Until recently [5], the recovery was only shown to hold for the orthogonal or incoherent dictionaries $\boldsymbol{\Psi}$. Consequently, much attention has been directed towards finding sensing matrices which makes M as small as possible given an incoherent dictionary [6, 7]. The dictionary $\boldsymbol{\Psi}$ in (1) consists of D column vectors $\{\boldsymbol{\psi}_d\}_{d=1}^D$ which are often referred to as atoms, and popular choices of the dictionary are the Fourier basis and a wavelet basis. Typically, the atoms can be represented by a parametric function $f(\boldsymbol{\phi})$, and each atom is constructed by selecting a specific value $\boldsymbol{\psi}_d = f(\boldsymbol{\phi}_d)$ for the parameter of this function. For example, the atoms of the incoherent Fourier basis is formed by sampling the frequency parameter $\phi = \omega$ on the Fourier grid $\phi_d = 2\pi d/D$ with $D = N$. For most real world signals, however, the parameter ϕ is a continuous parameter corresponding to highly coherent dictionaries with $D \to \infty$. A sparse decomposition with a finite dictionary is therefore in direct contradiction with the physics behind most signal models of the form

$$\boldsymbol{x} = \boldsymbol{A}(\boldsymbol{\phi})\boldsymbol{\alpha} + \boldsymbol{w} \qquad (2)$$

where $\boldsymbol{w} \in \mathbb{C}^N$ is a noise vector, and $\boldsymbol{A}(\boldsymbol{\phi}) \in \mathbb{C}^{N \times S}$ is parametrised by $\boldsymbol{\phi} \in \mathbb{C}^{K-S}$ and contains the S true atoms with the amplitudes $\boldsymbol{\alpha} \in \mathbb{C}^S$. Comparing the models in (2) and (1), we see that $\boldsymbol{A}(\boldsymbol{\phi})\boldsymbol{\alpha} \approx \boldsymbol{\Psi}\boldsymbol{s}$ with equality if the true atoms in $\boldsymbol{A}(\boldsymbol{\phi})$ are included in $\boldsymbol{\Psi}$. As demonstrated in [8], we obtain an inferior compression scheme by using the model in (1) rather than (2) when equality does not hold. CS has been developed under the assumption that equality holds. In other words, $\boldsymbol{\phi}$ is assumed to be a discrete parameter whose possible values are used to construct the atoms of the dictionary. When CS is viewed in this light, we may interpret the RIP as a requirement to the distance between adjacent values that $\boldsymbol{\phi}$ may take.

In this paper, we do not assume that $\boldsymbol{\phi}$ is a discrete parameter. For various popular sensing matrices [4,9], we instead investigate the accuracy with which we can estimate the continuous parameters of the model in (2) when we are giving $\boldsymbol{y}$ instead of $\boldsymbol{x}$. For the MUSIC algorithm, we noted a significant loss in estimation accuracy in [10]. Here, however, we do not consider a specific estimation algorithm, but only the best possible performance that any unbiased estimation algorithm can obtain. We therefore base the analysis on the Cramer-Rao lower bound (CRLB) which has previously [11, 12] been used to assess the estimation accuracy of the non-zero elements of $\boldsymbol{s}$, assuming a finite dictionary. In this paper, however, we work directly with the model in (2), corresponding to an infinite dictionary. The paper is organised as follows: In Sec. 2, we present the CRLB for the model in (2). The CRLB is used to benchmark the performance

of unbiased estimators, and we modify it to the situation in which CS is used in Sec. 3. In Sec. 4, we establish a connection between the CRLB with and without CS by deriving a lower bound on the expected CRLB for some of the popular sensing matrices. An illustrative simulation is presented in Sec. 5, and Sec. 6 concludes this paper.

2. CRAMER-RAO LOWER BOUND

Consider the general problem in which we observe N random data points $\boldsymbol{x}$ which we mathematically describe by a family of probability density functions $p(\boldsymbol{x}; \boldsymbol{\theta})$. Without loss of generality, we assume that this model is parametrised by the real[1] parameter vector $\boldsymbol{\theta}$ which we wish to estimate based on the data. In order to do this, we construct an unbiased estimator $\hat{\boldsymbol{\theta}}$ which maps the data into an estimate. For the covariance matrix $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}}$ of any unbiased estimator, the CRLB guarantees that $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} - \mathcal{I}^{-1}(\boldsymbol{\theta}) \geq \boldsymbol{0}$ where the inequality denotes positive semi-definitiveness. Thus, for the variance of the estimator for the k'th parameter, we have that

$$\operatorname{var}(\hat{\theta}_k) = [\boldsymbol{C}_{\hat{\boldsymbol{\theta}}}]_{kk} \geq \left[\mathcal{I}^{-1}(\boldsymbol{\theta})\right]_{kk} \tag{3}$$

where $[\cdot]_{kk}$ denotes the (k, k)'th element. The matrix $\mathcal{I}(\boldsymbol{\theta})$ is the Fisher information matrix (FIM), and it is given by [13]

$$\mathcal{I}(\boldsymbol{\theta}) = E\left\{\frac{\partial \ln p(\boldsymbol{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\boldsymbol{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T\right\} \tag{4}$$

where $(\cdot)^T$ denotes matrix transpose. It can be shown that if $\boldsymbol{x}$ has a multivariate complex normal distribution whose mean and covariance are parametrised by $\boldsymbol{\theta}$, i.e., $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{C}(\boldsymbol{\theta}))$, then the (k, l)'th element of the FIM is given by [13]

$$[\mathcal{I}(\boldsymbol{\theta})]_{kl} = 2\operatorname{Re}\left[\frac{\partial \boldsymbol{\mu}^H(\boldsymbol{\theta})}{\partial \theta_k} \boldsymbol{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_l}\right]$$
$$+ \operatorname{tr}\left[\boldsymbol{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{C}(\boldsymbol{\theta})}{\partial \theta_k} \boldsymbol{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{C}(\boldsymbol{\theta})}{\partial \theta_l}\right] . \tag{5}$$

where $(\cdot)^H$ denotes complex transposition, $\operatorname{tr}(\cdot)$ denotes matrix trace, and $\operatorname{Re}[\cdot]$ takes the real part of a complex number. For the model in (2) with $\boldsymbol{w} \sim \mathcal{CN}(\boldsymbol{0}, \sigma_w^2 \boldsymbol{I}_N)$, the $(K + 1)$-dimensional parameter vector is $\boldsymbol{\theta} \triangleq \begin{bmatrix} \boldsymbol{\phi}^T & \boldsymbol{\alpha}^T & \sigma_w^2 \end{bmatrix}^T$, and we have that

$$\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{A}(\boldsymbol{\phi})\boldsymbol{\alpha}, \sigma_w^2 \boldsymbol{I}_N) \tag{6}$$

where $\boldsymbol{I}_N$ is the N-dimensional identity matrix. Using (5) and (6), the FIM is given by

$$\mathcal{I}(\boldsymbol{\theta}) = \begin{bmatrix} 2\sigma_w^{-2}\operatorname{Re}(\boldsymbol{Q}^H \boldsymbol{Q}) & \boldsymbol{0} \\ \boldsymbol{0}^T & N\sigma_w^{-4} \end{bmatrix} \tag{7}$$

where we have defined

$$\boldsymbol{q}_k \triangleq \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_k} \quad \text{for } k = 1, 2, \ldots, K \tag{8}$$

$$\boldsymbol{Q} \triangleq \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_K \end{bmatrix} . \tag{9}$$

[1]If the model is parametrised by complex parameters $\boldsymbol{\xi} = \boldsymbol{\xi}_r + j\boldsymbol{\xi}_i$, say, then $\boldsymbol{\theta}$ is defined as $\boldsymbol{\theta} \triangleq \begin{bmatrix} \boldsymbol{\xi}_r^T & \boldsymbol{\xi}_i^T \end{bmatrix}^T$.

3. THE EXPECTED PROJECTION MATRIX

As mentioned in the introduction, we observe $\boldsymbol{y}$ instead of $\boldsymbol{x}$ in compressed sensing. Incorporating this into (2) yields

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} = \boldsymbol{\Phi}\boldsymbol{A}(\boldsymbol{\phi})\boldsymbol{\alpha} + \boldsymbol{\Phi}\boldsymbol{w} \tag{10}$$

from which we see that

$$\boldsymbol{y}|\boldsymbol{\Phi} \sim \mathcal{CN}(\boldsymbol{\Phi}\boldsymbol{A}(\boldsymbol{\phi})\boldsymbol{\alpha}, \sigma_w^2 \boldsymbol{\Phi}\boldsymbol{\Phi}^H) . \tag{11}$$

In the following sections, we investigate how the sensing matrix affects the CRLB derived in Sec. 2.

3.1. Fisher Information Matrix in Compressed Sensing

For $k, l \leq K$, we obtain from (5) and (11) that

$$[\mathcal{I}_{\mathrm{CS}}(\boldsymbol{\theta})]_{kl} = 2\sigma_w^{-2}\operatorname{Re}\left[\boldsymbol{q}_k^H \boldsymbol{\Phi}^H (\boldsymbol{\Phi}\boldsymbol{\Phi}^H)^{-1} \boldsymbol{\Phi}\boldsymbol{q}_l\right]$$
$$= 2\sigma_w^{-2}\operatorname{Re}\left[\boldsymbol{q}_k^H \boldsymbol{\Pi}\boldsymbol{q}_l\right] \tag{12}$$

where we have defined $\boldsymbol{\Pi} \triangleq \boldsymbol{\Phi}^H (\boldsymbol{\Phi}\boldsymbol{\Phi}^H)^{-1} \boldsymbol{\Phi}$ which is an $N \times N$ orthogonal projection matrix of rank M. For $k = l = K + 1$, we can rewrite (5) as

$$[\mathcal{I}_{\mathrm{CS}}(\boldsymbol{\theta})]_{kl} = \operatorname{tr}\left\{\sigma_w^{-2}(\boldsymbol{\Phi}\boldsymbol{\Phi}^H)^{-1}\boldsymbol{\Phi}\boldsymbol{\Phi}^H \sigma_w^{-2}(\boldsymbol{\Phi}\boldsymbol{\Phi}^H)^{-1}\boldsymbol{\Phi}\boldsymbol{\Phi}^H\right\}$$
$$= \operatorname{tr}\left\{\sigma_w^{-4}\boldsymbol{I}_M\right\} = M\sigma_w^{-4} . \tag{13}$$

Thus, the FIM is given by

$$\mathcal{I}_{\mathrm{CS}}(\boldsymbol{\theta}) = \begin{bmatrix} 2\sigma_w^{-2}\operatorname{Re}(\boldsymbol{Q}^H \boldsymbol{\Pi}\boldsymbol{Q}) & \boldsymbol{0} \\ \boldsymbol{0}^T & M\sigma_w^{-4} \end{bmatrix} . \tag{14}$$

Compared against the FIM in (7), we see that $\mathcal{I}_{\mathrm{CS}}(\boldsymbol{\theta})$ differs in terms of the scaling factor of the $(K + 1, K + 1)$'th element and the inclusion of the projection matrix $\boldsymbol{\Pi}$ inside the inner matrix product $\boldsymbol{Q}^H \boldsymbol{Q}$. The interpretation of the latter is straightforward; we project the columns of $\boldsymbol{Q}$ onto the subspace spanned by $\boldsymbol{\Phi}^H$. Therefore, the diagonal elements of $\mathcal{I}_{\mathrm{CS}}(\boldsymbol{\theta})$ decrease compared against the corresponding elements of $\mathcal{I}(\boldsymbol{\theta})$ unless $\boldsymbol{Q}$ is spanned by $\boldsymbol{\Phi}^H$.

3.2. Typical Sensing Matrices

As alluded to in the introduction, the choice of sensing matrix $\boldsymbol{\Phi}$ is vital in CS; we wish to find a sensing matrix that obeys the RIP for as few measurements M as possible. Perhaps surprisingly, stochastic sensing matrices have been shown to be nearly optimal for almost any choice of basis $\boldsymbol{\Psi}$ [14]. Some of the most popular choices are listed below [4].

1. Select the entries of $\boldsymbol{\Phi}$ as i.i.d. samples from a normal distribution with variance $1/M$.

2. Sample N N-dimensional i.i.d. vectors from a normal distribution with unit variance. Find an orthonormal basis of these N random vectors and select the rows of $\boldsymbol{\Phi}$ as M random rows from this orthonormal basis.

3. Sample N M-dimensional i.i.d. vectors uniformly at random from the unit sphere.

4. Select the entries of $\boldsymbol{\Phi}$ as i.i.d. samples from a symmetric Bernoulli distribution with outcomes $\pm 1/M$.

Once the sensing matrix has been selected, the FIM is easy to calculate. Since the sensing matrix is often selected at random, however, it is not particularly useful to say something about a specific realisation of the sensing matrix. Therefore, the next logical step in our analysis is to investigate the statistics of the inverse FIM when the sensing matrix is selected at random from some matrix variate distribution. Unfortunately, this is in general a very hard problem, and we therefore consider the simpler task of investigating the expected FIM for the various sensing matrices. We use this to derive a lower bound on the expected inverse FIM in Sec. 4.

Since the unknown parameters $\boldsymbol{\theta}$ are assumed to be deterministic variables, it readily follows from (14) that the expected FIM is given by

$$E\left\{\mathcal{I}_{\mathrm{CS}}(\boldsymbol{\theta})\right\} = \begin{bmatrix} 2\sigma_w^{-2}\mathrm{Re}[\boldsymbol{Q}^H E\{\boldsymbol{\Pi}\}\boldsymbol{Q}] & \mathbf{0} \\ \mathbf{0}^T & M\sigma_w^{-4} \end{bmatrix} . \quad (15)$$

Thus, in order to find the expected FIM, we have to find the expected projection matrix $E\{\boldsymbol{\Pi}\} = E\{\boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{-1}\boldsymbol{\Phi}\}$. For this purpose, we use the following theorem.

Theorem 1. *Let $\boldsymbol{\Phi}$ be a random $M \times N$ matrix with $M < N$, rank M almost everywhere, and the probability density function (pdf) $f_{\boldsymbol{\Phi}}(\boldsymbol{\Phi})$. Furthermore, let $\boldsymbol{\Pi} = \boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{-1}\boldsymbol{\Phi}$ be the $N \times N$ orthogonal projection matrix of rank M onto the subspace spanned by the rows of $\boldsymbol{\Phi}$ and denote the space of points corresponding to all such projection matrices by $P_{M,N-M}$. If $f_{\boldsymbol{\Phi}}(\boldsymbol{\Phi})$ is invariant under the right-orthogonal transformation $\boldsymbol{\Phi} \to \boldsymbol{\Phi}\boldsymbol{R}$ for any $N \times N$ orthogonal matrix $\boldsymbol{R}$, then the PDF of $\boldsymbol{\Pi}$ is uniform on $P_{M,N-M}$ and $E\{\boldsymbol{\Pi}\} = (M/N)\boldsymbol{I}_N$.*

Sketch of the Proof. It follows from [15, Th. 2.4.3] that $\boldsymbol{H} = (\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{-1/2}\boldsymbol{\Phi}$ is uniformly distributed on the Stiefel manifold $V_{M,N}$ if $f_{\boldsymbol{\Phi}}(\boldsymbol{\Phi})$ is invariant under any right-orthogonal transformation. It then follows from [15, Th. 2.2.1] that $\boldsymbol{Q}\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{Q}^T$ is uniformly distributed on $P_{M,N-M}$ for any orthogonal $\boldsymbol{Q}$ independent of $\boldsymbol{H}^T\boldsymbol{H}$. Now, since $\boldsymbol{\Pi} = \boldsymbol{H}^T\boldsymbol{H}$, it therefore follows that $\boldsymbol{\Pi}$ is uniformly distributed on $P_{M,N-M}$. Moreover, if follows that

$$E\{\boldsymbol{\Pi}\} = \boldsymbol{Q}E\{\boldsymbol{\Pi}\}\boldsymbol{Q}^T \implies E\{\boldsymbol{\Pi}\} \propto \boldsymbol{I}_N \quad (16)$$

$$\mathrm{tr}(E\{\boldsymbol{\Pi}\}) = E\{\mathrm{tr}(\boldsymbol{\Pi})\} = M \quad (17)$$

Eq. (17) follows from the fact that a projection matrix of rank M has exactly M ones and $N - M$ zeros as its eigenvalues. Combining (16) and (17) readily gives $E\{\boldsymbol{\Pi}\} = (M/N)\boldsymbol{I}_N$. $\square$

By use of Theorem 1, we can show that the sensing matrices of type 1 and 2 from above result in an expected projection matrix of $E\{\boldsymbol{\Pi}\} = (M/N)\boldsymbol{I}_N$. As shown in Sec. 5, empirical evidence also suggests that this is the case for the sensing matrices of type 3 and 4. Interestingly, for the simplest possible sensing matrix, the Kronecker sensing matrix which is the identity matrix with $N - M$ random rows removed, we obtain the same expected projection matrix. To see this, consider that the projection matrix corresponding to such a Kronecker sensing matrix $\boldsymbol{K}$ is given by $\boldsymbol{\Pi} = \boldsymbol{K}^T\boldsymbol{K}$ which is a diagonal matrix with M ones and $N-M$ zeros uniformly distributed on the diagonal. Thus, there are

$$_NC_M = \binom{N}{M} = \frac{N!}{M!(N-M)!} \quad (18)$$

distinct projection matrices each with probability $1/_NC_M$. The expected value of the projection matrix is therefore

$$E\{\boldsymbol{\Pi}\} = \frac{1}{_NC_M}\sum_{i=1}^{_NC_M} \boldsymbol{\Pi}_i = \frac{_{N-1}C_{M-1}}{_NC_M}\boldsymbol{I}_N = \frac{M}{N}\boldsymbol{I}_N \quad (19)$$

where the second equality follows from the fact that there are $_{N-1}C_{M-1}$ projection matrices $\boldsymbol{\Pi}$ with a one on the k'th diagonal element.

4. A BOUND ON THE EXPECTED CRLB

Inserting the expected projection matrix of $E\{\boldsymbol{\Pi}\} = (M/N)\boldsymbol{I}_N$ into (15) yields

$$E\left\{\mathcal{I}_{\mathrm{CS}}(\boldsymbol{\theta})\right\} = \frac{M}{N}\mathcal{I}(\boldsymbol{\theta}) . \quad (20)$$

Thus, on average, the elements of the FIM with CS is M/N times the elements of the FIM without CS. The expected inverse FIM can be bounded as

$$E\left\{\mathcal{I}_{\mathrm{CS}}^{-1}(\boldsymbol{\theta})\right\} \geq (E\left\{\mathcal{I}_{\mathrm{CS}}(\boldsymbol{\theta})\right\})^{-1} = \frac{N}{M}\mathcal{I}^{-1}(\boldsymbol{\theta}) . \quad (21)$$

Since $M < N$ in CS, the expected CRLB with CS increases as compared to the case without CS. Interestingly, the popular Gaussian and nearly orthogonal sensing matrices do not on average perform better than the Kronecker sensing matrix from an estimation theoretic point of view. Thus, on average, the Gaussian and the nearly orthogonal sensing matrices of the form outlined above decrease the estimation accuracy by an amount equal to the case where random samples are simply thrown away. Furthermore, the expected estimation accuracy decreases inversely proportional to at least the number of samples that we retain in the data acquisition step of CS. This was also demonstrated in [10] in which the direction-of-arrival estimation accuracy for a varying M was compared to the CRLB. Can we do any better than this? That is, can we select a sensing matrix such that the elements of the inverse FIM with CS are closer to the elements of the CRLB without CS? To answer this question, we first take a closer look at the orthogonal projection matrix $\boldsymbol{\Pi}$. From (17), we have that the trace of an expected projection matrix must be equal to M. All projection matrices must fulfil this constraint so if we wish to construct a diagonal expected projection matrix with equal elements, we therefore have that $E\{\boldsymbol{\Pi}\} = (M/N)\boldsymbol{I}_N$. This result is the same as the expected projection matrix for the popular sensing matrices presented above. However, as we saw in Sec. 3.1, it is possible to design a sensing matrix such that the CRLB is unaffected provided that the columns of $\boldsymbol{Q}$ are spanned by the rows of $\boldsymbol{\Phi}$. Unfortunately, since the design of such a sensing matrix requires that we know the parameters we wish to estimate, it is infeasible, unless we have a strong prior knowledge about the values of the missing parameters. In this case, however, it may be better to employ Bayesian inference methods which offer a unified way of incorporating prior knowledge.

5. SIMULATIONS

We demonstrate the validity of our analysis on a simple but well-known example. In the example, we consider a complex sinusoid in complex white Gaussian noise, i.e.,

$$x_n = \alpha e^{(j\omega n + j\varphi)} + w_n , \quad \text{for } n = 0, \dots, N-1 \quad (22)$$

where $\alpha > 0$, $\varphi \in [-\pi, \pi]$ and $\omega \in [-\pi, \pi]$ are the amplitude, phase and (angular) frequency, respectively. The noise variance is σ_w^2. The CRLB for this signal is well-known and given by [13]

$$\mathrm{var}(\hat{\alpha}) \geq \frac{\sigma_w^2}{2N} \qquad\qquad \mathrm{var}(\hat{\varphi}) \geq \frac{\sigma_w^2(2N-1)}{\alpha^2 N(N+1)}$$

$$\mathrm{var}(\hat{\omega}) \geq \frac{6\sigma_w^2}{\alpha^2 N(N^2-1)} \qquad \mathrm{var}(\hat{\sigma}_w^2) \geq \frac{\sigma_w^4}{N}$$
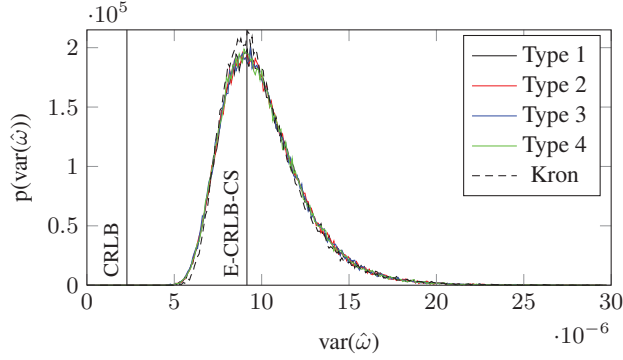
Fig. 1. Estimated pdf of the CRLB for the frequency parameter based on 100,000 Monte Carlo runs. In each run, sensing matrices of type 1-4 and the Kronecker matrix all of size 16×64 were generated.

	E-CRLB-CS	Type 1	Type 2	Type 3	Type 4	Kron
$E\{var(\hat{\alpha})\}$	3.125	3.363	3.366	3.365	3.371	3.125
$E\{var(\hat{\phi})\}$	12.21	13.42	13.41	13.40	13.41	13.31
$E\{var(\hat{\omega})\}$	0.0092	0.0101	0.0101	0.0101	0.0101	0.0101

Table 1. The mean value for the CRLB based on 100,000 Monte Carlo runs. In each run, sensing matrices of type 1-4 and the Kronecker matrix all of size 16×64 were generated. All values are scaled by a factor of 1000.

and the k'th row of $\boldsymbol{Q}$ is

$$[\boldsymbol{Q}]_{k:} = \begin{bmatrix} e^{(j\omega n + j\varphi)} & j\alpha e^{(j\omega n + j\varphi)} & j\alpha n e^{(j\omega n + j\varphi)} \end{bmatrix} . \quad (23)$$

For each of the four types of sensing matrices and for the Kronecker matrix, we ran 100,000 Monte Carlo runs in which we calculated the inverse FIM given by the inverse of (14). The size[2] of the sensing matrices was 16×64. For the diagonal elements of the inverse FIM corresponding to the amplitude, phase and frequency, we calculated their 500 bins normalised histograms and mean values. Fig. 1 shows the normalised histograms of the CRLB for the frequency parameter. Clearly, the histograms are almost coinciding with the exception of the histogram corresponding to the Kronecker sensing matrix. Fig. 1 also shows the CRLB without CS as well as the lower bound for the expected CRLB with CS. Table 1 lists the mean values of the CRLB with CM corresponding to the amplitude, phase and frequency. The lower bound for the expected CRLB is also listed. Again, we see the same pattern; the mean values for the type 1-4 sensing matrices were more or less the same while the mean value for the Kronecker sensing matrix was slightly different. All values were on or above the lower bound for the expected CRLB.

6. CONCLUSION

In this paper, we have analysed compressed sensing (CS) from an estimation theoretic point of view by use of the Cramer-Rao lower bound (CRLB). Not surprisingly, our analysis have shown that CS on average degrades our ability to estimate continuous parameters. For

some of the popular sensing matrices such as the Gaussian sensing matrix, we quantified the expected degradation by showing that the ratio between the expected CRLB with CS and the CRLB without CS is lower bounded by the ratio between the number of columns N and the number of rows M of the sensing matrix. Perhaps more surprisingly, we also showed that the bound is the same for the Kronecker sensing matrix. That is, from an estimation theoretic point of view some of the popular sensing matrices degrade on average our estimation accuracy by an amount equal to the situation in which we throw $N - M$ random samples away.

7. REFERENCES

[1] Stephane Mallat and Zhifeng Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[2] S. Chen and D. Donoho, "Basis pursuit," in *Rec. Asilomar Conf. Signals, Systems, and Computers*, 1994.

[3] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[4] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[5] E. J. Candès, Y. Eldar, and D. Needell, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 59–73, 2010.

[6] A. Hormati, A. Karbasi, S. Mohajer, and M. Vetterli, "An estimation theoretic approach for sparsity pattern recovery in the noisy setting," Unpublished manuscript, Nov. 2009.

[7] V. Abolghasemi, S. Ferdowsi, B. Makkiabadi, and S. Sanei, "On optimization of the measurement matrix for compressive sensing," in *Proc. European Signal Processing Conf.*, Aug. 2010, pp. 427–431.

[8] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," Unpulished manuscript, 2011.

[9] R. G. Baraniuk, "Compressive sensing," *Lecture Notes in IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–120, Jul. 2007.

[10] M. G. Christensen and J. K. Nielsen, "Joint direction-of-arrival and order estimation in compressed sensing using angles between subspaces," in *Proc. IEEE Workshop on Stat. Signal Process.*, Jun. 2011, pp. 449–452.

[11] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramer-Rao bound for noisy compressive sampling," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1233–1236, Mar. 2009.

[12] Z. Ben-Haim and Y. Eldar, "The Cramer-Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3384–3389, Jun. 2010.

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Prentice Hall PTR, Mar. 1993.

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

[15] Y. Chikuse, *Statistics on Special Manifolds*, Springer, 1 edition, Feb. 2003.

---

[2]As a rule of thumb, the value of M should be approximately four times the number of unknown parameters [4].