Int. J. Electron. Commun. (AEÜ) 69 (2015) 1354–1365

Contents lists available at ScienceDirect

International Journal of Electronics and Communications (AEÜ)

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

Information-theoretic analysis of support recovery from sparsely corrupted measurements

Wenbo Xu*, Zhilin Li, Yun Tian, Jiaru Lin

Key Lab of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China

ARTICLE INFO

Article history:
Received 21 January 2015
Accepted 29 May 2015

Keywords:
Compressed sensing
Information-theoretic bound
Sparse error
Support recovery

ABSTRACT

This paper studies the recovery of the support of sparse signal that is corrupted by both dense noise and gross error. The gross error is an unknown sparse vector whose nonzero entries maybe unbounded. This setup covers a wide range of applications, such as face recognition, inpainting and sensor networks. We derive the information-theoretic lower bounds on the sampling rate required to obtain a desired error rate, which depend on the properties of both the signal and the gross error. The investigations are given in the high-dimensional setting. Some illustrations are provided to further reveal the relationship of these bounds.

1. Introduction

Recent theory of compressed sensing (CS) states that a K-sparse signal $\mathbf{x} \in \mathbb{R}^N$ can be represented by its fewer measurements in the form of $\mathbf{y} = A\mathbf{x}$, where $\mathbf{y} \in \mathbb{R}^M$ is the measurement vector, and $A \in \mathbb{R}^{M \times N}$ is the measurement matrix ($K < M < N$) [1]. It has been one of the fastest growing research areas in signal processing, and has been studied in many application fields, e.g., imaging, channel estimation, radar, face recognition [2–6].

Define the support as the index set of the nonzero entries in $\mathbf{x}$, and one fundamental problem in CS is to recover the support of $\mathbf{x}$ from $\mathbf{y}$ based on the prior information of signal sparsity. Many papers, e.g., [7–14], have provided the theoretical analysis of support recovery when the measurements are corrupted by dense noise such that $\mathbf{y} = A\mathbf{x} + \mathbf{w}$, where $\mathbf{w} \in \mathbb{R}^M$ is the noise vector. [7–10] studied the sufficient and necessary conditions, [11,12] obtained the bounds on the tradeoff between the sampling rate and the fraction of detection errors, where the sampling rate stands for the ratio M/N, and we derived in [13,14] the probability of exact support recovery and partial support recovery, respectively.

In many practical applications, such as face recognition, inpainting and sensor networks, one needs to solve the problem of recovering the signal support from $\mathbf{y} = A\mathbf{x} + \mathbf{e} + \mathbf{w}$, where the measurements are also corrupted by a sparse error $\mathbf{e}$ [5,6]. $\mathbf{e}$ is usually called the gross error and its entries can have arbitrarily large magnitude. This problem and its variants have gained increasing attentions recently in terms of theoretical analysis [15–26]. [15,16] investigated the sparse signal recovery from $\mathbf{y} = A\mathbf{x} + \mathbf{e}$, and provided the recovery guarantees when $\mathbf{x}$ is recovered with high probability. [17–19] derived the deterministic recovery guarantees based on $\mathbf{y} = A\mathbf{x} + B\mathbf{e}$ where $\mathbf{x}$ and $\mathbf{e}$ are both perfectly sparse. [20] provided probabilistic recovery guarantees that improve upon the ones in [17–19], where varying degrees of knowledge of the signal support and sparse error support are available. Later in [21], Studer and Baraniuk extended the results in [17–19] to the case of approximately sparse signals and noisy measurements. To investigate the recovery from $\mathbf{y} = A\mathbf{x} + B\mathbf{e}$, [22] derived a general achievability bound for the compression rate in separating signals $\mathbf{x}$ and $\mathbf{e}$. In addition, [23,24] showed that the combination matrix $[A, B]$ obeys the restricted isometry property when A is a Gaussian distributed matrix, where B is an identity matrix in [24]. Nguyen et al. studied the problem of recovering sparse signal $\mathbf{x}$ from $\mathbf{y} = A\mathbf{x} + \mathbf{e} + \mathbf{w}$ [25,26]. [25] derived the recovery guarantees for random sub-sampled unitary matrices, and [26] proposed an extended optimization method and discussed its corresponding performance. In these existing works, [17–21] considered deterministic matrices A, [15,16,23–26] considered random matrices A, and [22] dealt with almost all matrices A.

This paper considers the model $\mathbf{y} = A\mathbf{x} + \mathbf{e} + \mathbf{w}$ with random matrix A. Though some papers have studied the related model [20–22] or even the same model [25,26], they mainly focused on the recovery guarantees or the performance of some reconstruction algorithms. By contrast, we discuss the relationship between the sampling rate and the recovery error. Specifically, we give the information-theoretic lower bounds on the sampling rate $\rho = M/N$ (i.e., necessary conditions) required to obtain a desired recovery error rate D. These bounds are given in terms of both the properties of the signal and the properties of the gross error. Our results

* Corresponding author. Tel.: +86 13810019644.
E-mail address: xuwb@bupt.edu.cn (W. Xu).

http://dx.doi.org/10.1016/j.aeue.2015.05.022
1434-8411/

are derived in the high-dimensional setting where the length N approaches to infinite, but with the sparsity rate and the signal-to-noise ratio (SNR) being finite constants. Here, sparsity rate means the fraction of nonzero entries in a sparse signal.

A summary of our contributions is as follows:

(i) Information-theoretic limits: we derive the lower bounds on the sampling rate required for any possible recovery algorithm when both dense noise and sparse gross error exist. The bounds, which are formulated differently for Gaussian distributed gross error and arbitrarily distributed gross error, are obtained by using the properties of mutual information, different entropy and entropy power. The results generalize those in [11], where the gross error is not considered.

(ii) Insights at high SNR: at high SNR, we show that the lower bound of the required sampling rate is dominated by the sum of sparsity rates of the signal and the gross error. In other words, if we regard the combination signal $\mathbf{x}_c = [\mathbf{x}\,;\mathbf{e}]$ as the integrated sparse signal, the support recovery requires the sampling rate to grow at least linearly with the sparsity rate of $\mathbf{x}_c$.

(iii) Characterization of the impact of gross error: in addition to showing the relationship between the required sampling rate and the properties of the signal itself, we reveal the dependence of the lower bounds on sparsity rate of gross error and the entropy power of its nonzero entries. Using some illustrations, we further demonstrate the impact of gross error in recovering the signal support.

The rest of the paper is organized as follows. Section 2 states the signal model and distortion measure. Section 3 develops the main results of this paper, and discusses the connections to previous work. Some illustrations for the derived bounds are provided in Section 4. Section 5 concludes this paper. Finally, some mathematical proofs are provided in Appendices.

Notations: A random variable and a random vector are denoted as X and $\mathbf{X}$ in the uppercase letters, and their realizations are denoted as x and $\mathbf{x}$ in the lowercase letters, respectively. For a random matrix $\mathbf{A}$, its realization is denoted A. For a set $L \subseteq \{1, 2, \ldots, N\}$, the cardinality is denoted by $|L|$, and L^c means its complement. For a vector $\mathbf{x}$, $\mathbf{x}_i$ is the ith entry, and $\mathbf{x}_S$ stands for the $|S|$-dimensional vector whose entries are selected from the index set S of $\mathbf{x}$. For a given $M \times N$ matrix A, its transpose is denoted by A^T, its Moore-Penrose inverse is $A^{\dagger}$, and its determinant is denoted by $|A|$. The matrix $A_L \in \mathbb{R}^{|L| \times N}$ is obtained from A by retaining the rows of A with indices in L, the matrix $A_{*,S} \in \mathbb{R}^{M \times |S|}$ is obtained by retaining the columns of A with indices in S, and the matrix $A_{L,S} \in \mathbb{R}^{|L| \times |S|}$ is obtained by retaining the rows and columns of A with indices in L and S, respectively. We use $\mathbf{I}_{n\times n}$ to refer to the identity matrix, and $\mathbf{0}_{M\times 1}$ is the $M \times 1$ vector with all entries being zero. The natural logarithm is referred to as log.

2. Problem statement

2.1. Signal model

Consider a K-sparse signal $\mathbf{X} \in \mathbb{R}^N$, and its noisy measurements in the form of

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{E} + \frac{1}{\sqrt{\mathrm{snr}}}\mathbf{W} \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{M\times N}$ is the random measurement matrix, $\mathbf{E} \in \mathbb{R}^M$ is the τ-sparse error, $\mathbf{W} \sim \mathcal{N}(0, \mathbf{I}_{M\times M})$ is the additive white Gaussian noise, and $\mathrm{snr} \in (0, \infty)$. Assume the vector $\mathbf{X}$, the matrix $\mathbf{A}$, the error $\mathbf{E}$, and the noise $\mathbf{W}$ are mutually independent. The supports of $\mathbf{X}$ and $\mathbf{E}$ are denoted as S and L, respectively, where the support is the index set of nonzero entries such that

$$S = \{i \in \{1, 2, \ldots, N\} : \mathbf{X}_i \neq 0\} \tag{2}$$

and

$$L = \{i \in \{1, 2, \ldots, N\} : \mathbf{E}_i \neq 0\} \tag{3}$$

Note that $K = |S|$ and $\tau = |L|$.

The problem investigated in this paper is the recovery of the support S in the high-dimensional setting (i.e., $N \to \infty$) when the decoder is given the vector $\mathbf{Y}$ and the matrix $\mathbf{A}$. To characterize the corresponding behavior, we give some assumptions.

Assumption 1. (Assumptions on signal $\mathbf{X}$):

(i) The nonzero positions of S are distributed uniformly within $\{1, 2, \ldots, N\}$ with a known size K, where $\lim_{N\to\infty} \frac{K}{N} = \kappa$ for some sparsity rate $\kappa \in (0, 1/2)$.

(ii) The probability distribution of a variable X is given by

$$p_X = (1-\kappa)\delta_0 + \kappa p_U \tag{4}$$

where δ_0 is a point-mass at $x = 0$, and p_U is the distribution of the nonzero entries in X.

Assumption 2. (Assumptions on measurement matrix $\mathbf{A}$):

(i) The distribution on $\mathbf{A}$ is independent of $\mathbf{X}$, $\mathbf{W}$ and $\mathbf{E}$.

(ii) The number of rows M obeys $\lim_{N\to\infty} \frac{M}{N} = \rho$ for some sampling rate $\rho \in (0, 1]$.[1]

(iii) Assume $E[\|\mathbf{A}\|_F^2] = M$ where $\|\cdot\|_F$ is the Frobenius norm. That is, each row has unit magnitude on average.

Assumption 3. (Assumptions of i.i.d. entries in matrix $\mathbf{A}$): The entries of $\mathbf{A}$ are i.i.d. with mean zero and variance $1/N$.

Assumption 4. (Assumptions on sparse error $\mathbf{E}$):

(i) The size of support L obeys $\lim_{N\to\infty} \frac{\tau}{N} = \varsigma$ for some sparsity rate $\varsigma \in (0, 1/2)$.

(ii) The probability distributions of a variable E and a variable E_L are denoted as p_E and p_{E_L}, respectively. Their definitions are similarly to p_X in Assumption 1.

Based on Assumption 4, if we define $\lim_{N\to\infty} \frac{\tau}{M} = \bar{\varsigma}$, then $\bar{\varsigma} = \varsigma/\rho$. Define the variances of the distributions p_X, p_E and p_{E_L} as V_X, V_E and $\bar{V}_E$, respectively.

2.2. Distortion measure

The quanlity of support recovery is usually measured by two quantities, i.e., missed detection rate (MDR) and false alarm rate (FAR). Define $\hat{S}$ as the recovered support. MDR is given by

$$\mathrm{MDR}(S, \hat{S}) = \frac{1}{|S|}\sum_{i=1}^{N} \mathbf{1}(i \in S, i \notin \hat{S}) \tag{5}$$

[1] In [11], $\rho \in (0, \infty)$, whereas we only consider the case of $\rho \leq 1$ that corresponds to the compressed sensing setting.