Signal Processing 139 (2017) 70–74

Contents lists available at ScienceDirect

Signal Processing

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

Short communication

Sparse-based estimation performance for partially known overcomplete large-systems

Guillaume Bouleux[a,*], Rémy Boyer[b]

[a] Univ Lyon, INSA-Lyon, UJM-Saint-Etienne, DISP, EA 4570, Villeurbanne 69621, France
[b] L2S laboratory University of Paris-Sud, CentraleSupelec, CNRS, France

ARTICLE INFO

Article history:
Received 4 July 2016
Revised 5 April 2017
Accepted 8 April 2017
Available online 13 April 2017

Keywords:
Overcomplete Bayesian linear model
Asymptotic estimation performance
Subspace prior-knowledge
Large-systems

ABSTRACT

We assume the direct sum $\langle\mathbf{A}\rangle\oplus\langle\mathbf{B}\rangle$ for the signal subspace. As a result of post-measurement, a number of operational contexts presuppose the a priori knowledge of the L_B-dimensional "interfering" subspace $\langle\mathbf{B}\rangle$ and the goal is to estimate the L_A amplitudes corresponding to subspace $\langle\mathbf{A}\rangle$. Taking into account the knowledge of the orthogonal "interfering" subspace $\langle\mathbf{B}\rangle\perp$, the Bayesian estimation lower bound is derived for the L_A-sparse vector in the doubly asymptotic scenario, i.e. N, L_A, $L_B \to \infty$ with a finite asymptotic ratio. By jointly exploiting the Compressed Sensing (CS) and the Random Matrix Theory (RMT) frameworks, closed-form expressions for the lower bound on the estimation of the non-zero entries of a sparse vector of interest are derived and studied. The derived closed-form expressions enjoy several interesting features: (i) a simple interpretable expression, (ii) a very low computational cost especially in the doubly asymptotic scenario, (iii) an accurate prediction of the mean-square-error (MSE) of popular sparse-based estimators and (iv) the lower bound remains true for any amplitudes vector priors. Finally, several idealized scenarios are compared to the derived bound for a common output signal-to-noise-ratio (SNR) which shows the interest of the joint estimation/rejection methodology derived herein.

1. Introduction

The Compressive Sampling or Compressed Sensing (CS) is an attractive domain which gives new trends for people interested in sampling theory of sparse signals [1–3]. The CS theory states that a sparse signal, *i.e.*, a signal that can be decomposed as few non-zero values in a given basis (Fourier, wavelets, etc.) can be sampled at a rate T_S lower than the one predicted by the Shannon's theory. This paradigm has been successfully exploited for solving ill-posed problems arising for instance in bio-medical analysis, RADAR detection, array processing, wireless communications and radioastronomy imaging. In the CS framework, it is well known that any matrix $\mathbf{H}$ of size $N \times L$ generated from an i.i.d. centered sub-Gaussian distribution with a variance of $1/N$ verifies the Restricted Isometry Property (RIP) [2] with a high probability [1]. On the other hand, the doubly asymptotic spectrum and the empirical moments of the product $\mathbf{H}^T\mathbf{H}$ have been extensively studied in the context of the Random Matrix Theory (RMT) [4].

In the literature, CS and RMT techniques are usually applied to the noisy linear model where there is no interfering signals. However, in a wide range of real life applications, the signal of interest is often corrupted by a partially known interfering signal and an additive noise (see [5–9] for instance). This context motivates this work. More specifically, the CS and the RMT frameworks will be associated to derive new analytical closed-form expressions for the Bayesian lower bound [10] on the estimation of a sparse amplitude vector [11] for the noisy linear model corrupted by a partially known interfering signal.

2. Compressed Sensing (CS) integrating an *a priori* knowledge

2.1. Definition of the CS model

Let $\mathbf{y}$ an observed vector of N measurements corrupted by an additive white centered zero-mean, Gaussian circular noise vector of variance σ^2. The standard CS model [1–3] is defined according to

$$\mathbf{y} = \mathbf{\Psi s} + \mathbf{n} = \mathbf{\Psi\Phi x} + \mathbf{n} \tag{1}$$

where $\mathbf{\Psi}$ is the known measurement matrix of size $N \times K$ with $N < K$, the vector $\mathbf{s} = \mathbf{\Phi x}$ of size $K \times 1$ admits an L-sparse representation, denoted by $\mathbf{x}$, in the basis $\mathbf{\Phi}$ (which could be Fourier basis, Wavelets basis, canonical basis, etc.) with $L < N$ and where $\mathbf{H} \overset{\text{def.}}{=} \mathbf{\Psi\Phi}$ is often called the overcomplete dictionary.

* Corresponding author.
E-mail addresses: guillaume.bouleux@insa-lyon.fr, guillaume.bouleux@gmail.com (G. Bouleux), remy.boyer@l2s.centralesupelec.fr (R. Boyer).

http://dx.doi.org/10.1016/j.sigpro.2017.04.010
0165-1684/

One of the main problems risen up by the theory of the Compressed Sensing relates to the minimum number of measurements N needed for retrieving the L-sparse vector $\mathbf{x}$. To address this problem, the authors of [1–3] have defined the Restricted Isometry Property (RIP). A standard strategy, called universal design strategy to ensure that dictionary $\mathbf{H}$ satisfies the RIP condition with high probability, is to generate the i.i.d. entries of dictionary matrix $\mathbf{H}$ following a sub-Gaussian distribution with zero mean and variance $1/N$ [1].

2.2. Exploiting the "interfering" subspace knowledge

In many real life applications, we do have the knowledge of information given by the physics of the context. Those useful information help in tailoring models that precisely take into account the knowledge of particular frequencies [12] for spectral analysis purpose, spatial angles for array processing [5] or RADAR processing, and have demonstrated their power through biomedical analysis or radioastronomy imaging. So, we adopt the following "signal+interference" model $\mathbf{s} = \mathbf{A}\boldsymbol{\alpha} + \mathbf{i}$ with $\mathbf{i} = \mathbf{B}\boldsymbol{\beta}$ where $[\mathbf{A}]_{k,\ell} = g(kT_S - \tau_\ell)$ with $1 \le \ell \le L_A$, $[\mathbf{B}]_{k,\ell'} = g(kT_S - \tilde{\tau}_{\ell'})$ with $1 \le {}'\ell \le L_B$ are the "steering matrices" parametrized by the regular discretization at rate T_S of a known waveform $g(t)$ along the time space. More precisely, $\mathcal{T} = \{\tau_\ell, 1 \le \ell \le L_A\}$ stands for the time-delays of the L_A sources of interest $\boldsymbol{\alpha}$ and $\tilde{\mathcal{T}} = \{\tilde{\tau}_\ell, 1 \le \ell \le L_B\}$ is associated to the L_B interfering sources $\boldsymbol{\beta}$. In the sequel, it is assumed that (i) $\langle\mathbf{A}\rangle$ and $\langle\mathbf{B}\rangle$ are two disjoint subspaces, meaning that there is no time overlapping between the sources of interest and of the interfering sources and (ii) $\langle\mathbf{B}\rangle$ is known or previously estimated (matrix $\mathbf{A}$ and $\langle\mathbf{A}\rangle$ are unknown). For instance, the learning of $\langle\mathbf{B}\rangle$ is based on pre-estimation of the clutter echo time-delays in RADAR processing or by known strongly shining "calibrator stars" in radioastronomy imaging. The problem of interest is to estimate vector $\boldsymbol{\alpha}$ based on a measurement vector where the contribution of $\mathbf{i}$ has been removed using the knowledge of $\langle\mathbf{B}\rangle$. The standard "signal+interference" model described by signal $\mathbf{s}$ can be extended in the CS framework of model (1) following a straightforward strategy. Let $\boldsymbol{\Phi}$ be a basis matrix such as $[\boldsymbol{\Phi}]_{k,k'} = g((k-k')T_S)$ where $1 \le k, k' \le K$. For a sufficiently fine partition, *i.e.*, for $K > N > L = L_A + L_B$, we have $\langle\mathbf{A}\rangle \oplus \langle\mathbf{B}\rangle \subset \langle\boldsymbol{\Phi}\rangle$. Let $\mathbf{U}_{\mathbf{B}^\psi}$ be a $N \times (N - L_B)$ orthonormal basis matrix such as $\langle\mathbf{U}_{\mathbf{B}^\psi}\rangle = \langle\mathbf{B}^\psi\rangle^\perp$. We have finally the deflated observation, defined according to

$$\bar{\mathbf{y}} = \mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{y} = \mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{H}\mathbf{x} + \bar{\mathbf{n}} \tag{2}$$

where $\bar{\mathbf{n}} = \mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{n}$ and $\mathbf{x}$ is a $(K - L_A)$-sparse such as $\mathbf{x}_{\mathcal{T}} = \boldsymbol{\alpha}$. The reader will find an illustration of the procedure in Fig. 1.

3. ECRB for projected measurements and a large random dictionary

3.1. Dealing with projected measurements

Let $\text{MSE} = \frac{1}{L_A}\mathbb{E}_{\bar{\mathbf{y}},\boldsymbol{\alpha}}[\|\hat{\boldsymbol{\alpha}}(\bar{\mathbf{y}}) - \boldsymbol{\alpha}\|^2]$ be the normalized Bayesian Mean Squared Error for an estimate $\hat{\boldsymbol{\alpha}}(\bar{\mathbf{y}})$ of $\boldsymbol{\alpha}$. The Expected Cramér-Rao Bound (ECRB) [10], denoted by $C_{\mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{A}^\psi}$ for the random amplitude vector $\boldsymbol{\alpha}$, of unspecified distribution $p(\boldsymbol{\alpha})$ given the observation model (2) fulfills relation $\text{MSE} \ge C_{\mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{A}^\psi} = \frac{\sigma^2}{L_A}\,\text{Tr}\{(\mathbf{A}^{\psi T}\mathbf{P}_{\mathbf{B}^\psi}^\perp \mathbf{A}^\psi)^{-1}\}$ where $\mathbf{A}^\psi = \boldsymbol{\Psi}\mathbf{A}$. Introduce model $(\mathcal{M})$: $\bar{\mathbf{y}}|\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = \mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{A}^\psi \boldsymbol{\alpha}$ and $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_{N-L_B}$ which is the covariance matrix of noise $\bar{\mathbf{n}}$. After some calculus, the ECRB admits the following expression:

$$C_{\mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{A}^\psi} = \frac{\sigma_\alpha^2}{\text{SNR}^{(\text{na})}} \frac{\text{Tr}\{\mathbf{A}^{\psi T}\mathbf{P}_{\mathbf{B}^\psi}^\perp \mathbf{A}^\psi\}}{N - L_B} \frac{\text{Tr}\left\{\left(\mathbf{A}^{\psi T}\mathbf{P}_{\mathbf{B}^\psi}^\perp \mathbf{A}^\psi\right)^{-1}\right\}}{L_A}, \tag{3}$$

where $\text{SNR}^{(\text{na})} = \frac{\mathbb{E}\|\boldsymbol{\mu}\|^2}{\text{Tr}\{\boldsymbol{\Sigma}\}} = \frac{\sigma_\alpha^2 \text{Tr}\{\mathbf{A}^{\psi T}\mathbf{P}_{\mathbf{B}^\psi}^\perp \mathbf{A}^\psi\}}{\sigma^2 (N - L_B)}$ is the output and non-asymptotic SNR.

3.2. Doubly asymptotic regime

The practical interest of CRB-type expressions have been exposed in [13,14] but we show in this work that expression (3) can be reduced to a very simple closed form expression with the advantage of remaining valid even for the low sample regime, using some powerful results extracted from the RMT [4] where it is assumed N, L_A, $L_B \to \infty$ with $N/L_A \to \rho$ and $L_B/L_A \to c$. Towards this goal, the following Lemma is provided.

Lemma 1. *Let $\mathbf{F} = \mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{A}^\psi \in \mathbb{R}^{(N-L_B)\times L_A}$ whose elements $\{F_{ij}\}_{i,j=1\ldots N-L_B,L_A}$ are zero mean and i.i.d. with variance $\frac{1}{N}$. Now, for N, L_A, $L_B \to \infty$, and $N/L_A \to \rho > 1$, $(N - L_B)/L_A \to \tilde{\rho} = \rho - c > 1$, then*

$$\frac{1}{L_A}\text{Tr}\left\{\left(\mathbf{F}^T\mathbf{F}\right)^{-1}\right\} \xrightarrow{a.s.} \frac{\rho}{\tilde{\rho} - 1} = \frac{\rho}{\rho - c - 1}, \tag{4}$$

$$\frac{1}{N - L_B}\text{Tr}\left\{\mathbf{F}^T\mathbf{F}\right\} \xrightarrow{a.s.} \frac{1}{\rho}, \tag{5}$$

where a.s. stands for the almost sure convergence.

Proof. See the appendix. □

Under the assumptions of Lemma 1 and using (3), a very compact expression of $C^\infty_{\mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{A}^\psi}$ is enunciated by the following.

Result 1. Assume that N, L_A, $L_B \to \infty$ and $N/L_A \to \rho > 1$, $(N - L_B)/L_A \to \tilde{\rho} > 1$, then, we have $C_{\mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{A}^\psi} \xrightarrow{a.s.} C^\infty_{\mathbf{U}_{\mathbf{B}^\psi}^T \mathbf{A}^\psi} = \frac{\sigma_\alpha^2}{\text{SNR}} \frac{1}{\tilde{\rho} - 1}$ where $\text{SNR} = \frac{\sigma_\alpha^2}{\sigma^2 \rho}$ is the almost sure doubly asymptotic equivalent of $\text{SNR}^{(\text{na})}$.

4. Benchmarking ECRBs and estimators

This section is devoted to give a relation of order between the ECRB given by (3) with respect to two other ECRBs viewed as benchmarks and to analyze the behavior of sparse-based estimators. Let $(\mathcal{M}_0) : \mathbf{y}_0|\boldsymbol{\alpha}, \boldsymbol{\beta} \sim \mathcal{N}\left(\mathbf{A}^\psi \boldsymbol{\alpha} + \mathbf{B}^\psi \boldsymbol{\beta}, \sigma_0^2 \mathbf{I}_N\right)$ and $(\mathcal{M}_1) : \mathbf{y}_1|\boldsymbol{\alpha} \sim \mathcal{N}\left(\mathbf{A}^\psi \boldsymbol{\alpha}, \sigma_1^2 \mathbf{I}_N\right)$. Model $\mathcal{M}_0$ is associated with the scenario where no ad-hoc strategy is developed to mitigate the corruption from the interference signals. In other words, the interference signals are wrongly interpreted as signals of interest. So, this bound does not solve the problem of interest and is given by

$$C_{[\mathbf{A}^\psi \mathbf{B}^\psi]} = \frac{\sigma_0^2}{L}\ \text{Tr}\left\{\left([\mathbf{A}^\psi \mathbf{B}^\psi]^T [\mathbf{A}^\psi \mathbf{B}^\psi]\right)^{-1}\right\} \tag{6}$$

$$= \frac{\sigma_\alpha^2}{\text{SNR}_0^{(\text{na})}} \frac{\text{Tr}\left\{\left([\mathbf{A}^\psi \mathbf{B}^\psi]^T [\mathbf{A}^\psi \mathbf{B}^\psi]\right)^{-1}\right\}}{L} \cdot \frac{\sigma_\alpha^2\ \text{Tr}\left\{\left(\mathbf{A}^{\psi T}\mathbf{A}^\psi\right)\right\} + \sigma_\beta^2\ \text{Tr}\left\{\left(\mathbf{B}^{\psi T}\mathbf{B}^\psi\right)\right\}}{N}, \tag{7}$$

where $\text{SNR}_0^{(\text{na})} = \frac{\sigma_\alpha^2\ \text{Tr}\{(\mathbf{A}^{\psi T}\mathbf{A}^\psi)\} + \sigma_\beta^2\ \text{Tr}\{(\mathbf{B}^{\psi T}\mathbf{B}^\psi)\}}{\sigma_0^2 N}$ and $\text{SIR} = \sigma_\alpha^2/\sigma_\beta^2$. The second model $\mathcal{M}_1$ is associated with the ideal free-interference scenario. This bound admits the following expression:

$$C_{\mathbf{A}^\psi} = \frac{\sigma_1^2}{L_A}\text{Tr}\left\{\left(\mathbf{A}^{\psi T}\mathbf{A}^\psi\right)^{-1}\right\} = \frac{\sigma_\alpha^2}{\text{SNR}_1^{(\text{na})}} \frac{\text{Tr}\left\{\left(\mathbf{A}^{\psi T}\mathbf{A}^\psi\right)\right\}}{N} \frac{\text{Tr}\left\{\left(\mathbf{A}^{\psi T}\mathbf{A}^\psi\right)^{-1}\right\}}{L_A}, \tag{8}$$