IJST, Transactions of Electrical Engineering, Vol. 37, No. E2, pp 101-120
Printed in The Islamic Republic of Iran, 2013


# SPARSE SIGNAL RECONSTRUCTION FROM COMPRESSED SENSING MEASUREMENTS BASED ON DETECTION THEORY*

H. AZAD, A. SHEIKHI,** AND M. A. MASNADI-SHIRAZI

Dept. of Computer and Electrical Engineering, Shiraz University, Shiraz, I. R. of Iran
Email: sheikhi@shirazu.ac.ir

**Abstract–** The problem of sparse signal reconstruction from the well-known Compressed Sensing measurement is considered in this paper. The measured signal is assumed to be corrupted with additive white Gaussian noise with zero mean and known variance. Based on detection theory, two iterative algorithms are developed for detection and estimation of nonzero elements of sparse signal. The principle of the proposed methods is based on applying composite multiple hypothesis test to the underlying problem at each iteration. Simulation results show the satisfactory performance of the proposed algorithms in sparse signal recovery. The proposed approach has the potential of being applied to other models for noise and signal.

**Keywords–** Sparse signal reconstruction, compressed sensing, detection theory, composite multiple hypothesis test

## 1. INTRODUCTION

Sparse representation of signals has received considerable attention in the communication community in recent years and has been applied to various fields such as image processing [1], wireless communication and channel estimation [2-4], antenna beamforming [5] and radar signal processing [6-7]. Among all related topics to sparse signal processing, Compressed Sensing (CS) has motivated much research recently. Using far fewer samples derived from linear random projection of main data, one would be able to recover the main sparse signal with high probability. To be more precise, according to basic CS theory [8-9], if a signal has a sparse representation in an overcomplete basis $\mathbf{\Psi}$ [12], it can be reconstructed from a small number of measurements resulting from projection on a random basis $\mathbf{\Phi}$ that is incoherent with $\mathbf{\Psi}$. Various reconstruction algorithms in literature are composed of two major categories [10-11]: convex or non-convex optimization methods and greedy ones. The first approach relies on solution of the optimization of a cost function which leads to more accurate results. Greedy methods [10] take advantage of the speed of greedy algorithms. However, the former is more complex while the latter has less precise results. Basis Pursuit (BP) and its noisy version called Basis Pursuit DeNoising (BPDN) are the examples of the first approach [12] while Orthogonal Matching Pursuit (OMP) [13] and its derived modified versions such as Stagewise OMP (StOMP) [14] and Regularized OMP (ROMP) [10] are some samples of greedy methods.

In this paper we confront the problem of reconstructing sparse signals measured from CS approach using detection theory. In particular, our basic method and its modified version are iterative ones which, at each iteration, the reconstruction problem is formulated as a multiple composite hypothesis test and an attempt is made to find the element of sparse signal, until the stopping criteria is satisfied. So it can be said

*Received by the editors April 10, 2013; Accepted October 26, 2013.
**Corresponding author

that our proposed methods are similar to some conventional methods such as OMP as both of them try to find one element of sparse signal at each iteration. However, it should be mentioned that in algorithms such as OMP, the estimated element of sparse signal and its corresponding column of basis, are subtracted from the main signal in each iteration (i.e. the effect of detected column of $\mathbf{\Phi\Psi}$ in measured data is eliminated), while in our proposed methods, we do not subtract the estimated element. Instead, the structure of assumed detection problem would be changed in such a way that at each new step compared to previous steps, a better model is assumed for noise vector in detection problem. Also, in spite of algorithms such as OMP which do not consider the noise and optimization methods which usually assume additive noise with limited norm, in our methods, the statistical properties of the additive noise are considered such that we are able to use various models of additive noise. Another main advantage of the proposed algorithms is that they utilize a new approach based on detection theory which enables us to use different detection approaches in our application. To the best of our knowledge there has been no reference using this approach till now, although there are some references on detection of existence of sparse signal in noise [15-16]. In such references, it has only been decided if there exists any sparse signal in noise (a binary test to decide between noise only hypothesis vs. noise plus sparse signal) but there is no method on estimation of sparse signal if it exists. In spite of this, in this paper the sparse signal's reconstruction has been considered based on detection theory.

The rest of this paper is organized as follows. A brief review of basic CS theory is given in section 2 Then in section 3, the concepts of the proposed algorithm are explained. A modified version of the proposed method is derived in Section 4. Section 5 presents the simulation results and finally, concluding remarks are given in section 6.

## 2. THEORY OF COMPRESSED SENSING

Compressed Sensing was initially stated in [8-9]. According to basic theory of CS, let's assume the desired signal $\mathbf{x}_{N\times 1} \in \mathbf{R}^N$ has a *k*-sparse representation in a basis $\mathbf{\Psi}_{N\times N}$. In other words, if we write $\mathbf{x}$ as $\mathbf{x} = \mathbf{\Psi} \times \boldsymbol{\alpha}_{N\times 1}$, only *k* elements of $\boldsymbol{\alpha}$ are nonzero, where $k << N$. The set of nonzero indices of nonzero elements of $\boldsymbol{\alpha}$ is called *support* of vector $\boldsymbol{\alpha}$ which is usually assumed to be unknown a prior. Also, *k* is called the sparsity level of $\boldsymbol{\alpha}$. This signal is measured through a measurement matrix $\mathbf{\Phi}_{m\times N}$ where $m < N$, which results in a measured signal $\mathbf{y}_{m\times 1}$. In other words, we have:

$$\begin{aligned} &\mathbf{y} = \mathbf{\Phi} \times \mathbf{x} = \mathbf{\Phi\Psi}\boldsymbol{\alpha} = \mathbf{H}\boldsymbol{\alpha} \\ &where \;\; \mathbf{H}_{m\times N} = \mathbf{\Phi\Psi} \\ &\mathbf{H} = [\mathbf{h}_1 \;\; \mathbf{h}_2 \;\; ... \;\; \mathbf{h}_N] \end{aligned} \tag{1}$$

The main problem is estimating the desired signal $\mathbf{x}$ using measured data $\mathbf{y}$. Since the number of equations is less than that of unknown variables, in general form there is no unique solution for $\mathbf{x}$. But since $\mathbf{x}$ has a sparse representation in an known basis $\mathbf{\Psi}_{N\times N}$, the basic problem can be converted to the problem of finding vector $\boldsymbol{\alpha}$ so that it can be estimated based on the assumption of sparsity, and then use

$$\mathbf{x} = \mathbf{\Psi} \times \boldsymbol{\alpha}.$$

Signal reconstruction under sparseness constraints can be formulated as an optimization problem of $\ell_0$-norm criterion which is unfortunately an NP-hard one. Due to the fact of not being numerically feasible, alternative approaches have been proposed that can be divided into two main groups of greedy algorithms and relaxed optimization ones. In greedy methods, one or more nonzero elements are estimated

in each iteration. Orthogonal Matching Pursuit (OMP) [13] and its modified versions such as Stagewise OMP (StOMP) [14] and Regularized OMP (ROMP) [10] can be categorized into this group. The iterative approach of greedy algorithms results in less computational complexity methods compared to that of the optimization ones. The latter algorithms make use of some relaxation approaches to $\ell_0$-norm optimization [8] in order to convert the main problem to a convex one (The term $\ell_0$-norm for a vector is simply the number of its nonzero elements). The most famous member of this category is Basis Pursuit (BP) [12], which results in an $\ell_1$-norm criterion that can be solved using Linear Programming (LP). To guarantee exact recovery of every k-sparse signal $\boldsymbol{\alpha}$ from m compressed measurements $\mathbf{y}$ in BP algorithm, the matrix $\boldsymbol{\Phi\Psi}$ should satisfy the Restricted Isometry Property (RIP) as below:

Restricted Isometry Property [10]: A matrix $\boldsymbol{\Phi\Psi}$ is said to satisfy the RIP of order k, if there exists a constant $\delta$ such that for any k-sparse vector $\boldsymbol{\alpha}$ we have:

$$(1-\delta)\|\boldsymbol{\alpha}\|_2^2 \le \|\boldsymbol{\Phi\Psi\alpha}\|_2^2 \le (1+\delta)\|\boldsymbol{\alpha}\|_2^2 \tag{2}$$

It is shown that RIP is satisfied with high probability by random matrices with i.i.d. Gaussian or Bernoulli elements.

In practical situations, the desired signal $\mathbf{x}$ is measured in the presence of additive noise $\mathbf{n}_{m\times 1}$; i.e. we have:

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{n} \tag{3}$$

The additive noise has usually been modeled as a limited norm signal while in some references, white Gaussian model has been assumed for the noise signal. Although the presence of noise converts the problem to a new one, most recovery methods such as OMP do not exploit these new assumptions in their formulation. But some relaxed optimization based methods, due to their special mathematical formulation, change their optimization criteria so that they can be applied to the case of sparse signals corrupted with limited norm noise. An example of such methods is BP DeNosing (BPDN) which is the noisy version of the most famous BP.

In this paper, we consider the special case of white Gaussian noise with distribution $\mathbf{n} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ with known $\sigma^2$. In the next sections, we address the problem of reconstruction of $\mathbf{x}$ from noisy measurement signal $\mathbf{y}$.

## 3. SPARSE SIGNAL RECOVERY BASED ON DETECTION THEORY

As it was mentioned, the main concept of the proposed methods is based on extraction of one nonzero element in each iteration. This approach is similar to the most famous OMP method in sparse signal estimation. The main difference is that in OMP method the effect of estimated nonzero element is subtracted from measured signal in each iteration. But in the proposed methods, a detection hypothesis test should be solved in each iteration. Although we know that the signal $\mathbf{x}$ is composed of $k \ge 1$ elements of basis $\boldsymbol{\Psi}$, at first step it has been assumed that there is only one nonzero element in $\boldsymbol{\alpha}$; i.e. $\mathbf{x}$ is constructed from one column of $\boldsymbol{\Psi}$. Based on this assumption, the resultant test has been solved and the first nonzero element has been detected. Then, instead of subtraction of this element (similar to OMP), a new hypothesis test has been applied in which the location of the first element has been considered based on solution of the first step and the location of the second one is unknown. Finding the second index, the test is modified to the problem of detecting the third index assuming the first two indices have been found in the previous steps. The same procedure has been considered for finding the other elements. Note that the number of nonzero elements (k) is unknown aprior, so that we cannot find these elements by solving only one multiple hypothesis test. Instead, we have tried to detect the nonzero elements one by one in each iteration.

As it has been described, at first step suppose that there is only one nonzero element in $\boldsymbol{\alpha}$; i.e. $\mathbf{x}$ is constructed from one column of $\boldsymbol{\Psi}$. So, the main problem of detecting the nonzero elements of $\boldsymbol{\alpha}$ can be modeled as selecting one of the following structures:

$$\begin{cases} H_1 : \mathbf{y} = \mathbf{h}_1 \times \alpha_1 + \mathbf{n}^{(1)} \\ H_2 : \mathbf{y} = \mathbf{h}_2 \times \alpha_2 + \mathbf{n}^{(1)} \\ \vdots \\ H_N : \mathbf{y} = \mathbf{h}_N \times \alpha_N + \mathbf{n}^{(1)} \end{cases} \tag{4}$$

where $\mathbf{h}_i$ is the *i*th column of $\mathbf{H}$ and $\alpha_i$ denotes the *i*th element of $\boldsymbol{\alpha}$ in (1), respectively.

Assuming there is only one nonzero element in $\boldsymbol{\alpha}$, the noise vector $\mathbf{n}^{(1)}$ has normal distribution $N\left(\mathbf{0}, \sigma^2 \mathbf{I}_m\right)$; i.e. there is no signal component in $\mathbf{n}^{(1)}$ (this model is approximately true for cases with high SNR for one nonzero element and low SNR for the others).

Selecting one of these structures with unknown values of $\alpha_1, \alpha_2, ..., \alpha_N$ is equivalent to a composite multiple hypothesis test [17] and the proper solution is *Generalized Maximum Likelihood Rule* (*GMLR)*; i.e. we should solve the following problem (assuming $p(H_i) = \frac{1}{N},\ i = 1,2,..., N$):

$$j_1 = \arg\ \max_i \left[ p\left(\mathbf{y} \mid \hat{\alpha}_i, H_i\right) \right], i = 1,2,..., N \tag{5}$$

where $\hat{\alpha}_i$ is maximum likelihood estimation for $\alpha_i$:

$$\begin{aligned} & p\left(\mathbf{y} \mid \alpha_i, H_i\right) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{m}{2}}} \times \exp\left[ -\frac{1}{2\sigma^2} \left(\mathbf{y} - \mathbf{h}_i \alpha_i\right)^T \left(\mathbf{y} - \mathbf{h}_i \alpha_i\right) \right] \\ & \Rightarrow \frac{d\ p\left(\mathbf{y} \mid \alpha_i, H_i\right)}{d\ \alpha_i} = 0 \Rightarrow \hat{\alpha}_i = \left(\mathbf{h}_i^T \mathbf{h}_i\right)^{-1} \mathbf{h}_i^T \mathbf{y} \end{aligned} \tag{6}$$

By replacing the ML estimation in (5), we have:

$$\begin{aligned} & j_1 = \arg\ \max_i \left[ p\left(\mathbf{y} \mid \hat{\alpha}_i, H_i\right) \right] \\ & \Rightarrow j_1 = \arg\ \max_i \left[ \ln\ p\left(\mathbf{y} \mid \hat{\alpha}_i, H_i\right) \right] \\ & \Rightarrow j_1 = \arg\ \max_i [ -\frac{m}{2} \ln\left(2\pi\sigma^2\right) \\ & \qquad\qquad -\frac{1}{2\sigma^2} \left(\mathbf{y} - \mathbf{h}_i \hat{\alpha}_i\right)^T \left(\mathbf{y} - \mathbf{h}_i \hat{\alpha}_i\right) ] \end{aligned}$$

$$\begin{aligned} & \hat{\alpha}_i = \left(\mathbf{h}_i^T \mathbf{h}_i\right)^{-1} \mathbf{h}_i^T \mathbf{y} \Rightarrow j_1 = \arg\ \max_i [ -\frac{m}{2} \ln\left(2\pi\sigma^2\right) \\ & \qquad\qquad -\frac{1}{2\sigma^2} \mathbf{y}^T \mathbf{y} + \frac{1}{2\sigma^2} \mathbf{y}^T \mathbf{h}_i \left(\mathbf{h}_i^T \mathbf{h}_i\right)^{-1} \mathbf{h}_i^T \mathbf{y} ] \\ & \Rightarrow j_1 = \arg\ \max_i \left[ \frac{1}{2\sigma^2} \mathbf{y}^T \mathbf{h}_i \left(\mathbf{h}_i^T \mathbf{h}_i\right)^{-1} \mathbf{h}_i^T \mathbf{y} \right] \end{aligned} \tag{7}$$

Now since the location of the first nonzero element is known, we can search for the second one. Similar to the assumptions of the first detection problem, and considering $j_1$ as the first nonzero element, the first detection problem can be modified to the second one as follows:

$$\begin{cases} H_1 : \mathbf{y} = \left[\mathbf{h}_{j_1}\ \mathbf{h}_1\right] \times \begin{bmatrix} \alpha_{j_1} \\ \alpha_1 \end{bmatrix} + \mathbf{n}^{(2)} \\ H_2 : \mathbf{y} = \left[\mathbf{h}_{j_1}\ \mathbf{h}_2\right] \times \begin{bmatrix} \alpha_{j_1} \\ \alpha_2 \end{bmatrix} + \mathbf{n}^{(2)} \\ \vdots \\ .H_{j_1-1} : \mathbf{y} = \left[\mathbf{h}_{j_1}\ \mathbf{h}_{j_1-1}\right] \times \begin{bmatrix} \alpha_{j_1} \\ \alpha_{j_1-1} \end{bmatrix} + \mathbf{n}^{(2)} \\ H_{j_1+1} : \mathbf{y} = \left[\mathbf{h}_{j_1}\ \mathbf{h}_{j_1+1}\right] \times \begin{bmatrix} \alpha_{j_1} \\ \alpha_{j_1+1} \end{bmatrix} + \mathbf{n}^{(2)} \\ \vdots \\ H_N : \mathbf{y} = \left[\mathbf{h}_{j_1}\ \mathbf{h}_N\right] \times \begin{bmatrix} \alpha_{j_1} \\ \alpha_N \end{bmatrix} + \mathbf{n}^{(2)} \end{cases} \tag{8}$$

Note that in the second detection problem, since one nonzero index of $\boldsymbol{\alpha}$ is found in the previous step, we should select one of the other N-1 indices (the first (N)- tuple hypothesis test changes to a (N-1)-tuple one). If it is assumed that there is only two nonzero elements in $\boldsymbol{\alpha}$, the noise vector $\mathbf{n}^{(2)}$ has normal distribution $N\left(\mathbf{0}, \sigma^2 \mathbf{I}_m\right)$. In this hypothesis test, $\alpha_{j_1}$ and $\alpha_i$ ($i = 1,2,..., j_1 - 1, j_1 + 1,..., N$) are all assumed to be unknown.

Similar to the previous test, GMLR can be used to detect the second index. By using the same procedure and assuming equal probable $H_i$, we have:

$$j_2 = \arg\ \max_i \left[p\left(\mathbf{y} \mid \left[\hat{\alpha}_{j_1}, \hat{\alpha}_i\right]^T, H_i\right)\right]\ i = 1,2,..., N, i \neq j_1$$

$$\begin{bmatrix} \hat{\alpha}_{j_1} \\ \hat{\alpha}_i \end{bmatrix} = \left(\left[\mathbf{h}_{j_1}\ \mathbf{h}_i\right]^T \times \left[\mathbf{h}_{j_1}\ \mathbf{h}_i\right]\right)^{-1} \times \left[\mathbf{h}_{j_1}\ \mathbf{h}_i\right]^T \times \mathbf{y} \tag{9}$$

$$\Rightarrow j_2 = \arg\ \max_i \left[\frac{1}{2\sigma^2} \mathbf{y}^T \left[\mathbf{h}_{j_1}\ \mathbf{h}_i\right] \times \left(\left[\mathbf{h}_{j_1}\ \mathbf{h}_i\right]^T \times \left[\mathbf{h}_{j_1}\ \mathbf{h}_i\right]\right)^{-1} \times \left[\mathbf{h}_{j_1}\ \mathbf{h}_i\right]^T \times \mathbf{y}\right]$$

Note that the second formulation in (8) is more accurate than the first one in (4) and represents a better form of the main problem, since the assumption of being a normal vector represents $\mathbf{n}^{(2)}$ better than $\mathbf{n}^{(1)}$.

Finding the second index, the test is modified to the problem of detecting the third index assuming the first two indices to be $j_1$ and $j_2$. Similar procedure can be applied to the new problem to find the next index, and then one should find the forth index, and so on. As a stopping criterion, $C_p$ should be calculated and compared to predefined threshold as follows.

$$\begin{aligned} &C_p = \left\| \mathbf{y} - \mathbf{H}_{\Gamma_p} \times \hat{\boldsymbol{\alpha}}_p \right\|^2 \\ &\Gamma_p : \text{set of detected indices upto iteration } p \\ &\mathbf{H}_{\Gamma_p} : m \times p \text{ matrix composed of } p \text{ columns} \\ &\qquad \text{of } \mathbf{H} \text{ corresponding to indices in set } \Gamma_p \\ &\hat{\boldsymbol{\alpha}}_p = \left(\mathbf{H}_{\Gamma_p}^T \mathbf{H}_{\Gamma_p}\right)^{-1} \times \mathbf{H}_{\Gamma_p}^T \mathbf{y} \end{aligned} \tag{10}$$

Once the sparse signal has been estimated, the desired vector $\mathbf{x}$ can be calculated using $\mathbf{x} = \mathbf{H}_{\Gamma_p} \times \hat{\boldsymbol{\alpha}}_p$. We call the proposed algorithm the Iterative Detection Based Compressed Sensing (IDBCS).

A brief review of the proposed algorithm can be seen in Table I. Note that to solve sparse signal reconstruction from CS measurements, the proposed approach can be used to develop various algorithms for different statistical models of noise or signal. At last, it should be mentioned that as it is described in appendix A, the matrix inversion of $\left(\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T \times \left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\right)^{-1}$ at step 2 can be achieved by a recursive method using $\left(\mathbf{H}_{\Gamma_{p-1}}{}^T \times \mathbf{H}_{\Gamma_{p-1}}\right)^{-1}$ from the previous iteration which decreases the computational complexity dramatically.

Table 1. Brief review of IDBCS

***Initialize algorithm with*** $\mathbf{H}_{m\times N} = \mathbf{\Phi\Psi}, \Gamma_0 = \phi, p = 0$

1. $p = p + 1$
2. $$j = \arg\max_i \left[\frac{1}{2\sigma^2}\mathbf{y}^T\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\times \left(\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T \times \left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\right)^{-1} \times \left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T \times \mathbf{y}\right]$$
   $i = 1,2,...,N\ \&\ i \notin \Gamma_{p-1}$
   $\mathbf{H}_{\Gamma_{p-1}}$ : $m \times (p-1)$ matrix composed of *(p*-1*)* columns of $\mathbf{H}$ corresponding to indices in set $\Gamma_{p-1}$
   $\mathbf{h}_i$ : *i*th column of matrix $\mathbf{H}$
3. $\Gamma_p = \Gamma_{p-1} \bigcup \{j\}$
4. $$C_p = \left\|\mathbf{y} - \mathbf{H}_{\Gamma_p} \times \hat{\boldsymbol{\alpha}}_p\right\|^2$$
   $$\hat{\boldsymbol{\alpha}}_p = \left(\mathbf{H}_{\Gamma_p}^T \mathbf{H}_{\Gamma_p}\right)^{-1} \times \mathbf{H}_{\Gamma_p}\mathbf{y}$$
   ***If $C_p$>threshold, go to step 1, else stop***
5. $\mathbf{x} = \mathbf{H}_{\Gamma_p} \times \hat{\boldsymbol{\alpha}}_p$

## 4. MODIFICATION OF THE PROPOSED ALGORITHM

As mentioned in the previous section, the main concept of the proposed method is detecting only one nonzero element at each iteration and simultaneously, trying to modify the detection problem formulation for the next step. Let's have an insight into the first step in (4). In this equation, it is assumed that there exists only one nonzero element in $\boldsymbol{\alpha}$. So the noise vector $\mathbf{n}^{(1)}$ in (4) has been modeled as a vector with normal distribution $N\left(\mathbf{0}, \sigma^2\mathbf{I}_m\right)$; i.e. the mean value of this noise has been assumed to be zero. But, there are some other nonzero elements that force the noise vector to have a nonzero mean. In other words, a better model for the noise vector is to assume a nonzero mean normal distribution; i.e. $\mathbf{n}^{(1)}$ in (4), should be modeled as a vector with normal distribution $N\left(\boldsymbol{\mu}, \sigma^2\mathbf{I}_m\right)$. Although the assumed distribution can be considered as a perfect one for the noise vector, the mean vector $\boldsymbol{\mu}$ is an unknown vector of length m. So the number of unknown parameters would be m+1 (m elements of the mean vector $\boldsymbol{\mu}$ and one nonzero element of $\boldsymbol{\alpha}$ which is $\alpha_i$). It can be shown that such an assumption leads to a linear system of equations with infinite solution in finding the ML estimation of $\boldsymbol{\mu}$. Also, the trivial solution for ML estimator is $\hat{\boldsymbol{\mu}} = \mathbf{y}$, which results in maximization of a constant term for all hypotheses, which is meaningless. So, instead of having a fully unknown mean vector we heuristically model the noise signal by a vector that has a structured mean vector. For the first iteration, the mean of all columns of matrix $\mathbf{H}$ multiplied by an unknown coefficient $\mu^{(1)}$ is considered as the noise mean. To be more precise, the noise vector $\mathbf{n}^{(1)}$ is

modeled as a vector with normal distribution $N\left(\mu^{(1)}\mathbf{d}^{(1)},\sigma^2\mathbf{I}_m\right)$, where $\mathbf{d}^{(1)}$ is the mean of all columns of matrix $\mathbf{H}$. The term "structured" refers to assuming that the noise vector has a mean equal to mean of all columns of matrix $\mathbf{H}$. In other words, it has been assumed that the mean cannot be any vector in m-dimensional space but instead it is equal to a known vector multiplied by an unknown coefficient.

With similar assumption to that of the previous section, the detection problem of (4) can be solved using the modified noise model as follows:

$$j_1=\arg\max_i\left[p\left(\mathbf{y}\mid\hat{\alpha}_i,\hat{\mu}^{(1)},H_i\right)\right]\ i=1,2,...,N \tag{11}$$

where $\hat{\alpha}_i$ and $\hat{\mu}^{(1)}$ are maximum likelihood estimations for $\alpha_i$ and $\mu^{(1)}$. Let's first find the ML estimation of $\alpha_i$'s:

$$\begin{aligned}p\left(\mathbf{y}\mid\alpha_i,\mu^{(1)},H_i\right)&=\frac{1}{\left(2\pi\sigma^2\right)^{\frac{m}{2}}}\times\\&\exp\left[-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{h}_i\alpha_i-\mu^{(1)}\mathbf{d}^{(1)}\right)^T\left(\mathbf{y}-\mathbf{h}_i\alpha_i-\mu^{(1)}\mathbf{d}^{(1)}\right)\right]\\\Rightarrow\frac{d\ p\left(\mathbf{y}\mid\alpha_i,\mu^{(1)},H_i\right)}{d\ \alpha_i}&=0\Rightarrow\hat{\alpha}_i=\left(\mathbf{h}_i^T\mathbf{h}_i\right)^{-1}\mathbf{h}_i^T\left(\mathbf{y}-\mu^{(1)}\mathbf{d}^{(1)}\right)\end{aligned} \tag{12}$$

Now, by replacing the ML estimation of $\alpha_i$, after some mathematical manipulations, the ML estimation of the second parameter which is $\mu^{(1)}$ can be calculated as follows:

$$\frac{d\ p\left(\mathbf{y}\mid\hat{\alpha}_i,\mu^{(1)},H_i\right)}{d\ \mu^{(1)}}=0\Rightarrow\hat{\mu}^{(1)}=\frac{\mathbf{y}^T\mathbf{h}_i\left(\mathbf{h}_i^T\mathbf{h}_i\right)^{-1}\mathbf{h}_i^T\mathbf{d}^{(1)}-\mathbf{y}^T\mathbf{d}^{(1)}}{\mathbf{d}^{(1)T}\mathbf{h}_i\left(\mathbf{h}_i^T\mathbf{h}_i\right)^{-1}\mathbf{h}_i^T\mathbf{d}^{(1)}-\mathbf{d}^{(1)T}\mathbf{d}^{(1)}} \tag{13}$$

Finding the ML estimation of unknown parameters, the solution to the considered detection problem is:

$$\begin{aligned}&j_1=\arg\max_i\left[p\left(\mathbf{y}\mid\hat{\alpha}_i,\hat{\mu}^{(1)},H_i\right)\right]\\&\Rightarrow j_1=\arg\max_i\left[\ln p\left(\mathbf{y}\mid\hat{\alpha}_i,\hat{\mu}^{(1)},H_i\right)\right]\\&\Rightarrow j_1=\arg\max_i[-\frac{m}{2}\ln\left(2\pi\sigma^2\right)\\&\qquad-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{h}_i\hat{\alpha}_i-\hat{\mu}^{(1)}\mathbf{d}^{(1)}\right)^T\left(\mathbf{y}-\mathbf{h}_i\hat{\alpha}_i-\hat{\mu}^{(1)}\mathbf{d}^{(1)}\right)]\end{aligned} \tag{14}$$

Solving the first detection problem, the location of the first nonzero element will be found. Then similar to the previous section, we can search for the second one. The second detection problem would be the same as (8) except that a non-zero mean noise model would be utilized in this section; i.e. the noise vector $\mathbf{n}^{(2)}$ is assumed to have a normal distribution of $N\left(\mu^{(2)}\mathbf{d}^{(2)},\sigma^2\mathbf{I}_m\right)$. The vector $\mathbf{d}^{(2)}$ denotes the mean of all columns of matrix $\mathbf{H}$ except the detected column in the first iteration ($\mathbf{h}_{j_1}$).

Similar to the previous test, GMLR can be used to detect the second index. By using the same procedure and assuming equal probable $H_i$, we have:

$$\begin{aligned}&j_2=\arg\max_i\left[p\left(\mathbf{y}\mid\left[\hat{\alpha}_{j_1},\hat{\alpha}_i\right]^T,\hat{\mu}^{(2)},H_i\right)\right]\ i=1,2,...,N,i\neq j_1\\&\begin{bmatrix}\hat{\alpha}_{j_1}\\\hat{\alpha}_i\end{bmatrix}=\left(\left[\mathbf{h}_{j_1}\ \ \mathbf{h}_i\right]^T\times\left[\mathbf{h}_{j_1}\ \ \mathbf{h}_i\right]\right)^{-1}\times\left[\mathbf{h}_{j_1}\ \ \mathbf{h}_i\right]^T\times\left(\mathbf{y}-\mu^{(2)}\mathbf{d}^{(2)}\right)\\&\hat{\mu}^{(2)}=\frac{\mathbf{y}^T\times\left[\mathbf{h}_{j_1}\ \ \mathbf{h}_i\right]\times\left(\left[\mathbf{h}_{j_1}\ \ \mathbf{h}_i\right]^T\times\left[\mathbf{h}_{j_1}\ \ \mathbf{h}_i\right]\right)^{-1}\times\left[\mathbf{h}_{j_1}\ \ \mathbf{h}_i\right]^T\times\mathbf{d}^{(2)}-\mathbf{y}^T\mathbf{d}^{(2)}}{\mathbf{d}^{(2)^T}\times\left[\mathbf{h}_{j_1}\ \ \mathbf{h}_i\right]\times\left(\left[\mathbf{h}_{j_1}\ \ \mathbf{h}_i\right]^T\times\left[\mathbf{h}_{j_1}\ \ \mathbf{h}_i\right]\right)^{-1}\times\left[\mathbf{h}_{j_1}\ \ \mathbf{h}_i\right]^T\times\mathbf{d}^{(2)}-\mathbf{d}^{(2)^T}\mathbf{d}^{(2)}}\\&j_2=\arg\max_i\left[\ln p\left(\mathbf{y}\mid\left[\hat{\alpha}_{j_1},\hat{\alpha}_i\right]^T,\hat{\mu}^{(2)},H_i\right)\right]\end{aligned}$$

$$\Rightarrow j_2 = \arg\max_i [-\frac{m}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left(\mathbf{y}-\left[\mathbf{h}_{j_1}\ \mathbf{h}_i\right]\times\begin{bmatrix}\hat{\alpha}_{j_1}\\ \hat{\alpha}_i\end{bmatrix}-\hat{\mu}^{(2)}\mathbf{d}^{(2)}\right)^T\left(\mathbf{y}-\left[\mathbf{h}_{j_1}\ \mathbf{h}_i\right]\times\begin{bmatrix}\hat{\alpha}_{j_1}\\ \hat{\alpha}_i\end{bmatrix}-\hat{\mu}^{(2)}\mathbf{d}^{(2)}\right)] \quad (15)$$

Using a similar procedure, the algorithm should be continued until the stopping criteria of (10) is satisfied. A brief review of this modified version of the proposed algorithm can be seen in Table 2. As it will be seen in the computer simulation section, the performance of this modified method is better than that of the previous one due to assuming nonzero mean for noise signal.

Table 2. Brief review of Modified IDBCS

***Initialize algorithm with*** $\mathbf{H}_{m\times N}=\mathbf{\Phi\Psi}, \Gamma_0=\phi, p=0$

1. $p=p+1$
2. $$j=\arg\max_i [-\frac{m}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left(\mathbf{y}-\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\times\begin{bmatrix}\hat{\alpha}_{\Gamma_{p-1}}\\ \hat{\alpha}_i\end{bmatrix}-\hat{\mu}^{(p)}\mathbf{d}^{(p)}\right)^T\left(\mathbf{y}-\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\times\begin{bmatrix}\hat{\alpha}_{\Gamma_{p-1}}\\ \hat{\alpha}_i\end{bmatrix}-\hat{\mu}^{(p)}\mathbf{d}^{(p)}\right)]$$
   $$\begin{bmatrix}\hat{\alpha}_{\Gamma_{p-1}}\\ \hat{\alpha}_i\end{bmatrix}=\left(\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\right)^{-1}\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T\times\left(\mathbf{y}-\hat{\mu}^{(p)}\mathbf{d}^{(p)}\right)$$
   $$\hat{\mu}^{(p)}=\frac{\mathbf{y}^T\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\times\left(\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\right)^{-1}\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T\times\mathbf{d}^{(p)}-\mathbf{y}^T\mathbf{d}^{(p)}}{\mathbf{d}^{(p)^T}\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\times\left(\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\right)^{-1}\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T\times\mathbf{d}^{(p)}-\mathbf{d}^{(p)^T}\mathbf{d}^{(p)}}$$
   $i=1,2,...,N\ \&\ i\notin\Gamma_{p-1}$
   $\mathbf{H}_{\Gamma_{p-1}}$ : $m\times(p-1)$ matrix composed of
   *(p-1)* columns of $\mathbf{H}$ corresponding to indices
   in set $\Gamma_{p-1}$
   $\mathbf{h}_i$ : *i*th column of matrix $\mathbf{H}$
   $\mathbf{d}^{(p)}$ : mean of all columns of matrix $\mathbf{H}$ except the ones
   corresponding to indices in set $\Gamma_{p-1}$
3. $\Gamma_p=\Gamma_{p-1}\bigcup\{j\}$
4. $C_p=\left\|\mathbf{y}-\mathbf{H}_{\Gamma_p}\times\hat{\boldsymbol{\alpha}}_p\right\|^2$
   $\hat{\boldsymbol{\alpha}}_p=\left(\mathbf{H}_{\Gamma_p}^T\mathbf{H}_{\Gamma_p}\right)^{-1}\times\mathbf{H}_{\Gamma_p}\mathbf{y}$
   ***If $C_p$>threshold, go to step 1, else stop***
5. $\mathbf{x}=\mathbf{H}_{\Gamma_p}\times\hat{\boldsymbol{\alpha}}_p$

## 5. ERROR PROBABILITY ANALYSIS

In this section, we derive a union bound for error probability of the assumed detection problem in IDBCS method [18]. At first step, consider the special binary case with the following detector structure. Note that this model can be considered as the binary form for the general M-ary detector at each iteration. In this model, $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$ are the vectors of unknown parameters which should be replaced by their maximum likelihood estimation. Compared to general M-ary form in $p$ th iteration, $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$ would be vectors of length $p$. So we have:

$$
\begin{aligned}
&\begin{cases} H_0 : \mathbf{y} = \mathbf{H}_0\boldsymbol{\alpha}_0 + \mathbf{n} \\ H_1 : \mathbf{y} = \mathbf{H}_1\boldsymbol{\alpha}_1 + \mathbf{n} \end{cases} \\
&\hat{\boldsymbol{\alpha}}_i = \left(\mathbf{H}_i^T\mathbf{H}_i\right)^{-1}\mathbf{H}_i^T\mathbf{y} \Rightarrow j = \arg\max_{i=0,1}[-\frac{m}{2}\ln\left(2\pi\sigma^2\right) \\
&\qquad\qquad -\frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{y} + \frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{H}_i\left(\mathbf{H}_i^T\mathbf{H}_i\right)^{-1}\mathbf{H}_i^T\mathbf{y}] \\
&\Rightarrow j = \arg\max_{i=0,1}\left[\frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{H}_i\left(\mathbf{H}_i^T\mathbf{H}_i\right)^{-1}\mathbf{H}_i^T\mathbf{y}\right] = \arg\max_{i=0,1}\left[\frac{1}{\sigma^2}\mathbf{y}^T\mathbf{W}_i\mathbf{y}\right] \\
&\Rightarrow j = \arg\max_{i=0,1} L_i \\
&L_i = \frac{1}{\sigma^2}\mathbf{y}^T\mathbf{W}_i\mathbf{y} \\
&\mathbf{W}_i = \mathbf{H}_i\left(\mathbf{H}_i^T\mathbf{H}_i\right)^{-1}\mathbf{H}_i^T \\
&\mathbf{W}_i : \textit{symmetric } m\times m \textit{ matrix} \\
&\mathbf{W}_i^2 = \mathbf{W}_i \ (\textit{idempotent})
\end{aligned}
\tag{16}
$$

To analyze the error probability, the decision rule can be rewritten as follows:

$$
\begin{aligned}
&L_1 - L_0 \underset{H_0}{\overset{H_1}{\gtrless}} 0 \Rightarrow \frac{1}{\sigma^2}\mathbf{y}^T\left(\mathbf{W}_1 - \mathbf{W}_0\right)\mathbf{y} \underset{H_0}{\overset{H_1}{\gtrless}} 0 \Rightarrow \mathbf{y}^T\mathbf{W}\mathbf{y} \underset{H_0}{\overset{H_1}{\gtrless}} 0 \Rightarrow Q \underset{H_0}{\overset{H_1}{\gtrless}} 0 \\
&Q = \mathbf{y}^T\mathbf{W}\mathbf{y} \\
&\mathbf{W} = \mathbf{W}_1 - \mathbf{W}_0
\end{aligned}
\tag{17}
$$

By utilizing this formulation, the error probability assuming that $H_1$ is the true hypothesis (i.e. $\mathbf{y} = \mathbf{H}_1\boldsymbol{\alpha}_1 + \mathbf{n}$) would be calculated as follows:

$$P_{error_1} = \int_{-\infty}^{0} f_Q(Q)dQ \tag{18}$$

The decision variable $Q$ has a quadratic form. To find its probability density function $f_Q(Q)$, the following theorem can be used.

**Theorem 1:** Let $\mathbf{z}_{m\times 1}$ be a vector with normal distribution $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is a positive definite matrix. Let $\mathbf{A}$ be a symmetric matrix and define $b = \mathbf{z}^T\mathbf{A}\mathbf{z}$. Let $\lambda_i (i = 1,2,...,m)$ be the eigenvalues of $\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{A}\boldsymbol{\Sigma}^{\frac{1}{2}}$ (or $\boldsymbol{\Sigma}\mathbf{A}$) and $\mathbf{P}$ be an orthogonal $m\times m$ matrix such that:

$$\mathbf{P}^T\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{A}\boldsymbol{\Sigma}^{\frac{1}{2}}\right)\mathbf{P} = diag(\lambda_1,\lambda_2,...,\lambda_m) \tag{19}$$

Finally, define $\mathbf{U}$ and $\boldsymbol{\xi}$ as:

$$
\begin{aligned}
&\mathbf{U} = \left[U_1, U_m, ..., U_m\right]^T = \mathbf{P}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\mathbf{z} - \boldsymbol{\mu}\right) \\
&\boldsymbol{\xi} = \left[\xi_1, \xi_m, ..., \xi_m\right]^T = \mathbf{P}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}
\end{aligned}
\tag{20}
$$

The random variables $U_i$'s are mutually independent and have standard normal distribution ($N(0,1)$). Then $b = \mathbf{z}^T\mathbf{A}\mathbf{z}$ can be expressed as $b = \sum_{i=1}^{m}\lambda_i\left(U_i + \xi_i\right)^2$.

Applying theorem 1 to our problem and assuming that $H_1$ is the true hypothesis, the decision variable $Q$ can be written as $Q = \sum_{i=1}^{m}\lambda_i\left(U_i + \xi_i\right)^2$ where:

$$\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$
$$\boldsymbol{\mu} = \mathbf{H}_1 \boldsymbol{\alpha}_1$$
$$\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$$
$$\lambda_i (i = 1,2,...,m) : \textit{eigenvalues of } \left(\sigma^2 \times \mathbf{W}\right) \tag{21}$$
$$\mathbf{P} \textit{ is an } m \times m \textit{ orthogonal matrix such that } \mathbf{P}^T \left(\sigma^2 \times \mathbf{W}\right)\mathbf{P} = diag(\lambda_1, \lambda_2, ..., \lambda_m)$$
$$\mathbf{U} = \left[U_1, U_m, ..., U_m\right]^T = \frac{\mathbf{P}^T \left(\mathbf{z} - \mathbf{H}_1 \boldsymbol{\alpha}_1\right)}{\sigma}$$
$$\boldsymbol{\xi} = \left[\xi_1, \xi_m, ..., \xi_m\right]^T = \frac{\mathbf{P}^T \mathbf{H}_1 \boldsymbol{\alpha}_1}{\sigma}$$

Since the random variables $U_i$'s are independent and have standard normal distribution, we can say that $Q$ is a linear combination of $m$ independent random variables with non-central chi-square distribution, all with one degree of freedom and noncentrality parameters $\xi_i^2$ ($i = 1,2,...,m$), respectively [19].

Finally, the probability density function for such a random variable ($Q$) can be expressed as follows [20-21]:

$$f_Q(Q) = \begin{cases} \displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\theta_i \theta'_j}{\Gamma(\tau + i)} b^{\frac{-(\tau+i+\tau'+j)}{2}} Q^{\frac{(\tau+i+\tau'+j-2)}{2}} \\ \qquad \times \exp\left(\frac{Q \times \left(\beta'^{-1} - \beta^{-1}\right)}{4}\right) \times W_{\frac{(\tau+i-\tau'-j)}{2}, \frac{(1-\tau'-j-\tau-i)}{2}}(bQ) \qquad \textit{for } Q > 0 \\ \displaystyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\theta_i \theta'_j}{\Gamma(\tau' + i)} b^{\frac{-(\tau+i+\tau'+j)}{2}} (-Q)^{\frac{(\tau+i+\tau'+j-2)}{2}} \\ \qquad \times \exp\left(\frac{Q \times \left(\beta'^{-1} - \beta^{-1}\right)}{4}\right) \times W_{\frac{(\tau'+j-\tau-i)}{2}, \frac{(1-\tau-i-\tau'-j)}{2}}(-bQ) \qquad \textit{for } Q < 0 \end{cases} \tag{22}$$

To define the parameters of this equation, suppose that $\lambda_i > 0$ for $i = 1,2,...,\rho$, $\lambda_i < 0$ for $i = \rho+1, \rho+2, ..., \rho+\varepsilon$, and $\lambda_i = 0$ for $i = \rho+\varepsilon+1, ..., m$ [21]. Note that some of $\lambda_i$'s for $i = 1,2,...,\rho$ may be equal. Let $\ell_j, j = 1,2,...,t$ denote the $t$ distinct positive eigenvalues among the $\lambda_i$'s and let $\tau_j$ denote the multiplicity of $\ell_j$ (in other words the number of $\lambda_i$'s which are equal to $\ell_j$). By similar definition for negative eigenvalues, we can write $Q$ as follows:

$$Q = \sum_{i=1}^{m} \lambda_i \left(U_i + \xi_i\right)^2 = T - V$$
$$T = \sum_{i=1}^{\rho} \lambda_i \left(U_i + \xi_i\right)^2 = \sum_{j=1}^{t} \ell_j \sum_{i=1}^{\tau_j} \left(U_{j_i} + \xi_{j_i}\right)^2 = \sum_{j=1}^{t} \ell_j T_j$$
$$d_j = \sum_{i=1}^{\tau_j} \xi_{j_i}^2 \qquad , \ j = 1,2,....,t \tag{23}$$
$$V = \sum_{i=\rho+1}^{\rho+\varepsilon} \left(-\lambda_i\right)\left(U_i + \xi_i\right)^2 = \sum_{j=t+1}^{t+w} \ell_j \sum_{i=1}^{\tau_j} \left(U_{j_i} + \xi_{j_i}\right)^2 = \sum_{j=t+1}^{t+w} \ell_j T_j$$
$$d_j = \sum_{i=1}^{\tau_j} \xi_{j_i}^2 \qquad , \ j = t+1,2,....,t+w$$

Then for positive eigenvalues we have:

$$d=\sum_{j=1}^{t} d_j \quad , \quad \tau=\sum_{j=1}^{t}\frac{\tau_j}{2} \quad , \quad c_j=1-\frac{\beta}{\ell_j} \ \ and \ \ \beta \ \ such\ that \left|1-\frac{\beta}{\ell_j}\right|<1 \ \ (for \ \ j=1,2,..,t)$$

$$i=1,2,.... \rightarrow \begin{cases} \theta_i=\dfrac{a_i}{(2\beta)^{\tau+i}} \\ a_0=e^{\frac{-d}{2}}\displaystyle\prod_{j=1}^{t}\left(\frac{\beta}{\ell_j}\right)^{\frac{\tau_j}{2}} \\ a_i=(2i)^{-1}\times\displaystyle\sum_{r=0}^{i-1} b_{i-r}a_r \\ b_i=i\times\beta\times\displaystyle\sum_{j=1}^{t}\left(\frac{d_j}{\ell_j}\right)c_j^{i-1}+\sum_{j=1}^{t}\tau_j c_j^{i} \end{cases} \tag{24}$$

And for negative eigenvalues we have:

$$\gamma=\sum_{j=t+1}^{t+w} d_j \quad , \quad \tau'=\sum_{j=t+1}^{t+w}\frac{\tau_j}{2} \quad , \quad c'_j=1-\frac{\beta'}{\ell_j} \ \ and \ \ \beta' \ \ such\,that \left|1-\frac{\beta'}{\ell_j}\right|<1 \ \ (for \ \ j=t+1,..,t+w)$$

$$j=1,2,.... \rightarrow \begin{cases} \theta'_j=\dfrac{a'_j}{(2\beta')^{\tau'+j}} \\ a'_0=e^{\frac{-\gamma}{2}}\displaystyle\prod_{i=t+1}^{t+w}\left(\frac{\beta'}{\ell_j}\right)^{\frac{\tau_j}{2}} \\ a'_j=(2j)^{-1}\times\displaystyle\sum_{r=0}^{j-1} b'_{j-r}a'_r \\ b'_j=j\times\beta'\times\displaystyle\sum_{i=t+1}^{t+w}\left(\frac{d_j}{\ell_j}\right)c_i'^{\,j-1}+\sum_{i=t+1}^{t+w}\tau_j c_i'^{\,j} \end{cases} \tag{25}$$

The parameter $b$ is defined as $\frac{(\beta^{-1}+\beta'^{-1})}{2}$ and $\Gamma(x)$ represents the well known Gamma function. Also, $W_{a,b}(x)$ denotes Whittaker's function, defined as follows:

$$W_{a,b}(x)=\frac{\Gamma(-2b)}{\Gamma(\frac{1}{2}-b-a)}x^{\left(b+\frac{1}{2}\right)}e^{\frac{-x}{2}}{}_1F_1\left(b-a+\frac{1}{2},2b+1;x\right)+ \frac{\Gamma(2b)}{\Gamma(\frac{1}{2}+b-a)}x^{\left(-b+\frac{1}{2}\right)}e^{\frac{-x}{2}}{}_1F_1\left(-b-a+\frac{1}{2},-2b+1;x\right)$$

$$where: \tag{26}$$

$${}_1F_1(c,d;x)=\sum_{i=1}^{\infty}\frac{(c)_i x^i}{(d)_i i!}$$

$$(e)_i=\frac{\Gamma(e+i)}{\Gamma(e)}$$

The density function of (22) is written based on the assumption that $\tau$ and $\tau'$ both are not nonnegative integers plus $\frac{1}{2}$. The probability density function for the special case of $\tau$ and $\tau'$ are both nonnegative integers plus $\frac{1}{2}$ is also given in [21].

By similar procedure, $P_{error_0}$, which denotes the error probability assuming that $H_0$ is the true hypothesis (i.e. $\mathbf{y} = \mathbf{H}_0\boldsymbol{\alpha}_0 + \mathbf{n}$) can be calculated and finally, by assuming equal probable hypothesis (i.e. $P(H_0) = P(H_1) = \frac{1}{2}$) we have:

$$P_{error} = \frac{1}{2}P_{error_0} + \frac{1}{2}P_{error_1} \tag{27}$$

The exact expression for the total error probability in general M-ary case cannot be easily evaluated. It includes the joint pdf of $L_i$'s which, because of not being independent, cannot be written using the marginal pdf of each variable. Instead, we can apply the well known union bound approach [18]. For the *p*th iteration, suppose that $H_i$ is the true hypothesis, so that the error probability can be written as:

$$P_{error_i} = P\left(L_1 > L_i \ or \ ... \ or \ L_{i-1} > L_i \ or \ L_{i+1} > L_i \ or \ ... \ or \ L_{N-p+1} > L_i \mid H_i\right) \tag{28}$$

Based on the theory of statistics, we have:

$$P(A \ or \ B) = P(A \bigcup B) \le \left(P(A) + P(B)\right) \tag{29}$$

Applying the inequality of (29) to error expression in (28), we can find a bound for $P_{error_i}$

$$P_{error_i} \le \sum_{\substack{b=1 \\ b \ne i}}^{N-p+1} P(L_b > L_i \mid H_i) \tag{30}$$

By writing similar expression for error probability in the case of assuming each of the other hypothesis being the true one, the total error probability can be written as follows (for equal probable hypothesis, i.e. $P(H_i) = \frac{1}{N-p+1}$):

$$\begin{aligned} P_{error} &= \sum_{i=1}^{N-p+1} P_{error_i} \times P(H_i) = \frac{1}{N-p+1}\sum_{i=1}^{N-p+1} P_{error_i} \\ \Rightarrow P_{error} &= \frac{1}{N-p+1}\sum_{i=1}^{N-p+1} P_{error_i} \le \frac{1}{N-p+1}\sum_{i=1}^{N-p+1}\sum_{\substack{b=1 \\ b \ne i}}^{N-p+1} P(L_b > L_i \mid H_i) \end{aligned} \tag{31}$$

In this equation, $P(L_b > L_i \mid H_i)$ can be considered as the error probability of binary case assuming $H_i$ is the true hypothesis which can be calculated using the results in (18) and (22).
For the modified IDBCS method, the probability density function of decision variable ($L_i$) is so complicated and is not theoretically tractable.

## 6. COMPUTER SIMULATION

To evaluate the performance of the proposed IDBCS algorithm and its modified version, we have exploited the test models of [10]. In all cases here, "N" is selected to be 256 and number of trials is set to be 1000. Also, basis matrix $\boldsymbol{\Psi}$ is assumed to be the identity matrix $\mathbf{I}_N$. For the first simulation, we use the Gaussian measurement matrix. Noise vector is assumed to be white Gaussian and normalized to have a

norm equal to 0.5. For each trial, a random $\mathbf{\Phi}$ and a random noise vector is used. In addition, we generate the binary signals with a support uniformly selected at random similar to simulations in [10]. Table 3 shows the values of the probability of correctly detecting nonzero indices of vector $\boldsymbol{\alpha}$ ($P_d$) versus various values of sparsity level (k) and number of measurements (m). As it is expected, by increasing the number of measurements, $P_d$ increases. Further, for greater values of k, more measurements should be used to detect the nonzero elements correctly. Since the approach of the proposed methods is similar to greedy ones (which one or more nonzero element would be estimated at each iteration), the results are compared to those of two well-known greedy algorithms: OMP method [13] and also ROMP [10]. As it has been claimed in [10], ROMP is the first algorithm to provide both benefits of optimization-based methods and greedy ones. Also, OMP is known as a basic algorithm in greedy methods group. Note that since we have simulated the noisy case, as it has been suggested in [10], the stopping criteria for ROMP has been changed by allowing the algorithm iterates at most *k* times.

As it can be seen in this table and the forthcoming tables, the performance of this new approach is slightly better than the OMP and ROMP results.

Table 3. Probability of correct detection of nonzero elements of sparse signal vs. various values of sparsity level (k) and number of measurements (m) for Gaussian measurement matrix

| | | OMP | ROMP | IDBCS | Modified IDBCS | | | OMP | ROMP | IDBCS | Modified IDBCS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| k = 2 | m=5 | 0 | 0 | 0.001 | 0.128 | k = 12 | m=5 | 0 | 0 | 0 | 0 |
| | m=25 | 0.83 | 0.62 | 0.919 | 0.97 | | m=25 | 0 | 0 | 0 | 0 |
| | m=50 | 1 | 0.992 | 1 | 1 | | m=50 | 0.01 | 0 | 0.06 | 0.099 |
| | m=75 | 1 | 1 | 1 | 1 | | m=75 | 0.616 | 0.02 | 0.735 | 0.826 |
| | m=125 | 1 | 1 | 1 | 1 | | m=125 | 0.997 | 0.398 | 0.999 | 1 |
| | m=256 | 1 | 1 | 1 | 1 | | m=256 | 1 | 1 | 1 | 1 |
| k = 4 | m=5 | 0 | 0 | 0 | 0 | k = 20 | m=5 | 0 | 0 | 0 | 0 |
| | m=25 | 0.27 | 0.03 | 0.393 | 0.62 | | m=25 | 0 | 0 | 0 | 0 |
| | m=50 | 0.989 | 0.498 | 0.986 | 0.998 | | m=50 | 0 | 0 | 0 | 0 |
| | m=75 | 1 | 0.886 | 1 | 1 | | m=75 | 0.014 | 0 | 0.055 | 0.093 |
| | m=125 | 1 | 1 | 1 | 1 | | m=125 | 0.941 | 0.188 | 0.964 | 0.967 |
| | m=256 | 1 | 1 | 1 | 1 | | m=256 | 1 | 1 | 1 | 1 |

Similar simulation has been applied to the Bernoulli measurement matrix and the results are presented in Table 4. As it can be seen, the Bernoulli case needs a few more measurements to have a similar $P_d$ as compared with Gaussian case, especially in greater values of k.

Table 4. Probability of correct detection of nonzero elements of sparse signal vs. Various values of sparsity level (k) and number of measurements (m) for Bernoulli measurement matrix

| | | OMP | ROMP | IDBCS | Modified IDBCS | | | OMP | ROMP | IDBCS | Modified IDBCS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| k = 2 | m=5 | 0 | 0 | 0 | 0 | k = 12 | m=5 | 0 | 0 | 0 | 0 |
| | m=25 | 0.874 | 0.664 | 0.886 | 0.916 | | m=25 | 0 | 0 | 0 | 0 |
| | m=50 | 1 | 1 | 1 | 1 | | m=50 | 0.037 | 0 | 0.043 | 0.05 |
| | m=75 | 1 | 1 | 1 | 1 | | m=75 | 0.708 | 0.04 | 0.718 | 0.748 |
| | m=125 | 1 | 1 | 1 | 1 | | m=125 | 1 | 0.526 | 1 | 1 |
| | m=256 | 1 | 1 | 1 | 1 | | m=256 | 1 | 1 | 1 | 1 |
| k = 4 | m=5 | 0 | 0 | 0 | 0 | k = 20 | m=5 | 0 | 0 | 0 | 0 |
| | m=25 | 0.27 | 0.012 | 0.265 | 0.345 | | m=25 | 0 | 0 | 0 | 0 |
| | m=50 | 0.982 | 0.516 | 0.981 | 0.981 | | m=50 | 0 | 0 | 0 | 0 |
| | m=75 | 1 | 0.902 | 1 | 1 | | m=75 | 0.024 | 0 | 0.021 | 0.025 |
| | m=125 | 1 | 1 | 1 | 1 | | m=125 | 0.942 | 0.203 | 0.945 | 0.945 |
| | m=256 | 1 | 1 | 1 | 1 | | m=256 | 1 | 1 | 1 | 1 |

Since the results for ROMP method show poorer performance than OMP, for incoming simulations, only OMP method has been considered to compare with the proposed algorithms.

The curves of required number of measurements to have a $P_d$ equal to 99% for various values of sparsity level for Gaussian measurement matrix can be seen in Fig. 1. These curves show the required size of measurement vector so that in 99% of trials, the indices of nonzero elements are detected correctly. As it is expected, the greater the sparsity level, the higher the number of measurements needed.

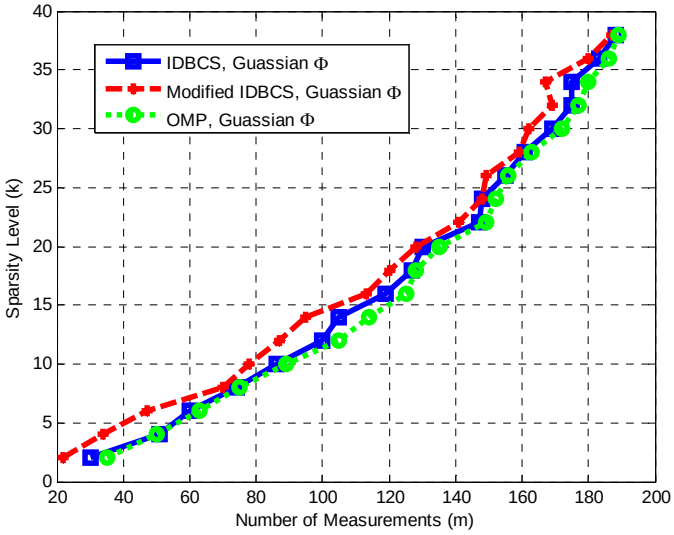


Fig. 1. Required number of measurements (m) to have a $P_d$ equal to 99% for various values of sparsity level (k)

Another parameter that affects the performance of reconstruction algorithms is the SNR of nonzero elements of sparse vector $\boldsymbol{\alpha}$. To have a better insight to the effect of SNR, a simulation with the previous conditions is performed, except that the noise vector is assumed to have normal distribution with zero mean and $\sigma^2 = 1$. Also, the amplitude of nonzero elements is taken equal and assigned a value so that the predetermined SNR is satisfied. The SNR for each nonzero element is defined as:

$$SNR = \frac{|\alpha_i|^2}{\sigma^2} \tag{32}$$

The resulting values for Fix sparsity level equal to 12 and different number of measurements versus various values of SNR can be seen in Tables 5 and 6. Also, for two values of $m$, the curves of probability of detection vs. SNR can be seen in Figs. 2 and 3. As expected, increasing the SNR of nonzero elements increases the probability of correctly detecting their indices.

Table 5. Probability of correct detection of nonzero elements of sparse signal vs. various values of SNR and number of measurements (m) for Gaussian measurement matrix (k=12)

| | | OMP | IDBCS | Modified IDBCS | | | OMP | IDBCS | Modified IDBCS |
|---|---|---|---|---|---|---|---|---|---|
| m = 51 | SNR=-6dB | 0.000 | 0.000 | 0.050 | m = 71 | SNR=-6dB | 0.009 | 0.013 | 0.091 |
| | SNR=1dB | 0.061 | 0.076 | 0.105 | | SNR=1dB | 0.656 | 0.647 | 0.700 |
| | SNR=6dB | 0.138 | 0.151 | 0.162 | | SNR=6dB | 0.758 | 0.771 | 0.811 |
| | SNR=11dB | 0.170 | 0.181 | 0.193 | | SNR=11dB | 0.799 | 0.815 | 0.847 |
| | SNR=14dB | 0.194 | 0.221 | 0.223 | | SNR=14dB | 0.839 | 0.842 | 0.849 |
| | SNR=16dB | 0.210 | 0.250 | 0.262 | | SNR=16dB | 0.853 | 0.847 | 0.856 |
| m = 61 | SNR=-6dB | 0.000 | 0.000 | 0.026 | m = 81 | SNR=-6dB | 0.033 | 0.036 | 0.103 |
| | SNR=1dB | 0.307 | 0.336 | 0.410 | | SNR=1dB | 0.866 | 0.857 | 0.874 |
| | SNR=6dB | 0.502 | 0.511 | 0.538 | | SNR=6dB | 0.924 | 0.930 | 0.968 |
| | SNR=11dB | 0.547 | 0.590 | 0.626 | | SNR=11dB | 0.941 | 0.947 | 0.956 |
| | SNR=14dB | 0.565 | 0.566 | 0.580 | | SNR=14dB | 0.949 | 0.948 | 0.951 |
| | SNR=16dB | 0.593 | 0.603 | 0.612 | | SNR=16dB | 0.951 | 0.950 | 0.968 |

Table 6. Probability of correct detection of nonzero elements of sparse signal vs. various values of SNR and number of measurements (m) for Bernoulli measurement matrix (k=12)

| | | OMP | IDBCS | Modified IDBCS | | | OMP | IDBCS | Modified IDBCS |
|---|---|---|---|---|---|---|---|---|---|
| m = 51 | SNR=-6dB | 0.000 | 0.000 | 0.007 | m = 71 | SNR=-6dB | 0.002 | 0.003 | 0.025 |
| | SNR=1dB | 0.031 | 0.027 | 0.085 | | SNR=1dB | 0.522 | 0.535 | 0.543 |
| | SNR=6dB | 0.084 | 0.064 | 0.067 | | SNR=6dB | 0.649 | 0.658 | 0.675 |
| | SNR=11dB | 0.098 | 0.091 | 0.101 | | SNR=11dB | 0.714 | 0.731 | 0.755 |
| | SNR=14dB | 0.101 | 0.098 | 0.116 | | SNR=14dB | 0.738 | 0.737 | 0.756 |
| | SNR=16dB | 0.103 | 0.123 | 0.138 | | SNR=16dB | 0.741 | 0.744 | 0.755 |
| m = 61 | SNR=-6dB | 0.000 | 0.000 | 0.027 | m = 81 | SNR=-6dB | 0.021 | 0.032 | 0.035 |
| | SNR=1dB | 0.217 | 0.236 | 0.289 | | SNR=1dB | 0.767 | 0.776 | 0.831 |
| | SNR=6dB | 0.357 | 0.346 | 0.371 | | SNR=6dB | 0.884 | 0.879 | 0.885 |
| | SNR=11dB | 0.425 | 0.433 | 0.434 | | SNR=11dB | 0.891 | 0.884 | 0.903 |
| | SNR=14dB | 0.435 | 0.439 | 0.444 | | SNR=14dB | 0.900 | 0.903 | 0.918 |
| | SNR=16dB | 0.440 | 0.441 | 0.465 | | SNR=16dB | 0.907 | 0.912 | 0.925 |

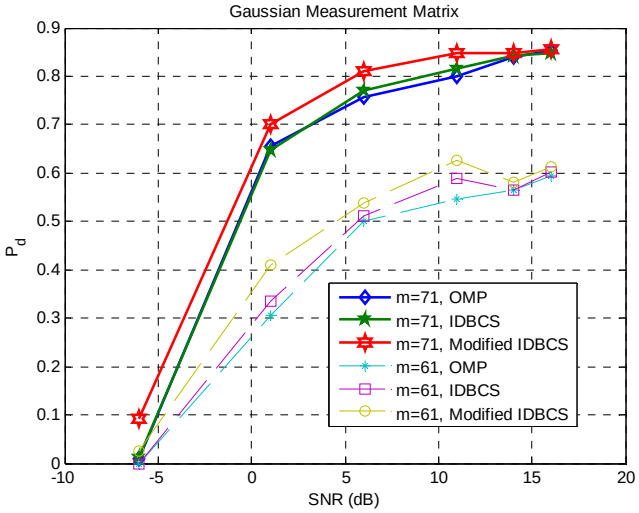


Fig. 2. Probability of correct detection of nonzero elements of sparse signal vs. various values of SNR and number of measurements (m=61 and 71) for Gaussian measurement matrix (k=12)

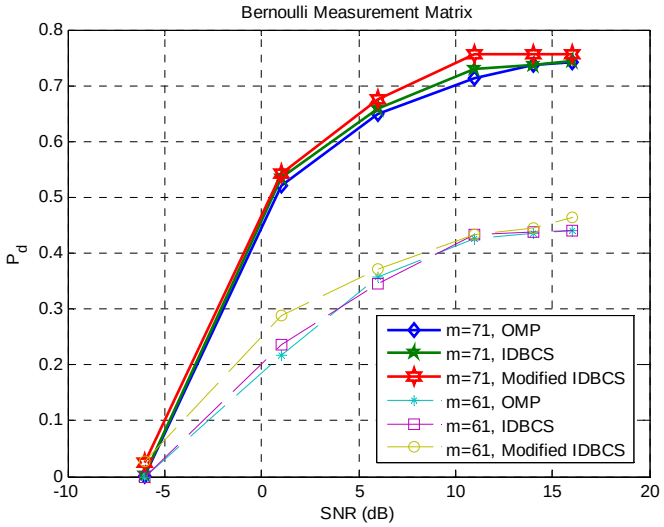


Fig. 3. Probability of correct detection of nonzero elements of sparse signal vs. various values of SNR and number of measurements (m=61 and 71) for Bernoulli measurement matrix (k=12)

The proposed detection methods can be compared with a complete detection theoretic approach at least for small N. For example, for N=3 we have a 7-ary hypothesis test in which 3 hypotheses assume

sparse signal of order one, 3 hypothesis assume order two (i.e. 011,101, and 110) and one hypothesis assumes order 3. In order to run simulations for this test, the following values for N, m and k have been considered.

Table 7. The simulation parameters for comparing the results of proposed method with a complete detection theoretic approach

| |
|---|
| N=3, m=3, k=1 |
| N=3, m=3, k=2 |
| N=3, m=2, k=1 |
| N=3, m=2, k=2 |
| N=3, m=1, k=1 |

As mentioned in [17], since this problem is a nested one, the following maximization (which represents a similar approach to GLRT in binary test) would always decide hypothesis with order 3 as the true one (assuming equal probable hypothesis).

$$j = \arg \max_i \left[ p\left(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_i, H_i\right) \right], i = 1,2,...,7 \tag{33}$$

In this test, the vector $\hat{\boldsymbol{\theta}}_i$ represents the maximum likelihood estimation of the vector of unknown nonzero elements under hypothesis $H_i$. Also, by using nested test we mean the hypothesis with order 1 and 2 are subsets of the hypothesis with order 3. Instead of (34), the following maximization should be done [17]:

$$j = \arg \max_i \left[ \ln p\left(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_i, H_i\right) - \frac{1}{2} \ln \det\left(\mathbf{I}(\hat{\boldsymbol{\theta}}_i)\right) \right], i = 1,2,...,7 \tag{34}$$

where $\mathbf{I}(.)$ denotes Fisher matrix. Using this approach, the simulation results would be as follows.

Table 8. The probability of correct detection of nonzero elements of sparse signal for simulations of comparison of OMP and IDBCS with a complete detection theoretic approach

| | OMP | IDBCS | Complete Detection Theoretic Approach |
|---|---|---|---|
| N=3, m=3, k=1 | 1 | 1 | 0.816 |
| N=3, m=3, k=2 | 1 | 1 | 0.918 |
| N=3, m=2, k=1 | 0.628 | 0.638 | 0.015 |
| N=3, m=2, k=2 | 0.316 | 0.395 | 0.02 |
| N=3, m=1, k=1 | 0.317 | 0.346 | 0.014 |

Note that, as it is mentioned in the texts of detection theory, GLRT approach (and similarly the above maximization) is not an optimum one. As it can be seen in simulation results, in this special problem, OMP and IDBCS show better performance compared to GLRT approach.

In section 5, the union bound for error probability of IDBCS has been derived. To justify the derived equations, the computer simulation results have been shown in this part. In order to run this simulation, we assume a vector with sparsity level equal to 3 (k=3), N=256 and various values for number of measurements (m). Also, we consider the final iteration in IDBCS and assume that the first 2 nonzero elements have been correctly detected in previous steps. The number of trials has been set to 10000. In each trial, the third index for nonzero element has been selected randomly and with equal probability from indices other than the previous 2 indices. Then the IDBCS algorithm has been run and the detected index has been compared to the true one. If these two indices were different, that trial has been marked as a trail with error occurrence. Finally, the number of these kinds of trials divided by 10000 gives the total error probability. The comparison between the simulation results and the union bound from theoretical equation in (31) can be seen in Fig. 4.

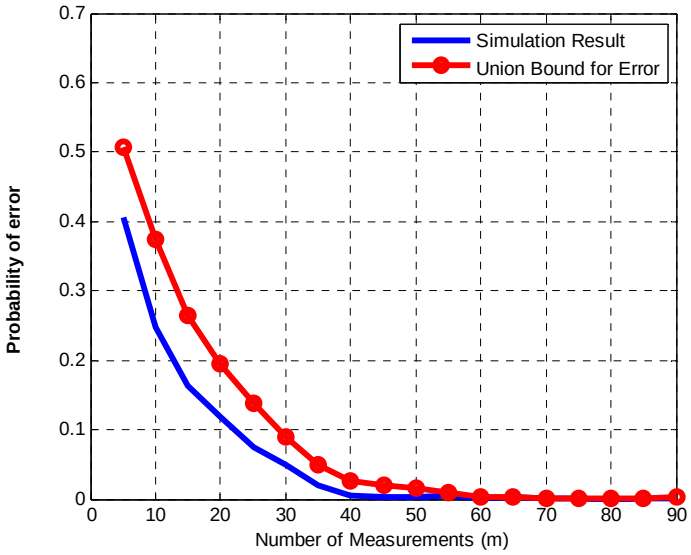


Fig. 4. Comparison between the analytical union bound for error probability with simulation results in IDBCS algorithm for various values of number of measurements (m) for Gaussian measurement matrix (k=3, N=256)

As can be concluded from the presented results, the performance of the proposed algorithms has slightly better performance in comparison with OMP. The performance of these algorithms can be improved by an estimation-confirmation approach for finding and estimating nonzero elements. Instead of detecting one nonzero element in each step, although there is the possibility of existence of some other nonzero elements in the noise term in the considered test, one can guess either the existence or non existence of potential indices. In other words, the potential nonzero elements can be estimated using some simple inaccurate approaches and then for confirmation step, the detection theory can be utilized for selection between these probable indices. In such an algorithm, for the detection test, the possibility for the noise vector not including signal terms would be higher (since nearly all the signal terms have been extracted with high probability in the first step). Such an approach is our trend for future works. Although in modified IDBCS, an attempt has been made for the deficiency in noise model to be covered by assuming a mean vector for noise based on the mean of all signal terms.

In addition to comparison between performance of the proposed methods, OMP and ROMP in detecting nonzero elements of sparse vector, in this part the computational load of algorithms has been compared. For OMP method in each iteration, there is a sorting and selection through N elements, multiplication by $\mathbf{H}_{m\times N}^{T}$ and finally, a least squares problem is solved. These matrix manipulations can be done in $\mathrm{O}(N)$, $\mathrm{O}(mN)$ and $\mathrm{O}(k^2N)$ respectively [10]. So, the overall cost for OMP in each iteration is of order $\mathrm{O}((k^2+m)N)$. Besides these operations, there is a regularization step in ROMP with total cost of $\mathrm{O}(k)$ in each iteration. So the total cost of ROMP in each iteration is of order $\mathrm{O}((k^2+m)N)$ similar to OMP [10].

Based on Appendix 1 and Table 1, the computational load for calculating $\left(\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\right)^{-1}$ is of order $O\left((p-1)^2\times m\right)$. The inner product of measured data $\mathbf{y}_{m\times 1}$ and each column of $\mathbf{H}_{m\times N}$ can be calculated and saved before running IDBCS method so that the terms $\mathbf{y}^T\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]$ can be substituted using the results in saved memory. Other than these parts, the complexity for matrix multiplications and also sorting and selection in Table 1 is of order $O\left(p^2+p\right)$ and $O(N)$, respectively. Since $m$ is the number of measurement which is much greater than 1, the total complexity load for IDBCS can be written of order $\mathrm{O}(N+m(p-1)^2)$ where $p$ denotes the iteration number. Assume that both OMP and IDBCS algorithms iterate k times until all $k$ nonzero elements have been estimated, it can be said that the

maximum computational load for IDBCS occurs in *k*th iteration and is of order $\mathrm{O}(N+m(k-1)^2)$. For modified IDBCS, the equation for $\hat{\mu}^{(p)}$ can be written as shown below:

$$\hat{\mu}^{(p)}=\frac{\mathbf{y}^T\times\left(\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\times\left(\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\right)^{-1}\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T-\mathbf{I}\right)\times\mathbf{d}^{(p)}}{\mathbf{d}^{(p)^T}\times\left(\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\times\left(\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\right)^{-1}\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T-\mathbf{I}\right)\times\mathbf{d}^{(p)}} \tag{35}$$

Based on Appendix 1, the computational load for $\left(\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\right)^{-1}$ is of order $O\left((p-1)^2\times m\right)$. The matrix multiplication in $\left(\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\times\left(\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\right)^{-1}\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T-\mathbf{I}\right)$ is of order $O\left(mp^2+m^2p\right)$. Considering the vector-matrix multiplications in nominator and denominator of $\hat{\mu}^{(p)}$, the total complexity load for calculating $\hat{\mu}^{(p)}$ is of order $O\left(mp^2+m^2p+m^2+m\right)$ which can be approximately written as $O\left(mp^2+m^2p\right)$. Also, by substituting the ML estimation of nonzero elements, the term $\left(\mathbf{y}-\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\times\begin{bmatrix}\hat{\alpha}_{\Gamma_{p-1}}\\ \hat{\alpha}_i\end{bmatrix}-\hat{\mu}^{(p)}\mathbf{d}^{(p)}\right)$ in Table 2 can be written as follows:

$$-\left(\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\times\left(\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\right)^{-1}\times\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]^T-\mathbf{I}\right)\times\left(\mathbf{y}-\hat{\mu}^{(p)}\mathbf{d}^{(p)}\right) \tag{36}$$

The first term has been calculated before, so that calculation of $\left(\mathbf{y}-\left[\mathbf{H}_{\Gamma_{p-1}}\ \mathbf{h}_i\right]\times\begin{bmatrix}\hat{\alpha}_{\Gamma_{p-1}}\\ \hat{\alpha}_i\end{bmatrix}-\hat{\mu}^{(p)}\mathbf{d}^{(p)}\right)$ only includes a matrix-vector multiplication which can be calculated with order $O\left(m^2\right)$. Finally, the criteria can be calculated with order $O(m)$ and sorting can be done with $O(N)$. So for modified IDBCS, the computational load would be of order $\mathrm{O}(N+mp^2+pm^2)$ in each iteration, so that maximum load for modified algorithm is of order $\mathrm{O}(N+mk^2+km^2)$.

## 7. CONCLUSION

Two new algorithms for reconstruction of sparse signals measured by compressed sensing have been developed. The new approaches are based on detection theory. By applying composite multiple hypothesis tests, the recovery problem has been solved based on an iterative method. It has been assumed that the signal is corrupted with additive normal noise with known parameters. The union bound for the error probability of one detector has been analytically calculated. Also, computer simulations have been provided which show acceptable performance for the proposed methods in reconstruction of sparse signals. Due to the general structure of the proposed approach, it can be developed for other noise models with known or unknown parameters which are our trends for future research.

***Acknowledgement*:** We wish to thank the Iran Telecommunication Research Center (ITRC) for supporting this work financially.

## REFERENCES

1. Chen, Y., Nasrabadi, N. M. & Tran, T. D. Hyperspectral image classification via kernel sparse representation, *IEEE Tarns. On Geoscience and Remote Sensing, Signal Processing Mag*., Vol. 51, No. 1, pp. 217-231.
2. Nikoofar, H. R. & Sharafat, A. R. (2010). Modulation classification for burst-mode QAM signals in multipath fading channels. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, Vol. 34, No. B3, pp. 257-274.

3. Barbotin, Y., Hormati, A., Rangan, S. & Vetterli, M. (2012). Estimation of sparse MIMO channels with common support. *IEEE Trans. On Communications*, Vol. 60, No. 12, pp. 3705-3716.
4. Al-Safadi, E. B. & Al-Naffouri, T. Y. (2012). Peak reduction and clipping mitigation in OFDM by augmented compressive sensing. *IEEE Transactions. on Signal Processing*, Vol. 60, No. 7, pp. 3834-3839.
5. Tang, Z., Blacquiere G. & Leus, G. (2011). Aliasing-free wideband beamforming using sparse signal representation. *IEEE Trans. On Signal Processing*, Vol. 59, No. 7, pp. 3464-3469.
6. Sen, S. (2013). OFDM radar space-time adaptive processing by exploiting spatio-temporal sparsity. *IEEE Transactions on Signal Processing*, Vol. 61, No. 1, pp. 118-130.
7. Du, X., Duan, C. & Hu, W. (2013). Sparse representation based autofocusing technique for ISAR images. *IEEE Tarns. On Geoscience and Remote Sensing*, Vol. 51, Issue 3, pp. 1826-1835.
8. Candès, E., Romberg, J. & Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, Vol. 52, No. 2, pp. 489–509.
9. Donoho, D. (2006). Compressed sensing. *IEEE Trans. Inform. Theory*, Vol. 52, No. 4, pp. 1289-1306.
10. Needel, D. (2009). Topics in compressed sensing. PhD dissertation, Dept. of Mathematics, California Univ., Davis, CA.
11. Kamali, A., Aghabozorgi Sahaf, M. R., Doost Hosseini, A. M. & Tadaion, A. A. (2013). Block subspace pursuit for block-sparse signal reconstruction. Iranian Journal of Science and Technology, Transactions of Electrical Engineering, Vol. 37, No. E1, pp. 1-16.
12. Chen, S., Donoho, D. & Saunders, M. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput*., Vol. 20, No. 1, pp. 33-61.
13. Tropp, J. A. & Gilbert, A. C. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Information Theory*, Vol. 53, No. 12, pp. 4655-4666.
14. Donoho, D. L., Tsaig, Y., Drori, I. & Starck, J. L. (2007). Sparse solution of underdetermined linear equations by Stagewise orthogonal matching pursuit. *IEEE Trans. Information Theory*, Vol. 58, No. 2, pp. 1094–1121.
15. Paris, S., Mary, D. & Ferrari, A. (2013). Detection tests using sparse models, with application to hyperspectral data. *IEEE Trans. on Signal Processing*, Vol. 61, No. 6, pp.1481-1494.
16. Rama, B., Rao, S. Chatterjee & Ottersten, B. (2012). Detection of sparse random signals using compressive measurements. *Proc. of 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3257–3260.
17. Kay, S. T. (1998). *Fundamentals of statistical signal processing: Detection theory*. Prentice Hall , Inc.
18. Proakis, J. G. (1998). *Digital communications*. McGraw Hill, 4th edition.
19. Mathai, A. M. & Provost, S. B. (1992). Quadratic forms in random variables. M. D. Inc.
20. Gradshteyn, I. S. & Ryzhik, I. M. (2007). *Table of integrals, series, and products*. Academic Press, Seventh Edition.
21. Provost, S. B. & Rudiuk, E. M. (1996). The exact distribution of indefinite quadratic forms in noncentral normal vectors. *Ann. Ins. Statist. Math.*, Vol. 48, No. 2, pp. 381-394.
22. Horn, R. A. & Johnson, C. R. (1985). *Matrix analysis*. Cambridge University Press.


## APPENDIX A: NOTE ON COMPUTATIONAL COMPLEXITY

The most computational load of the proposed algorithms at $p$ th iteration is the matrix inversion of $\left(\left[\mathbf{H}_{\Gamma_{p-1}} \ \mathbf{h}_i\right]^T \times \left[\mathbf{H}_{\Gamma_{p-1}} \ \mathbf{h}_i\right]\right)^{-1}$ which is a matrix inversion of size $p \times p$. However, by utilizing the following lemma, this inversion can be done using $\left(\mathbf{H}_{\Gamma_{p-1}}{}^T \times \mathbf{H}_{\Gamma_{p-1}}\right)^{-1}$ which is calculated in the previous iteration.

*Lemma 1:* For proper dimensions of matrices, the following equality (inverse of partitioned matrix) can be proved if all inverses exist [22]:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{K}\mathbf{C}\mathbf{A}^{-1}) & (-\mathbf{A}^{-1}\mathbf{B}\mathbf{K}) \\ (-\mathbf{K}\mathbf{C}\mathbf{A}^{-1}) & \mathbf{K} \end{bmatrix}$$
$$\mathbf{K} = \left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \tag{A.1}$$

Now, let's rewrite our matrix inversion problem as follows:

$$\left(\left[\mathbf{H}_{\Gamma_{p-1}} \ \mathbf{h}_i\right]^T \times \left[\mathbf{H}_{\Gamma_{p-1}} \ \mathbf{h}_i\right]\right)^{-1} = \begin{bmatrix} \mathbf{H}_{\Gamma_{p-1}}^T \mathbf{H}_{\Gamma_{p-1}} & \mathbf{H}_{\Gamma_{p-1}}^T \mathbf{h}_i \\ \mathbf{h}_i^T \mathbf{H}_{\Gamma_{p-1}} & \mathbf{h}_i^T \mathbf{h}_i \end{bmatrix}^{-1} \tag{A.2}$$

Using lemma 1, we can write:

$$\begin{aligned} \mathbf{A} &= \mathbf{H}_{\Gamma_{p-1}}^T \mathbf{H}_{\Gamma_{p-1}} \\ \mathbf{B} &= \mathbf{H}_{\Gamma_{p-1}}^T \mathbf{h}_i \\ \mathbf{C} &= \mathbf{h}_i^T \mathbf{H}_{\Gamma_{p-1}} \\ \mathbf{D} &= \mathbf{h}_i^T \mathbf{h}_i \end{aligned} \tag{A.3}$$

So after some mathematical manipulations, we have:

$$\mathbf{F}_p = \left(\left[\mathbf{H}_{\Gamma_{p-1}} \ \mathbf{h}_i\right]^T \times \left[\mathbf{H}_{\Gamma_{p-1}} \ \mathbf{h}_i\right]\right)^{-1} = \begin{bmatrix} \mathbf{f}_{11} & \mathbf{f}_{21} \\ \mathbf{f}_{12} & f_{22} \end{bmatrix}$$

*where*:

$$\begin{aligned} \mathbf{f}_{11} &= \left(\mathbf{H}_{\Gamma_{p-1}}^T \mathbf{H}_{\Gamma_{p-1}}\right)^{-1} + \left(\mathbf{H}_{\Gamma_{p-1}}^T \mathbf{H}_{\Gamma_{p-1}}\right)^{-1} \mathbf{H}_{\Gamma_{p-1}}^T \mathbf{h}_i \times \left(\mathbf{h}_i^T \mathbf{h}_i - \mathbf{h}_i^T \mathbf{H}_{\Gamma_{p-1}} \left(\mathbf{H}_{\Gamma_{p-1}}^T \mathbf{H}_{\Gamma_{p-1}}\right)^{-1} \mathbf{H}_{\Gamma_{p-1}}^T \mathbf{h}_i\right)^{-1} \times \mathbf{h}_i^T \mathbf{H}_{\Gamma_{p-1}} \left(\mathbf{H}_{\Gamma_{p-1}}^T \mathbf{H}_{\Gamma_{p-1}}\right)^{-1} \\ \mathbf{f}_{12} &= -\left(\mathbf{H}_{\Gamma_{p-1}}^T \mathbf{H}_{\Gamma_{p-1}}\right)^{-1} \mathbf{H}_{\Gamma_{p-1}}^T \mathbf{h}_i \times \left(\mathbf{h}_i^T \mathbf{h}_i - \mathbf{h}_i^T \mathbf{H}_{\Gamma_{p-1}} \left(\mathbf{H}_{\Gamma_{p-1}}^T \mathbf{H}_{\Gamma_{p-1}}\right)^{-1} \mathbf{H}_{\Gamma_{p-1}}^T \mathbf{h}_i\right)^{-1} \\ \mathbf{f}_{21} &= -\left(\mathbf{h}_i^T \mathbf{h}_i - \mathbf{h}_i^T \mathbf{H}_{\Gamma_{p-1}} \left(\mathbf{H}_{\Gamma_{p-1}}^T \mathbf{H}_{\Gamma_{p-1}}\right)^{-1} \mathbf{H}_{\Gamma_{p-1}}^T \mathbf{h}_i\right)^{-1} \times \mathbf{h}_i^T \mathbf{H}_{\Gamma_{p-1}} \left(\mathbf{H}_{\Gamma_{p-1}}^T \mathbf{H}_{\Gamma_{p-1}}\right)^{-1} \\ f_{22} &= \left(\mathbf{h}_i^T \mathbf{h}_i - \mathbf{h}_i^T \mathbf{H}_{\Gamma_{p-1}} \left(\mathbf{H}_{\Gamma_{p-1}}^T \mathbf{H}_{\Gamma_{p-1}}\right)^{-1} \mathbf{H}_{\Gamma_{p-1}}^T \mathbf{h}_i\right)^{-1} \end{aligned} \tag{A.4}$$

If we define $\mathbf{F}_{p-1} = \left(\mathbf{H}_{\Gamma_{p-1}}^T \mathbf{H}_{\Gamma_{p-1}}\right)^{-1}$ which has been calculated in $(p-1)$th iteration, and also $\mathbf{G} = \mathbf{F}_{p-1} \mathbf{H}_{\Gamma_{p-1}}^T \mathbf{h}_i$ then we have:

$$\begin{aligned} \mathbf{f}_{11} &= \mathbf{F}_{p-1} + \mathbf{G} \times \left(\mathbf{h}_i^T \mathbf{h}_i - \mathbf{h}_i^T \mathbf{H}_{\Gamma_{p-1}} \mathbf{G}\right)^{-1} \times \mathbf{G}^T \\ \mathbf{f}_{12} &= -\mathbf{G} \times \left(\mathbf{h}_i^T \mathbf{h}_i - \mathbf{h}_i^T \mathbf{H}_{\Gamma_{p-1}} \mathbf{G}\right)^{-1} \\ \mathbf{f}_{21} &= -\left(\mathbf{h}_i^T \mathbf{h}_i - \mathbf{h}_i^T \mathbf{H}_{\Gamma_{p-1}} \mathbf{G}\right)^{-1} \times \mathbf{G}^T \\ f_{22} &= \left(\mathbf{h}_i^T \mathbf{h}_i - \mathbf{h}_i^T \mathbf{H}_{\Gamma_{p-1}} \mathbf{G}\right)^{-1} \end{aligned} \tag{A.5}$$

The computational complexity of $\mathbf{F}_p$ by calculating $\left(\left[\mathbf{H}_{\Gamma_{p-1}} \ \mathbf{h}_i\right]^T \times \left[\mathbf{H}_{\Gamma_{p-1}} \ \mathbf{h}_i\right]\right)^{-1}$ and (A.5) is of order $O(m^3)$ and $O\left((p-1)^2 \times m\right)$, respectively. Since $p$ is the iteration number (which is much less than $m$), by using (A.5), the computational complexity of $\mathbf{F}_p$ dramatically decreases compared to calculation of $\left(\left[\mathbf{H}_{\Gamma_{p-1}} \ \mathbf{h}_i\right]^T \times \left[\mathbf{H}_{\Gamma_{p-1}} \ \mathbf{h}_i\right]\right)^{-1}$.