

Research Article

A Novel Detection Scheme with Multiple Observations for Sparse Signal Based on Likelihood Ratio Test with Sparse Estimation

Hongbo Zhao,¹ Lei Chen,¹ Wenquan Feng,¹ and Chuan Lei²

¹School of Electronic and Information Engineering, Beihang University, Beijing 100191, China

²China Academy of Electronics and Information Technology, Beijing 100041, China

Correspondence should be addressed to Lei Chen; chenleilei@buaa.edu.cn

Received 4 May 2016; Revised 5 July 2016; Accepted 11 July 2016

Academic Editor: Zhike Peng

Copyright © 2016 Hongbo Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, the problem of detecting unknown and arbitrary sparse signals has attracted much attention from researchers in various fields. However, there remains a peck of difficulties and challenges as the key information is only contained in a small fraction of the signal and due to the absence of prior information. In this paper, we consider a more general and practical scenario of multiple observations with no prior information except for the sparsity of the signal. A new detection scheme referred to as the likelihood ratio test with sparse estimation (LRT-SE) is presented. Under the Neyman-Pearson testing framework, LRT-SE estimates the unknown signal by employing the l_1 -minimization technique from compressive sensing theory. The detection performance of LRT-SE is preliminarily analyzed in terms of error probabilities in finite size and Chernoff consistency in high dimensional condition. The error exponent is introduced to describe the decay rate of the error probability as observations number grows. Finally, these properties of LRT-SE are demonstrated based on the experimental results of synthetic sparse signals and sparse signals from real satellite telemetry data. It could be concluded that the proposed detection scheme performs very close to the optimal detector.

1. Introduction

Nowadays, with network and information boom, there is massive data influx in every walk of life including military detection, industrial monitoring, IT industry, entertainment industry, and many other fields. These big data of high-volume, high-velocity, and high-variety information demand innovative forms of information processing for enhanced insight, decision-making, and signal detection [1]. However, it is increasingly difficult to detect and extract the truly useful portion from massive data.

Plenty of state-of-the-art methods have been put forward to detect targets and signals for big data, for example, utilizing artificial neural networks or deep neural networks to do fault diagnosis with massive data [2, 3]. As the research on the sparsity of signal further develops, it is also a cutting-edge method to employ sparse representation to characterize the

massive data because these data are always full of redundant information. A signal could be regarded as a sparse signal if the coefficient of the vector itself, or under certain basis, contains only a few nonzero components [4].

In this paper, the sparse signal detection problem which is to determine whether a sparse signal exists in a small fraction of the background noise has been discussed. This problem has attracted growing attention as sparse signals are usually closely interrelated to unpredictable changes, abnormality, and danger, for example, fault detection and diagnosis of satellite navigation, network anomaly detection, industrial process supervision, and prediction of natural disasters. It is of great significance to avoid serious consequences induced by these undiscoverable early warnings.

However, there are many challenges to distinguish and separate these sparse signals from noise because useful information takes up only small part of the entries in the

signal. In addition, prior information is almost unavailable due to the diversity and indeterminism of the abnormal sparse signals.

It is a common way to formulate the detection problem into composite hypothesis testing. A classical approach to detect signals with unknown parameter is the generalized likelihood ratio test (GLRT) [5]. The basic idea of GLRT is to replace the unknown parameter with its maximum likelihood (ML) estimates and then continue employing likelihood ratio test (LRT). The test has the form $\Lambda_g(\mathbf{x}) = \max_{\theta_1 \in \Theta_1} P_{r|\theta_1}(\mathbf{x} | \theta_1) / \max_{\theta_0 \in \Theta_0} P_{r|\theta_0}(\mathbf{x} | \theta_0) \underset{H_0}{\overset{H_1}{\geq}} \gamma$, where \mathbf{x} is the observations vector and θ_1 and θ_0 are the parameter vectors to be determined. Frequently, we work with the logarithm $\ln P_{r|\theta_i}(\mathbf{x} | \theta_i)$. If the maximum of the log likelihood function is interior to the range of θ_i and $\ln P_{r|\theta_i}(\mathbf{x} | \theta_i)$ has a continuous first derivative, then a necessary condition on the ML estimation is obtained by differentiating $\ln P_{r|\theta_i}(\mathbf{x} | \theta_i)$ with respect to θ_i . Therefore, it is fairly important to divide the parameter space into two partitions, which has been proved to be a crucial factor of the asymptotic optimality of GLRT [6]. Unfortunately, the problem in this paper could hardly meet the necessary condition of GLRT. In addition, it is proved by Hartigan [7] and Bickel [8] that the usual GLRT has nonstandard behavior within the detection boundary of LRT, as the maximized ratio tends to ∞ under H_0 . Moreover, it is time-consuming and almost infeasible to calculate GLRT when the signal is of high dimension.

Inspired by the previous study, a novel scheme of detecting unknown and arbitrary signals against noise under the name of LRT-SE is proposed in this paper. LRT-SE integrates the hypothesis test under Neyman-Pearson framework and l_1 -minimization technique from CS theory. There exist notable differences between this new method and GLRT owing to the fact that the new method estimates the unknown sparse signal utilizing the l_1 -minimization which takes full advantage of the signal's sparsity. Preliminary analyses are provided to characterize the performance of LRT-SE in both finite and high dimension. It could be revealed that the prospective near-optimal performance is achievable by the experiment results of both synthetic sparse signals and sparse satellite telemetry data.

The rest of this paper is organized as follows. Section 2 provides a review of some related work. Section 3 introduces the background on Neyman-Pearson test and CS theory. In Section 4, the application scenario and the formulation of the problem are firstly introduced. The scheme of LRT-SE is then presented with the corresponding flow chart. In addition, theoretical analyses are discussed, preliminarily characterizing the properties of LRT-SE. These properties have been verified by the results in Section 5 based on the experiments on synthetic signals and satellite telemetry data. Section 6 discusses the results in detail. Finally, Section 7 concludes the full text and points out the research direction in the future.

2. Related Work

Detection of unknown sparse signal mixed with noise is a long-standing problem of composite hypothesis testing that arises in many scientific applications such as signal processing [9], wireless sensor networks [10], and remote sensing systems. A variety of methods and schemes have been developed to solve this universal yet intricate problem.

Except for the classical GLRT method, most methods so far concentrate on making assumptions of prior information about the signals of interest to alleviate the dilemma. On this premise, some researchers turned over to seek test statistics of the signal's sparsity adopting the recently prevalent compressive sensing (CS) framework. For instance, Duarte et al. estimated the relevant sufficient statistics for the test by directly extracting them from a small number of random projections without ever reconstructing the signal. They proved that the CS framework is information scalable to a much wider range of statistical inference tasks [11]. In the same year, Davenport et al. further demonstrated how to solve a variety of signal detection problems including sparse signal detection and estimation problems and gave the measurements without ever reconstructing the signals themselves [12]. Another paper written by Haupt and Nowak proposed a detector that collects a set of universal samples obtained without prior knowledge of the signal structure and examined the performance of CS for the problem of signal detection. However, some auxiliary channels are required to provide the so-called "future knowledge" [13]. Later in 2008, Wang et al. introduced a set of detectors called subspace compressive detectors aiming at solving the problem that the compressive measurements are not efficient at gathering signal energy [14]. In a recent article, the problem of sparse signal detection based on partial support set estimation with compressive measurements in a distributed network was discussed by Wimalajeewa and Varshney. The basic idea of this article resembled the work of [11] but intensively studied how to determine the minimum fraction of the support to be estimated so that the detection would perform optimally [15].

Nonetheless, it is less universal to impose on the aforementioned schemes because they more or less required prior knowledge about the sparse signals as some schemes exploited greedy algorithms or had other auxiliary conditions to improve performance.

On the other hand, the detection boundary in terms of sparsity and signal strength is another research priority. The sum of two types of error probabilities tends to zero or one depending on whether the mean of sparse signal, denoted by μ , exceeds this detection boundary. The sparsity exponent β is defined as the proportion of zero components to the signal dimension n , while the smallest possible signal strength r is defined as a function of the sparsity exponent, denoted by

$$r^*(\beta) = \begin{cases} \beta - \frac{1}{2} & \frac{1}{2} < \beta \leq \frac{3}{4} \\ (1 - \sqrt{1 - \beta})^2 & \frac{3}{4} < \beta < 1. \end{cases} \quad (1)$$

In [16], if the mean of signal obeys $\mu = \sqrt{2r \log n}$, it will show that the likelihood ratio test could reach

asymptotic power when $r > r^*(\beta)$, while conversely if $r < r^*(\beta)$ the test is asymptotically indistinguishable as the total error probabilities would converge to one. Based on this conclusion, Donoho and Jin demonstrated that this boundary still stands for the Higher Criticism which is also called second-level significance testing [17]. Afterward, the authors of [16] extended the existing detection boundary phenomenon to high dimensional linear regression and further established the boundaries under the condition that the variance of noise is unknown [18]. Moreover, [19] considered the different detection boundaries among the analysis of variance (ANOVA), the Max test, and the Higher Criticism when detecting sparse alternatives. It proved that ANOVA could obtain optimal performance under moderate levels of sparsity, that is, $\beta \in [0, 1/2]$, requiring μ to grow as a power of n . On the contrary, reliable detection is possible when employing the Max test only if the sparsity is very strong, namely, $\beta \in [3/4, 1]$, requiring μ to be on the order of $\sqrt{\log n}$, while the detection threshold for Higher Criticism remains unchanged. Inspired by the former papers, the authors of [20] discussed the detection problem under a general sparse mixture model and derived an explicit expression for the detection boundary under mild regularity conditions. To sum up, these papers work out the problem of when the sparse signal is detectable.

In light of these prior works and obeying the detection boundary, we are aiming to put forward a more efficient and practical detection scheme of unknown sparse signal mixture model. Detailed theoretical analyses about the properties of the new scheme are also introduced in this paper. Ultimately, we give out plentiful experiment results generated by detecting synthetic sparse signals and telemetry data.

3. Theoretical Background

The detection problem discussed in this paper mainly employs hypothesis test as framework. The null hypothesis is assumed to be white Gaussian noise, while the alternative hypothesis is an unknown sparse signal with no prior information except for its sparsity. As is well known, under likelihood ratio test framework, N-P criterion could achieve the largest detection probability among various decision criteria. Furthermore, the sparse estimation technique from Compressive Sampling is incorporated into the framework to estimate the unknown parameter of likelihood function.

3.1. Neyman-Pearson Test. Statistical hypothesis testing is a crucial method to detect and classify signals. Based on the statistical theory, there exist a variety of decision criteria for hypothesis testing, for example, the Bayes Criterion, minimum total error probability criterion, Neyman-Pearson criterion, and Maximin criterion. These criteria have been widely applied to radar signal detection, multisensor nondestructive testing, robot control, medical diagnosis, and other regions.

Instead of calculating the risk as Bayes Criterion requires, Neyman and Pearson [21] were inclined to work out a criterion making P_F (false alarm probability) as small as

possible while making P_D (detection probability) as large as possible. Afterward, the corresponding lemma which suggests the most powerful testing criterion was put forward referring to as Neyman-Pearson lemma. The test based on this lemma perfectly matches the situation in this paper when no prior information is available and the hypothesis is composite. In addition, the lemma tells us exactly how to find the appropriate threshold separating acceptance and rejection regions for the test [22].

Neyman-Pearson Lemma [22]. Let $\Theta = \Theta_0 \cup \Theta_1$ be a disjoint covering of the parameter space. Let H_0 and H_1 denote the hypothesis that $\theta \in \Theta_0$ and $\theta \in \Theta_1$, respectively.

Assume that X is a vector of observations from observation space with distribution $F_\theta(X)$. Constrain $P_F = \alpha' \leq \alpha$ and then the likelihood ratio test

$$l(x) = \frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} \underset{H_0}{\overset{H_1}{\geq}} \eta \quad (2)$$

is proved to be the most powerful test where the threshold η is decided by

$$P_F = \int_{\eta}^{\infty} f_{\theta_0}(l) dl = \alpha. \quad (3)$$

The receiver operating characteristic (ROC) figure plotting P_D versus P_F is commonly adopted to describe the performance of the test as a function of the parameter of interest. The larger the area under the ROC curves, the better the performance the test possesses.

3.2. l_1 -Minimization of Compressed Sensing. Compressed Sensing, firstly proposed by Donoho [23], has attracted widespread attention from signal processing, image processing, computer vision, and pattern recognition fields [24]. As a cutting-edge technique, the critical observation of CS is that, given the prior knowledge of a signal's sparsity, it is possible to efficiently and accurately reconstruct the full-length signal from the small amount of collected data with sampling rate far below Nyquist rate.

As one of the hottest research spots of CS, the recovery algorithms put forward so far could be roughly divided into three categories [25]: greedy pursuits (matching pursuit), basis pursuits (or l_1 -minimization), and combinatorial algorithms. Each type of algorithm has its merits and applicable scenarios. l_1 -minimization algorithms succeed with the least measurements and dispensing of prior information at the expense of relatively heavy computational burden. Matching pursuit shows an advantage in computation speed yet it is not the global optimal solution. For combinatorial algorithms, it requires a large number of unusual samples that may not be easily acquired [25]. In reality, strong background noise always comes along with the signal. Robust recovery could be achieved by the l_1 -error version of basis pursuit even when SNR is fairly low. However, greedy pursuits may perform estimation successfully only if all nonzero components of the signal are somewhat larger than the noise level [26].

l_1 -minimization is employed in this paper as it best fits the problem in this paper and could provide the relatively

accurate estimation value of the signal. The use of l_1 norm as a sparsity-promoting function dates back to several decades because minimizing subject to linear equality constraints can be easily transformed to a linear program which could be developed using more efficient solution algorithms [27]. Then, we may first formulate a sensing problem and then introduce l_1 -minimization.

A signal $\mathbf{x} \in \mathbf{R}^n$ can be sparsely represented by a linear combination of S vectors from Ψ . Denote S -sparse vector $\mathbf{s} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\}$ to express the coefficients of \mathbf{x} under Ψ . Then, one has sample \mathbf{x} with a matrix $\Phi \in \mathbf{R}^{m \times n}$, $m \ll n$, where rows of Φ are incoherent with columns of Ψ . Eventually, it is easy to obtain an incoherent observation as

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \mathbf{s} = \mathbf{A} \mathbf{s}, \quad \mathbf{y} \in \mathbf{R}^m, \quad (4)$$

where \mathbf{A} is defined as the measurement matrix.

After getting the sampled data \mathbf{y} , recovering \mathbf{x} from (4) is an underdetermined linear problem because $m \ll n$. It is nearly impossible to determine a definite unique solution directly. l_0 -minimization method which exhausts all sparse subsets is also an N-P hard problem and is impractical [28]. Innovatively, Chen et al. put forward an algorithm called *basis pursuit* (or l_1 -minimization) that selects l_1 -norm instead of l_0 -norm to alleviate the dilemma [29].

l_1 -Minimization. One has

$$\begin{aligned} \mathbf{s}^\# &= \arg \min \quad \|\mathbf{s}\|_1 \\ \text{s.t.} \quad & \Phi \mathbf{x} = \mathbf{A} \mathbf{s} = \mathbf{y}. \end{aligned} \quad (5)$$

In the case that

$$m \geq C \cdot S \cdot \log\left(\frac{n}{S}\right), \quad (6)$$

for some positive constant C , the solution to the former equation could be exact with overwhelming probability.

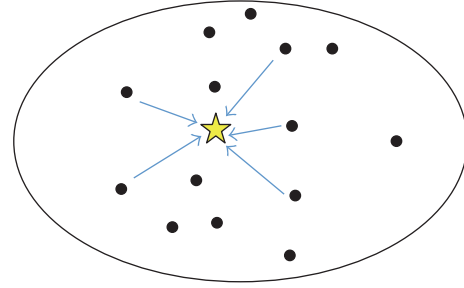
In 2005, Candès et al. further demonstrated that, under restricted isometry property (RIP), l_0 -norm and l_1 -norm optimization have the same and exact solutions [30].

Restricted Isometry Property (RIP) [30]. Let \mathbf{A}_T , $T \subset \{1, 2, \dots, n\}$, be the $m \times |T|$ submatrix generated by randomly extracting T columns out of matrix \mathbf{A} . Then, define the S -restricted isometry constant δ_S of \mathbf{A} as the smallest quantity satisfying the following inequality:

$$(1 - \delta_S) \|\mathbf{x}\|_2^2 \leq \|\mathbf{A}_T \mathbf{x}\|_2^2 \leq (1 + \delta_S) \|\mathbf{x}\|_2^2, \quad (7)$$

for all subsets T with $|T| \leq S$ and all S -sparse vectors \mathbf{x} . We may loosely say that a matrix \mathbf{A} obeys the RIP of order S if δ_S is not too close to one. As previously described, an approach to generate a matrix \mathbf{A} with the RIP of high order is to use random matrices such as random Gaussian matrix [31].

Accordingly, there exists the denoising form of l_1 -minimization. Suppose there is an additive noise $\mathbf{n} \in \mathbf{R}^m$



● Sensor node
★ Sparse event

FIGURE 1: Multiple-observation scenario.

contaminated observation $\mathbf{y} = \mathbf{A} \mathbf{s} + \mathbf{n}$; the error version of l_1 -minimization becomes as follows:

Basis Pursuit Denoising (BPDN) [29] (Also Known as Lasso). This is represented as follows:

$$\begin{aligned} \mathbf{s}^\# &= \arg \min \quad \|\mathbf{s}\|_1 \\ \text{s.t.} \quad & \|\mathbf{A} \mathbf{s} - \mathbf{y}\|_2 \leq \varepsilon, \end{aligned} \quad (8)$$

where ε is the size of \mathbf{n} .

4. LRT-SE Algorithm

As has been stated in the preceding sections, the problem in this paper is more general and practical, for no prior information about the signal is assumed apart from its sparsity. Applying the discussed background theories, the general strategy of LRT-SE is the so-called joint estimation and detection technique, which estimates the parameter value of likelihood function that is the sparse signal by utilizing l_1 -minimization. It is important to declare that we are more interested in making a decision rather than estimating the exact value of the sparse signal. In addition, we consider a multiple-observation scenario here in terms of multiple sensors or a receiver receiving signals for multiple times. Consequently, we give an overall judgment of whether the sequence of observations contains sparse signals or only noise.

4.1. Problem Model Setup. We considered a scenario where a total of t sensors are employed as shown in Figure 1 or a receiver is receiving signals for t times in a system, monitoring an event which could be described as a signal \mathbf{s} . Nevertheless, this signal is S -sparse in time domain which means apparent impulses only take place at S certain points while in other times the signal remains zero. Meanwhile, our cognition about the signal is limited to its sparsity. The occurrence time and the amplitude of each impulse can take any real number within range. Moreover, the detection suffers from white Gaussian noise which is a harmful effect for monitoring. Thereby, the problem addressed could be formulated as a binary composite hypothesis test.

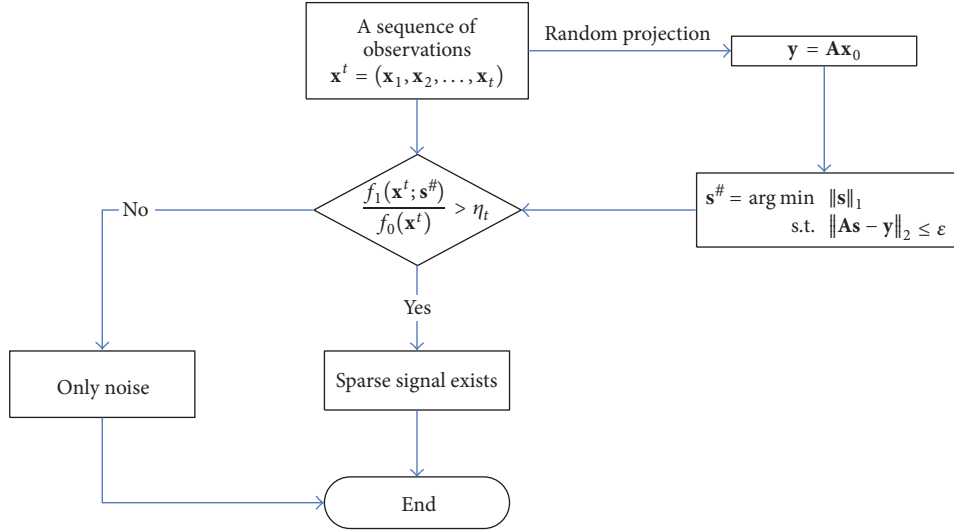


FIGURE 2: The structure of LRT-SE.

Let $(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_t)$ ($t \geq 1$) be a sequence of i.i.d. observations from t sensors measuring the event. According to the description of the problem, the binary hypothesis could be formulated as follows:

$$\begin{aligned} H_0 : \mathbf{x}_i &= \mathbf{n} \\ H_1 : \mathbf{x}_i &= \mathbf{s} + \mathbf{n}, \end{aligned} \quad (9)$$

where $\mathbf{x}_i, \mathbf{s}, \mathbf{n} \in \mathbb{R}^n$. \mathbf{s} is an unknown S -sparse signal with $\|\mathbf{s}\|_0 \leq S$, and the noise \mathbf{n} is assumed to be Gaussian white noise $\mathbf{n} \sim N(0, \sigma^2 \mathbf{I}_n)$. Therefore, the distributions of the two hypotheses obey

$$\begin{aligned} H_0 : \mathbf{x}_i &\stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2 \mathbf{I}_n) \quad 0 \leq i \leq 1 \\ H_1 : \mathbf{x}_i &\stackrel{\text{i.i.d.}}{\sim} N(\mathbf{s}, \sigma^2 \mathbf{I}_n) \quad 0 \leq i \leq 1. \end{aligned} \quad (10)$$

Next, a novel scheme under the name LRT-SE is put forward to solve the problem.

4.2. Scheme of LRT-SE. Based on the problem we set above, it is obvious that all tests could never be better than a simple hypothesis test in which the receiver first measures \mathbf{s} perfectly and then designs the optimal likelihood ratio test. However, there exists no uniformly most powerful (UMP) test as the likelihood ratio test for every possible vector of \mathbf{s} cannot be completely defined without knowledge of \mathbf{s} .

A reasonable approach we propose here is to estimate \mathbf{s} assuming that H_1 was true by applying l_1 -minimization and then substitute the estimated signal into the likelihood function as if it was the real signal.

LRT-SE works as follows:

- (1) For $i = 0$, estimate \mathbf{s} from \mathbf{x}_0 assuming that H_1 was true. Denote the estimate $\mathbf{s}^\# = \mathbf{s}^\#(\mathbf{x}_0)$ by employing the l_1 -minimization algorithm in accord with (7):

$$\begin{aligned} \mathbf{s}^\# &= \arg \min \|\mathbf{s}\|_1 \\ \text{s.t. } &\|\mathbf{A}\mathbf{s} - \mathbf{A}\mathbf{x}_0\|_2 \leq \varepsilon, \end{aligned} \quad (11)$$

where ε is the size of noise, $\|\mathbf{n}\|_2 \leq \varepsilon$.

- (2) For $i = 1, 2, \dots, t$ ($t \geq 1$), the likelihood functions are calculated as

$$\begin{aligned} f_0(\mathbf{x}_i) &= \prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_{ij}^2}{2\sigma^2}\right) \quad \text{for } H_0, \\ f_1(\mathbf{x}_i; \mathbf{s}^\#) &= \prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_{ij} - \mathbf{s}_j^\#)^2}{2\sigma^2}\right) \quad \text{for } H_1, \end{aligned} \quad (12)$$

where $\mathbf{s}_j^\#$ is the j th entry of $\mathbf{s}^\#$ and x_{ij} is the j th entry of \mathbf{x}_i .

- (3) The likelihood ratio for $\mathbf{x}^t = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t)$ is

$$T(\mathbf{x}^t) = \frac{f_1(\mathbf{x}^t; \mathbf{s}^\#)}{f_0(\mathbf{x}^t)}. \quad (13)$$

- (4) Set up a reasonable test level α ($0 < \alpha < 1$) and choose the matching threshold η_t according to N-P lemma (3).
- (5) Ultimately, the detector of hypotheses testing is

$$\varphi(\mathbf{x}^t; \eta) = \begin{cases} 1, & T(\mathbf{x}^t) > \eta_t \\ 0, & T(\mathbf{x}^t) < \eta_t. \end{cases} \quad (14)$$

The structure of the scheme is displayed in Figure 2.

4.3. Sparse Estimation Accuracy. Different from ML estimation which lacks specific analyses on estimation accuracy, researchers have already reached some conclusions about the reconstruction performance of l_1 -minimization. On the basis of the aforementioned principle in theoretical background, if the measurement matrix \mathbf{A} obeys RIP (7) and the vector \mathbf{s} is sufficiently sparse, then Theorem 1 describes the relationship between the level of noise perturbing the measurements and the accuracy of the sparse estimation.

Theorem 1 (see [30]). *Let S be such that $\delta_{3S} + 3\delta_{4S} < 2$. Then, for any signal \mathbf{s} supported on T with $|T| \leq S$ and any perturbation $\mathbf{e} = \mathbf{A}\mathbf{n}$ with $\|\mathbf{e}\|_2 \leq \varepsilon$, the solution $\mathbf{s}^\#$ to (8) obeys*

$$\|\mathbf{s}^\# - \mathbf{s}\|_2 \leq C_S \cdot \varepsilon, \quad (15)$$

where the positive constant C_S only depends on δ_{4S} , for example, typically $C_S \approx 8.82$ for $\delta_{4S} = 1/5$.

Later on, the Coherence-Based Guarantee for BPDN has been found by Ben-Haim et al. under the assumption of stochastic noise [26] as Theorem 2 describes.

Theorem 2 (see [26]). *Suppose that $\mathbf{n} \sim N(0, \sigma^2 \mathbf{I})$ is a random noise vector. With overwhelming probability, the solution $\mathbf{s}^\#$ to (8) is unique and the following inequality holds:*

$$\|\mathbf{s}^\# - \mathbf{s}\|_2^2 \leq \left(\sqrt{3} + 3\sqrt{2(1+\alpha)\log(n-S)} \right)^2 S\sigma^2, \quad (16)$$

for some fairly small $\alpha > 0$.

This theorem presents a stronger performance guarantee because of the assumption of random noise and the characteristic of the estimator behavior for typical noise values.

Utilizing the result from Theorem 1, we proposed a corollary which states the upper bound of noise level under which the estimation could be highly correlated with the original signal. In addition, it provides the prerequisite that we could take N-P test as the UMP test when detecting the sparse signal in a composite hypothesis framework.

Corollary 3. *If $\sigma < (\|\mathbf{s}\|_2/C_S)(m + \lambda\sqrt{2m})^{-1/2}$, then $\mathbf{s}^T \mathbf{s}^\#$ is bounded above some positive constant ζ .*

Proof. It is easy to discover that the size of perturbation $\mathbf{e} = \mathbf{A}\mathbf{n} \in \mathbf{R}^m$ is still white Gaussian noise with $\mathbf{e} \sim N(0, \sigma^2 \mathbf{I}_m)$.

From inequality (15), it is straightforward to obtain

$$\mathbf{s}^T \mathbf{s}^\# \geq \frac{1}{2} \left(\|\mathbf{s}\|_2^2 + \|\mathbf{s}^\#\|_2^2 - C_S \varepsilon^2 \right). \quad (17)$$

From the triangle inequality and the guarantee of the results from Theorem 1, we could get

$$\|\mathbf{s}\|_2 - \|\mathbf{s}^\#\|_2 \leq \|\mathbf{s}^\#\|_2 - \|\mathbf{s}\|_2 \leq \|\mathbf{s}^\# - \mathbf{s}\|_2 \leq C_S \varepsilon. \quad (18)$$

Then, it becomes as follows:

$$\begin{aligned} \|\mathbf{s}^\#\|_2 &\geq \|\mathbf{s}\|_2 - C_S \varepsilon, \\ \mathbf{s}^T \mathbf{s}^\# &\geq \frac{1}{2} \left[\|\mathbf{s}\|_2^2 + (\|\mathbf{s}\|_2 - C_S \varepsilon)^2 - C_S^2 \varepsilon^2 \right] \\ &= \|\mathbf{s}\|_2^2 - \|\mathbf{s}\|_2 C_S \varepsilon. \end{aligned} \quad (19)$$

Therefore, if $\|\mathbf{s}\|_2 > C_S \varepsilon$, denoting $\zeta = \|\mathbf{s}\|_2^2 - \|\mathbf{s}\|_2 C_S \varepsilon$ and according to the abovementioned inequality, it yields that

$$\mathbf{s}^T \mathbf{s}^\# \geq \zeta > 0. \quad (20)$$

For Gaussian noise $\mathbf{e} \sim N(0, \sigma^2 \mathbf{I}_m)$, $\varepsilon^2 = \sigma^2(m + \lambda\sqrt{2m})$ is established for large probability when $\|\mathbf{e}\|_2 \leq \varepsilon$. The upper bound could be obtained where

$$\sigma^2 < \frac{\|\mathbf{s}\|_2^2}{C_S^2} (m + \lambda\sqrt{2m})^{-1}. \quad (21)$$

Usually, $\lambda = 2$ or 3 is a reasonable choice and other choices are also possible. \square

4.4. Preliminary Analyses in Finite Dimension. In practice, we are on target to detect the sparse signal even though it is weak and the noise is substantial. Therefore, the detection scheme is required to be sensitive to the target objective and to be robust against interference. In this paper, detection sensitivity and robustness are measured in terms of two kinds of error probabilities P_F and P_M . Next, finite dimensional preliminary analyses will be introduced focusing on these two properties.

False alarm probability P_F and miss detection probability P_M could reflect the sensitivity and robustness of the detection scheme to the target object, respectively. However, these two properties contradict each other. By adjusting the parameter and increasing the decision threshold, the robustness of the detection system would improve as P_F would decrease while the sensitivity gets worse with P_M rising up. On the contrary, lowering the decision threshold makes the system more sensitive since P_M decreases yet P_F gets bigger. In reality, $1 - P_M$, which is also called detection probability P_D , is usually utilized to measure the sensitivity of the detection scheme. As adopted in this paper, Neyman-Pearson test could obtain the largest P_D under certain constrained P_F . Nevertheless, we could obtain an upper bound of P_M employing the corollary mentioned before.

Lemma 4. *Under the test of (13), supposing that $t = 1$, if the false alarm probability is given as $P_F = \alpha$, then the following inequality establishes*

$$\begin{aligned} P_M &\leq \frac{C_2}{\sqrt{\|\mathbf{s}\|_2 \cos \theta / \sigma - C_1 + C_3}} \\ &\cdot \exp \left(-\frac{(\|\mathbf{s}\|_2 \cos \theta - C_1 \sigma)^2}{2\sigma^2} \right), \end{aligned} \quad (22)$$

where $\cos \theta = \mathbf{s}^T \mathbf{s}^\# / (\|\mathbf{s}\|_2 \times \|\mathbf{s}^\#\|_2)$ and θ is the angle between \mathbf{s} and $\mathbf{s}^\#$.

Proof. In the case that there is only one observation, by taking a logarithm, we obtain that test of (13) could be equivalent to testing on $l = (1/t) \sum_{i=1}^t \mathbf{x}_i^T \mathbf{s}^\#$. Also, the threshold changes correspondingly as

$$\begin{aligned} & l \underset{H_0}{\overset{H_1}{\gtrless}} \tau_t, \\ & \tau_t = \frac{\sigma^2}{t} \log \eta_t + \frac{1}{2}. \end{aligned} \quad (23)$$

The distributions of the test l become as follows:

$$\begin{aligned} H_0 : l & \sim N\left(0, \frac{\sigma^2 \|\mathbf{s}^\#\|_2}{t}\right) \\ H_1 : l & \sim N\left(\mathbf{s}^T \mathbf{s}^\#, \frac{\sigma^2 \|\mathbf{s}^\#\|_2}{t}\right). \end{aligned} \quad (24)$$

As a result, for a test level of α ($0 < \alpha < 1$), the threshold should satisfy the following equation:

$$P_F = P_r(l > \tau_t | H_0) = 1 - \Phi\left(\frac{\sqrt{t}\tau_t}{\sigma \|\mathbf{s}^\#\|_2}\right) = \alpha, \quad (25)$$

where Φ is the CDF of standard normal distribution.

As seen there, it yields that

$$\tau_t = \frac{\sigma \|\mathbf{s}^\#\|_2 \Phi^{-1}(1 - \alpha)}{\sqrt{t}}, \quad (26)$$

$$P_M = P_r(l < \tau_t | H_1) = \Phi\left(\frac{(\tau_t - \mathbf{s}^T \mathbf{s}^\#) \sqrt{t}}{\sigma \|\mathbf{s}^\#\|_2}\right). \quad (27)$$

Then, utilizing the inequality for the tail probability of standard normal distribution [32],

$$1 - \Phi(x) \leq \frac{\phi(x)}{\sqrt{\max\{x, x^2\} + 2/\pi}} \quad \forall x : x \geq 0. \quad (28)$$

It is not hard to get

$$\begin{aligned} P_M & \leq \frac{C_2}{\sqrt{\|\mathbf{s}\|_2 \cos \theta / \sigma - C_1 + C_3}} \\ & \cdot \exp\left(-\frac{(\|\mathbf{s}\|_2 \cos \theta - C_1 \sigma)^2}{2\sigma^2}\right). \end{aligned} \quad (29)$$

Under the condition of high SNR, acting up to similar technique of Corollary 3 that $\cos \theta > 1 - 2C_s \varepsilon / \|\mathbf{s}\|_2$, we could work out the following inequality:

$$\begin{aligned} P_M & \leq \frac{C_2}{\sqrt{\|\mathbf{s}\|_2 / \sigma - 2C_s \varepsilon / \sigma - C_1 + C_3}} \\ & \cdot \exp\left(-\frac{1}{2} \left(\frac{\|\mathbf{s}\|_2}{\sigma} - \frac{2C_s \varepsilon}{\sigma} - C_1\right)^2\right). \end{aligned} \quad (30)$$

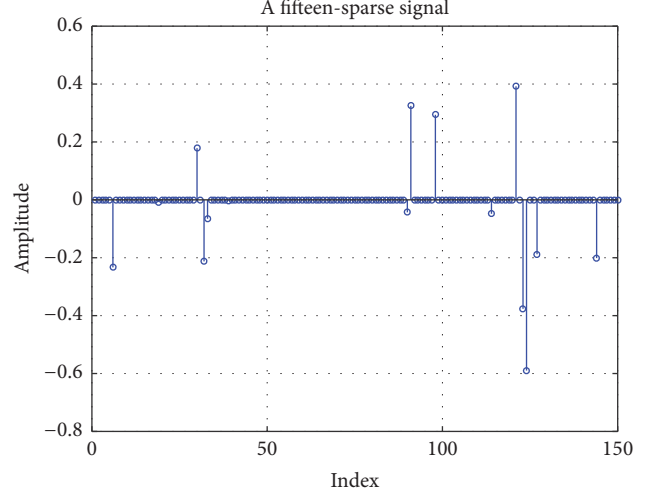


FIGURE 3: Synthetic sparse signal example.

This lemma indicates that the miss detection probability decays exponentially with the SNR and the accuracy of sparse estimation in terms of the angle θ . Considering the factor of exponent decay rate, $\|\mathbf{s}\|_2 / \sigma$ is the square root of the SNR of a matched filter when \mathbf{s} is available, which is totally different from input signal-to-noise ratio, denoted by $\text{SNR} = \|\mathbf{s}\|_2 / n\sigma$. The latter declines linearly with the increase of signal dimension n . Another factor of the decay rate is $C_s \varepsilon$ which shows the cost incurred by the unknown signal nature where ε is the error of sparse estimation. \square

5. Experiments and Results

A great deal of simulation experiments of LRT-SE have been undertaken using synthetic sparse signals and typical sparse signal samples from satellite telemetry to verify the properties of LRT-SE. The setup of simulation will be explained in Section 5.1, including the explanation of the sparse signals and the setting value of n, m , and other parameters. Section 5.2 illustrates plentiful experimental results and compares the performance of LRT-SE against the optimal LRT and other detectors.

5.1. Sparse Signals and Parameters Setting. For the sake of evaluating the proposed detection scheme, both synthetic and real sampled sparse signals are applied to experiments. The synthetic sparse signal $\mathbf{s} \in \mathbf{R}^n$ is created by randomly choosing its support set $\text{supp}(\mathbf{s}) := \{i \mid \mathbf{s}_i \neq 0\}$ with sparsity S . And the nonzero components are i.i.d. drawn from standard normal distributions. Empirically, energy estimation could be regarded as an insignificant factor in finite settings, which enables us to normalize the energy of \mathbf{s} so that $\|\mathbf{s}\|_2 = 1$. Figure 3 presents a typical randomly generated sparse signal. It is because the sparse signal is randomly generated that during the simulation \mathbf{s} could be deterministic and unknown by the detectors of LRT-SE, while in the case of LRT scheme it is known.

TABLE I: Simulation parameters.

Parameter	Value
n	150
m	120
S	Changes accordingly, $S \ll n$
\mathbf{s}	(1) $\text{supp}(\mathbf{s}) := \{i \mid \mathbf{s}_i \neq 0\}$ is randomly chosen with sparsity $S(2) \{s_i \mid i \in \text{supp}(\mathbf{s})\}$ are i.i.d. drawn from standard normal distributions (3) Normalization of energy $\ \mathbf{s}\ _2 = 1$
SNR	$\text{SNR} = 10 \log_{10} (\ \mathbf{s}\ _2 / n\sigma^2)$, $-20 \text{ dB} \sim 0 \text{ dB}$
σ^2	Changes with SNR
\mathbf{n}	$\mathbf{n} \sim N(0, \sigma^2 \mathbf{I}_n)$
t	$t \geq 1$
\mathbf{A}	Gaussian random matrix $\mathbf{A} \in \mathbf{R}^{m \times n}$
λ	2
ε	$\varepsilon^2 = \sigma^2(m + \lambda \sqrt{2m})$
α	0.01~1

The specific parameters of the simulation experiments on synthetic sparse signal are tabulated in Table I.

The satellite telemetry data employed in this paper comes from a high altitude satellite consisting of a 1024-dimensional synthetic angular momentum of wheel vector, a 512-dimensional yaw angle vector, and a 512-dimensional pitch angle vector, as Figure 4 shows. These data all possess the characteristics of high dimensionality, huge amount, sparsity in time domain, and noise contamination. Moreover, according to index of downlink signal, the signal power of telemetry data is -110 dBm while the noise power is -100.7 dBm . After amplification, the final SNR is approximately -9.3 dB . Also, as has been declared, there exists no prior information about the telemetry data except for their sparsity which accords with the application condition of LRT-SE.

5.2. Simulation Results. Subsequently, a large number of simulation experiments have been conducted and several plots are generated as the result figures show.

Figures 5–7 are results from simulation experiments on synthetic sparse signals. Figure 5 depicts the ROC curves of LRT-SE and LRT. The variation trend of detection probability with respect to SNR is displayed in Figure 6. Figure 7 illustrates the ROC curve of LRT-SE as a function of the number of measurements. These figures come from the average of experiment results corresponding to 10000 test signals mixed with Gaussian noise vectors.

Figure 8 shows the ROC curves of LRT-SE detecting telemetry data. Moreover, the experiment is carried out on 500 signals extracted from the satellite downlink signal files.

- (i) *ROC Curves of Detecting Synthetic Signals.* See Figure 5.
- (ii) *Detection Probability versus SNR.* See Figure 6.
- (iii) *ROC Curves versus Measurements Number.* See Figure 7.
- (iv) *ROC Curves of Detecting Telemetry Data.* See Figure 8.

6. Discussion

In this section, a detailed discussion on the results presented in previous sections is firstly carried out. Then, we further investigate the high dimensional properties of LRT-SE and discuss the corresponding result figures.

6.1. Analytical Discussion

(i) *ROC Curves of Detecting Synthetic Signals.* From Figure 5, it can be observed that the proposed LRT-SE has approximate performance to the optimal LRT under relatively low SNR. By choosing appropriate P_F , such as 0.05, the detection rate could reach 0.9 which is acceptable in most cases. The difference of the detection probabilities between the two detectors is obvious when P_F is less than 0.1 due to the cost of no prior information.

(ii) *Detection Probability versus SNR.* Similar to Figure 5, in Figure 6, the curve of LRT performs as an ideal detector when SNR is fairly low while LRT-SE and GLRT approach nearly perfect detection when SNR exceeds -5 dB with the sparsity $S = 15$. This is mainly because when noise is repressed, $\mathbf{s}^T \mathbf{s}^\#$ would enlarge according to Corollary 3 and bring about improvement of detection probability of likelihood test based on sparse estimation. It should be pointed out that although GLRT seems to own better asymptotic optimality compared with LRT-SE, it is an N-P hard optimization problem utilizing GLRT if the signal is compressed by measurement matrix in advance [33]. GLRT has to search every set satisfying $\{\mathbf{s} : \|\mathbf{s}\|_0 \leq S\}$ making it impossible to calculate for high dimensional signals in practice. Conversely, LRT-SE could estimate the unknown and arbitrary sparse signal with relatively low computational complexity. In general, in the case of strong sparsity, the asymptotic performance of LRT-SE is in the vicinity of GLRT while LRT-SE could be more easily implemented.

(iii) *ROC Curves versus Measurements Number.* As plotted in Figure 7, it could be discovered that, for a fixed sparsity level of 2, SNR of -5 dB , and $t = 1$, the detection probability increases as more measurements have been undertaken. The compression ratio is defined as m/n . It is not hard to expect such a result for the larger the compression ratio is, the more the crucial information may be lost inducing difficulty in accurately reconstructing the sparse signal. Although compression ratio is not specifically demanded in this paper, there exists a lower bound of measurements where $m \geq C \cdot S \cdot \log(n/S)$ to guarantee the matrix obeying the restricted isometry property, just as Section 3.2 mentioned.

(iv) *ROC Curves of Detecting Telemetry Data.* Every figure of Figure 8 depicts three curves with $t = 1, 5, 10$, respectively, because the satellite telemetry system could provide at most ten observations. From Figure 8(a), it is notable that LRT-SE reaches nearly a hundred percent detection rate with P_F less than 0.05 even if only one observation is provided. It performs near optimally with larger t .

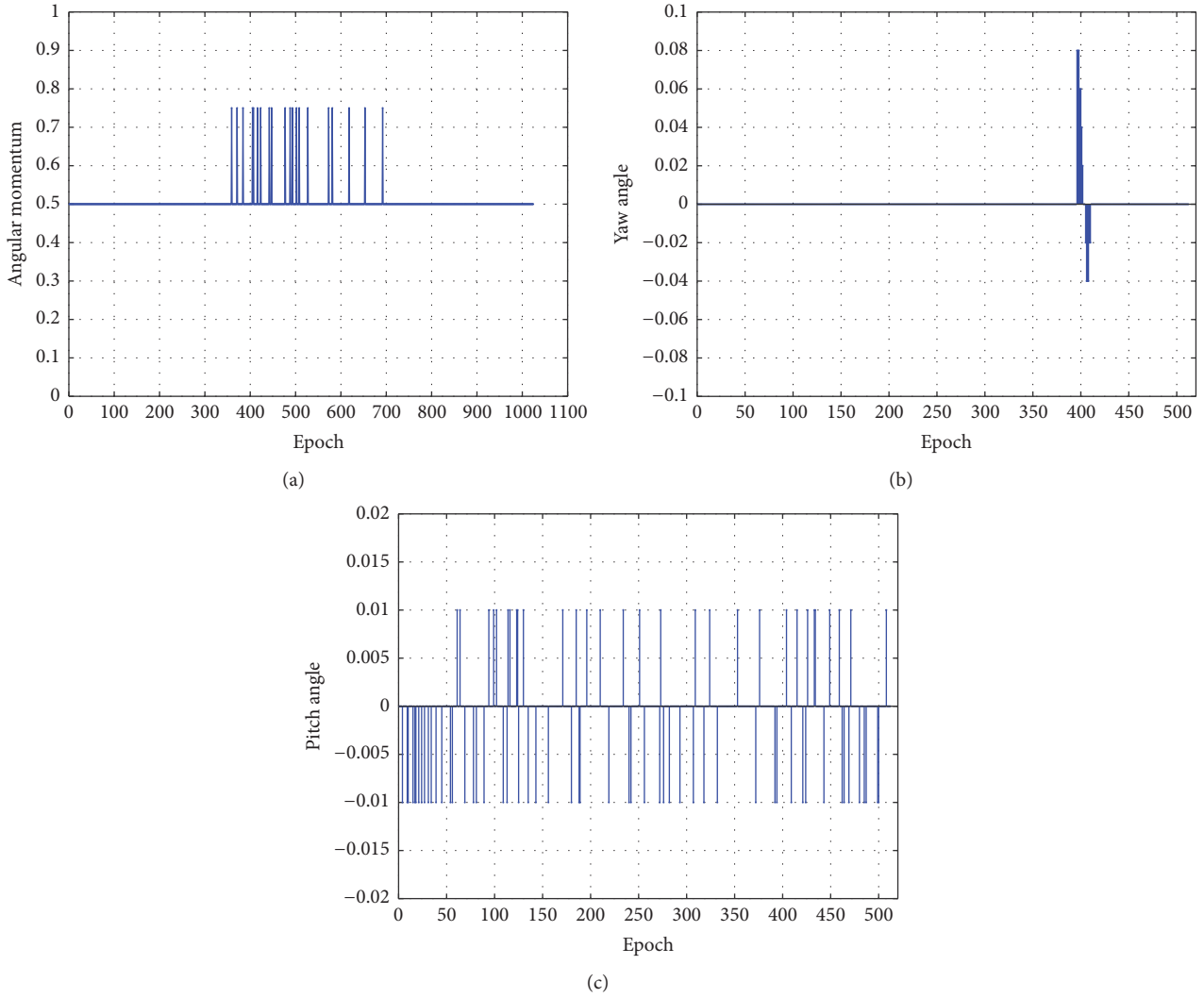


FIGURE 4: Snapshots of the sparse telemetry data: (a) original data of synthetic angular momentum of wheel; (b) original data of yaw angle; and (c) original data of pitch angle.

However, the curves of Figures 8(b) and 8(c) are not as good as in Figure 8(a) which is partly because of the sparsity and the signal amplitude as shown in Figures 4(b) and 4(c). Comparing Figures 4(a) and 4(b), the amplitudes of yaw angle data are less uniform than of angular momentum which is reflected in poorer detection probability. Likewise, it could be indicated that Figure 8(c) shows worse performance because the sparsity of pitch angle data is weaker than of angular momentum data. This phenomenon is worth researching in the future works.

Nonetheless, the loss caused by weak sparsity and irregular amplitudes could be compensated by obtaining more observations as the arrows in Figures 8(b) and 8(c) suggest. If possible, under certain P_F and SNR, the more the observations available, the better the detection rate that LRT-SE could approach.

6.2. Further Discussion. When discussing the high dimensional condition, asymptotic properties or inferences, for

example, consistency, are usually adopted to describe the performance of point estimator and hypothesis test. Further investigation has been conducted under the condition that $n \rightarrow \infty, t \rightarrow \infty$.

(i) *Chernoff Consistency.* Consistency is a property describing the convergence of error probabilities to zero in some sense, as $t \rightarrow \infty$. Among different definitions of consistency, Chernoff consistency requires that both P_F and P_M converge to 0 with $t \rightarrow \infty$ [34].

Definition 5. The detector $\varphi(\mathbf{x}^t; \eta)$ of the hypothesis test is called Chernoff-consistent if and only if there exists a criterion η_t such that the hypothesis test (13) can perfectly distinguish samples from the binary hypotheses when $t \rightarrow \infty$. That is,

$$(1) \lim_{t \rightarrow \infty} P_D[\varphi(\mathbf{x}^t; \eta)] = 1, \text{ for any } \mathbf{x}^t \text{ under } H_1.$$

$$(2) \lim_{t \rightarrow \infty} P_F[\varphi(\mathbf{x}^t; \eta)] = 0, \text{ for any } \mathbf{x}^t \text{ under } H_0.$$

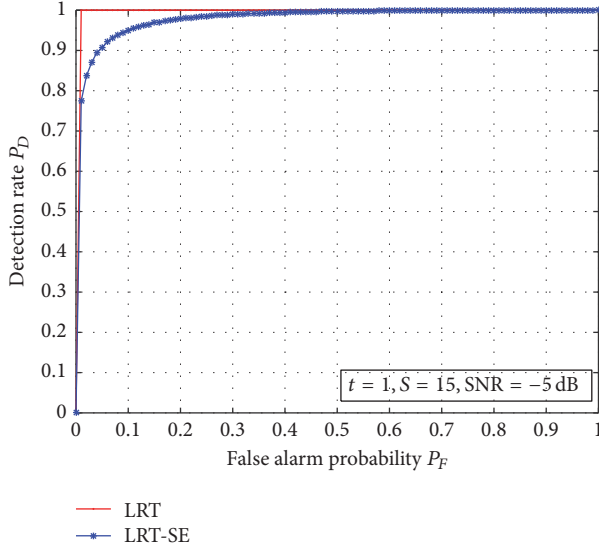


FIGURE 5: ROC curves of LRT-SE compared to LRT.

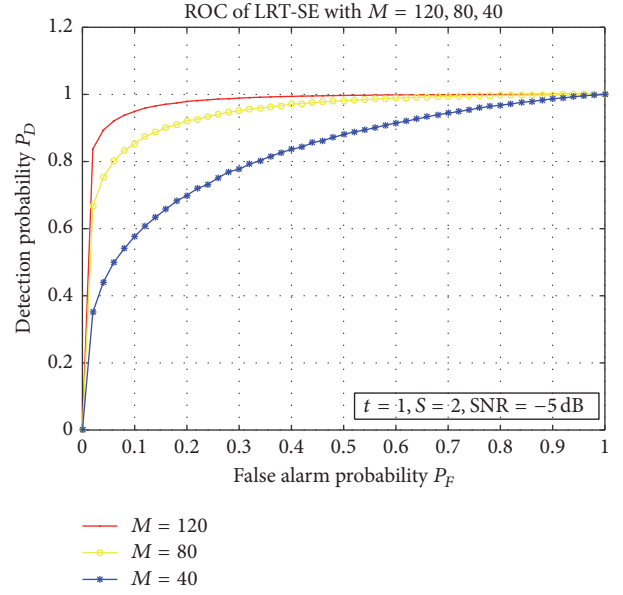


FIGURE 7: ROC curve of LRT-SE under different measurement numbers.

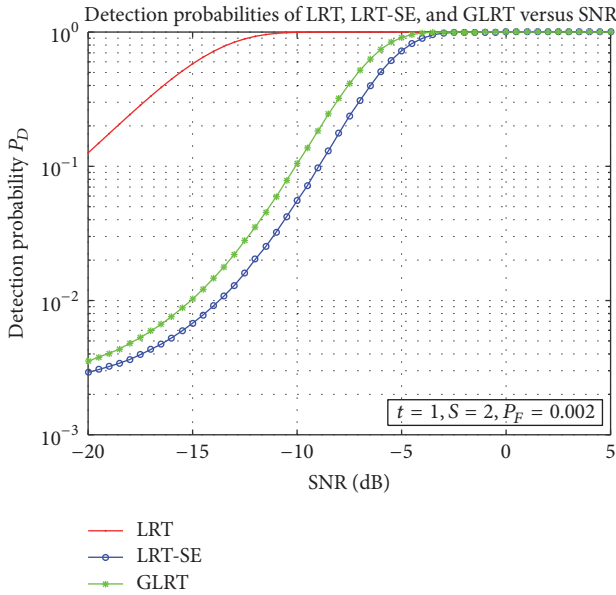


FIGURE 6: Detection probability versus SNR.

The next theorem describes that only if the noise level is controlled within a certain range could a reliable and consistent detector be possible.

Theorem 6. One has $\mathbf{A} \in \mathbf{R}^{m \times n}$, and \mathbf{s} is the S -sparse unknown signal of interest where S is a function of \mathbf{A} . Let $\{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_t\}$ ($t \geq 1$) be a sequence of i.i.d. observations sampled from one of the two hypotheses in (13). Let $\varphi(\mathbf{x}^t; \eta)$ be a detector defined in (14).

If the standard deviation obeys $\sigma < (\|\mathbf{s}\|_2/C_s)(m + \lambda\sqrt{2m})^{-1/2}$, then $\varphi(\mathbf{x}^t; \eta)$ is Chernoff-consistent, where C_s is a positive constant from (15).

Proof. Given $\sigma < (\|\mathbf{s}\|_2/C_s)(m + \lambda\sqrt{2m})^{-1/2}$ and applying Corollary 3, apparently we have $\mathbf{s}^T \mathbf{s}^\# \geq \zeta > 0$. This condition guarantees that the reconstructed signal is sufficiently good to separate H_0 and H_1 .

As has been obtained in (27) that $P_M(\varphi(\mathbf{x}^t; \eta)) = \Phi(\Phi^{-1}(1 - \alpha) - \sqrt{t} \mathbf{s}^T \mathbf{s}^\# / \sigma \|\mathbf{s}^\#\|_2)$, note that the test level is α for all t . Further, let $[\alpha(t)] \subset (0, 1)$ be a sequence satisfying $\alpha(t) \rightarrow 0$ and $\sqrt{t} \alpha(t) \rightarrow \infty$. Then, $\varphi(\mathbf{x}^t; \eta)$ has test level $\alpha(t)$ for each t , and therefore $P_F(\varphi(\mathbf{x}^t; \eta)) = \alpha(t) \rightarrow 0$.

At the same time, as $\alpha(t) \rightarrow 0$, it is obvious that $\Phi^{-1}(1 - \alpha(t)) \rightarrow \infty$. Considering that $\sqrt{t} \mathbf{s}^T \mathbf{s}^\# / \sigma \|\mathbf{s}^\#\|_2$ is relatively small, we could come to a conclusion that $P_M(\varphi(\mathbf{x}^t; \eta))$ goes to zero with $\alpha(t)$ in place of α .

Summing up the above, the detector $\varphi(\mathbf{x}^t; \eta)$ of LRT-SE is Chernoff-consistent.

(i) *Result Figure of Chernoff Consistency.* For the sake of verifying the Chernoff consistency of LRT-SE, experiments under a noise level which is a little bit larger than the upper bound introduced in Theorem 6 are carried out. Particularly, from the inequality $\sigma^2 < \|\mathbf{s}\|_2^2 / C_s^2 (m + \lambda\sqrt{2m})$, it is defined that [17] $C_s = (4/\sqrt{3})(1/(\sqrt{1 - \delta_{4s}} - \sqrt{1/3}\sqrt{1 + \delta_{3s}}))$. We could get $C_s > 4/\sqrt{3}$ and enforce $\sigma^2 = \|\mathbf{s}\|_2^2 / (m + \lambda\sqrt{2m})$. In addition, during the experiment, three sparsity levels of $S = 2, S = 15, S = 30$ are chosen with fixed SNR under 1000 tests.

We use the smallest total error probability $P_F + P_M$ across all possible thresholds to characterize the Chernoff consistency and check if it would reach zero ultimately. The result is plotted in Figure 9 which identifies with the theoretical derivation. When observations are plenty enough, the sum of false alarm and miss detection probabilities decreases to zero for all three sparsity levels. What is more, the stronger the sparsity of the signal is, the quicker the test approaches zero error. This indicates that, in the case of fixed

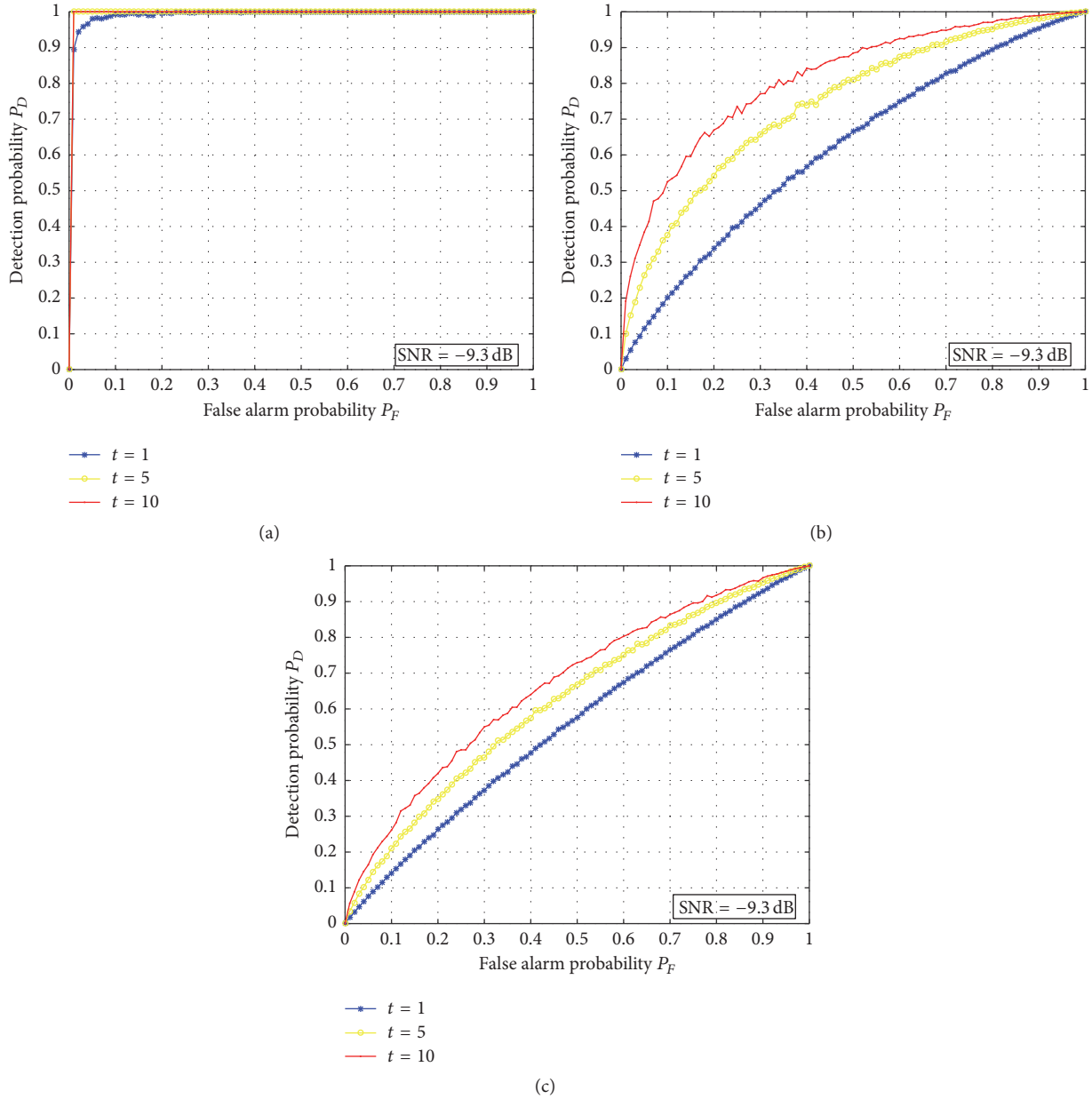


FIGURE 8: LRT-SE performance of detecting telemetry data: (a) angular momentum detection performance; (b) yaw angle detection performance; and (c) pitch angle detection performance.

SNR, the less sparse the signal is, the harder it is for detectors to separate it from background noise.

(ii) *Error Exponent*. Error exponent is the rate at which the error probabilities of the detector decrease as the number of observations increases. It is practical to discuss the error exponent because it provides us with a rough estimate of how many t are needed to control the error probabilities to a fairly small scale. Consequently, the error exponent for a Neyman-Pearson LRT-SE detector is defined as

$$E_{\text{LRT-SE}}^{\text{NP}} = -\lim_{t \rightarrow \infty} \frac{1}{t} \log P_M, \quad (31)$$

describing the rate at which the miss probability decays exponentially when $t \rightarrow \infty$.

Note that the test statistic in (23) could be the sum of the of i.i.d. random variables $t(x) = x^T \mathbf{s}^\#$, whose distributions are

$$\begin{aligned} h_0 : t(x) &\sim N(0, \sigma^2), \\ h_1 : t(x) &\sim N(\mathbf{s}^T \mathbf{s}^\#, \sigma^2). \end{aligned} \quad (32)$$

The test has been simplified to better characterize the error exponent. By applying the following lemma, the ideal error exponent could be obtained. \square

Lemma 7 (Chernoff-Stein lemma [35]). *Let $\{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_t\}$ ($t \geq 1$) be i.i.d. drawn from one of the hypotheses in (10). Denote the sparse reconstruction from \mathbf{x}_0 by $\mathbf{s}^\#$. Consider that the test on each observation obeys h_0 and h_1 in (32) under*

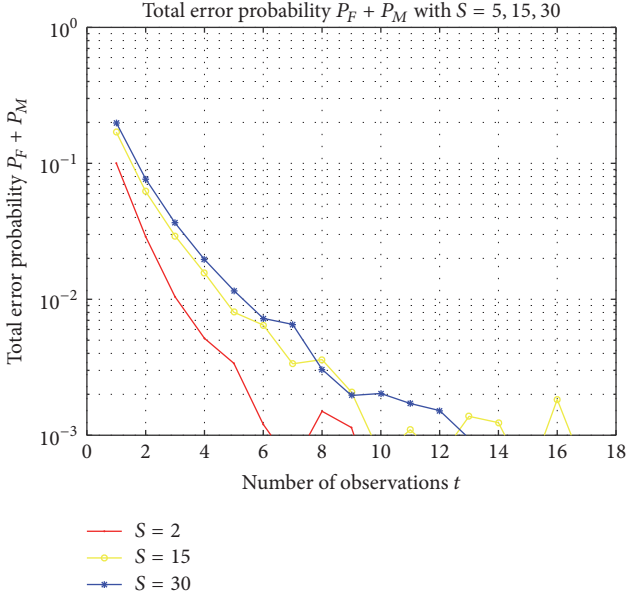


FIGURE 9: Total error probability based on 1000 tests with fixed SNR.

H_0 and H_1 , respectively. Let $A_t \subseteq \mathcal{T}^t$ be an acceptance region for hypothesis H_0 , where $\mathcal{T}^t = t(\mathbf{x}_1) \times t(\mathbf{x}_2) \times \dots \times t(\mathbf{x}_t)$ is the Cartesian product of $t(\mathbf{x}_1)$ through $t(\mathbf{x}_t)$. Then, let the false alarm and miss probabilities corresponding to A_t be

$$\begin{aligned} P_F^t &= h_0^t(A_t^c), \\ P_M^t &= h_1^t(A_t). \end{aligned} \quad (33)$$

For $0 < \alpha < 1/2$, define

$$\begin{aligned} P_M^{t,\alpha} &= \min P_M^t, \\ A_t &\subseteq \mathcal{T}^t, \\ P_F^t &< \alpha, \end{aligned} \quad (34)$$

where the minimization is over all possible acceptance regions in \mathcal{T}^t having a false alarm probability less than α . Then

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log P_M^{t,\alpha} = -D(h_0 \parallel h_1), \quad (35)$$

where $D(h_0 \parallel h_1) < \infty$ is the Kullback-Leibler distance between h_0 and h_1 .

Lemma 7 gives us a quantitative description of error exponent. Afterwards, Theorem 8 introduces the theoretical value error exponent utilizing Lemma 7.

Theorem 8. One has

$$E_{LRT-SE}^{NP} = \frac{1}{2} \cdot \frac{\|\mathbf{s}\|_2^2 \cos^2 \theta}{\sigma^2}, \quad (36)$$

where θ is the angle between \mathbf{s} and $\mathbf{s}^\#$.

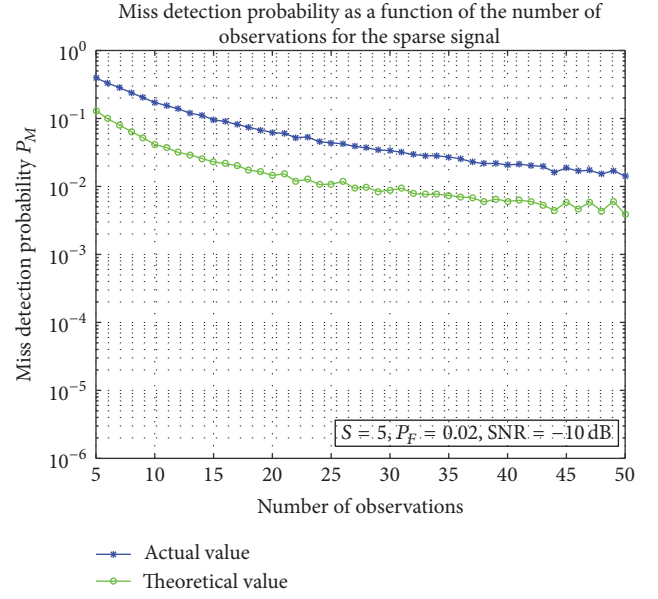


FIGURE 10: Miss detection probability against number of observations.

Proof. The Kullback-Leibler distance is defined as

$$D(h_0 \parallel h_1) = \int_{-\infty}^{\infty} h_0(x) \ln \frac{h_0(x)}{h_1(x)} dx, \quad (37)$$

where $h_0(x)$ and $h_1(x)$ are distributions of (32).

After simple calculation of the integral, it can be derived that

$$\begin{aligned} E_{LRT-SE}^{NP} &= -\lim_{t \rightarrow \infty} \frac{1}{t} \log P_M = D(h_0 \parallel h_1) \\ &= \frac{1}{2} \cdot \frac{\|\mathbf{s}\|_2^2 \cos^2 \theta}{\sigma^2}, \end{aligned} \quad (38)$$

where $\cos \theta = \mathbf{s}^T \mathbf{s}^\# / (\|\mathbf{s}\|_2 \times \|\mathbf{s}^\#\|_2)$. \square

(i) *Result Figure of Error Exponent.* Error exponent is presented in Figure 10 by plotting the miss detection probability against number of observations in exponent coordinates. The green curve depicts the theoretical decay curve of LRT-SE as Theorem 8 indicates, while the blue curve displays the actual miss detection probabilities of LRT-SE. It could be revealed that both actual and theoretical curves decay exponentially as the observation numbers increase because they are approximated to a straight line. Nevertheless, there exists an eternal gap between these two curves because the analysis takes no account of constant term. Beyond the result of this figure, it could be inferred and easily verified that the optimal LRT detector also possesses error exponent which could reach 10^{-10} order with very few observations.

From further discussion, we could reach a conclusion that the performance of LRT-SE could get promoted when n and t are big enough via asymptotic inference analysis. In order to satisfy the sufficient condition of consistency, the noise level should be under a certain limit. In addition, it is expected that

the more the observations available, the faster the decay of the detection error probability.

7. Conclusions and Future Works

In this paper, a new method under the name LRT-SE which integrates the likelihood ratio test and sparse estimation for sparse signal detection is proposed. As long as the background noise level is limited within a relative broad range, LRT-SE could perform in close proximity to the existing asymptotic optimal methods in both finite dimension and high dimension. It is revealed that the miss detection probability has an upper bound in direct proportion to the angle between estimated signal and original signal. But as the number of observations increases and approaches infinity, the false alarm probability and miss detection probability would eventually reach zero. Moreover, the exponential decay rate of miss detection probability could be defined as the error exponent and calculated by large deviation. In short, the LRT-SE scheme provides a novel detection method which could achieve high availability and robustness of detecting sparse signal with no prior information as well as relatively low computational complexity compared with GLRT.

In the future, there are two promising improvement directions for us to do further research. Firstly, the accuracy of sparse estimation has the potential to be improved by employing a better reconstruction algorithm. Secondly, the application of LRT-SE in distributed network deserves deeper research where the distributed network could be a wireless sensor network or in a telemetry system.

Competing Interests

The authors declare no competing interests.

Authors' Contributions

Hongbo Zhao and Lei Chen developed the overall algorithm and wrote the paper. Wenquan Feng performed the simulation and experiments. Chuan Lei analyzed the experimental data and reviewed the paper.

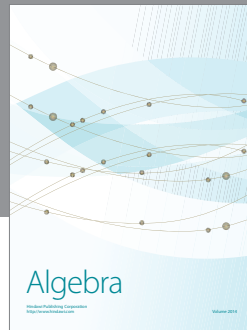
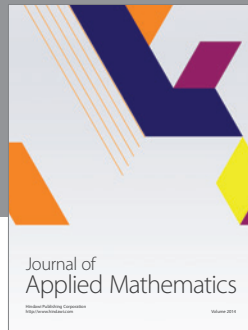
Acknowledgments

This research was supported by the aerospace fund of China under Contract no. 2015ZD51033.

References

- [1] S. Yin and O. Kaynak, "Big data for modern industry: challenges and trends," *Proceedings of the IEEE*, vol. 103, no. 2, pp. 143–146, 2015.
- [2] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.
- [3] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, vol. 72–73, pp. 303–315, 2016.
- [4] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, *Introduction to Compressed Sensing*, vol. 93 of *Lecture Note*, 2012.
- [5] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, John Wiley & Sons, 2004.
- [6] O. Zeitouni, J. Ziv, and N. Merhav, "When is the generalized likelihood ratio test optimal?" *IEEE Transactions on Information Theory*, vol. 38, no. 5, pp. 1597–1602, 1992.
- [7] J. A. Hartigan, "A failure of likelihood asymptotics for normal mixtures," in *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, vol. 2, pp. 807–810, Belmont, Calif, USA, 1985.
- [8] P. J. Bickel, "Asymptotic distribution of the likelihood ratio statistic in a prototypical non regular problem," in *Statistics and Probability: A Raghu Raj Bahadur Festschrift*, 1993.
- [9] R. L. Dobrusin, "A statistical problem arising in the theory of detection of signals in the presence of noise in a multi-channel system and leading to stable distribution laws," *Theory of Probability & Its Applications*, vol. 3, no. 2, pp. 161–173, 1958.
- [10] J. Meng, H. Li, and Z. Han, "Sparse event detection in wireless sensor networks using compressive sensing," in *Proceedings of the 43rd Annual Conference on Information Sciences and Systems (CISS '09)*, pp. 181–185, Baltimore, Md, USA, March 2009.
- [11] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "Sparse signal detection from incoherent projections," in *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 3, p. 3, IEEE, Toulouse, France, 2006.
- [12] M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "Detection and estimation with compressive measurements," Tech. Rep., Department of ECE, Rice University, 2006.
- [13] J. Haupt and R. Nowak, "Compressive sampling for signal detection," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, vol. 3, pp. III-1509–III-1512, Honolulu, Hawaii, USA, April 2007.
- [14] Z. Wang, G. R. Arce, and B. M. Sadler, "Subspace compressive detection for sparse signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 3873–3876, IEEE, Las Vegas, Nev, USA, April 2008.
- [15] T. Wimalajeewa and P. K. Varshney, "Sparse signal detection with compressive measurements via partial support set estimation," <http://arxiv.org/abs/1602.06230>.
- [16] Y. I. Ingster, "Minimax detection of a signal for l^n -balls," *Mathematical Methods of Statistics*, vol. 7, no. 4, pp. 401–428, 1998.
- [17] D. Donoho and J. Jin, "Higher criticism for detecting sparse heterogeneous mixtures," *Annals of Statistics*, vol. 32, no. 3, pp. 962–994, 2004.
- [18] Y. I. Ingster, A. B. Tsybakov, and N. Verzelen, "Detection boundary in sparse regression," *Electronic Journal of Statistics*, vol. 4, pp. 1476–1526, 2010.
- [19] E. Arias-Castro, E. J. Candès, and Y. Plan, "Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism," *The Annals of Statistics*, vol. 39, no. 5, pp. 2533–2556, 2011.
- [20] T. T. Cai and Y. Wu, "Optimal detection of sparse mixtures against a given null distribution," *IEEE Transactions on Information Theory*, vol. 60, no. 4, pp. 2217–2232, 2014.

- [21] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," in *Breakthroughs in Statistics*, pp. 73–108, Springer, New York, NY, USA, 1992.
- [22] L. L. Scharf, *Statistical Signal Processing*, vol. 98, Addison-Wesley, Reading, Mass, USA, 1991.
- [23] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [24] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: algorithms and applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.
- [25] 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.
- [26] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5030–5043, 2010.
- [27] R. Baraniuk, M. A. Davenport, M. F. Duarte, and C. Hegde, *An Introduction to Compressive Sensing. Connexions e-Textbook*, 2011.
- [28] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, New York, NY, USA, 2010.
- [29] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.
- [30] E. J. Candès, 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] G. Pope, *Compressive sensing: a summary of reconstruction algorithms [Ph.D. dissertation]*, ETH, Swiss Federal Institute of Technology Zurich, Department of Computer Science, Zurich, Switzerland, 2009.
- [32] W. Abu-Dayyeh and M. S. Ahmed, "A double inequality for the tail probability of standard normal distribution," *Information and Optimization Sciences*, vol. 14, no. 2, pp. 155–159, 1993.
- [33] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.
- [34] J. Shao, *Mathematical Statistics*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1999.
- [35] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 2012.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

