Sparsity-Based Estimation Bounds With Corrupted Measurements

Remy Boyer, Pascal Larzabal

Sparsity-Based Estimation Bounds With Corrupted Measurements

Rémy Boyer[a,*], Pascal Larzabal[b]

[a] *University of Paris-Sud, Laboratoire des Signaux et Systèmes (L2S), France*
[b] *ENS-Cachan, SATIE, France*

Abstract

In typical Compressed Sensing operational contexts, the measurement vector $\mathbf{y}$ is often partially corrupted. The estimation of a sparse vector acting on the entire support set exhibits very poor estimation performance. It is crucial to estimate set $\mathcal{I}_{uc}$ containing the indexes of the uncorrupted measures. As $\mathcal{I}_{uc}$ and its cardinality $|\mathcal{I}_{uc}| < N$ are unknown, each sample of vector $\mathbf{y}$ follows an i.i.d. Bernoulli prior of probability P_{uc}, leading to a Binomial-distributed cardinality. In this context, we derive and analyze the performance lower bound on the Bayesian Mean Square Error (BMSE) on a $|\mathcal{S}|$-sparse vector where each random entry is the product of a continuous variable and a Bernoulli variable of probability P and $|\mathcal{S}|\big||\mathcal{I}_{uc}|$ follows a hierarchical Binomial distribution on set $\{1, \ldots, |\mathcal{I}_{uc}| - 1\}$. The derived lower bounds do not belong to the family of "oracle" or "genie-aided" bounds since our *a priori* knowledge on support $\mathcal{I}_{uc}$ and its cardinality is limited to probability P_{uc}. In this context, very compact and simple expressions of the Expected Cramer-Rao Bound (ECRB) are proposed. Finally, the proposed lower bounds are compared to standard estimation strategies robust to an impulsive (sparse) noise.

Keywords: Compressed sensing, corrupted measurements, Cramér-Rao

Bound, statistical priors for support sets of random cardinalities, Gaussian measurement matrix

---

1. Introduction

In the Compressed Sensing (CS) framework [1, 2, 3], it is assumed that the signal of interest can be linearly decomposed into few basis vectors. By exploiting this property, CS allows for using sampling rates lower [4] than the Shannon's sampling rate [5]. As a result, CS methods have found a plethora of applications in numerous areas, e.g. array processing [6, 7], wireless communications [8, 9], video processing [10] or in MIMO radar [11, 12, 13].

A fundamental problem is to derive the estimation performance of sparse signal [14]. To reach this goal, the lower bounds on the mean-square error (MSE) are useful as a benchmark against any estimators can be compared [15, 16]. They have been investigated for deterministic sparse vector estimation in [17, 18, 19, 20] and for the Bayesian linear model in [21, 22, 23, 24].

In realistic contexts, the estimation accuracy in terms of the Bayesian MSE (BMSE), of standard sparse-based estimator collapses in presence of a corrupted measurements [25, 26, 27, 28, 29]. In this work, our aim is to study the estimation performance limit in presence of corrupted measurements. CS with corrupted measurements [30, 31, 32] plays a central role in numerous applications, such as the restoration of signals from impulse noise, strong narrowband interference, bursts of high noise (*e.g.*, hardware power-supply spikes), measurements dropped during transmission, malfunctioning sensors in network, *etc.* In practice, the indexes, *i.e.*, the support $\mathcal{I}_{\mathrm{uc}}$, constituted by the uncorrupted measurements and its cardinality, denoted by $|\mathcal{I}_{\mathrm{uc}}|$, are unknown. So, to take into account this uncertain knowledge, the support $\mathcal{I}_{\mathrm{uc}}$ is modelized as a collection of i.i.d. Bernoulli-distributed random variables with a probability $1 - P_{\mathrm{uc}}$ to be corrupted. Thus, in our framework, the proposed lower bounds do not belong to the family of "oracle" or "genie-aided" bounds since only the knowledge of probability P_{uc} is assumed to be *a priori* known. As a consequence, the unknown

cardinality $|\mathcal{I}_{\mathrm{uc}}|$ follows a Binomial prior in set $\{1, \ldots, N-1\}$. So, our goal is to derive a lower bound on the BMSE for the estimation of a $|\mathcal{S}|$-sparse amplitude vector for (i) a Gaussian measurement matrix and (ii) for random support, $\mathcal{S}$, and cardinality, assuming that each entry of the vector of interest is modeled as the product of a continuous random variable and a Bernoulli-distributed random variable indicating that the current entry is non-zero with probability P. To ensure the model identifiability contraint, we must have $|\mathcal{S}| < |\mathcal{I}_{\mathrm{uc}}|$, meaning that $|\mathcal{S}| \big| |\mathcal{I}_{\mathrm{uc}}|$ follows a hierarchical Binomial distribution confined in the set $\{1, \ldots, |\mathcal{I}_{\mathrm{uc}}| - 1\}$. This work proposes several new contributions regarding the state of art on the lower bounds for the estimation of sparse signal. Contrary to [17, 18, 20], the proposed lower bounds do not assume the knowledge of the support and its cardinality. Regarding the references [21, 22, 23], the proposed lower bounds remain true for any continuous prior on the non-zero entries of interest. Our framework differs from [24] since the derived results are obtained in the non-asymptotic scenario. We can note that to the best of our knowledge the derivation of an Bayesian lower bound with corrupted measurements has not been proposed in the literature. Finally, the proposed lower bounds are illustrated in the context of the standard estimation strategies robust to an impulsive (sparse) noise.

This work is composed by two main parts. The first one presents the Expected Cramer-Rao Bound (ECRB) based on a complete measurement vector scenario meaning that $P_{\mathrm{uc}} = 1$. This section has been partially presented during the IEEE SSP'16 conference [33]. The second part presents the major contribution of this work, Specifically, the more challenging corrupted measurement vector scenario is tackled.

<u>*Notations*</u>: The symbols $(\cdot)^T$, $(\cdot)^\dagger$, $\mathrm{Tr}(\cdot)$ and $(\cdot)!$ denote the transpose, the pseudo-inverse, the trace operator and the factorial, respectively. Furthermore, $\mathcal{N}(\mu, \sigma^2)$ stands for the real Gaussian probability density function (pdf) with mean μ and variance σ^2. $\mathrm{Bernou}(P)$ stands for the Bernoulli distribution of probability of success P. $\mathrm{Binomial}(N, P)$ stands for the Binomial distribu-

tion in $\{0, \ldots, N\}$ with a success probability P [34]. The binomial coefficient is $\binom{a}{b} = \frac{a!}{b!(a-b)!}$. $|\cdot|$ is the cardinality of the set given as an argument. $1_{\mathcal{X}}(x)$ is the indicator function with respect to the set $\mathcal{X}$, *i.e.*, $1_{\mathcal{X}}(x) = 1$ if $x \in \mathcal{X}$ and 0 otherwise. $O(\cdot)$ is the Big-O notation [35]. $\mathbb{E}_X$ (*resp.* $\mathbb{E}_{X|Y}$) denotes the mathematical (*resp.* conditional) expectation. log is the logarithm function and ∂ is the partial derivative symbol. A function in C^1 is continuously differentiable. $p(\cdot)$ denotes a probability density function (pdf) and $\Pr(\cdot)$ denotes the probability mass function (pmf).

2. CS model and recovery requirements

Let $\mathbf{y}$ be a $N \times 1$ noisy measurement vector in the (real) Compressed Sensing (CS) model [1, 2, 3]:

$$\mathbf{y} = \boldsymbol{\Psi}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{n}$ is a (zero-mean) white Gaussian noise vector with component variance σ^2 and $\boldsymbol{\Psi}$ is the $N \times K$ sensing/measurement matrix with $N < K$. The vector $\mathbf{s}$ is given by $\mathbf{s} = \boldsymbol{\Phi}\boldsymbol{\theta}$, where $\boldsymbol{\Phi}$ is a $K \times K$ orthonormal matrix and $\boldsymbol{\theta}$ is a $K \times 1$ amplitude vector. With this definition eq. (1) can be recast as

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{n} \tag{2}$$

where the overcomplete $N \times K$ matrix $\mathbf{H} = \boldsymbol{\Psi}\boldsymbol{\Phi}$ is commonly referred to as the dictionary. The amplitude vector θ_k are assumed to be random with an unspecified pdf. Let $\boldsymbol{P}$ be a $K \times K$ diagonal matrix composed by K random binary entries. This matrix modelizes the mechanism to randomly "sparsify" the dense random vector $\boldsymbol{\theta}'$ on the support set $\mathcal{S}$. This set is composed by the collection of indices of the non-zero θ_k. The cardinality of the support is denoted by $|\mathcal{S}|$. So, the $K \times 1$ vector $\boldsymbol{\theta} = \boldsymbol{P}\boldsymbol{\theta}'$ is $|\mathcal{S}|$-sparse, with $|\mathcal{S}| < N < K$. Under this assumption and using the property $\boldsymbol{P}^2 = \boldsymbol{P}$, we can rewrite the first summand in eq. (2) as

$$\mathbf{H}\boldsymbol{\theta} = \mathbf{H}\boldsymbol{P}\boldsymbol{\theta}' = \mathbf{H}\boldsymbol{P}^2\boldsymbol{\theta}' = [\mathbf{H}\boldsymbol{P}]_{\mathcal{S}}[\boldsymbol{P}\boldsymbol{\theta}']_{\mathcal{S}} = \mathbf{H}_{\mathcal{S}}\boldsymbol{\theta}_{\mathcal{S}}$$

with $\mathbf{H}_{\mathcal{S}} = \mathbf{\Psi}\mathbf{\Phi}_{\mathcal{S}}$ and the $N \times |\mathcal{S}|$ matrix $\mathbf{\Phi}_{\mathcal{S}}$ is built up with the $|\mathcal{S}|$ columns of $\mathbf{\Phi}$ having their indices in $\mathcal{S}$ and the $|\mathcal{S}| \times 1$ vector $\boldsymbol{\theta}_{\mathcal{S}}$ is composed by the non-zero entries in $\boldsymbol{\theta}'$ randomly selected thanks to matrix $\boldsymbol{P}$. Fig. 1 illustrates the considered model.

2.1. Statistical priors

2.1.1. Universal design strategy of matrix $\mathbf{H}$

Determining whether the dictionary $\mathbf{H} = \mathbf{\Psi}\mathbf{\Phi}$ satisfies the the concentration inequality is combinatorially complex but the so-called universal design strategy has been introduced for instance in [3, 2]. Assume that matrix $\mathbf{\Phi}$ is an orthonormal basis and the measurement matrix $\mathbf{\Psi}$ is drawn from an independent and identically distributed Gaussian entries of zero mean and variance $1/N$. For $0 < \epsilon < 1$, the concentration probability for dictionary $\mathbf{H}$ is

$$\Pr\left(\left|||\mathbf{H}\boldsymbol{\theta}||^2 - ||\boldsymbol{\theta}||^2\right| \geq \epsilon||\boldsymbol{\theta}||^2\right) = \Pr\left(\left|||\mathbf{\Psi}\mathbf{s}||^2 - ||\mathbf{s}||^2\right| \geq \epsilon||\mathbf{s}||^2\right)$$

since $||\mathbf{s}||^2 = ||\boldsymbol{\theta}||^2$ thanks to $\mathbf{\Phi}^T\mathbf{\Phi} = \mathbf{I}$. So, according to the above equality, we can see that the concentration probability for $\mathbf{H}$ with an orthonormal $\mathbf{\Phi}$ is characterized by the concentration probability for the measurement matrix $\mathbf{\Psi}$. According to [36, 37, 38, 39, 40], it is well known that Gaussian matrices satisfy with high probability the concentration inequality:

$$\Pr\left(\left|||\mathbf{\Psi}\mathbf{s}||^2 - ||\mathbf{s}||^2\right| \geq \epsilon||\mathbf{s}||^2\right) \leq e^{-cN\epsilon^2}$$

where c is a given positive constant. Note that this statistical guaranty ensures that practical estimators can successfully recover a $|\mathcal{S}|$-sparse amplitude vector from noisy measurements with high probability for a number of measurements $N = O(|\mathcal{S}|\log(K/|\mathcal{S}|))$. Note that the number of measurements is smaller than the classical sampling theory [5].

2.1.2. Design of the selection matrix $\boldsymbol{P}$

Definition 2.1 (Guaranty on the non-singularity of the Fisher information).
Define the deterministic set $\mathcal{I} \subset \{1, \ldots, K\}$ of cardinality $|\mathcal{I}| = N - 1 < K$.

Given $|\mathcal{I}|$ available measurements, the Fisher information associated to model of eq. (2) is said to be non-singular if the degree of freedom satisfies $|\mathcal{I}| - |\mathcal{S}| \geq 0$. In the estimation point of view, considering more parameters of interest than the number available measurements leads to a rank deficient Fisher Information Matrix (FIM).

In the context of Definition 2.1, it cannot exist an estimator with finite variance [41, 42, 43, 44, 24].

Random cardinalities with the FIM non-singularity guaranty. For $1 \leq k \leq K$, we have two possible cases:

$$\begin{cases} \theta_{k \in \mathcal{I}} & \neq \quad 0 \quad \text{with probability } P, \\ \theta_{k \in \mathcal{I}} & = \quad 0 \quad \text{with probability } 1 - P. \end{cases}$$

The above formulation can be compactly expressed according to

$$[\boldsymbol{P}]_{k,k} = 1_{\mathcal{S}}(k) 1_{\mathcal{I}}(k) \tag{3}$$

where $1_{\mathcal{I}}(k)$ enforces the FIM non-singularity guaranty and

$$1_{\mathcal{S}}(k) \sim \text{Bernou}(P)$$

for a probability of success given by $P = L/(N-1)$ and $L = \mathbb{E}|\mathcal{S}|$.

By definition the cardinality of $\mathcal{S}$ conditionally to a given set $\mathcal{I}$ is

$$|\mathcal{S}| \,\Big|\, \mathcal{I} = \text{Tr}\boldsymbol{P} = \sum_{k=1}^{K} 1_{\mathcal{S}}(k) 1_{\mathcal{I}}(k) = \sum_{k \in \mathcal{I}} 1_{\mathcal{S}}(k)$$

with $|\mathcal{I}| = N - 1$. So, $|\mathcal{S}| \,\big|\, \mathcal{I}$ is the sum of $|\mathcal{I}|$ i.i.d. Bernoulli-distributed variables. As a consequence,

$$|\mathcal{S}| \sim \text{Binomial}(|\mathcal{I}|, P).$$

2.2. Bayesian lower bound on the BMSE

2.2.1. Definition of the performance criterion

We define the conditional BMSE given $\mathcal{S}$ and $|\mathcal{S}|$ to be

$$\text{BMSE}_{\mathcal{S},|\mathcal{S}|} = \frac{1}{N} \mathbb{E}_{\mathbf{y},\mathbf{H},\boldsymbol{\theta}|\mathcal{S},|\mathcal{S}|} \left\| \boldsymbol{\theta}_{\mathcal{S}} - \hat{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{H}_{\mathcal{S}}) \right\|^2$$
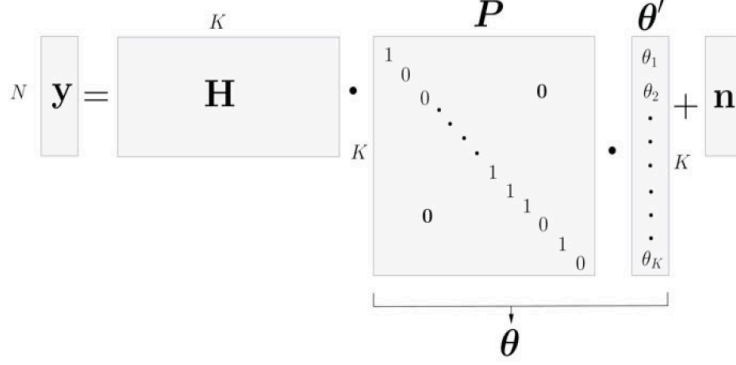
6

Figure 1: CS framework with random support and cardinality

where $\hat{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{H}_{\mathcal{S}})$ is an estimate of $\boldsymbol{\theta}_{\mathcal{S}}$ that knows the support $\mathcal{S}$ and respects the FIM non-singularity guaranty described in the previous section. Averaging the conditional $\text{BMSE}_{\mathcal{S},|\mathcal{S}|}$ over the random quantities, *i.e.*, $\mathcal{S}$ and $|\mathcal{S}|$ yields:

$$\text{BMSE} = \mathbb{E}_{|\mathcal{S}|\big|\mathcal{I}}\mathbb{E}_{\mathcal{S}\big||\mathcal{S}|}\text{BMSE}_{\mathcal{S},|\mathcal{S}|}.$$

Remark 2.2.   • *The* BMSE *is a natural limit performance criterion for an estimator unaware of the support $\mathcal{S}$ and its cardinality $|\mathcal{S}|$.*

• *The* BMSE *can easily take into account Definition 2.1.*

2.2.2. Which Bayesian CRB-type bound ?

In the family of the Bayesian lower bounds [15] based on the Cramér-Rao Bound (CRB) framework, it exists the VanTree's lower bound and the Expected CRB.

Remark 2.3. *We focus our effort on the ECRB for the two following reasons:*

• *The* ECRB *is based on the deterministic/Bayesian connexion [45] and is simple to derive.*

• *The* ECRB *is the tightest Bayesian lower bound in the low noise regime regarding the considered Bayesian linear model [46]. As a price to pay, the* ECRB *inherits from the regularity conditions of the deterministic CRB.*

7

In particular, to derive bound ECRB, *the non-singularity of the Fisher information described in Definition 2.1 cannot be violated.*

The deterministic/Bayesian connexion principle is briefly recalled here. The MSE conditioned to $\{\mathbf{H}, \boldsymbol{\theta}, \mathcal{S}, |\mathcal{S}|\}$ is defined as

$$\mathrm{MSE} = \frac{1}{N} \mathbb{E}_{\mathbf{y}\big|\mathbf{H},\boldsymbol{\theta},\mathcal{S},|\mathcal{S}|} ||\boldsymbol{\theta}_{\mathcal{S}} - \hat{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{H}_{\mathcal{S}})||^2$$

and enjoys the following fundamental inequality

$$\mathrm{MSE} \geq \mathrm{CRB} = \frac{1}{N} \mathrm{Tr}\left[\mathbf{F}^{-1}\right] \tag{4}$$

where CRB is the normalized trace CRB with $\mathbf{F}$ the FIM [16] given by

$$\mathbf{F} = \mathbb{E}_{\mathbf{y}\big|\mathbf{H},\boldsymbol{\theta},\mathcal{S},|\mathcal{S}|}\left[\boldsymbol{\zeta}\boldsymbol{\zeta}^T\right] \tag{5}$$

where $\boldsymbol{\zeta}$ is the score function defined as the first-order derivative of the log-likelihood function $p(\mathbf{y}|\mathbf{H}, \boldsymbol{\theta}, \mathcal{S}, |\mathcal{S}|) \in C^1$ with respect to $\boldsymbol{\theta}_{\mathcal{S}}^T$. The BMSE is defined according to

$$\mathrm{BMSE} = \mathbb{E}_{\mathbf{H},\boldsymbol{\theta},\mathcal{S},|\mathcal{S}|\big|\mathcal{I}} \mathrm{MSE}.$$

Using eq. (4) in the above definition, we obtain the definition of the ECRB such as

$$\mathrm{BMSE} \geq \mathrm{ECRB} = \mathbb{E}_{|\mathcal{S}|\big|\mathcal{I}} \mathbb{E}_{\mathcal{S}\big||\mathcal{S}|} \underbrace{\underbrace{\mathbb{E}_{\mathbf{H},\boldsymbol{\theta}\big|\mathcal{S},|\mathcal{S}|} \mathrm{CRB}}_{\mathrm{ECRB}_{\mathcal{S},|\mathcal{S}|}}}_{\mathrm{ECRB}_{|\mathcal{S}|}}.$$

The final expression of the ECRB conditionally to set $\mathcal{I}$ is

$$\begin{aligned}
\mathrm{ECRB} &= \mathbb{E}_{|\mathcal{S}|\big|\mathcal{I}} \mathrm{ECRB}_{|\mathcal{S}|} \\
&= \sum_{\ell=1}^{|\mathcal{I}|} \mathrm{Pr}(|\mathcal{S}| = \ell)\, \mathrm{ECRB}_{\ell}.
\end{aligned} \tag{6}$$

Case of the Bayesian linear model with Gaussian noise. As $\log p(\mathbf{y}|\mathbf{H}, \boldsymbol{\theta}, \mathcal{S}, |\mathcal{S}|)$ follows a log-normal distribution denoted by $\log \mathcal{N}(\mathbf{H}_{\mathcal{S}}\boldsymbol{\theta}_{\mathcal{S}}, \sigma^2 \mathbf{I})$, we use the well-known Slepian-Bangs formula [16]. Consequently, a simple derivation based on eq. (5) leads to

$$\mathrm{CRB} = \frac{\sigma^2}{N}\mathrm{Tr}\left[(\mathbf{H}_{\mathcal{S}}^T\mathbf{H}_{\mathcal{S}})^{-1}\right].$$

Remark 2.4. *Due to the linear relation between the measurement and the amplitude vector, we have the following properties:*

- *The lower bound* CRB *is not a function of* $\boldsymbol{\theta}$, *thus*

$$\mathrm{ECRB}_{\mathcal{S},|\mathcal{S}|} = \mathbb{E}_{\mathbf{H},\boldsymbol{\theta}|\mathcal{S},|\mathcal{S}|}\mathrm{CRB} = \mathbb{E}_{\mathbf{H}|\mathcal{S},|\mathcal{S}|}\mathrm{CRB}.$$

  This means that the derived ECRB will be valid for any amplitude prior $p(\boldsymbol{\theta})$.

- *The amplitude prior could not be in* C^1 *as for instance the uniform distribution. This implies relaxed constraint on the choice of* $p(\boldsymbol{\theta})$.

Given the above remark, the conditional ECRB relatively to $|\mathcal{S}|$ is given by

$$\mathrm{ECRB}_{|\mathcal{S}|} = \frac{\sigma^2}{N}\mathbb{E}_{\mathcal{S}||\mathcal{S}|}\mathbb{E}_{\mathbf{H}|\mathcal{S},|\mathcal{S}|}\mathrm{Tr}\left[(\mathbf{H}_{\mathcal{S}}^T\mathbf{H}_{\mathcal{S}})^{-1}\right]. \tag{7}$$

Case of a Gaussian measurement matrix.

Lemma 2.5. *For a Gaussian measurement matrix, the lower bound given in eq. (7) reads*

$$\mathrm{ECRB}_{|\mathcal{S}|} = \sigma^2 \frac{|\mathcal{S}|}{N - |\mathcal{S}|}. \tag{8}$$

Proof The proof is straightforward by using eq. (7) and the property of the Wishart matrices [47] for $|\mathcal{S}| < N$. Let $\mathbf{Z}_{\mathcal{S}} = \sqrt{N}\mathbf{H}_{\mathcal{S}}$. Observe that the entries of matrix $\mathbf{Z}_{\mathcal{S}}$ have now a unit variance. Thus $\mathbb{E}_{\mathbf{z}|\mathcal{S},|\mathcal{S}|}\mathrm{Tr}\left[\left(\mathbf{Z}_{\mathcal{S}}^T\mathbf{Z}_{\mathcal{S}}\right)^{-1}\right]$ is given by eq. (8). Now, remark that the above expression is not a function of the support $\mathcal{S}$, then using the above expression in eq. (7) provides eq. (8). $\square$

Remark 2.6. *For a Gaussian measurement matrix, the ECRB is only a function of the random cardinality $|\mathcal{S}|$ but not of the instantiations of $\mathcal{S}$.*

Result 2.1. *Using Lemma 2.5, the ECRB in eq. (6) is*

$$\mathrm{ECRB} = \sigma^2 \sum_{\ell=1}^{|\mathcal{I}|} \frac{\ell}{N-\ell} \mathrm{Pr}(|\mathcal{S}| = \ell).$$

Case of random supports and cardinalities.

Result 2.2. *For any amplitude vector prior, for $L < N - 1$ where $L = \mathbb{E}|\mathcal{S}|$ and for a probability of success given by $P = L/(N-1)$, the ECRB verifies the following inequality:*

$$\mathrm{BMSE} \geq \mathrm{ECRB} = \sigma^2 \frac{P}{1-P} \left(1 - P^{N-1}\right). \tag{9}$$

Proof See Appendix 6.1. $\square$

Remark 2.7. *For a large number of measurements, i.e., $N \gg 1$[1], we can give the following approximation:*

$$\mathrm{ECRB} \approx \sigma^2 \frac{P}{1-P} = \sigma^2 \frac{L}{N-L}.$$

3. CS with corrupted measurements scenario

3.1. Model formalism

Assume that there exists an $N \times N$ unknown random selection matrix $\bar{\boldsymbol{P}}$. The aim is to estimate the amplitudes in vector $\boldsymbol{\theta}_{\mathcal{S}}$, based on the reduced-size measurement vector:

$$\mathbf{y}_{\mathcal{I}_{\mathrm{uc}}} = [\bar{\boldsymbol{P}}\mathbf{y}]_{\mathcal{I}_{\mathrm{uc}}}$$

where $\mathcal{I}_{\mathrm{uc}}$ is the unknown random set of indexes of the measurements classified as uncorrupted of mean cardinality given by $\mathbb{E}|\mathcal{I}_{\mathrm{uc}}| = N_{\mathrm{uc}}$ and the full measurement vector $\mathbf{y}$ has been defined by eq. (2). An equivalent expression is

$$\mathbf{y}_{\mathcal{I}_{\mathrm{uc}}} = [\bar{\boldsymbol{P}}\mathbf{H}\boldsymbol{\theta}]_{\left\{\mathcal{I}_{\mathrm{uc}}, \{1,\ldots,K\}\right\}} + \mathbf{n}_{\mathcal{I}_{\mathrm{uc}}}$$

---

[1]Note that it is assumed that L is not neglected with respect to N.

where $\mathbf{n}_{\mathcal{I}_{uc}} = [\bar{\boldsymbol{P}}\mathbf{n}]_{\mathcal{I}_{uc}}$ and

$$[\bar{\boldsymbol{P}}\mathbf{H}\boldsymbol{\theta}]_{\left\{\mathcal{I}_{uc},\{1,\ldots,K\}\right\}} = [\bar{\boldsymbol{P}}\mathbf{H}\boldsymbol{P}^2\boldsymbol{\theta}']_{\left\{\mathcal{I}_{uc},\{1,\ldots,K\}\right\}}$$
$$= [\bar{\boldsymbol{P}}\mathbf{H}\boldsymbol{P}]_{\{\mathcal{I}_{uc},\mathcal{S}\}}\boldsymbol{\theta}_{\mathcal{S}}. \tag{10}$$

Remark that eq. (10) shows a double random selections acting on the rows and columns of the dictionary $\mathbf{H}$ as illustrated in Fig. 2.
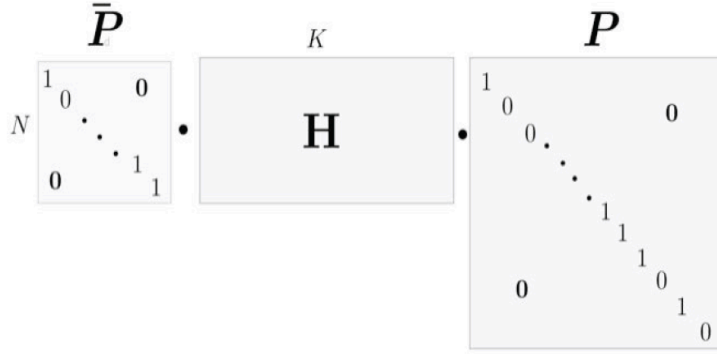

Figure 2: Random selections acting on the rows and columns of the dictionary

Note $\mathbf{H}_{\mathcal{P}} = [\bar{\boldsymbol{P}}\mathbf{H}\boldsymbol{P}]_{\{\mathcal{I}_{uc},\mathcal{S}\}}$ with $\mathcal{P} = \{\mathcal{I}_{uc}, \mathcal{S}\}$, then the CS model with corrupted measurements is given by

$$\mathbf{y}_{\mathcal{I}_{uc}} = \mathbf{H}_{\mathcal{P}}\boldsymbol{\theta}_{\mathcal{S}} + \mathbf{n}_{\mathcal{I}_{uc}}. \tag{11}$$

Remark 3.1. *CS with corrupted measurements takes formally a similar formulation as the used one in the previous section with the following substitutions:*

$$\mathcal{S} \leftrightarrow \mathcal{P},$$

$$\mathbf{y} \leftrightarrow \mathbf{y}_{\mathcal{I}_{uc}}$$

$$\mathbf{H}_{\mathcal{S}} \leftrightarrow \mathbf{H}_{\mathcal{P}}.$$

3.2. Characterization of the selection matrices

Definition 3.2. *We recall that the the non-singularity of the Fisher information constraint means that we cannot estimate more amplitude parameters than the number of available measurements. So, the FIM non-singularity guaranty*

for the corrupted measurement scenario can be naturally generalized from the definition given in Definition 2.1. Specifically, given $|\mathcal{I}_{\mathrm{uc}}|$ the random number of uncorrupted measurements, we must have $|\mathcal{S}| \leq |\mathcal{I}_{\mathrm{uc}}|$ to ensure a positive degree of freedom of the system.

Remark 3.3. *CS with corrupted measurements has a major difference with respect to the full measurements case since the number of rows, $|\mathcal{I}_{\mathrm{uc}}|$, and of columns, $|\mathcal{S}|$, of dictionary $\mathbf{H}_{\mathcal{P}}$ are now two dependent randoms variables. In addition, due to the FIM non-singularity guaranty, the random cardinality $|\mathcal{S}|$ has to be defined in a hierarchical framework, meaning that its pmf will be a function of the random cardinality $|\mathcal{I}_{\mathrm{uc}}|$. The adopted model is illustrated in Fig. 3.*
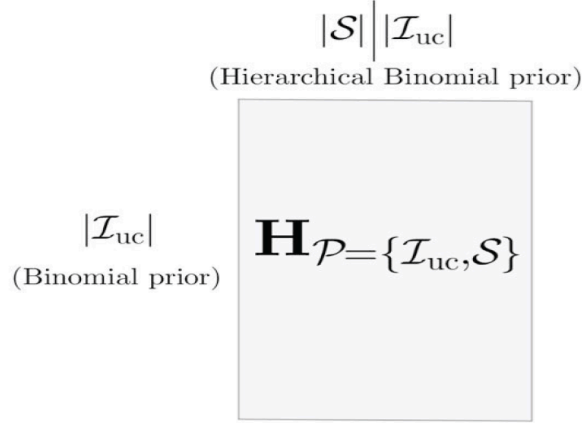

$$|\mathcal{S}|\Big|\,|\mathcal{I}_{\mathrm{uc}}|$$
(Hierarchical Binomial prior)

$$|\mathcal{I}_{\mathrm{uc}}|$$
(Binomial prior)

$$\mathbf{H}_{\mathcal{P}=\{\mathcal{I}_{\mathrm{uc}},\mathcal{S}\}}$$

Figure 3: Dictionary random sizes for CS with corrupted measurements

3.2.1. Selection matrix $\bar{\boldsymbol{P}}$

For $1 \leq n \leq N$, a convenient choice for the diagonal selection matrix $\bar{\boldsymbol{P}}$ is

$$[\bar{\boldsymbol{P}}]_{n,n} = 1_{\mathcal{I}_{\mathrm{uc}}}(n) \sim \mathrm{Bernou}(P_{\mathrm{uc}}) \text{ with } P_{\mathrm{uc}} = \frac{N_{\mathrm{uc}}}{N-1}.$$

Consequently, the cardinality $|\mathcal{I}_{\mathrm{uc}}|$ is given by

$$|\mathcal{I}_{\mathrm{uc}}| = \mathrm{Tr}\bar{\boldsymbol{P}} = \sum_{n=1}^{|\mathcal{I}|} 1_{\mathcal{I}_{\mathrm{uc}}}(n) \sim \mathrm{Binomial}(N-1, P_{\mathrm{uc}}). \tag{12}$$

3.2.2. New construction of the selection matrix $\boldsymbol{P}$

Following the same formalism as eq. (3), we have

$$[\boldsymbol{P}]_{k,k} = 1_{\mathcal{S}}(k)1_{\mathcal{I}_{\mathrm{uc}}}(k)$$

where $1_{\mathcal{I}_{\mathrm{uc}}}(k)$ enforces the FIM non-singularity guaranty. The cardinality of $\mathcal{S}$ is given by

$$|\mathcal{S}| \; \Big| \; \mathcal{I}_{\mathrm{uc}} = \mathrm{Tr}\boldsymbol{P} = \sum_{k=1}^{K} 1_{\mathcal{S}}(k)1_{\mathcal{I}_{\mathrm{uc}}}(k) = \sum_{k \in \mathcal{I}_{\mathrm{uc}}} 1_{\mathcal{S}}(k).$$

As the above expression is only a function of the random cardinality $|\mathcal{I}_{\mathrm{uc}}|$ and is the sum of $|\mathcal{I}_{\mathrm{uc}}| - 1$ i.i.d. independent Bernoulli variables, we conclude

$$|\mathcal{S}| \; \Big| \; |\mathcal{I}_{\mathrm{uc}}| \sim \mathrm{Binomial}(|\mathcal{I}_{\mathrm{uc}}| - 1, \bar{P}). \tag{13}$$

Note that $|\mathcal{S}|$ conditionally to $|\mathcal{I}_{\mathrm{uc}}|$ follows a hierarchical [48] Binomial distribution since $|\mathcal{I}_{\mathrm{uc}}|$ is also random. Marginalizing over $|\mathcal{I}_{\mathrm{uc}}|$ allows to derive the probability $\bar{P}$. More precisely, let $\bar{L} = \mathbb{E}_{|\mathcal{S}|,|\mathcal{I}_{\mathrm{uc}}|}|\mathcal{S}|$ be the mean number of non-zero amplitudes, we have

$$\bar{L} = \mathbb{E}_{|\mathcal{I}_{\mathrm{uc}}|}\mathbb{E}_{|\mathcal{S}| \; \big| \; |\mathcal{I}_{\mathrm{uc}}|}|\mathcal{S}|$$

$$= (\mathbb{E}_{|\mathcal{I}_{\mathrm{uc}}|}(|\mathcal{I}_{\mathrm{uc}}|) - 1)\bar{P}$$

$$= (N_{\mathrm{uc}} - 1)\bar{P}$$

and thus

$$\bar{P} = \frac{\bar{L}}{N_{\mathrm{uc}} - 1}. \tag{14}$$

3.3. ECRB based on corrupted measurements

3.3.1. Definition of a new performance criterion

Let $\hat{\boldsymbol{\theta}}(\mathbf{y}_{\mathcal{I}_{\mathrm{uc}}}, \mathbf{H}_{\mathcal{P}})$ be an estimator of $\boldsymbol{\theta}_{\mathcal{S}}$. The BMSE conditionally to $\mathcal{P}$ and $|\mathcal{P}| = \{|\mathcal{I}_{\mathrm{uc}}|, |\mathcal{S}|\}$ in case of the new model according to

$$\mathrm{BMSE}_{\mathcal{P},|\mathcal{P}|}^{\mathrm{uc}} = \frac{1}{N}\mathbb{E}_{\mathbf{y},\mathbf{H},\boldsymbol{\theta}|\mathcal{P},|\mathcal{P}|}||\boldsymbol{\theta}_{\mathcal{S}} - \hat{\boldsymbol{\theta}}(\mathbf{y}_{\mathcal{I}_{\mathrm{uc}}}, \mathbf{H}_{\mathcal{P}})||^2,$$

$$\mathrm{BMSE}_{|\mathcal{P}|}^{\mathrm{uc}} = \mathbb{E}_{\mathcal{P}||\mathcal{P}|}\mathrm{BMSE}_{\mathcal{P},|\mathcal{P}|}^{\mathrm{uc}},$$

$$\mathrm{BMSE}^{\mathrm{uc}} = \sum_{n=1}^{|\mathcal{I}|}\sum_{\ell=0}^{n-1}\Pr(|\mathcal{P}| = \{\ell, n\})\,\mathrm{BMSE}_{\{n,\ell\}}^{\mathrm{uc}}.$$

3.3.2. Derivation of the ECRB

Lemma 3.4. *The ECRB for the non-zero amplitudes conditionally to $|\mathcal{I}_{\mathrm{uc}}|$ and for $|\mathcal{S}| < |\mathcal{I}_{\mathrm{uc}}|$ is given by*

$$\mathrm{ECRB}_{|\mathcal{P}|}^{\mathrm{uc}} = \sigma^2 \frac{|\mathcal{S}|}{|\mathcal{I}_{\mathrm{uc}}| - |\mathcal{S}|}.$$

Proof The proof is similar to Lemma 2.5 with $\mathbf{Z}_{\mathcal{P}} = \sqrt{\frac{N}{|\mathcal{I}_{\mathrm{uc}}|}}\mathbf{H}_{\mathcal{P}}$. $\square$

Now, we can give the following result.

Result 3.1. *Based on Lemma 3.4 and the modelization introduced in Section 3.2.1, the ECRB defined by*

$$\mathrm{ECRB}^{\mathrm{uc}} = \sum_{n=1}^{N-1}\sum_{\ell=0}^{n-1}\Pr(|\mathcal{P}| = \{\ell, n\})\,\mathrm{ECRB}_{n,\ell}^{\mathrm{uc}}$$

is given by

$$\mathrm{ECRB}^{\mathrm{uc}} = \frac{\sigma^2}{1 - \bar{P}}\left(\bar{P} + (1 - \bar{P})(1 - P_{\mathrm{uc}})^{N-1} - (1 - P_{\mathrm{uc}}(1 - \bar{P}))^{N-1}\right) \quad (15)$$

where P_{uc} and $\bar{P}$ are given by eq. (12) and eq. (14), respectively.

Proof See Appendix 6.2. $\square$

Result 3.2. *For large N, eq. (15) in the previous result can be approximated according to*

$$\mathrm{ECRB}^{\mathrm{uc}} \approx \sigma^2 \frac{\bar{P}}{1 - \bar{P}} = \sigma^2 \frac{\bar{L}}{N_{\mathrm{uc}} - \bar{L}}.$$

14

Proof Using Result 3.1, the proof is straightforward by noticing that $(1 - P_{\mathrm{uc}})^{N-1}$ and $(1 - P_{\mathrm{uc}}(1 - \bar{P}))^{N-1}$ vanish for large N. We can neglige these two terms as long as N_{uc} is not too small with respect to N, *i.e.*, for a not too small P_{uc}. $\square$

4. Application to corrupted measurements due to an impulsive (sparse) noise

To illustrate the effect of the corrupted measurements, Fig. 4 shows the ratio $\frac{\mathrm{ECRB}_{\mathrm{uc}}}{\mathrm{ECRB}}$ with respect to, P_{uc}, the probability of each measurement to be uncorrupted for several values of P which is the probability for each amplitude to be non-zero. First, we can note that the $\mathrm{ECRB}_{\mathrm{uc}}$ can be around 10^2 times higher than the ECRB for small a P_{uc}. So, we observe the drastic degradation of the corruption of the measurement on estimation accuracy limit.
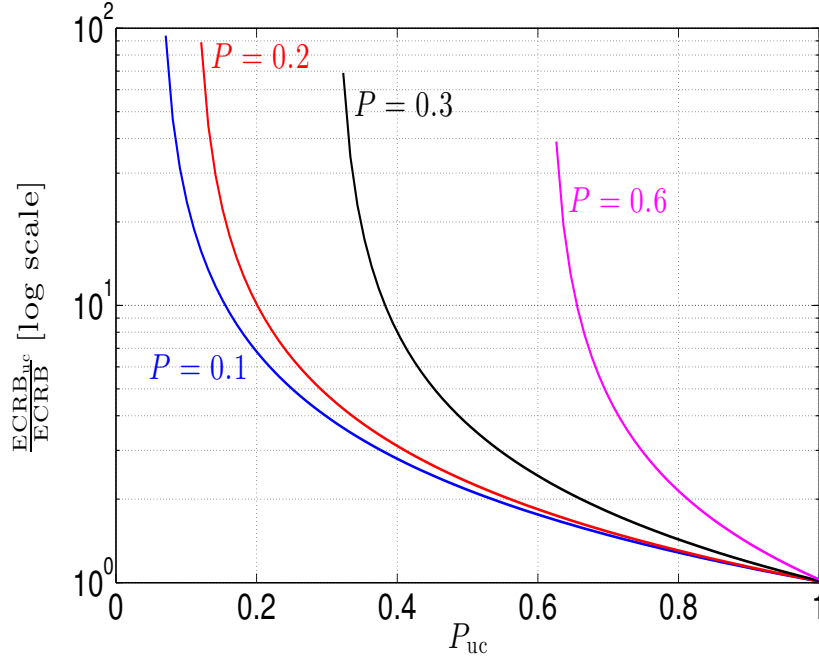

Figure 4: ratio *vs.* P_{uc}, SNR $= 30$ dB

In this section, we apply the proposed lower bounds in the context of the presence of an impulsive (sparse) noise. Specifically, the considered model is the following [25, 26, 27, 28]:

$$\tilde{\mathbf{y}} = \mathbf{y} + \mathbf{e} = \mathbf{H}\boldsymbol{\theta} + \mathbf{e} + \mathbf{n}$$

where $\mathbf{y}$ has been defined in eq. (2), $\mathbf{e}$ is a impusive random noise such as $\mathbf{e}_{\mathcal{I}_{\mathrm{uc}}} = [\bar{\boldsymbol{P}}\mathbf{e}]_{\mathcal{I}_{\mathrm{uc}}} = \mathbf{0}$. Each non-zero entry in $\mathbf{e}$ corresponds to a corrupted measurement in vector $\tilde{\mathbf{y}}$. We have

$$\tilde{\mathbf{y}}_{\mathcal{I}_{\mathrm{uc}}} = [\bar{\boldsymbol{P}}\tilde{\mathbf{y}}]_{\mathcal{I}_{\mathrm{uc}}} = \mathbf{y}_{\mathcal{I}_{\mathrm{uc}}} + [\bar{\boldsymbol{P}}\mathbf{e}]_{\mathcal{I}_{\mathrm{uc}}} = \mathbf{H}_{\mathcal{P}}\boldsymbol{\theta}_{\mathcal{S}} + \mathbf{e}_{\mathcal{I}_{\mathrm{uc}}} + \mathbf{n}_{\mathcal{I}_{\mathrm{uc}}} = \mathbf{H}_{\mathcal{P}}\boldsymbol{\theta}_{\mathcal{S}} + \mathbf{n}_{\mathcal{I}_{\mathrm{uc}}}.$$

Remark that the model is formally equivalent to the one given by eq. (11). We assume the two following standard strategies.

(A) In many operational contexts as for instance Radar processing [49], source localisation [50, 51], array calibration for radio-interferometers [52, 53], a set of measurements free from the signal of interest is available. Let $\mathbf{y}^0 = \mathbf{e} + \mathbf{n}$ be this secondary set of measurements. The estimation of the support $\mathcal{I}_{\mathrm{uc}}$ is described according to

$$\mathcal{A}(\mathbf{y}^0, \mathbf{I}_N) \rightarrow \hat{\mathcal{I}}_{\mathrm{uc}}$$

where $\mathcal{A}(\cdot, \cdot)$ denotes any sparse-based estimators [54, 55, 56, 57, 58, 59]. Given this estimated support, it is possible to remove the corrupted measurements to focus the estimation only on the vector of interest $\boldsymbol{\theta}_{\mathcal{S}}$. Finally, the entire process is described according to

$$\mathcal{A}\left(\tilde{\mathbf{y}}_{\hat{\mathcal{I}}_{\mathrm{uc}}}, [\mathbf{H}]_{\left\{\hat{\mathcal{I}}_{\mathrm{uc}}, \{1, \ldots, K\}\right\}}\right) \rightarrow (\hat{\mathcal{S}}, \hat{\boldsymbol{\theta}}_{\hat{\mathcal{S}}}).$$

(B) A second strategy is based on the principle of "democracy policy" or "Justice Pursuit" [27, 26]. In this context, the sparse noise $\mathbf{e}_{\mathcal{I}_{\mathrm{uc}}}$ is viewed as a signal of interest and has to be estimated jointly with the vector of interest $\boldsymbol{\theta}_{\mathcal{S}}$. The considered model is

$$\tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{H} & \mathbf{I}_N \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{e} \end{bmatrix} + \mathbf{n}.$$

16

Note that in this strategy, a larger dictionary is constituted by the union of dictionaries $\mathbf{H}$ and $\boldsymbol{\Omega}$. Specifically, we have

$$\mathcal{A}(\tilde{\mathbf{y}}, \begin{bmatrix} \mathbf{H} & \mathbf{I}_N \end{bmatrix}) \rightarrow \left(\begin{bmatrix} \hat{\mathcal{S}} \\ \hat{\mathcal{I}}_{\mathrm{uc}} \end{bmatrix}, \begin{bmatrix} \hat{\boldsymbol{\theta}}_{\hat{\mathcal{S}}} \\ \hat{\mathbf{e}}_{\hat{\mathcal{I}}_{\mathrm{uc}}} \end{bmatrix} \right).$$

In the conducted simulations, estimator $\mathcal{A}(\cdot, \cdot)$ is the Orthogonal Matching Pursuit (OMP) [56, 57]. The OMP adapted to strategy *(B)* follows the acronym OMP-DS. In addition, the impulsive noise is assumed sparse in the sample domain as usually done [27, 26]. On Fig. 5, the BMSE is drawn for a wide range of SNR with $N = 100$, $K = 200$ and $\bar{L} = L = 10$. We first observe that the BMSE of the OMP in presence of corrupted measurements is saturated. This is true even for a large N_{uc} as in Fig. 5 or for a smaller N_{uc} as in Fig. 6. Another observation is that for a large N_{uc}, the strategies *(A)* and *(B)* show very close estimation accuracies. In addition, we can also note the accurate prediction proposed by the derived lower bounds. We can also remark that bounds ECRB and ECRB$_{\mathrm{uc}}$ are almost identical since only a few number of measurements given by $N - N_{\mathrm{uc}} = 5$ are corrupted.

On Fig. 6, the BMSE is drawn for a wide range of SNR but for a larger number of corrupted measurements of $N - N_{\mathrm{uc}} = 35$. This implies that P and $\bar{P}$ take different values and the two bounds ECRB and ECRB$_{\mathrm{uc}}$ can clearly be distinguished. In this difficult context, the strategy *(B)*, through estimator OMP-DS, seems ineffective. At the same time, the proposed bound ECRB$_{\mathrm{uc}}$ predicts with a high accuracy the estimation performance of the strategy *(A)*.

Finally, the BMSE is drawn on Fig. 7 with respect to the probability P_{uc}. We recall as much P_{uc} is small as much the number of corrupted measurements is large. We can observe that the OMP is ineffective even for P_{uc} close to one. The OMP-DS corresponding to the strategy *(B)* reaches the bound ECRB$_{\mathrm{uc}}$ for a small number of corrupted measurements. At contrary, the strategy *(A)* is able to meet the bound ECRB$_{\mathrm{uc}}$ for P_{uc} approximately larger than 0.5.
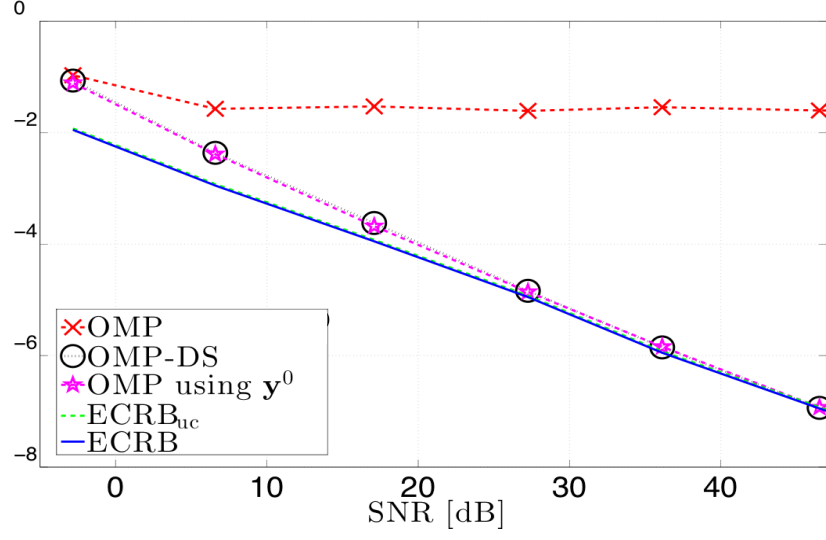
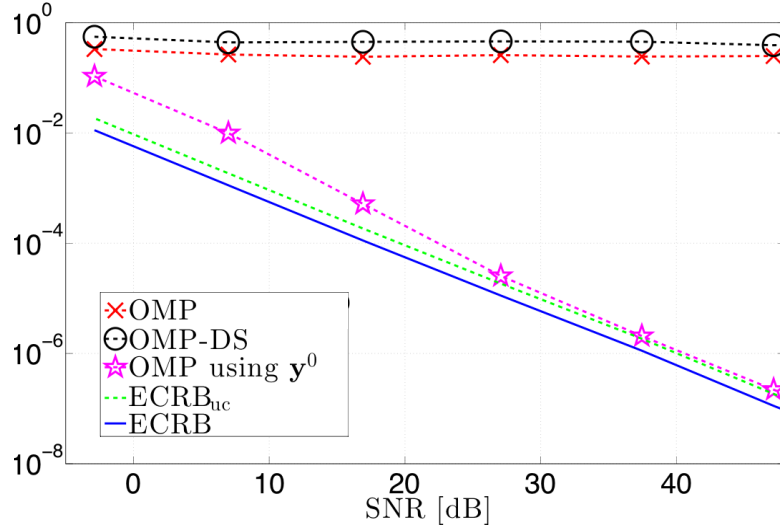Figure 5: BMSE *vs.* SNR with $N_{\mathrm{uc}} = 95$ and $P \approx \bar{P} \approx 0.1$

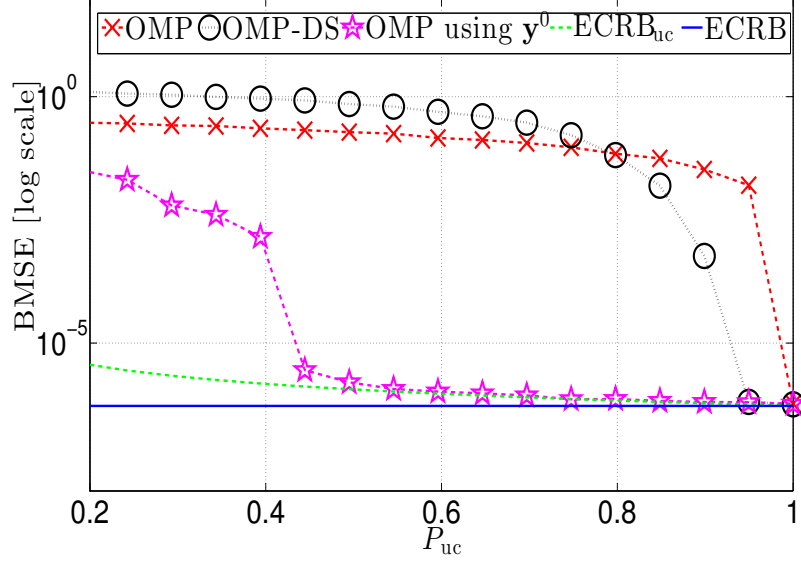Figure 6: BMSE *vs.* SNR with $N_{\mathrm{uc}} = 65$, $P \approx 0.1$ and $\bar{P} \approx 0.16$

18

Figure 7: BMSE *vs.* P_{uc} with $L = 5$, $19 \leq N_{\mathrm{uc}} \leq N - 1$, $P \approx 0.1$ and $0.05 \leq \bar{P} \leq 0.28$, SNR = 30 dB

5. Conclusion

CS with corrupted measurements is a timely and important research topic. In practice, we have often to face to the presence of corrupted measurement samples in a vector $\mathbf{y}_{\mathcal{I}_{\mathrm{uc}}}$ extracted from the complete measurement vector $\mathbf{y}$. To take into account of an uncertain knowledge of $\mathcal{I}_{\mathrm{uc}}$ and its cardinality, it is assumed that each measurement sample in $\mathbf{y}$ has the probability $1 - P_{\mathrm{uc}}$ to be corrupted. As a consequence, the cardinality $|\mathcal{I}_{\mathrm{uc}}|$ follows a Binomial prior. In this context, the Expected CRB (ECRB) which is fundamental lower bound on the BMSE of any estimators is derived for the estimation of a $|\mathcal{S}|$-sparse amplitude vector where each of its entry is the product of a continuous random variable of unspecified pdf and a Bernoulli random variable of probability P to be in support $\mathcal{S}$. As, we focus our effort to the Bayesian linear model with non-singular Fisher Information Matrix, the cardinal $|\mathcal{S}|\big||\mathcal{I}_{\mathrm{uc}}|$ has to follow a hierarchical Binomial distribution on set $\{1, \ldots, |\mathcal{I}_{\mathrm{uc}}| - 1\}$. In this framework and for a Gaussian measurement matrix, very compact and simple expressions

19

of the ECRB are proposed. Finally, the proposed lower bounds are illustrated in the context of two standard estimation strategies robust to an impulsive (sparse) noise.

6. Appendix

6.1. Proof of Result 2.2

Using in eq. (6), Lemma 2.5, Remark 2.6, the ECRB is given by

$$\text{ECRB} = \sigma^2 \sum_{\ell=1}^{N-1} \frac{\ell}{N-\ell} \binom{N-1}{\ell} P^\ell (1-P)^{N-\ell-1}$$
$$= \frac{\sigma^2}{N(1-P)} \left(\mathbb{E}G - NP^N \right)$$

where $G \sim \text{Binomial}(N, P)$. Using the first moment of the Binomial variable G [34] given by $\mathbb{E}G = NP$, we obtain eq. (9).

6.2. Proof of Result 3.1

Using eq. (13), the ECRB for corrupted measurements is given by

$$\text{ECRB}^{\text{uc}} = \sigma^2 \sum_{n=1}^{N-1} \text{Pr}(|\mathcal{I}_{\text{uc}}| = n) \sum_{\ell=0}^{n-1} \frac{\ell}{n-\ell} \text{Pr}(|\mathcal{S}| = \ell \big| |\mathcal{I}_{\text{uc}}| = n)$$
$$= \frac{\sigma^2}{1-\bar{P}} \left(\bar{P}\gamma_1 - \gamma_2 \right) \tag{16}$$

where

$$\gamma_1 = \sum_{n=1}^{N-1} \text{Pr}(|\mathcal{I}_{\text{uc}}| = n)$$
$$= 1 - \text{Pr}(|\mathcal{I}_{\text{uc}}| = 0) = 1 - (1 - P_{\text{uc}})^{N-1}$$
$$\gamma_2 = \sum_{n=1}^{N-1} \text{Pr}(|\mathcal{I}_{\text{uc}}| = n)\bar{P}^n$$
$$= \sum_{n=0}^{N-1} \text{Pr}(|\mathcal{I}_{\text{uc}}| = n)\bar{P}^n - \text{Pr}(|\mathcal{I}_{\text{uc}}| = 0)$$
$$= (\bar{P}P_{\text{uc}} + 1 - P_{\text{uc}})^{N-1} - (1 - P_{\text{uc}})^{N-1}.$$

Inserting the two above expressions in eq. (16) provides the desired result.

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

[2] R. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, 2007.

[3] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

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

[5] M. Unser, "Sampling-50 years after shannon," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, 2000.

[6] Y. Wang, G. Leus, and A. Pandharipande, "Direction estimation using compressive sampling array processing," in *Statistical Signal Processing, 2009. SSP'09. IEEE/SP 15th Workshop on.* IEEE, 2009, pp. 626–629.

[7] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[8] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proceedings of the 5th international conference on Information processing in sensor networks.* ACM, 2006, pp. 134–142.

[9] K. Hayashi, M. Nagahara, and T. Tanaka, "A user's guide to compressed sensing for communications systems," *IEICE transactions on communications*, vol. 96, no. 3, pp. 685–712, 2013.

[10] V. Stankovic, L. Stankovic, and S. Cheng, "Compressive video sampling," in *Signal Processing Conference, 2008 16th European.* IEEE, 2008, pp. 1–5.

[11] Y. Yu, A. P. Petropulu, and H. V. Poor, "Mimo radar using compressive sampling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 1, pp. 146–163, 2010.

[12] M. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2275–2284, June 2009.

[13] M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Spatial compressive sensing for mimo radar," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 419–430, 2014.

[14] E. J. Candes and M. A. Davenport, "How well can we estimate a sparse vector?" *Applied and Computational Harmonic Analysis*, vol. 34, no. 2, pp. 317 – 323, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520312001388

[15] H. L. Van Trees and K. L. Bell, "Bayesian bounds for parameter estimation and nonlinear filtering/tracking," *AMC*, vol. 10, p. 12, 2007.

[16] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Pearson/Prentice Hall Upper Saddle River, NJ, 2005.

[17] 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, 2012.

[18] 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, 2009.

[19] R. Boyer, B. Babadi, N. Kalouptsidis, and V. Tarokh, "Corrections to Asymptotic achievability of the Cramér–Rao bound for noisy compressive sampling," *IEEE Transactions on Signal Processing*, 2017.

[20] Z. Ben-Haim and Y. Eldar, "The Cramér-Rao bound for estimating a sparse parameter vector," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3384–3389, June 2010.

[21] A. Wiesel, Y. Eldar, and A. Yeredor, "Linear regression with gaussian model uncertainty: Algorithms and bounds," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2194–2205, June 2008.

[22] M. N. El Korso, R. Boyer, P. Larzabal, and B.-H. Fleury, "Estimation performance for the Bayesian hierarchical linear model," *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 488–492, 2016.

[23] R. Prasad and C. R. Murthy, "Cramér-Rao-type bounds for sparse Bayesian learning," *IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 622–632, 2013.

[24] R. Boyer, R. Couillet, B. H. Fleury, and P. Larzabal, "Large-system estimation performance in noisy compressed sensing with random support of known cardinality - a Bayesian analysis," *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5525–5535, 2016.

[25] M. Sundin, S. Chatterjee, and M. Jansson, "Combined modeling of sparse and dense noise improves Bayesian RVM," in *22nd European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, 2014, pp. 1841–1845.

[26] J. N. Laska, M. A. Davenport, and R. G. Baraniuk, "Exact signal recovery from sparsely corrupted measurements through the pursuit of justice," in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers.* IEEE, 2009, pp. 1556–1560.

[27] J. N. Laska, P. T. Boufounos, M. A. Davenport, and R. G. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 31, no. 3, pp. 429–443, 2011.

[28] V. Ollier, R. Boyer, M. N. El Korso, and P. Larzabal, "Bayesian lower bounds for dense or sparse (outlier) noise in the RMT framework," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, invited article, 2016.

[29] Q. Wan, J. Fang, H. Duan, H. Li, and Z. Xing, "Robust bayesian compressed sensing with outliers," *Signal Processing*, 2017.

[30] E. J. Candes, J. K. 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, 2006.

[31] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 310–316, 2010.

[32] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, "Recovery of sparsely corrupted signals," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3115–3130, 2012.

[33] R. Boyer, P. Larzabal, and B. Fleury, "Oracle performance estimation of Bernoulli-distributed sparse vectors," in *2016 IEEE International Workshop on Statistical Signal Processing*, Palma de Majorque, Spain, Jun. 2016. [Online]. Available: https://hal-supelec.archives-ouvertes.fr/hal-01313460

[34] N. L. Johnson, A. W. Kemp, and S. Kotz, *Univariate discrete distributions*. John Wiley & Sons, 2005, vol. 444.

[35] T. K. Moon and W. C. Stirling, *Mathematical methods and algorithms for signal processing*. Prentice hall New York, 2000, vol. 1.

[36] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Springer, 2013, vol. 1, no. 3.

[37] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589–592, 2008.

[38] M. Davenport, P. Boufounos, M. Wakin, and R. Baraniuk, "Signal processing with compressive measurements," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 445–460, April 2010.

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

[40] M. A. Davenport, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "A simple proof that random matrices are democratic," *arXiv preprint arXiv:0911.0736*, 2009.

[41] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 87–90, 2001.

[42] Z. Ben-Haim and Y. C. Eldar, "On the constrained Cramér-Rao bound with a singular Fisher Information Matrix," *IEEE Signal Processing Letters*, vol. 16, no. 6, pp. 453–456, 2009.

[43] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1285–1301, 1990.

[44] P. Stoica and B. C. Ng, "On the Cramér-Rao bound under parametric constraints," *Signal Processing Letters, IEEE*, vol. 5, no. 7, pp. 177–179, 1998.

[45] Z. Ben-Haim and Y. Eldar, "A lower bound on the bayesian mse based on the optimal bias function," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5179–5196, 2009.

[46] J. H. Goulart, M. Boizard, R. Boyer, G. Favier, and P. Comon, "Tensor CP decomposition with structured factor matrices: Algorithms and Performance," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 757–769, 2016.

[47] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*, 1st ed. New York, NY, USA: Cambridge University Press, 2011.

[48] A. Gelman, "Prior distributions for variance parameters in hierarchical models," *Bayesian analysis*, vol. 1, no. 3, pp. 515–534, 2006.

[49] J. Li and P. Stoica, "Mimo radar signal processing."

[50] R. Boyer and G. Bouleux, "Oblique projections for direction-of-arrival estimation with prior knowledge," *IEEE Trans. Signal Processing*, vol. 56, no. 4, pp. 1374–1387, 2008.

[51] G. Bouleux, P. Stoica, and R. Boyer, "An optimal prior-knowledge-based DOA estimation method," in *17th European Signal Processing Conf.*, Glasgow, United Kingdom, 2009, pp. 869–873.

[52] A.-J. van der Veen and S. J. Wijnholds, "Signal processing tools for radio astronomy," in *Handbook of Signal Processing Systems*. Springer, 2013, pp. 421–463.

[53] V. Ollier, M. N. El Korso, R. Boyer, P. Larzabal, and M. Pesavento, "Relaxed concentrated MLE for robust calibration of radio interferometers," in *EUSIPCO*, Budapest, Hungary, Aug. 2016. [Online]. Available: https://hal-centralesupelec.archives-ouvertes.fr/hal-01335481

[54] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[55] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[56] J. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[57] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Conference Record of The Twenty-Seventh Asilomar Conference*. IEEE, 1993, pp. 40–44.

[58] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[59] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.