This is very close to the final official version of the paper.

Compressive Detection of Sparse Signals in Additive White Gaussian Noise without Signal Reconstruction

Alireza Hariri, Massoud Babaie-Zadeh

School of Electrical Engineering, Sharif University of Technology, Tehran, Iran

Abstract

The main motivation behind compressive sensing is to reduce the sampling rate at the input of a digital signal processing system. However, if for processing the sensed signal one requires to reconstruct the corresponding Nyquist samples, then the data rate will be again high in the processing stages of the overall system. Therefore, it is preferred that the desired processing task is done directly on the compressive measurements, without the need for the reconstruction of the Nyquist samples. This paper addresses the case in which the processing task is "detection" (the existence) of a sparse signal in additive white Gaussian noise, with applications e.g. in radar systems. Moreover, we will propose two estimators for estimating the degree of sparsity of the detected signal. We will show that one of the estimators attains the Cramér-Rao lower bound of the problem.

Keywords: compressed sensing, detection-estimation, compressive sensing radar, generalized likelihood ratio test

1. Introduction

Detection of a sparse signal in additive white Gaussian noise (AWGN) from a small number of compressive measurements without signal reconstruction is an interesting subject that has received recent consideration [3], [11], [16], [2], [17]. This interest arises from a desire to avoid the difficulties caused by reconstructing the original signal for detection. Note that by using compressed sensing, the data rate is greatly reduced, that is, the rate of measurements is much lower

than the Nyquist rate. However, if for processing the measurements, one needs to reconstruct the original signal, (i.e. reconstruct the Nyquist samples) then the data rate is again very high. In other words, compressed sensing will not be very useful if the Nyquist samples are needed again to be reconstructed in some part of the system, and it is desirable to do the required process (sparse signal detection and parameter estimation) directly based on compressed measurements [4].

To the best of our knowledge, this problem has not yet been modeled using the sparse signal as a random process with an appropriate *a priori* probability density function (PDF). In [3], [2], [17], the signal is not random and there is no specific property in the signal modeling to express its sparsity. Authors of [16] exploit sparsity by performing the detection for all possible support patterns with fixed size on average. Therefore, the degree of sparsity (or equivalently, the support size) must be known exactly and this is difficult in practice.

In some studies (e.g. [11], [16]), it is assumed that the random signal has a Gaussian PDF, which is not a suitable distribution for modeling sparsity. Here, in contrast to these works, a sparsity-enforcing *a priori* PDF is assumed for the signal, and a new detection algorithm along with a sparsity degree estimation is proposed.

The paper is organized as follows. In Section 2, the problem of detecting a sparse signal in AWGN is modeled mathematically and a solution is provided for this model. In fact, by using a random model for the sparse signal and assuming that the number of Nyquist samples is large, the distribution of the measurements vector when the signal exits, can be approximated by a Gaussian distribution. This is achieved by using the central limit theorem (CLT). Moreover, since the sparsity degree of the signal is unknown, it is needed to rely on a generalized likelihood ratio test (GLRT) approach. In Section 3, the relationship between the signal to noise ratio (SNR) at the input of the detector and the SNR of the presented detector statistic is studied for high input SNRs. In Section 4, the performance of the proposed estimator for the sparsity degree is investigated and it is shown that the estimator attains the Cramér-Rao lower

bound (CRLB), that is, it is an efficient estimator. However, there is no guarantee that the estimated probability of activity is less than 1. Therefore, another estimator is proposed that is guaranteed to estimate this probability less than 1 but it does not attain the CRLB. Nevertheless, the variance of the estimator is close to the CRLB. Finally, Section 5 presents simulation results for detection of the sparse signal and estimation of the degree of sparsity.

Notations: For any vector $\mathbf{y}$, its transpose is $\mathbf{y}^T$. For any square matrix $\mathbf{A}$, its determinant is denoted by $|\mathbf{A}|$. The probability of an event A is represented by $\mathbb{P}\{A\}$. The expectation and covariance matrix of a random vector $\mathbf{y}$ are denoted by $\mathbb{E}\{\mathbf{y}\}$ and $\text{cov}\{\mathbf{y}\}$, respectively. Moreover, $\mathbf{I}_M$ stands for the $M \times M$ identity matrix.

2. Our problem modeling and solution

Detecting a sparse signal in AWGN is mathematically expressed as the hypothesis testing problem

$$\begin{cases} H_0 : \mathbf{y} = \mathbf{\Phi n} \\ H_1 : \mathbf{y} = \mathbf{\Phi}(\mathbf{s} + \mathbf{n}), \end{cases} \tag{1}$$

where $\mathbf{y}_{M\times 1}$ is the measurements vector, $\mathbf{\Phi}_{M\times N}$ is the compressed sensing measurement matrix where elements are drawn from a random Gaussian matrix with independent identically distributed (i.i.d.) elements having zero mean and unit variance (it should be noted that at the detector side, the matrix $\mathbf{\Phi}$ is randomly chosen and then kept fixed), and $M \ll N$. Moreover, $\mathbf{n}_{N\times 1}$ is the AWGN samples vector at the receiver with a normal distribution, having zero mean and variance σ_n^2 as $\mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I}_N)$, and $\mathbf{s}_{N\times 1}$ is the samples vector of the sparse signal. In this paper, σ_n^2 is *a priori* known (note that in practice, it can be measured in the absence of the sparse signal). In [4], detection of a target is studied for a simple radar application in which the radar signal is sparse.

To model the sparse signal, we assume that each element is independent of the others, and that it is "inactive" (i.e. nearly zero) with probability $1 - p$

having a normal distribution $\mathcal{N}(0, \sigma^2_{\text{off}})$, and it is "active" with probability p having a normal distribution $\mathcal{N}(0, \sigma^2_{\text{on}})$, where $\sigma^2_{\text{on}} \gg \sigma^2_{\text{off}}$. Here p is the sparsity degree of the signal and therefore is close to 0. The above assumptions result in $\mathbf{s}$ being an i.i.d. vector with elements each having a distribution $(1-p)\mathcal{N}(0, \sigma^2_{\text{off}}) + p\mathcal{N}(0, \sigma^2_{\text{on}})$. This is called the i.i.d. Bernoulli-Gaussian distribution, which has been frequently used in the literature to model sparsity [8], [13], [6], [12]. In this paper, p is not *a priori* known but σ_{on} and σ_{off} are. In fact, σ_{on} and σ_{off} can be pre-estimated by knowing the nature of the signal. For most of the sparse signals $\sigma_{\text{off}} = 0$ but for some of them (for example seismic signals [8]) this is not the case.

Let p_0 denote the *a priori* probability of the H_0 hypothesis. In order to derive the maximum *a posteriori* (MAP) detector, the likelihood ratio test (LRT) should be computed.

Observe that the distribution under hypothesis H_1 of test (1) is difficult to be obtained analytically for small number of samples. Fortunately, if we assume that N is sufficiently large, this distribution can be approximated by a multivariate Gaussian. Since $\mathbf{y}$ under H_1 is a linear transformation of a large random vector, the Lyapounov CLT can be used [1, Section 27] (for the sake of readability, this theorem is reviewed in Appendix A). This allows us to state the following theorem:

Theorem 1. *If it is assumed that the number of Nyquist samples (N) is sufficiently large, then the logarithm of the likelihood ratio test (LLRT) can be approximated by*

$$LLRT(\mathbf{y}|p) = \frac{M}{2} \ln\Big(\frac{\sigma_n^2}{\sigma_n^2 + \sigma^2(p)}\Big) + \frac{\sigma^2(p)}{2\sigma_n^2(\sigma_n^2 + \sigma^2(p))} \mathbf{y}^T \mathbf{A}^{-1} \mathbf{y} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{p_0}{1-p_0}, \quad (2)$$

where $\sigma^2(p) \triangleq p\sigma^2_{on} + (1-p)\sigma^2_{off}$ *and* $\mathbf{A} \triangleq \mathbf{\Phi}\mathbf{\Phi}^T$.

Proof. See Appendix B. □

As p is unknown, a GLRT should be used, i.e. $\text{LLRT}(\mathbf{y}|p)$ should be maximized over p to find the optimum value p_{opt}, and then compare $\text{GLR}(\mathbf{y}) =$

$\mathrm{LLRT}(\mathbf{y}|p_{\mathrm{opt}})$ with an appropriate threshold. After some straightforward manipulations (taking the derivative of $\mathrm{LLRT}(\mathbf{y}|p)$ with respect to p and finding its root), p_{opt} can be derived as

$$p_{\mathrm{opt}} = \frac{1}{\sigma_{\mathrm{on}}^2 - \sigma_{\mathrm{off}}^2}\Big(\frac{1}{M}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y} - \sigma_n^2 - \sigma_{\mathrm{off}}^2\Big). \tag{3}$$

It can be verified that at p_{opt}, the second derivative of the log-likelihood function is negative. In fact, it can be written as

$$\frac{d^2\big(\mathrm{LLRT}(\mathbf{y}|p)\big)}{dp^2}\bigg|_{p=p_{\mathrm{opt}}} = -\frac{M(\sigma_{\mathrm{on}}^2 - \sigma_{\mathrm{off}}^2)^2}{2(\sigma_n^2 + p_{\mathrm{opt}}\sigma_{\mathrm{on}}^2 + (1-p_{\mathrm{opt}})\sigma_{\mathrm{off}}^2)^2} < 0. \tag{4}$$

Finally, the GLRT detector is $\mathrm{GLR}(\mathbf{y}) \gtrless_{H_0}^{H_1} \frac{p_0}{1-p_0}$ which can be simplified as the following detector

$$t \underset{H_0}{\overset{H_1}{\gtrless}} \mathrm{Th}, \tag{5}$$

in which

$$t = \frac{1}{M\sigma_n^2}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y} - \ln(\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}), \tag{6}$$

$$\mathrm{Th} = \frac{2}{M}\ln\Big(\frac{p_0}{1-p_0}\Big) - \ln(M\sigma_n^2) + 1. \tag{7}$$

If the Cholesky factorization is used for the matrix $\mathbf{A}^{-1}$ as $\mathbf{A}^{-1} = \mathbf{L}\mathbf{L}^T$, $\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}$ can be rewritten as $\mathbf{w}^T\mathbf{w}$ where $\mathbf{w} \triangleq \mathbf{L}^T\mathbf{y}$. Therefore, (5) can be restated as

$$t_1 \underset{H_0}{\overset{H_1}{\gtrless}} \mathrm{Th}_1, \tag{8}$$

in which

$$t_1 = -\frac{M}{2}\ln(\mathbf{w}^T\mathbf{w}) - \frac{1}{2\mathbf{w}^T\mathbf{w}}\mathbf{w}^T\mathbf{w} - \Big(-\frac{M}{2}\ln(\sigma_n^2) - \frac{1}{2\sigma_n^2}\mathbf{w}^T\mathbf{w}\Big), \tag{9}$$

$$\mathrm{Th}_1 = \ln\Big(\frac{p_0}{1-p_0}\Big) - \frac{M}{2}\ln M + \frac{M-1}{2}. \tag{10}$$

The test statistic in (9), is a likelihood ratio between the variable that has its observed variance with respect to that variable having a variance σ_n^2, a classical detection in noise scenario, using the observed variance instead of an *a priori* known variance under H_1.

2.1. Detector after reconstruction

We present here, the detector based on the reconstructed signal, since later we will compare its performance with the detector obtained from test (1). In fact, after reconstruction, the hypothesis test will be

$$\begin{cases} H_0 : \hat{\mathbf{y}} = \mathbf{n} \\ H_1 : \hat{\mathbf{y}} = \mathbf{s} + \mathbf{n}. \end{cases} \tag{11}$$

where $\hat{\mathbf{y}}$ is the vector of the reconstructed measurements. Therefore

$$f(\hat{\mathbf{y}}|H_0) = \frac{1}{(2\pi\sigma_n^2)^{\frac{N}{2}}} \exp\Big\{ - \frac{\|\hat{\mathbf{y}}\|_2^2}{2\sigma_n^2} \Big\}. \tag{12}$$

Under H_1 hypothesis, each element of the vector $\hat{\mathbf{y}}$ is $\hat{y}_i = s_i + n_i$ for $i = 1, \ldots, N$ where s_i and n_i are the elements of the vectors $\mathbf{s}$ and $\mathbf{n}$, respectively. s_i and n_i are independent of each other. s_i has a normal distribution $\mathcal{N}(0, \sigma_{\text{on}}^2)$ with probability p and a normal distribution $\mathcal{N}(0, \sigma_{\text{off}}^2)$ with probability $1 - p$. n_i has a normal distribution $\mathcal{N}(0, \sigma_n^2)$. Therefore, $\hat{y}_i$ has a normal distribution $\mathcal{N}(0, \sigma_{\text{on}}^2 + \sigma_n^2)$ with probability p and has a normal distribution $\mathcal{N}(0, \sigma_{\text{off}}^2 + \sigma_n^2)$ with probability $1 - p$. In other words, $\hat{y}_i$ has a distribution $p\mathcal{N}(0, \sigma_{\text{on}}^2 + \sigma_n^2) + (1-p)\mathcal{N}(0, \sigma_{\text{off}}^2 + \sigma_n^2)$. On the other hand, $\hat{y}_i$ for $i = 1, \ldots, N$ are independent of each other because, the vectors $\mathbf{s}$ and $\mathbf{n}$ are both i.i.d. and independent of each other. Hence

$$\begin{aligned} f(\hat{\mathbf{y}}|H_1) = \prod_{i=1}^{N} \Bigg(& \frac{p}{\sqrt{2\pi(\sigma_{\text{on}}^2 + \sigma_n^2)}} \exp\Big\{ - \frac{|\hat{y}_i|^2}{2(\sigma_{\text{on}}^2 + \sigma_n^2)} \Big\} \\ & + \frac{1-p}{\sqrt{2\pi(\sigma_{\text{off}}^2 + \sigma_n^2)}} \exp\Big\{ - \frac{|\hat{y}_i|^2}{2(\sigma_{\text{off}}^2 + \sigma_n^2)} \Big\} \Bigg). \end{aligned} \tag{13}$$

Consequently, the logarithm of the likelihood ratio test is

$$\begin{aligned} \ln\Big(\frac{f(\hat{\mathbf{y}}|H_1)}{f(\hat{\mathbf{y}}|H_0)}\Big) = \sum_{i=1}^{N} \ln \Bigg(& \frac{p}{\sqrt{\sigma_{\text{on}}^2 + \sigma_n^2}} \exp\Big\{ - \frac{|\hat{y}_i|^2}{2(\sigma_{\text{on}}^2 + \sigma_n^2)} \Big\} \\ & + \frac{1-p}{\sqrt{\sigma_{\text{off}}^2 + \sigma_n^2}} \exp\Big\{ - \frac{|\hat{y}_i|^2}{2(\sigma_{\text{off}}^2 + \sigma_n^2)} \Big\} \Bigg) \\ & + N \ln \sigma_n + \frac{\|\hat{\mathbf{y}}\|_2^2}{2\sigma_n^2}. \end{aligned} \tag{14}$$

3. The SNR of the presented detector

It seems that obtaining a closed form formula for the receiver operating characteristic (ROC) of the presented detector would be too tricky. This is mainly due to the presence of the natural logarithm term in (6). Therefore, as a measure of the detector performance, the relationship between the SNR at the input of the detector and the SNR at the threshold comparison is computed. The signal at the input of the detector is $\mathbf{y}_{\text{in}} = \mathbf{s} + \mathbf{n}$, in which $\mathbf{s}$ is the signal term and $\mathbf{n}$ is the noise term. Therefore the signal power is $\mathbb{E}\{\|\mathbf{s}\|^2\} = N\sigma^2(p)$ and the noise power is $\mathbb{E}\{\|\mathbf{n}\|^2\} = N\sigma_n^2$. Hence, the input SNR will be

$$\text{SNR}_{\text{in}} = \frac{N\sigma^2(p)}{N\sigma_n^2} = \frac{\sigma^2(p)}{\sigma_n^2}. \tag{15}$$

The SNR at the threshold comparison is given in the following theorem.

Theorem 2. *If the input SNR is sufficiently high, the SNR at the threshold comparison is*

$$SNR_{det} \triangleq \frac{\mathbb{E}\{\textit{signal part in the detector statistic}\}}{\mathbb{E}\{\textit{noise part in the detector statistic}\}} = \frac{SP}{NP} = \frac{\sigma^2(p)}{\sigma_n^2}, \tag{16}$$

which is the same as the input SNR.

Proof. See Appendix C. □

Remark: In the above theorem, by sufficiently high (say above 7dB[1]), we mean so high that $\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}$ is large enough such that t (defined in (6)) can be approximated by its first term, that is, $t \cong \frac{1}{M\sigma_n^2}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}$.

4. Performance evaluation of the sparsity degree estimator

From (3), the proposed estimator for the sparsity degree is

$$\hat{p} = \frac{1}{\sigma_{\text{on}}^2 - \sigma_{\text{off}}^2}\Big(\frac{1}{M}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y} - \sigma_n^2 - \sigma_{\text{off}}^2\Big), \tag{17}$$

[1]For example, in radar applications, the SNR above 12dB is usually considered high.

where $\mathbf{y} = \mathbf{\Phi}(\mathbf{s} + \mathbf{n})$. Note that, if the following two parameters are defined as $\sigma_\Delta^2 \triangleq \sigma_{\text{on}}^2 - \sigma_{\text{off}}^2$, and $\sigma_\eta^2 \triangleq \sigma_{\text{off}}^2 + \sigma_n^2$, it is clear that $\sigma^2 + \sigma_n^2 = \sigma_\eta^2 + p\sigma_\Delta^2$ and hence, $\hat{p} = \frac{\frac{1}{M}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y} - \sigma_\eta^2}{\sigma_\Delta^2}$.

Theorem 3. *The proposed estimator* (17) *is unbiased and it attains the CRLB of the problem which equals* $\frac{2(\sigma^2(p)+\sigma_n^2)^2}{M(\sigma_{on}^2-\sigma_{off}^2)^2} = \frac{2}{M}\left(\frac{\sigma_\eta^2}{\sigma_\Delta^2} + p\right)^2$.

Proof. See Appendix D. □

From the above theorem, it is obvious that a small CRLB is obtained for large number of measurements (M), small sparsity degree (p) or when the difference between the active and inactive variances (σ_{on}^2 and σ_{off}^2) is much larger than the total noise variance (σ_η^2).

In the simulations, the estimator after reconstruction is needed for comparison. In fact, the reconstructed signal is $\hat{\mathbf{y}} = \mathbf{s}+\mathbf{n}$. As it was discussed at the end of Section 2, the reconstructed signal is a Gaussian mixture, since each element of the vector $\hat{\mathbf{y}}$ has the distribution $p\mathcal{N}(0, \sigma_{\text{on}}^2 + \sigma_n^2) + (1-p)\mathcal{N}(0, \sigma_{\text{off}}^2 + \sigma_n^2)$. Therefore, the parameters of the mixture including the mixture weights, i.e. p and $1-p$ can be estimated by using the expectation-maximization algorithm [7, Section 2.8].

Remark: There is no guarantee that $\hat{p} \leq 1$. However, as it is shown in Appendix E, we have $\hat{p} \leq 1$ with a probability very close to 1. In fact, in Appendix E a new estimator is proposed that is guaranteed to be less than 1. The new estimator is $\hat{p}_{\text{new}} = \frac{\min\{1,\hat{p}\}-1+\alpha}{\alpha}$ where α is the probability that $\hat{p} \leq 1$, that is, $\alpha \triangleq \mathbb{P}\{\hat{p} \leq 1\}$. The variance of the new estimator is $\sigma_{\hat{p}_{\text{new}}}^2 = \frac{1-\alpha}{\alpha}(1-p)^2 + \frac{1}{\alpha}\text{CRLB}$ in which CRLB is as given in Theorem 3. It is observed that the new proposed estimator does not attain the CRLB. However, it is a just little larger than the CRLB, since α is very close to 1 (for example for $p = 0.1$, $\sigma_{\text{on}} = 1$, $\sigma_{\text{off}} = 0.1$, $\text{SNR}_{\text{in}} = 7\text{dB}$, and $M = 66$, if $F(x, M)$ is defined as the cumulative density function (CDF) of a chi-square random variable with M degrees of freedom, we have $\alpha = F(515.76, 66)$ which is calculated as 1 under MATLAB precision and therefore $\sigma_{\hat{p}_{\text{new}}}^2 \cong \text{CRLB} = -32.77\text{dB}$). Another estimator can be $\hat{p}_1 = \min\{1, \hat{p}\}$, however, as it is shown in Appendix E, it is a

biased estimator, and its bias is equal to $\mathrm{B}(p) = \mathbb{E}\{\min\{1,\hat{p}\}\} - p = (1-\alpha)(1-p)$. Since α is very close 1, the bias is nearly zero. For the above mentioned biased estimator, the CRLB can be computed as $\Big(1 + \frac{d}{dp}\mathrm{B}(p)\Big)^2 / I(p)$ where $I(p) \triangleq -\mathbb{E}\Big\{\partial^2 \ln f(\mathbf{y};p)/\partial p^2\Big\}$ and $f(\mathbf{y};p)$ is the PDF of the measurements under the H_1 hypothesis which is a function of p [15, Problem 2.4.17]. $I(p)$ has been calculated in Appendix D (see (D.8)), from which, the CRLB for the biased estimator is obtained as

$$\mathrm{CRLB}_{\mathrm{biased}} \triangleq \frac{\Big(1 + \frac{d}{dp}\mathrm{B}(p)\Big)^2}{I(p)} = \frac{2\alpha^2(\sigma^2(p) + \sigma_n^2)^2}{M(\sigma_{\mathrm{on}}^2 - \sigma_{\mathrm{off}}^2)^2}. \tag{18}$$

Again since α is very close to 1, the above $\mathrm{CRLB}_{\mathrm{biased}}$ is very close to the CRLB given in Theorem 3. Moreover, from Appendix E, the variance and the mean square error (MSE) of the biased estimator can be calculated as

$$\begin{aligned}\mathrm{var}\{\hat{p}_1\} \triangleq \mathbb{E}\{\hat{p}_1^2\} - \mathbb{E}^2\{\hat{p}_1\} &= \mathbb{E}\{\min\{1,\hat{p}\}^2\} - \mathbb{E}^2\{\min\{1,\hat{p}\}\} \\ &= \alpha(1-\alpha)(1-p)^2 + \frac{2\alpha(\sigma^2(p) + \sigma_n^2)^2}{M(\sigma_{\mathrm{on}}^2 - \sigma_{\mathrm{off}}^2)^2}, \end{aligned} \tag{19}$$

$$\mathrm{MSE}\{\hat{p}_1\} \triangleq \mathrm{var}\{\hat{p}_1\} + \mathrm{B}^2(p) = (1-\alpha)(1-p)^2 + \frac{2\alpha(\sigma^2(p) + \sigma_n^2)^2}{M(\sigma_{\mathrm{on}}^2 - \sigma_{\mathrm{off}}^2)^2}. \tag{20}$$

As α is very close to 1, the variance and the MSE of the biased estimator is very close to the $\mathrm{CRLB}_{\mathrm{biased}}$, and consequently to the CRLB.

5. Results and simulations

In this section, seven simulations are conducted to experimentally evaluate the performance of the proposed algorithm for sparse signal detection and sparsity degree estimation. In all of these simulations, the following parameters are fixed

$$p = 0.1, \sigma_{\mathrm{off}} = 0.1, \sigma_{\mathrm{on}} = 1, p_0 = 0.5, N = 1320.$$

In what follows, for each ROC figure, the area under the curve (AUC) is given in a table. Moreover, by the compression ratio (CR) we mean the fraction $\frac{N}{M}$.

Experiments 1 and 2. performance evaluation for the low input SNRs

In these experiments, the presented detector performance is studied and compared with the traditional detector at low input SNRs (by "traditional detector", we mean the detector that first reconstructs the signal from its compressive measurements and then uses the GLRT traditional detector explained in Eqs. (11) to (14)). The GLRT detector statistic in (6) is probably too tricky to be simplified and obtain a closed form formula for the ROC. This is due to the presence of the natural logarithm term in (6). Therefore, two simulations are performed to evaluate the detector performance at low input SNRs empirically.

At first, the ROC curves[2] are plotted empirically for the CRs 20 and 80 at the input SNR 3dB in Fig. 1. As is observed, the proposed detector performance is acceptable for the CRs lower than around 20: the signal can be detected with a probability near 1, while using only about 5% of the number of measurements as for the Nyquist samples. For comparison, the ROC curves for the CS method with reconstruction at $\mathrm{SNR}_{\mathrm{in}} = 3\mathrm{dB}$ are also plotted. The OMP algorithm [14] is used for CS reconstruction. As can be seen, again the detection is good for the CRs lower than around 20 (the detection probability is near 0.9 for all false alarm probabilities). Because the input SNR is low, the reconstruction algorithm does not work as well as the proposed detector. Therefore, the performance of the detector after reconstruction is degraded and to have a better performance, the CR should be decreased.

In the second simulation, the ROC curves are derived for the low input SNRs 0dB and 6dB at CR = 40 empirically and depicted in Fig. 2. From the figure, it is obvious that the proposed detector performance is reasonable (the detection probability is near 1) for the input SNR 6dB. While this is a good result, the CR must be decreased if it is necessary to perform the detection at the

[2]In the simulations, the ROC curves are obtained as follows: For a fixed probability of false alarm (p_f), a threshold (Th) is obtained for the detector in (5). From the determined threshold, the value of the probability of detection (p_d) is calculated. Therefore, by sweeping p_f, different values of p_d and hence, multiple points in the ROC curve are achieved.

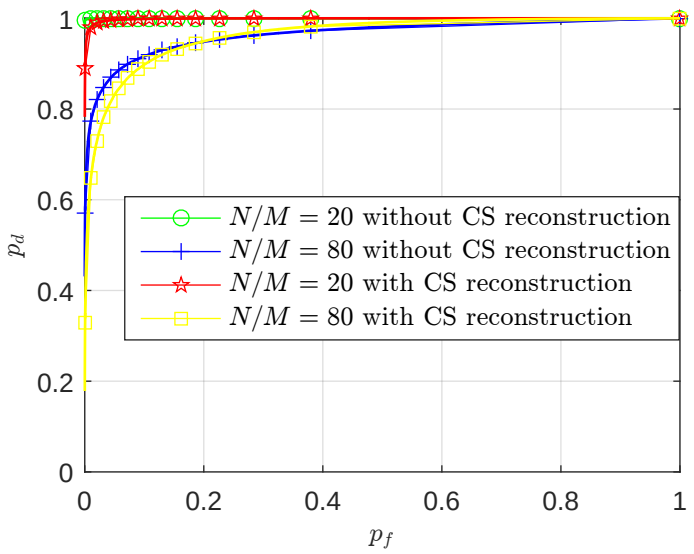

Figure 1: The empirical ROC curves for the CRs 20 and 80 at fixed $\mathrm{SNR_{in}} = 3$dB (the low input SNR).

Curve	AUC
$N/M = 20$ without CS reconstruction	0.9999
$N/M = 80$ without CS reconstruction	0.9653
$N/M = 20$ with CS reconstruction	0.9989
$N/M = 80$ with CS reconstruction	0.9635

Table 1: AUC for the curves in Fig. 1

lower input SNRs. In fact, there is a tradeoff between the maximum reachable CR and the minimum acceptable input SNR for a tolerable detector quality. Therefore, these factors should be chosen carefully if the desired performance is to be achieved. For comparison, the ROC curves for the CS method with reconstruction at CR = 40 are also plotted. It can again be seen that, to have an acceptable detection performance, the input SNR should be increased.

Experiments 3, 4 and 5. performance evaluation for the high input SNRs

In these experiments, the presented detector performance is studied and

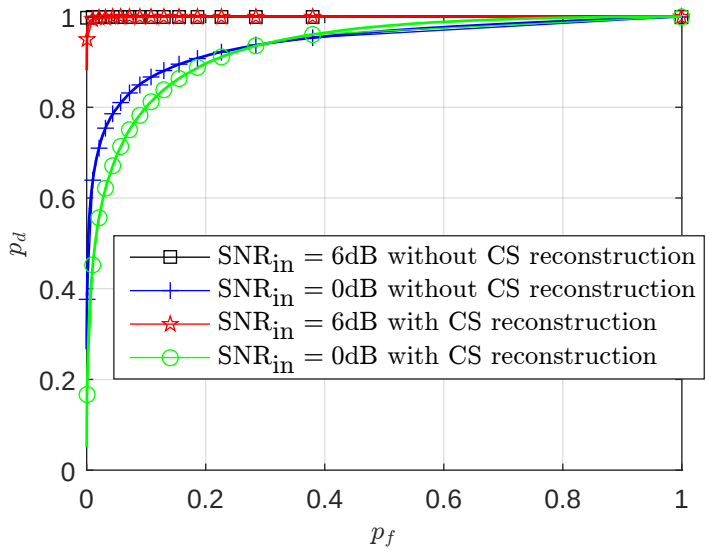

Figure 2: The empirical ROC curves for the low input SNRs 0dB and 6dB for CR = 40.

Curve	AUC
$\mathrm{SNR_{in}}$ = 6dB without CS reconstruction	0.9999
$\mathrm{SNR_{in}}$ = 0dB without CS reconstruction	0.9425
$\mathrm{SNR_{in}}$ = 6dB with CS reconstruction	0.9995
$\mathrm{SNR_{in}}$ = 0dB with CS reconstruction	0.9321

Table 2: AUC for the curves in Fig. 2

compared with the traditional detector at the high input SNRs.

In the first experiment, the ROC curves are plotted for the CRs 80 and 320 at the fixed input SNR 11dB (Fig. 3). It can be seen that for CR = 80, the proposed detector has a high performance quality, i.e. the detection probability is always near 1 even at small probabilities of false alarm. Also, the ROC curves are plotted for the same settings, when CS reconstruction is used. As seen, for small false alarm probabilities the detector performance is degraded. This is mainly because, the number of measurements is small. Moreover, for the low input SNRs (Fig. 1), although the number of measurements is high, at small false alarm probabilities the performance is degraded even more, but this is due

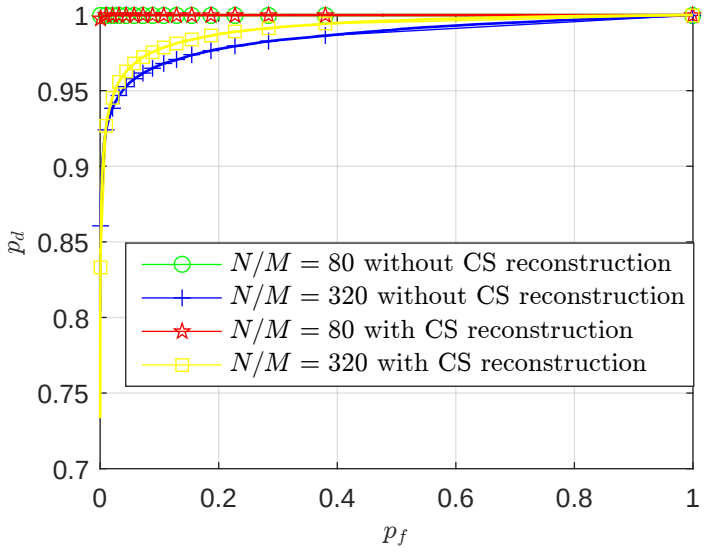

Figure 3: The ROC curves for the CRs 80 and 320 at fixed $\text{SNR}_{\text{in}} = 11\text{dB}$ (the high input SNR).

Curve	AUC
$N/M = 80$ without CS reconstruction	0.99990
$N/M = 320$ without CS reconstruction	0.9854
$N/M = 80$ with CS reconstruction	0.99989
$N/M = 320$ with CS reconstruction	0.9911

Table 3: AUC for the curves in Fig. 3

to the low-SNR reconstruction.

The second simulation is performed to examine the ROC curves for the high input SNRs 8dB and 14dB at fixed CR = 160. The results are shown in Fig. 4. As expected, for the proposed detector the detection probability is above 0.95 for the input SNR 14dB. In fact, by increasing the input SNR above around 14dB, the performance of the detector does not improve much. The reason is the high compression ratio. For the same settings but with CS reconstruction, the ROC curves are also plotted. It is again seen that, for all of the SNRs, the detector performance at low probabilities of false alarm deteriorates. In fact, in

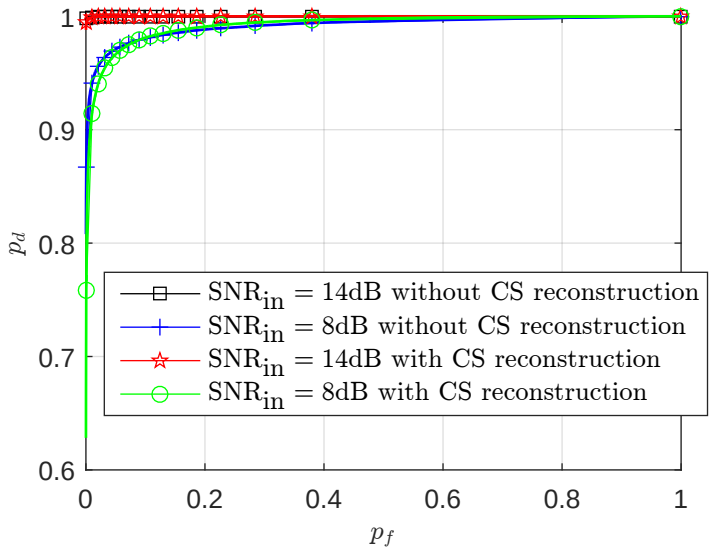

Figure 4: The ROC curves for the high input SNRs 8dB and 14dB for CR = 160.

Curve	AUC
$\mathrm{SNR_{in}}$ = 14dB without CS reconstruction	0.999843
$\mathrm{SNR_{in}}$ = 8dB without CS reconstruction	0.9919
$\mathrm{SNR_{in}}$ = 14dB with CS reconstruction	0.999839
$\mathrm{SNR_{in}}$ = 8dB with CS reconstruction	0.9924

Table 4: AUC for the curves in Fig. 4

order to improve the performance of the detector after reconstruction, both the input SNR and the number of measurements should be increased. Note that, the tradeoff between the input SNR and the CR is again seen in Experiments 3 and 4.

In the third experiment, the behavior of the detection probability versus the CR is investigated for different input SNRs. Here the false alarm probability is fixed as $p_f = 10^{-4}$. From Fig. 5, it is seen that, for the proposed detector, in order to have a probability of detection greater than 0.9, for input SNRs lower than 12dB, the CR should be lower than 20, and for input SNRs greater than 6dB, the CR should be lower than 320. In fact, as the input SNR de-

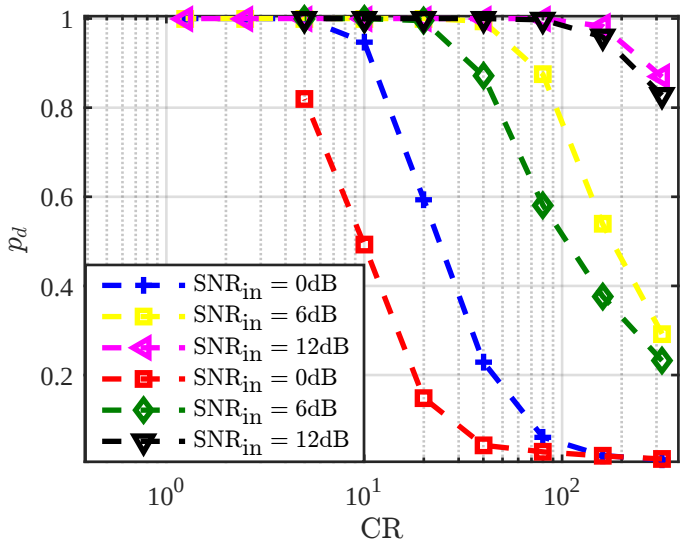

Figure 5: The probability of detection (p_d) versus the CR for different input SNRs at fixed probability of false alarm $p_f = 10^{-4}$. The first three legends are related to the simulated curves without CS reconstruction and the second three legends are related to the simulated curves with CS reconstruction.

creases, much more measurements is needed to have an acceptable probability of detection. Overall, the detector performance is so good that for the input SNRs greater than 0dB the number of measurements can be decreased to as low as about 2.5% of the Nyquist samples and still produce a high performance quality. For comparison, SNR curves are depicted for the same settings, when CS reconstruction is used. It is seen that at high input SNRs and at low CRs, the performances of the detector after reconstruction and the proposed detector are similar.

Experiment 6. Computational cost

In this experiment, the computational cost of the presented detector is studied and compared with the detector after reconstruction (here, for the reconstruction stage, two algorithms have been used, namely OMP as a greedy algorithm and basis pursuit de-noising (BPDN) as a penalization algorithm). In fact, the averaged CPU time over 100 simulations is used as a rough measure

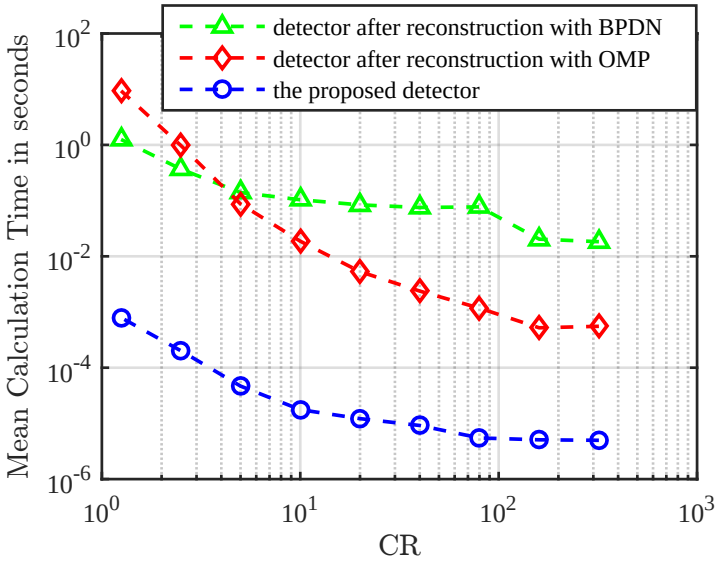

Figure 6: The mean calculation time in seconds versus the CR for the presented detector and also the detector after reconstruction.

of the detection algorithm complexity for different CRs.[3]

The result is depicted in Fig. 6. It is seen that the presented detector is always 2 degrees of magnitude better than the detector after reconstruction with either OMP or BPDN algorithm. This is because, the reconstruction process is highly computational. It should be noted that, for low CRs, the difference in the mean calculation time between the detectors is larger because more measurements are obtained.

We re-emphasize here that the main added-value of our approach is not decreasing the computational load, but it is keeping the data rate small in all parts of the system (because we are working directly with compressive measurements, without the need for reconstructing high rate Nyquist samples).

[3]The simulations are performed in MATLAB R2015b environment using Intel (R) Core (TM) i7-4720HQ @ 2.60GHz processor with 16GB of memory, and under 64 bit Microsoft Windows 10 operating system.

Experiment 7. Sparsity estimation accuracy

In this simulation, the accuracy of the estimation of the signal sparsity is studied. Here, the simulation is run 100 times. The MSE of the estimated sparsity for the two proposed estimators and also for the estimators after reconstruction with OMP and BPDN, and the CRLB are computed for the two input SNRs 3dB and 11dB. As shown in Figs. 7 and 8, the proposed estimators approximately attain the CRLB (It is notable that the two proposed estimators have nearly the same MSE). Moreover, by increasing the SNR and decreasing the CR, the MSE of the estimation decreases. This can be deduced from (D.2) in which by increasing the number of measurements (M) and decreasing the noise variance (σ_n^2), the variances of the estimators decrease. Therefore, although the detection performance is of high quality for the high CRs, more measurements are needed to provide a good estimation of signal sparsity. Also from Figs. 7 and 8, it is observed that, the MSE of the estimators after reconstruction is nearly 2 dB and 1 dB greater than the CRLB for the low and high input SNRs, respectively. Moreover, for very high CRs, the difference between the MSE and the CRLB is more. This is because, for very high CRs the reconstruction is not as efficient as for the low CRs.

6. Conclusion

In this paper, it has been shown that the detection of a sparse signal in AWGN and the estimation of its degree of sparsity can be accomplished using compressive measurements without signal reconstruction. In fact, performing the detection and estimation directly in the compressive space allows to keep the rate of measurements low through all the parts of the digital processing system. Here, the proposed approach for deriving the detector and the estimators is related to the signal sparsity implicitly. In fact, in order to maintain the information of the signal in the dimensionality reduction with random projections, the signal should be sparse. For the proposed detector, by using a CR as high as 20 for the low SNRs and as high as 80 for the high SNRs, the *detection*

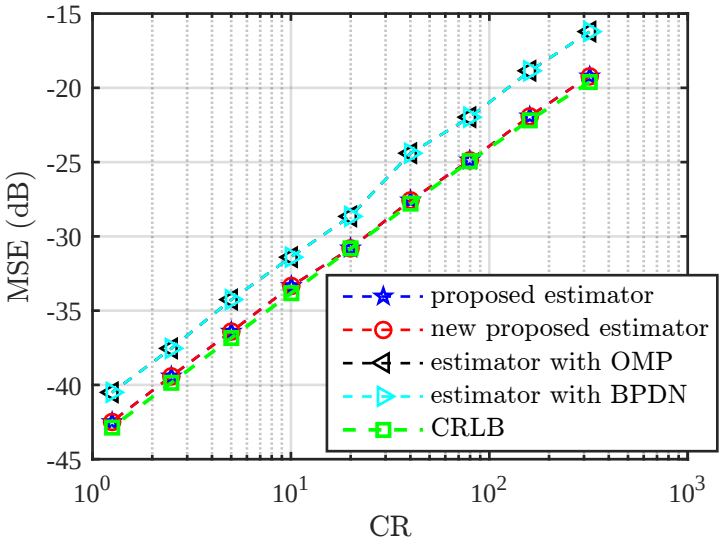

Figure 7: The mean square error of the estimated sparsity for the two proposed estimators and the CRLB in dB versus the CR for the input SNR 3dB (the low input SNR).

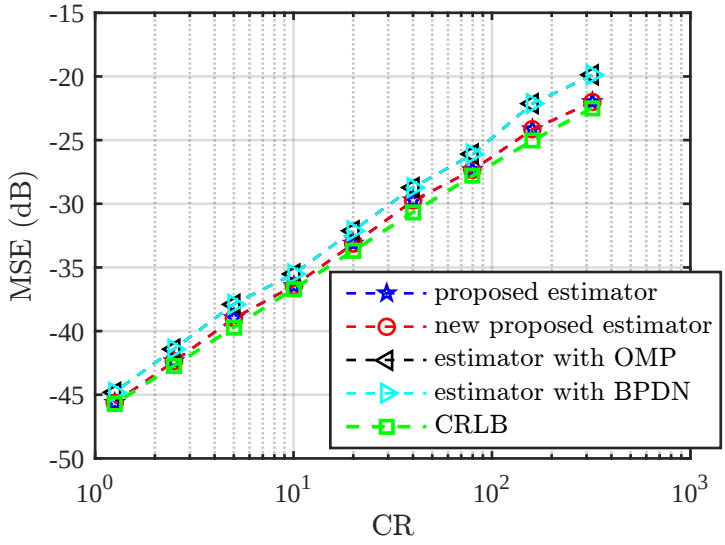

Figure 8: The mean square error of the estimated sparsity for the two proposed estimators and the CRLB in dB versus the CR for the input SNR 11dB (the high input SNR).

performance is very high (the probability of detection is always near 1). This is in contrast to what is observed in standard detection of sparse signals where, for

both the low and high SNRs, acceptable results are obtained only for the CRs less than 20. Moreover, in regard to the complexity, the presented detector is at least 2 degrees of magnitude better than the detector after reconstruction. For the proposed estimators, however, in order to produce an acceptable *estimation* quality for both the low and high SNRs, a CR less than 20 is needed.

Appendix A. Lyapounov CLT [1, Section 27]

Suppose that, the sequence of random variables $\{x_1, x_2, \dots\}$ satisfies the following two conditions:

Condition 1: All of these random variables are independent, each with finite mean μ_i and variance σ_i^2.

Condition 2: For some $\delta > 0$:

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}\big\{|x_i - \mu_i|^{2+\delta}\big\} = 0, \tag{A.1}$$

where $s_n^2 \triangleq \sum_{i=1}^{n} \sigma_i^2$.

Then, the sum of $\frac{x_i - \mu_i}{s_n}$ converges in distribution to a standard normal random variable as n goes to infinity, i.e.

$$\frac{1}{s_n} \sum_{i=1}^{n} (x_i - \mu_i) \to \mathcal{N}(0, 1). \tag{A.2}$$

In practice, it is sufficient to check the second condition for $\delta = 1$.

Appendix B. Proof of Theorem 1

To calculate the LRT, the PDFs of the measurements vector should be found conditioned on the H_0 and H_1 hypotheses. This is done in the following two lemmas.

Lemma 1. *The measurements vector distribution conditioned on H_0 is normal $\mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{\Phi}\mathbf{\Phi}^T)$, i.e.*

$$f(\mathbf{y}|H_0) = \frac{1}{(2\pi\sigma_n^2)^{\frac{M}{2}} |\mathbf{A}|^{\frac{1}{2}}} \exp\Big\{ -\frac{1}{2\sigma_n^2} \mathbf{y}^T \mathbf{A}^{-1} \mathbf{y} \Big\}, \tag{B.1}$$

where $\mathbf{A} \triangleq \mathbf{\Phi}\mathbf{\Phi}^T$.

Proof. Because the noise vector has a normal distribution, the measurements vector distribution conditioned on H_0 is also normal with covariance matrix $\mathbf{A} = \mathbf{\Phi}\mathbf{\Phi}^T$ [9, Section 6.2]. □

Lemma 2. *If it is assumed that the number of Nyquist samples (N) is sufficiently large, then the PDF of the measurements vector conditioned on the* H_1 *hypothesis will be a Gaussian distribution with zero mean and covariance matrix* $(\sigma^2(p) + \sigma_n^2)\mathbf{A}$ *where* $\sigma^2(p) \triangleq p\sigma_{on}^2 + (1-p)\sigma_{off}^2$, *i.e.*

$$f(\mathbf{y}|H_1, p) = \mathcal{N}(\mathbf{0}, (\sigma^2(p) + \sigma_n^2)\mathbf{A}). \tag{B.2}$$

Proof. In order to find the PDF of the measurements vector conditioned on the H_1 hypothesis, i.e. $f(\mathbf{y}|H_1)$, at first $f(\mathbf{z}|H_1)$ where $\mathbf{z} \triangleq \mathbf{\Phi}\mathbf{s}$ should be computed. Let $z_i, i = 1, 2, \ldots, M$ and $s_i, i = 1, 2, \ldots, N$ denote the i-th element of $\mathbf{z}$ and $\mathbf{s}$, respectively. Then,

$$z_k = \sum_{i=1}^{N} \phi_{ki} s_i = \sum_{i=1}^{N} t_{ki}, \tag{B.3}$$

in which ϕ_{ki} stands for the (k, i)-th element of Φ, and $t_{ki} \triangleq \phi_{ki} s_i$. The mean and variance of each term in the above summation can be computed as

$$\mu_{ki} \triangleq \mathbb{E}\{t_{ki}\} = \phi_{ki}\mathbb{E}\{s_i\} = 0 \tag{B.4}$$

$$\begin{aligned} \sigma_{ki}^2 \triangleq \mathrm{var}\{t_{ki}\} &= \phi_{ki}^2\big(p\sigma_{\mathrm{on}}^2 + (1-p)\sigma_{\mathrm{off}}^2\big) \\ &= \phi_{ki}^2\sigma^2(p), \end{aligned} \tag{B.5}$$

where $\sigma^2(p) \triangleq p\sigma_{\mathrm{on}}^2 + (1-p)\sigma_{\mathrm{off}}^2$.

It is now verified that the random variables t_{ki} fulfill both conditions of Lyapounov CLT theorem (expressed in Appendix A): $\{t_{ki}\}_{i=1}^N$ are independent because $\{s_i\}_{i=1}^N$ are independent, moreover, for each i, the mean and variance of t_{ki} are finite (condition 1). By defining x_{nk} as $x_{nk}^2 \triangleq \sum_{i=1}^n \sigma_{ki}^2$ for each $n = 1, 2, \ldots$, it can be verified that for $\delta = 1$, condition 2 is satisfied, i.e.

$\lim_{n\to\infty} \frac{1}{x_{nk}^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}\{|t_{ki}-\mu_{ki}|^{2+\delta}\} = 0$. In other words, for $\delta = 1$,

$$\begin{aligned}\mathbb{E}\{|t_{ki}-\mu_{ki}|^{2+\delta}\} &= |\phi_{ki}|^3 \mathbb{E}\{|s_i|^3\} \\ &= 2\sqrt{\frac{2}{\pi}}|\phi_{ki}|^3\big(p\sigma_{\text{on}}^3 + (1-p)\sigma_{\text{off}}^3\big),\end{aligned} \tag{B.6}$$

where the third order moments of normal random variables have been calculated using [9, Section 5.4]. Therefore

$$\begin{aligned}&\lim_{n\to\infty} \frac{1}{x_{nk}^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}\{|t_{ki}-\mu_{ki}|^{2+\delta}\} \\ &\quad= 2\sqrt{\frac{2}{\pi}} \frac{p\sigma_{\text{on}}^3 + (1-p)\sigma_{\text{off}}^3}{\sigma^3(p)} \lim_{n\to\infty} \frac{\sum_{i=1}^{n} |\phi_{ki}|^3}{\big(\sum_{i=1}^{n} |\phi_{ki}|^2\big)^{\frac{3}{2}}} \qquad \text{(B.7)} \\ &\quad= 2\sqrt{\frac{2}{\pi}} \frac{p\sigma_{\text{on}}^3 + (1-p)\sigma_{\text{off}}^3}{\sigma^3(p)} \lim_{n\to\infty} \frac{n\big(2\sqrt{\frac{2}{\pi}}\big)}{n^{\frac{3}{2}}} = 0, \qquad \text{(B.8)}\end{aligned}$$

where the law of large numbers (LLN) [9, Section 7.4] has been used for the calculation of the limit in the above equation.

Since the two conditions of the Lyapounov CLT are satisfied, the sum of $\frac{t_{ki}-\mu_{ki}}{x_{nk}}$ converges in distribution to a standard normal random variable as n goes to infinity which means that $\sum_{i=1}^{n} t_{ki} \xrightarrow{n\to\infty} \mathcal{N}(0, x_{nk}^2)$. Therefore, from the assumption that N is sufficiently large, we have

$$\lim_{n\to\infty} \sum_{i=1}^{n} t_{ki} \approx \sum_{i=1}^{N} t_{ki} = z_k, \tag{B.9}$$

and

$$p(z_k) = \mathcal{N}(0, x_{Nk}^2), \tag{B.10}$$

in which

$$x_{Nk}^2 \triangleq \sum_{i=1}^{N} \sigma_{ki}^2 = \sigma^2(p) \sum_{i=1}^{N} \phi_{ki}^2 = \sigma^2(p)\|\mathbf{\Phi}_k\|^2, \tag{B.11}$$

and $\mathbf{\Phi}_k^T$ is the k-th row of $\mathbf{\Phi}$. Since each z_k is a normal random variable, the vector $\mathbf{z}$ has a jointly normal distribution. Therefore, in order to derive its PDF, it is enough to find its mean vector and covariance matrix as

$$\begin{aligned}&\mathbb{E}\{z_k\} = \sum_{i=1}^{N} \mathbb{E}\{t_{ki}\} = 0 \quad \text{for} \quad k = 1, \ldots, M \\ &\Rightarrow \mathbb{E}\{\mathbf{z}\} = \mathbf{0},\end{aligned} \tag{B.12}$$

and

$$\mathbf{R} = [r_{jl}]_{M\times M} \triangleq \text{cov}\{\mathbf{z}\} = \mathbb{E}\{\mathbf{z}\mathbf{z}^T\}, \tag{B.13}$$

where

$$\begin{aligned} r_{jl} = \mathbb{E}\{z_j z_l\} &= \sum_{i=1}^{N}\sum_{u=1}^{N} \phi_{ji}\phi_{lu}\mathbb{E}\{s_i s_u\} \\ &= \sum_{i=1}^{N}\sum_{u=1}^{N} \phi_{ji}\phi_{lu}\sigma^2(p)\delta(i-u) = \sigma^2(p)\mathbf{\Phi}_j^T\mathbf{\Phi}_l, \end{aligned} \tag{B.14}$$

in which δ is the Kronecker delta function.

From (B.13) and (B.14), it can be deduced that $\mathbf{R} = \sigma^2(p)\mathbf{A}$, which means that $p(\mathbf{z}|H_1) = \mathcal{N}(\mathbf{0}, \sigma^2(p)\mathbf{A})$. Conditioned on H_1, $\mathbf{y} = \mathbf{z} + \mathbf{\Phi}\mathbf{n}$. As $\mathbf{z}$ and $\mathbf{n}$ are independent of each other and both are normal with zero mean, $\mathbf{y}$ is also normal with zero mean, and covariance matrix $\text{cov}\{\mathbf{y}\} = \text{cov}\{\mathbf{z}\} + \text{cov}\{\mathbf{\Phi}\mathbf{n}\} = \sigma^2(p)\mathbf{A} + \sigma_n^2\mathbf{A} = (\sigma^2(p) + \sigma_n^2)\mathbf{A}$. Therefore, $f(\mathbf{y}|H_1, p)$ will be [9, Section 6.2]

$$\begin{aligned} f(\mathbf{y}|H_1,p) = &\frac{1}{\left(2\pi(\sigma_n^2+\sigma^2(p))\right)^{\frac{M}{2}}|\mathbf{A}|^{\frac{1}{2}}} \\ &\times \exp\Big\{-\frac{1}{2(\sigma_n^2+\sigma^2(p))}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}\Big\}. \end{aligned} \tag{B.15}$$

□

From Lemma 1 and Lemma 2, the LLRT can be computed as

$$\begin{aligned} \text{LLRT}(\mathbf{y}|p) &= \ln\Big(\frac{f(\mathbf{y}|H_1,p)}{f(\mathbf{y}|H_0)}\Big) \\ &= \ln\Bigg(\Big(\frac{\sigma_n^2}{\sigma_n^2+\sigma^2(p)}\Big)^{\frac{M}{2}} \\ &\quad\times\exp\Big\{\Big(\frac{1}{2\sigma_n^2}-\frac{1}{2(\sigma_n^2+\sigma^2(p))}\Big)\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}\Big\}\Bigg) \\ &= \frac{M}{2}\ln\Big(\frac{\sigma_n^2}{\sigma_n^2+\sigma^2(p)}\Big) + \frac{\sigma^2(p)}{2\sigma_n^2(\sigma_n^2+\sigma^2(p))}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{p_0}{1-p_0}. \end{aligned} \tag{B.16}$$

Appendix C. Proof of Theorem 2

As derived in (6), the output statistic for GLRT is

$$t = \frac{1}{M\sigma_n^2}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y} - \ln(\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}). \tag{C.1}$$

If the input SNR is sufficiently high, $\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}$ is large. Hence, in the above equation, the first term is dominant. Therefore, it can be written as

$$\begin{aligned} t \cong \frac{1}{M\sigma_n^2}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y} &= \frac{1}{M\sigma_n^2}(\mathbf{\Phi}(\mathbf{s}+\mathbf{n}))^T\mathbf{A}^{-1}\mathbf{\Phi}(\mathbf{s}+\mathbf{n}) \\ &= \frac{1}{M\sigma_n^2}\big((\mathbf{\Phi}\mathbf{s})^T\mathbf{A}^{-1}\mathbf{\Phi}\mathbf{s} + 2(\mathbf{\Phi}\mathbf{s})^T\mathbf{A}^{-1}\mathbf{\Phi}\mathbf{n} \\ &\quad + (\mathbf{\Phi}\mathbf{n})^T\mathbf{A}^{-1}\mathbf{\Phi}\mathbf{n}\big). \end{aligned} \tag{C.2}$$

In the above equation, the signal term is $\frac{1}{M\sigma_n^2}(\mathbf{\Phi}\mathbf{s})^T\mathbf{A}^{-1}\mathbf{\Phi}\mathbf{s}$ and the noise term is $\frac{1}{M\sigma_n^2}\big(2(\mathbf{\Phi}\mathbf{s})^T\mathbf{A}^{-1}\mathbf{\Phi}\mathbf{n} + (\mathbf{\Phi}\mathbf{n})^T\mathbf{A}^{-1}\mathbf{\Phi}\mathbf{n}\big)$. Therefore, the signal power (SP) is

$$\begin{aligned} \text{SP} &= \mathbb{E}\Big\{\frac{1}{M\sigma_n^2}(\mathbf{\Phi}\mathbf{s})^T\mathbf{A}^{-1}\mathbf{\Phi}\mathbf{s}\Big\} \\ &= \frac{\sigma^2(p)}{M\sigma_n^2}\mathbb{E}\Big\{\frac{1}{\sigma^2(p)}\mathbf{z}^T\mathbf{A}^{-1}\mathbf{z}\Big\}, \end{aligned} \tag{C.3}$$

in which $\mathbf{z} \triangleq \mathbf{\Phi}\mathbf{s}$. If $\mathbf{x}_{K\times 1}$ is $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, then $\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}$ has a chi-square distribution with K degrees of freedom denoted by $\chi^2(x, K)$ [10]. It is shown in Appendix B that $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \sigma^2(p)\mathbf{A})$. Therefore, $\frac{1}{\sigma^2(p)}\mathbf{z}^T\mathbf{A}^{-1}\mathbf{z}$ has a chi-square distribution with M degrees of freedom. The mean and variance of a random variable c having a chi-square distribution with m degrees of freedom are [9, Section 5.5]

$$\mathbb{E}\{c\} = m, \tag{C.4}$$

$$\sigma_c^2 = 2m. \tag{C.5}$$

Hence, (C.3) can be computed as

$$\text{SP} = \frac{\sigma^2(p)}{M\sigma_n^2}M = \frac{\sigma^2(p)}{\sigma_n^2}. \tag{C.6}$$

Moreover, the noise power (NP) can be calculated as

$$\begin{aligned} \text{NP} = \frac{1}{M\sigma_n^2}\Big(&2\mathbb{E}\{(\mathbf{\Phi}\mathbf{s})^T\mathbf{A}^{-1}\mathbf{\Phi}\mathbf{n}\} \\ &+ \mathbb{E}\{(\mathbf{\Phi}\mathbf{n})^T\mathbf{A}^{-1}\mathbf{\Phi}\mathbf{n})\}\Big). \end{aligned} \tag{C.7}$$

In the above equation, two expectations should be computed. The first expectation (E_1) can be calculated as

$$\begin{aligned}\mathrm{E}_1 &= \mathbb{E}\{(\mathbf{\Phi s})^T\mathbf{A}^{-1}\mathbf{\Phi n}\}\\ &= \mathbb{E}\{\mathbf{s}^T\}\mathbf{\Phi}^T\mathbf{A}^{-1}\mathbf{\Phi}\mathbb{E}\{\mathbf{n}\} = 0.\end{aligned} \tag{C.8}$$

For computing the second expectation in (C.7), that we denote by E_2, it should be noted that $p(\mathbf{\Phi n}) = \mathcal{N}(\mathbf{0}, \sigma_n^2\mathbf{A})$. Therefore, $\frac{1}{\sigma_n^2}(\mathbf{\Phi n})^T\mathbf{A}^{-1}\mathbf{\Phi n}$ has a chi-square distribution with M degrees of freedom [10]. Hence, from (C.4), E_2 can be computed as

$$\begin{aligned}\mathrm{E}_2 &= \mathbb{E}\Big\{(\mathbf{\Phi n})^T\mathbf{A}^{-1}\mathbf{\Phi n}\Big\}\\ &= M\sigma_n^2.\end{aligned} \tag{C.9}$$

The total noise power will be

$$\mathrm{NP} = 1. \tag{C.10}$$

From (C.6) and (C.10), it can be concluded that the SNR at the threshold comparison is

$$\mathrm{SNR}_{\mathrm{det}} \triangleq \frac{\mathrm{SP}}{\mathrm{NP}} = \frac{\sigma^2(p)}{\sigma_n^2}, \tag{C.11}$$

which is the same as the input SNR in (15).

Appendix D. Proof of Theorem 3

It was shown in Section 2 that $p(\mathbf{y}) = \mathcal{N}(\mathbf{0}, (\sigma^2(p) + \sigma_n^2)\mathbf{A})$. Therefore, $\frac{1}{\sigma^2(p)+\sigma_n^2}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}$ has a chi-square distribution with M degrees of freedom [10]. In the following equations, (C.4) and (C.5) have been used.

In order to evaluate the estimator performance, it should be noted that it is unbiasd because,

$$\begin{aligned}\mathbb{E}\{\hat{p}\} &= \frac{1}{\sigma_{\mathrm{on}}^2 - \sigma_{\mathrm{off}}^2}\Big(\frac{1}{M}\mathbb{E}\{\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}\} - \sigma_n^2 - \sigma_{\mathrm{off}}^2\Big)\\ &= \frac{1}{\sigma_{\mathrm{on}}^2 - \sigma_{\mathrm{off}}^2}\Big(\frac{1}{M}(\sigma^2(p) + \sigma_n^2)M - \sigma_n^2 - \sigma_{\mathrm{off}}^2\Big)\\ &= \frac{\sigma^2(p) - \sigma_{\mathrm{off}}^2}{\sigma_{\mathrm{on}}^2 - \sigma_{\mathrm{off}}^2} = p.\end{aligned} \tag{D.1}$$

In what follows, the variance of the estimator is computed.

$$
\begin{aligned}
\sigma_{\hat{p}}^2 &= \mathbb{E}\{\hat{p}^2\} - p^2 \\
&= \frac{1}{(\sigma_{\text{on}}^2 - \sigma_{\text{off}}^2)^2}\Big(\frac{1}{M^2}\mathbb{E}\{(\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y})^2\} + (\sigma_n^2 + \sigma_{\text{off}}^2)^2 \\
&\qquad\qquad - \frac{2}{M}(\sigma_n^2 + \sigma_{\text{off}}^2)\mathbb{E}\{\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}\}\Big) - p^2 \\
&= \frac{1}{(\sigma_{\text{on}}^2 - \sigma_{\text{off}}^2)^2}\Big(\frac{(\sigma^2(p) + \sigma_n^2)^2}{M^2}(M^2 + 2M) + (\sigma_n^2 + \sigma_{\text{off}}^2)^2 \\
&\qquad\qquad - \frac{2}{M}(\sigma_n^2 + \sigma_{\text{off}}^2)(\sigma^2(p) + \sigma_n^2)M\Big) - p^2 \\
&= \frac{1}{(\sigma_{\text{on}}^2 - \sigma_{\text{off}}^2)^2}\Big((\sigma^2(p) + \sigma_n^2)^2 + (\sigma_n^2 + \sigma_{\text{off}}^2)^2 \\
&\qquad\qquad - 2(\sigma_n^2 + \sigma_{\text{off}}^2)(\sigma^2(p) + \sigma_n^2) \\
&\qquad\qquad + \frac{2}{M}(\sigma^2(p) + \sigma_n^2)^2\Big) - p^2 \\
&= \frac{1}{(\sigma_{\text{on}}^2 - \sigma_{\text{off}}^2)^2}\Big(\Big((\sigma^2(p) + \sigma_n^2) - (\sigma_n^2 + \sigma_{\text{off}}^2)\Big)^2 \\
&\qquad\qquad + \frac{2}{M}(\sigma^2(p) + \sigma_n^2)^2\Big) - p^2 \\
&= \frac{1}{(\sigma_{\text{on}}^2 - \sigma_{\text{off}}^2)^2}\Big((\sigma^2(p) - \sigma_{\text{off}}^2)^2 + \frac{2}{M}(\sigma^2(p) + \sigma_n^2)^2\Big) - p^2 \\
&= \frac{(p(\sigma_{\text{on}}^2 - \sigma_{\text{off}}^2))^2}{(\sigma_{\text{on}}^2 - \sigma_{\text{off}}^2)^2} + \frac{2(\sigma^2(p) + \sigma_n^2)^2}{M(\sigma_{\text{on}}^2 - \sigma_{\text{off}}^2)^2} - p^2 \\
&= \frac{2(\sigma^2(p) + \sigma_n^2)^2}{M(\sigma_{\text{on}}^2 - \sigma_{\text{off}}^2)^2}.
\end{aligned}
\tag{D.2}
$$

In order to evaluate the performance of the estimator, the CRLB can be found. Here, the observation is $\mathbf{y} = \mathbf{\Phi}(\mathbf{s} + \mathbf{n})$ and the parameter to be estimated is p. From the CRLB theorem [5, Section 3.4], the PDF $f(\mathbf{y}; p)$ must satisfy the *regularity* condition

$$
\mathbb{E}\Big\{\frac{\partial \ln f(\mathbf{y}; p)}{\partial p}\Big\} = 0 \quad \text{for all} \quad p, \tag{D.3}
$$

where the expectation is taken with respect to $f(\mathbf{y}; p)$. Here, $f(\mathbf{y}; p) = \mathcal{N}(\mathbf{0}, (\sigma^2(p)+$

$\sigma_n^2)\mathbf{A}$) where $\sigma^2(p)$ is a function of p. Therefore

$$\begin{aligned}\ln f(\mathbf{y};p) &= \ln\Bigg(\frac{1}{(2\pi(\sigma^2(p)+\sigma_n^2))^{\frac{M}{2}}|\mathbf{A}|^{\frac{1}{2}}} \\ &\qquad \times \exp\Big\{-\frac{1}{2(\sigma^2(p)+\sigma_n^2)}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}\Big\}\Bigg) \\ &= -\frac{M}{2}\ln(2\pi) - \frac{M}{2}\ln(\sigma^2(p)+\sigma_n^2) - \frac{1}{2}\ln|\mathbf{A}| \\ &\quad - \frac{1}{2(\sigma^2(p)+\sigma_n^2)}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}. \end{aligned} \tag{D.4}$$

Taking the first derivative

$$\begin{aligned}\frac{\partial \ln f(\mathbf{y};p)}{\partial p} &= \frac{\partial \ln f(\mathbf{y};p)}{\partial \sigma^2(p)}\frac{\partial \sigma^2(p)}{\partial p} \\ &= \Big(-\frac{M}{2(\sigma^2(p)+\sigma_n^2)} \\ &\qquad + \frac{1}{2(\sigma^2(p)+\sigma_n^2)^2}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}\Big) \\ &\qquad \times (\sigma_{\text{on}}^2 - \sigma_{\text{off}}^2). \end{aligned} \tag{D.5}$$

Subsequently, taking the expectation

$$\begin{aligned}&\mathbb{E}\Big\{\frac{\partial \ln f(\mathbf{y};p)}{\partial p}\Big\} \\ &\quad = (\sigma_{\text{on}}^2 - \sigma_{\text{off}}^2)\Big(-\frac{M}{2(\sigma^2(p)+\sigma_n^2)} + \frac{M}{2(\sigma^2(p)+\sigma_n^2)}\Big) = 0. \end{aligned} \tag{D.6}$$

Hence, $f(\mathbf{y};p)$ satisfies the *regularity* condition (D.3). Then the CRLB is $-1/\mathbb{E}\Big\{\partial^2 \ln f(\mathbf{y};p)/\partial p^2\Big\}$. The derivative is evaluated at the true value of p and the expectation is taken with respect to $f(\mathbf{y};p)$.

Taking the second derivative

$$\begin{aligned}\frac{\partial^2 \ln f(\mathbf{y};p)}{\partial p^2} &= \frac{\partial}{\partial \sigma^2(p)}\Big(\frac{\partial \ln f(\mathbf{y};p)}{\partial p}\Big)\frac{\partial \sigma^2(p)}{\partial p} \\ &= \Big(\frac{M}{2(\sigma^2(p)+\sigma_n^2)^2} - \frac{1}{(\sigma^2(p)+\sigma_n^2)^3}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}\Big) \\ &\qquad \times (\sigma_{\text{on}}^2 - \sigma_{\text{off}}^2)^2. \end{aligned} \tag{D.7}$$

Subsequently, taking the expectation of the above equation

$$\mathbb{E}\Big\{\frac{\partial^2 \ln f(\mathbf{y};p)}{\partial p^2}\Big\} = (\sigma^2_{\text{on}} - \sigma^2_{\text{off}})^2\Big(\frac{M}{2(\sigma^2(p)+\sigma^2_n)^2} - \frac{M}{(\sigma^2(p)+\sigma^2_n)^2}\Big)$$
$$= -\frac{M(\sigma^2_{\text{on}} - \sigma^2_{\text{off}})^2}{2(\sigma^2(p)+\sigma^2_n)^2}. \tag{D.8}$$

Therefore, the CRLB is $\frac{2(\sigma^2(p)+\sigma^2_n)^2}{M(\sigma^2_{\text{on}}-\sigma^2_{\text{off}})^2}$ which is the same as (D.2). This indicates that the proposed estimator attains the CRLB, that is, it is an efficient estimator.

Note that the CRLB can be rewritten in terms of σ^2_η and σ^2_Δ, previously defined in Section 4, as $\text{CRLB} = \frac{2}{M}\Big(\frac{\sigma^2_\eta}{\sigma^2_\Delta} + p\Big)^2$.

Appendix E. A new estimator for the sparsity degree guaranteed to be less than 1

As it was stated in Section 4, there is no guarantee that the estimator proposed in (17) ($\hat{p}$) is less than 1. Therefore, here a new estimator is proposed that is guaranteed to be less than 1 but it is shown that the new estimator does not attain the CRLB. Nevertheless, the variance of the estimator is very close to the CRLB.

The new estimator is proposed as $\hat{p}_{\text{new}} = \frac{\min\{1,\hat{p}\}-1+\alpha}{\alpha}$ where α is the probability that $\hat{p} \leq 1$, i.e. $\alpha \triangleq \mathbb{P}\{\hat{p} \leq 1\}$. It is obvious that $\hat{p}_{\text{new}}$ is less than 1, since $\min\{1,\hat{p}\}$ is less than 1.

At first, α should be computed as

$$\alpha = \mathbb{P}\{\hat{p} \leq 1\} = \mathbb{P}\Big\{\frac{1}{\sigma^2(p)+\sigma^2_n}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y} \leq M\frac{\sigma^2_{\text{on}}+\sigma^2_n}{\sigma^2(p)+\sigma^2_n}\Big\}. \tag{E.1}$$

As it was stated in Appendix D, $\frac{1}{\sigma^2(p)+\sigma^2_n}\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}$ has a chi-square distribution with M degrees of freedom. Hence,

$$\alpha = F\Big(M\frac{\sigma^2_{\text{on}}+\sigma^2_n}{\sigma^2(p)+\sigma^2_n}, M\Big), \tag{E.2}$$

where $F(x,M)$ is the CDF of a chi-square random variable with M degrees of freedom. For the selected parameters as $M = 66$, $\sigma_{\text{on}} = 1$, $p = 0.1$, $\sigma_{\text{off}} = 0.1$,

and $\mathrm{SNR}_{\mathrm{in}} = 7\mathrm{dB}$ we have $\alpha = F(515.76, 66)$ which is calculated as 1 under MATLAB precision.

Now, it is shown that the new estimator is unbiased. In fact,

$$\mathbb{E}\{\hat{p}_{\mathrm{new}}\} = \frac{1}{\alpha}(\mathbb{E}\{\min\{1, \hat{p}\}\} - 1 + \alpha). \tag{E.3}$$

The expectation at the right-hand side of the above equation can be calculated as

$$\begin{aligned}\mathbb{E}\{\min\{1, \hat{p}\}\} &= \mathbb{E}\{\min\{1, \hat{p}\}|\hat{p} \leq 1\}\mathbb{P}\{\hat{p} \leq 1\} \\ &\quad + \mathbb{E}\{\min\{1, \hat{p}\}|\hat{p} > 1\}\mathbb{P}\{\hat{p} > 1\} \\ &= \mathbb{E}\{\hat{p}\}.\alpha + 1.(1-\alpha) = p\alpha + 1 - \alpha,\end{aligned} \tag{E.4}$$

where in the last equation, (D.1) has been used. Therefore,

$$\mathbb{E}\{\hat{p}_{\mathrm{new}}\} = \frac{1}{\alpha}(p\alpha + 1 - \alpha - 1 + \alpha) = p. \tag{E.5}$$

In what follows, the variance of the estimator is computed as

$$\begin{aligned}\sigma^2_{\hat{p}_{\mathrm{new}}} &= \mathbb{E}\{\hat{p}^2_{\mathrm{new}}\} - p^2 \\ &= \frac{1}{\alpha^2}(\mathbb{E}\{\min\{1, \hat{p}\}^2\} + (1-\alpha)^2 \\ &\quad -2(1-\alpha)\mathbb{E}\{\min\{1, \hat{p}\}\}) - p^2 \\ &= \frac{1}{\alpha^2}(\mathbb{E}\{\min\{1, \hat{p}\}^2\} + (1-\alpha)^2 \\ &\quad - 2(1-\alpha)(p\alpha + 1 - \alpha)) - p^2,\end{aligned} \tag{E.6}$$

where in the last equation, (E.4) has been used. The last expectation in the above equation can be calculated as

$$\begin{aligned}\mathbb{E}\{\min\{1, \hat{p}\}^2\} &= \mathbb{E}\{\min\{1, \hat{p}\}^2|\hat{p} \leq 1\}\mathbb{P}\{\hat{p} \leq 1\} \\ &\quad + \mathbb{E}\{\min\{1, \hat{p}\}^2|\hat{p} > 1\}\mathbb{P}\{\hat{p} > 1\} \\ &= \mathbb{E}\{\hat{p}^2\}.\alpha + 1.(1-\alpha) \\ &= \Big(p^2 + \frac{2(\sigma^2(p) + \sigma_n^2)^2}{M(\sigma^2_{\mathrm{on}} - \sigma^2_{\mathrm{off}})^2}\Big)\alpha + 1 - \alpha,\end{aligned} \tag{E.7}$$

where in the last equation, (D.2) has been used. By substituting (E.7) in (E.6) and after some straightforward manipulations, the variance of the new proposed estimator is

$$\sigma^2_{\hat{p}_{\text{new}}} = \frac{1-\alpha}{\alpha}(1-p)^2 + \frac{1}{\alpha}\text{CRLB}, \tag{E.8}$$

where CRLB has been given in Appendix D. As is observed, since α is very close to 1, the variance of the new proposed estimator is very close to the CRLB. Hence, the new estimator is nearly efficient.

Acknowledgements

The authors would like to thank anonymous reviewers for their comments and suggestions.

References

[1] Billingsley, P., 2013. Convergence of probability measures. John Wiley & Sons.

[2] Cao, J., Lin, Z., 2014. Bayesian signal detection with compressed measurements. Information Sciences 289, 241–253.

[3] Davenport, M., Boufounos, P.T., Wakin, M.B., Baraniuk, R.G., et al., 2010. Signal processing with compressive measurements. Selected Topics in Signal Processing, IEEE Journal of 4, 445–460.

[4] Hariri, A., Babaie-Zadeh, M., 2015. Joint compressive single target detection and parameter estimation in radar without signal reconstruction. IET Radar, Sonar & Navigation .

[5] Kay, S.M., 1993. Fundamentals of statistical signal processing: estimation theory. Prentice-Hall, Inc.

[6] Lavielle, M., 1993. Bayesian deconvolution of bernoulli-gaussian processes. Signal processing 33, 67–79.

[7] McLachlan, G., Peel, D., 2004. Finite mixture models. John Wiley & Sons.

[8] Mendel, J.M., 1990. Maximum Likelihood Deconvolution. Springer.

[9] Papoulis, A., Pillai, S.U., 2002. Probability, random variables, and stochastic processes. Tata McGraw-Hill Education.

[10] Petersen, K.B., Pedersen, M.S., Chapter 8, Section 8.3, Subsection 8.3.2, p. 43, 2008. The matrix cookbook.

[11] Rao, B.S.M.R., Chatterjee, S., Ottersten, B., 2012. Detection of sparse random signals using compressive measurements, in: Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, IEEE. pp. 3257–3260.

[12] Santamaría-Caballero, I., Pantaleón-Prieto, C.J., Artés-Rodríguez, A., 1996. Sparse deconvolution using adaptive mixed-gaussian models. Signal Processing 54, 161–172.

[13] Taylor, H.L., Banks, S.C., McCoy, J.F., 1979. Deconvolution with the l1 norm. Geophysics 44, 39–52.

[14] Tropp, J.A., Gilbert, A.C., 2005. Signal recovery from partial information via orthogonal matching pursuit, in: IEEE Trans. on Inform. Theory, Citeseer.

[15] Van Trees, H.L., 2001. Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory. Wiley.

[16] Wimalajeewa, T., Chen, H., Varshney, P.K., 2010. Performance analysis of stochastic signal detection with compressive measurements, in: Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conference on, IEEE. pp. 813–817.

[17] Zahedi, R., Pezeshki, A., Chong, E.K., 2012. Measurement design for detecting sparse signals. Physical Communication 5, 64–75.