Signal Processing 143 (2018) 86–93

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Sparsity-based estimation bounds with corrupted measurements

Rémy Boyer[a,1,*], Pascal Larzabal[b]

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

ARTICLE INFO

Article history:
Received 28 February 2017
Revised 23 June 2017
Accepted 4 August 2017
Available online 16 August 2017

Keywords:
Compressed sensing
Corrupted measurements
Cramér–Rao Bound
Statistical priors for support sets of random cardinalities
Gaussian measurement matrix

ABSTRACT

In typical Compressed Sensing operational contexts, the measurement vector **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 $\mathcal{I}_{\rm uc}$ containing the indexes of the uncorrupted measures. As $\mathcal{I}_{\rm uc}$ and its cardinality $|\mathcal{I}_{\rm uc}| < N$ are unknown, each sample of vector **y** follows an i.i.d. Bernoulli prior of probability $P_{\rm uc}$, 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 $|\mathcal{S}|||\mathcal{I}_{\rm uc}|$ follows a hierarchical Binomial distribution on set $\{1,\dots,|\mathcal{I}_{\rm uc}|-1\}$. The derived lower bounds do not belong to the family of "oracle" or "genie-aided" bounds since our *a* priori knowledge on support $\mathcal{I}_{\rm uc}$ and its cardinality is limited to probability $P_{\rm uc}$. In this context, very compact and simple expressions of the Expected Cramér–Rao Bound (ECRB) are proposed. Finally, the proposed lower bounds are compared to standard estimation strategies robust to an impulsive (sparse) noise.

1. Introduction

In the Compressed Sensing (CS) framework [1–3], 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 [11–13].

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 [17–20] and for the Bayesian linear model in [21–24].

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 [25–29]. In this work, our aim is to study the estimation performance limit in presence of corrupted measurements. CS with corrupted measurements [30–32] 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 $\mathcal{I}_{\rm uc}$, constituted by the uncorrupted measurements and its cardinality, denoted by $|\mathcal{I}_{\rm uc}|$, are unknown. So, to take into account this uncertain knowledge, the support $\mathcal{I}_{\rm uc}$ is modelized as a collection of i.i.d. Bernoulli-distributed random variables with a probability $1-P_{\rm uc}$ 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 $P_{\rm uc}$ is assumed to be *a* priori known. As a consequence, the unknown cardinality $|\mathcal{I}_{\rm uc}|$ follows a Binomial prior in set $\{1,\dots,N-1\}$. 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 constraint, we must have $|\mathcal{S}| < |\mathcal{I}_{\rm uc}|$, meaning that $|\mathcal{S}|||\mathcal{I}_{\rm uc}|$ follows a hierarchical Binomial distribution confined in the set $\{1,\dots,|\mathcal{I}_{\rm uc}|-1\}$. This work proposes several new contributions regarding the state of art on the lower bounds

* Corresponding author.
E-mail addresses: remy.boyer@l2s.centralesupelec.fr, remy.boyer@lss.supelec.fr (R. Boyer), pascal.larzabal@satie.ens-cachan.fr (P. Larzabal).

[1] Remy Boyer and Pascal Larzabal are members of the EURASIP SAT SPMuS. This work was partly supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM# (Grant agreement no. 318306). Rémy Boyer and Pascal Larzabal are involved in projects MAGELLAN (ANR-14-CE23-0004-01).

http://dx.doi.org/10.1016/j.sigpro.2017.08.004
0165-1684/

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 [21–23], 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 Cramér–Rao Bound (ECRB) based on a complete measurement vector scenario meaning that $P_{\rm uc}=1$. 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.

Notations: 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. Bernou(P) stands for the Bernoulli distribution of probability of success P. Binomial(N, P) stands for the Binomial distribution in $\{0,\dots,N\}$ 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 [1–3]:

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

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 (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 (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 $\mathbf{H}_{\mathcal{S}}=\boldsymbol{\Psi}\boldsymbol{\Phi}_{\mathcal{S}}$ and the $N\times|\mathcal{S}|$ matrix $\boldsymbol{\Phi}_{\mathcal{S}}$ is built up with the $|\mathcal{S}|$ columns of $\boldsymbol{\Phi}$ 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 $\boldsymbol{H}$

Determining whether the dictionary $\mathbf{H}=\boldsymbol{\Psi}\boldsymbol{\Phi}$ satisfies the concentration inequality is combinatorially complex but the so-called universal design strategy has been introduced for instance in [2,3]. Assume that matrix $\boldsymbol{\Phi}$ is an orthonormal basis and the measurement matrix $\boldsymbol{\Psi}$ 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

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

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

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

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 $\mathbf{P}$

Definition 2.1 (Guaranty on the non-singularity of the Fisher information)**.** Define the deterministic set $\mathcal{I}\subset\{1,\dots,K\}$ of cardinality $|\mathcal{I}|=N-1<K$. Given $|\mathcal{I}|$ available measurements, the Fisher information associated to model of (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 [24,41–44].

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

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

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

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

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

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

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

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