Structure Based Bayesian Sparse Reconstruction Using Non-Gaussian Prior

Ahmed A. Quadeer, Syed Faraz Ahmed, and Tareq Y. Al-Naffouri
Electrical Engineering Department
King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia
Email: {aquadeer, sfaraz, naffouri}@kfupm.edu.sa

***Abstract*—In this paper, we present a fast Bayesian method for sparse signal recovery that makes a collective use of the sparsity information, a priori statistical properties, and the structure involved in the problem to obtain near optimal estimates at very low complexity. Specifically, we utilize the rich structure present in the sensing matrix encountered in many signal processing applications to develop a fast reconstruction algorithm when the statistics of the sparse signal are non-Gaussian or unknown. The proposed method outperforms the widely used convex relaxation approaches as well as greedy matching pursuit techniques all while operating at a much lower complexity.**

I. Introduction

Compressive Sensing/Compressed Sampling (CS) is a fairly new field of research that is finding many applications in statistics and signal processing [1]. CS has been utilized in numerous applications where the signal of interest is sparse in nature, for example, peak to average power ratio reduction in OFDM [2], [3], impulse noise estimation and cancellation in power-line communication and digital subscriber lines (DSL) [4], [**?**], magnetic resonance imaging (MRI) [6], channel estimation in communication systems [7], ultra-wideband (UWB) channel estimation [8], [9], and radar design [10].

The CS problem can be set up as follows. Let $\mathbf{x} \in \mathbb{C}^N$ be a K-sparse signal (i.e. a signal that can be represented by $K << N$ coefficients over an N-dimensional space) in some domain and $\mathbf{y} \in \mathbb{C}^M$ be the observation vector given by

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

Here $\mathbf{\Psi}$ is an $M \times N$ sensing matrix that is assumed to be incoherent with the domain in which $\mathbf{x}$ is sparse and $\mathbf{n}$ is the complex additive white Gaussian noise $\mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$. When $M << N$, we have an under-determined system of equations that has infinite solutions. However, if the signal of interest is sparse, this problem can be solved theoretically by posing it as an ℓ_0 minimization problem. Unfortunately, solving such an ℓ_0 minimization problem is NP-hard [11]. Thus, different sub-optimal approaches, categorized under Compressive Sensing, have been presented in literature to solve this problem. In [11]-[14], it has been shown that $\mathbf{x}$ can be reconstructed with high probability in polynomial time by using convex relaxation approaches with a penalty on the number of measurements. This is done by solving a relaxed ℓ_1 minimization problem using linear programming instead of ℓ_0 minimization. For ℓ_1 minimization to accurately reconstruct the sparse signal, the sensing matrix $\mathbf{\Psi}$ must obey the restricted isometry property (RIP) (see [15], [16] for details). However, convex relaxation approaches also have some drawbacks. The linear programming technique used to solve the ℓ_1 minimization problem makes these methods computationally quite complex ($\mathcal{O}(M^2 N^{3/2})$ when interior point methods are used [17]). Many fast greedy/matching pursuit algorithms [18]-[20] have been proposed to reduce this complexity. These methods are also not able to use any a priori statistical information (apart from sparsity information) about the signal support and additive noise. The a priori statistical information can be utilized to refine the output obtained by using convex relaxation methods but the performance will still be bottle-necked by the ability of these methods to recover the sparse signals support. Convex relaxation methods cannot utilize the structure present in the sensing matrix (e.g. a partial DFT and Toeplitz matrices) as well. In fact, they require the sensing matrix to be as random as possible to obtain best results.

This work was partially supported by SABIC through an internally funded project from DSR, KFUPM (Project No. SB101006) and partially by the National Science, Technology, and Innovation Plan (Project No. 09-ELE763-04). The work of Tareq Y. Al-Naffouri was also supported by the Fullbright Scholar Program.

A. Motivation and Paper Organization

This paper employs a low-complexity Bayesian approach that takes collective advantage of the a priori statistical & sparsity information, and the structure of the sensing matrix $\mathbf{\Psi}$ while pursuing a direct statistical approach. To the best of our knowledge, this is the first paper that makes use of the structure of the sensing matrix for sparsity recovery. In addition, the paper also demonstrates how this approach tackles the issues like non-Guassian prior and others stated above.

The paper organization is as follows. The signal model is described in Section II followed by Section III which details the MMSE estimation of $\mathbf{x}$. Section IV explains how the structure of the sensing matrix can be used for the sparse signal recovery in a divide and conquer manner. Section V details the proposed sparse reconstruction algorithm. In

Section VI, the algorithm is applied to Discrete Fourier Transform (DFT) matrix as an example. The paper ends with numerical simulations in Section VII that compare the performance of the proposed algorithm with other sparse reconstruction algorithms present in literature and by our conclusion in Section VIII.

II. Signal Model

In this paper, we are interested in recovery a sparse vector modelled as

$$\mathbf{x} = \mathbf{x}_B \odot \mathbf{x}_{NG}$$

where the entries of $\mathbf{x}_B$ are i.i.d. Bernoulli with success probability p and the entries of $\mathbf{x}_{NG}$ are drawn identically and independently from some non-Gaussian or unknown distribution. When the support $\mathcal{S}$ of $\mathbf{x}$ is known, we can equivalently write (1) as

$$\mathbf{y}|\mathcal{S} = \mathbf{\Psi}_\mathcal{S}\mathbf{x}_\mathcal{S} + \mathbf{n}_\mathcal{S} \tag{2}$$

where $\mathbf{\Psi}_\mathcal{S}$ denotes the sub-matrix formed by columns $\{\boldsymbol{\psi}_i : i \in \mathcal{S}\}$, indexed by the support $\mathcal{S}$.

III. MMSE Estimation of $\mathbf{x}$

Our task is to obtain the optimum estimate of $\mathbf{x}$ given the observation $\mathbf{y}$ and the sensing matrix $\mathbf{\Psi}$. The MMSE estimate of $\mathbf{x}$ given the observation $\mathbf{y}$ can be expressed as

$$\hat{\mathbf{x}}_{\text{MMSE}} = \mathbb{E}[\mathbf{x}|\mathbf{y}] = \sum_\mathcal{S} p(\mathcal{S}|\mathbf{y})\mathbb{E}[\mathbf{x}|\mathbf{y},\mathcal{S}] \tag{3}$$

where the sum is over the possible support set $\mathcal{S}$ of $\mathbf{x}$. For large values of N, the computational complexity for evaluating this expression becomes very high (as the sum needs to be evaluated over 2^N terms). The idea is to approximate the expression in (3) in such a way that the inherent computational complexity is reduced. Let's start by demonstrating how to calculate the various terms in (3).

1) Calculating $\mathbb{E}[\mathbf{x}|\mathbf{y},\mathcal{S}]$*:* Recall that the relationship between $\mathbf{y}$ and $\mathbf{x}$ is linear (see (1)). When $\mathbf{x}_\mathcal{S}$ is non-Gaussian or when its statistics are unknown, it is difficult or even impossible to calculate the expectation $\mathbb{E}[\mathbf{x}|\mathbf{y},\mathcal{S}]$. Thus we replace it by the best linear unbiased estimate (BLUE), i.e.

$$\mathbf{x}_\mathcal{S} = (\mathbf{\Psi}_\mathcal{S}^\mathsf{H}\mathbf{\Psi}_\mathcal{S})^{-1}\mathbf{\Psi}_\mathcal{S}^\mathsf{H}\mathbf{y} \tag{4}$$

2) Calculating $p(\mathcal{S}|\mathbf{y})$*:* Using Bayes rule, we can write

$$p(\mathcal{S}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathcal{S})p(\mathcal{S})}{\sum_\mathcal{S} p(\mathbf{y}|\mathcal{S})p(\mathcal{S})} \tag{5}$$

Here, the probability $p(\mathcal{S})$ is given by

$$p(\mathcal{S}) = p^{|\mathcal{S}|}(1-p)^{N-|\mathcal{S}|} \tag{6}$$

as the elements of $\mathbf{x}$ follow the Bernoulli process with success probability p. Note that $\mathbf{x}$ is an unknown arbitrary with support $\mathcal{S}$. Therefore, given the support $\mathcal{S}$, all we can say about $\mathbf{y}$ is that it is formed by a vector in the subspace spanned by the columns of $\mathbf{\Psi}_\mathcal{S}$, plus a white Gaussian noise vector $\mathbf{n}$. The conditional density of $\mathbf{y}$ given $\mathcal{S}$ is proportional to the negative exponential of the projection of $\mathbf{y}$ on the orthogonal complement of the span of the columns of $\mathbf{\Psi}_\mathcal{S}$. It follows that the corresponding MAP metric can be approximated by

$$p(\mathbf{y}|\mathcal{S}) \sim \exp\left(-\frac{1}{\sigma_n^2}\|\mathbf{P}_\mathcal{S}^\perp\mathbf{y}\|^2\right) \tag{7}$$

where

$$\mathbf{P}_\mathcal{S}^\perp = \mathbf{I} - \mathbf{\Psi}_\mathcal{S}\left(\mathbf{\Psi}_\mathcal{S}^\mathsf{H}\mathbf{\Psi}_\mathcal{S}\right)^{-1}\mathbf{\Psi}_\mathcal{S}^\mathsf{H} \tag{8}$$

is the orthogonal projector onto the orthogonal complement of the subspace spanned by the columns of $\mathbf{\Psi}_\mathcal{S}$.

3) Evaluation over $\mathcal{S}$*:* Note that the two summations that appear in (3) and (5) need to be evaluated over all possibilities of $\mathcal{S}$ (there will be 2^N such sets). Instead of this exhaustive search approach, we can limit the MMSE evaluation over the most probable support of $\mathbf{x}$, $\mathcal{S}_r$. The search space would then reduce to $2^{|\mathcal{S}_r|}$ points. There are two techniques to limit the search space.

1. Convex Relaxation: Starting from (1), we can use the standard convex relaxation tools [11]-[14] to find the support of the sparse vector $\mathbf{x}$. This is done by solving the ℓ_1 minimization problem and retaining some largest P non-zero values.

2. FBMP: A low-complexity Bayesian technique was presented in [23] that finds the dominant support and the MMSE estimate of $\mathbf{x}$ jointly by using the Gaussian prior. It involves an intelligent greedy search over all the possible combinations of the supports in pursuit of the dominant ones. The algorithm starts with zero support size and selects the best support of size one after evaluating the MAP-Gaussian likelihood for all the N positions. The next support of size two is then selected based on the support of size one selected in the previous step. This greedy search is concluded when the support of size P is reached where P is selected such that $\text{P}(\mathcal{S} > P)$ is very small.[1] The above procedure is repeated again by starting with the second best support of size one. The algorithm performs D such greedy searches in the same manner and performs MMSE estimation over the reduced search space (PD supports explored during the whole procedure). We can also make use of other greedy algorithms [18]-[20].

IV. A Structure Based Bayesian Recovery Approach

Neither of the two methods mentioned in Section III-3 make use of the structure of the sensing matrix $\mathbf{\Psi}$. Let us investigate how this structure can help us substantially reduce the computational complexity involved in evaluating the MMSE estimate. While in most CS literature the sensing matrix $\mathbf{\Psi}$ is assumed to be drawn from a random constellation [11]-[14], in most signal processing and communication

[1] As $\mathcal{S}$ is a binomial distribution $\sim \text{B}(N,p)$, it can be approximated by a Gaussian distribution $\sim \mathcal{N}(Np, Np(1-p))$ when $Np > 5$. Thus in this case, $\text{P}(\mathcal{S} > P) = \frac{1}{2}\text{erfc}\left(\frac{P-N(1-p)}{\sqrt{(2Np(1-p))}}\right)$ [23].

applications this matrix is highly structured. Thus, $\mathbf{\Psi}$ could be a partial DFT matrix (for example in OFDM applications including PAPR reduction [2], [3], and impulse noise cancellation in DSL [4], [**?**]) or a Toeplitz matrix (encountered in many convolution applications [7] and UWB [8], [9]) or a DCT matrix (as in image compression [24]).

Since $\mathbf{\Psi}$ is a wide matrix $(M << N)$, it is rank-deficient and thus its columns are not orthogonal. However, the M orthogonal columns that span the columns space of $\mathbf{\Psi}$ can usually be identified, for instance, in impulse noise cancellation in DSL and the above mentioned applications. The remaining $N - M$ columns of $\mathbf{\Psi}$ group around these orthogonal columns to form semi-orthogonal clusters (see Section VI for an example, see also [9], [5], and [25]).

Now we discuss how to obtain the MMSE estimate of $\mathbf{x}$ using the orthogonality of clusters. The MMSE estimate of $\mathbf{x}$ in (3) can be written as

$$\begin{aligned}\hat{\mathbf{x}}_{\text{MMSE}} &= \begin{bmatrix} \hat{\mathbf{x}}_1 \\ \hat{\mathbf{x}}_2 \\ \vdots \\ \hat{\mathbf{x}}_P \end{bmatrix} = \begin{bmatrix} \mathbb{E}[\mathbf{x}_1|\mathbf{y}] \\ \mathbb{E}[\mathbf{x}_2|\mathbf{y}] \\ \vdots \\ \mathbb{E}[\mathbf{x}_P|\mathbf{y}] \end{bmatrix} \\ &= \begin{bmatrix} \sum_{\mathcal{S}} p(\mathcal{S}|\mathbf{y})\mathbb{E}[\mathbf{x}_1|\mathbf{y},\mathcal{S}] \\ \sum_{\mathcal{S}} p(\mathcal{S}|\mathbf{y})\mathbb{E}[\mathbf{x}_2|\mathbf{y},\mathcal{S}] \\ \vdots \\ \sum_{\mathcal{S}} p(\mathcal{S}|\mathbf{y})\mathbb{E}[\mathbf{x}_P|\mathbf{y},\mathcal{S}] \end{bmatrix} \end{aligned} \tag{9}$$

Let $\mathcal{S}^o$ be a possible support of $\mathbf{x}$. The columns of $\mathbf{\Psi}_{\mathcal{S}}$ in (2) can be grouped into semi-orthogonal columns as follows

$$\mathbf{\Psi}_{\mathcal{S}^o} = [\mathbf{\Psi}_{\mathcal{S}_1}\ \mathbf{\Psi}_{\mathcal{S}_2}\ \cdots\ \mathbf{\Psi}_{\mathcal{S}_P}]$$

Due to orthogonality, we have

$$\mathbf{\Psi}_{\mathcal{S}_i}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{S}_j} \simeq 0$$

This can be used to show that (see equations (10)-(14) on the next page)

$$\mathbb{E}[\mathbf{x}|\mathbf{y},\mathcal{S}] = \begin{bmatrix} \mathbb{E}[\mathbf{x}_1|\mathbf{y},\mathcal{S}_1] \\ \mathbb{E}[\mathbf{x}_2|\mathbf{y},\mathcal{S}_2] \\ \vdots \\ \mathbb{E}[\mathbf{x}_P|\mathbf{y},\mathcal{S}_P] \end{bmatrix} \tag{15}$$

Due to orthogonality, we can also show that the likelihood $p(\mathbf{y}|\mathcal{S})$ can be factored as (see Appendix for details)

$$p(\mathbf{y}|\mathcal{S}) = \exp\left(\frac{P-1}{\sigma_n^2}\|\mathbf{y}\|^2\right)\prod_{i=1}^{P} p(\mathbf{y}|\mathcal{S}_i) \tag{16}$$

from which we can omit the common factor, $\exp\left(\frac{P-1}{\sigma_n^2}\|\mathbf{y}\|^2\right)$. This decomposition coupled with a similar decomposition

$$p(\mathcal{S}) = \prod_{i=1}^{P} p(\mathcal{S}_i) \tag{17}$$

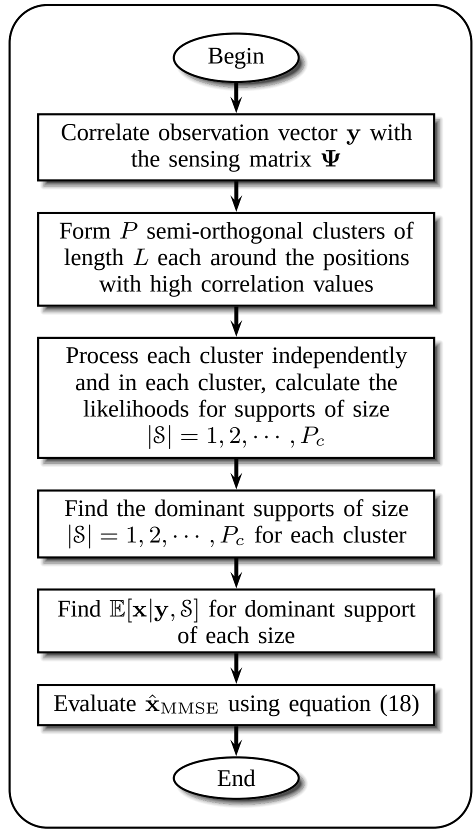


Fig. 1: Flowchart of the OC algorithm

together with (15) allows us to eventually show that

$$\begin{aligned}\hat{\mathbf{x}}_{\text{MMSE}} &= \sum_{\mathcal{S}} p(\mathcal{S}|\mathbf{y})\mathbb{E}[\mathbf{x}|\mathbf{y},\mathcal{S}] \\ &= \begin{bmatrix} \sum_{\mathcal{S}} p(\mathcal{S}|\mathbf{y})\mathbb{E}[\mathbf{x}_1|\mathbf{y},\mathcal{S}] \\ \sum_{\mathcal{S}} p(\mathcal{S}|\mathbf{y})\mathbb{E}[\mathbf{x}_2|\mathbf{y},\mathcal{S}] \\ \vdots \\ \sum_{\mathcal{S}} p(\mathcal{S}|\mathbf{y})\mathbb{E}[\mathbf{x}_P|\mathbf{y},\mathcal{S}] \end{bmatrix} \\ &= \begin{bmatrix} \sum_{\mathcal{S}_1} p(\mathcal{S}_1|\mathbf{y})\mathbb{E}[\mathbf{x}_1|\mathbf{y},\mathcal{S}_1] \\ \sum_{\mathcal{S}_2} p(\mathcal{S}_2|\mathbf{y})\mathbb{E}[\mathbf{x}_2|\mathbf{y},\mathcal{S}_2] \\ \vdots \\ \sum_{\mathcal{S}_P} p(\mathcal{S}_P|\mathbf{y})\mathbb{E}[\mathbf{x}_P|\mathbf{y},\mathcal{S}_P] \end{bmatrix} \end{aligned} \tag{18}$$

In other words, orthogonality allows us to approach the problem of sparse reconstruction in a divide and conquer manner.

V. Orthogonal Clustering (OC) Algorithm for Sparse Reconstruction

We summarize the main steps of the proposed OC algorithm for sparse signal recovery in Figure 1. In the following, we discuss these steps in detail.

A. Determine dominant positions

Consider the model given in (1). By correlating the observation vector $\mathbf{y}$ with the columns of the sensing matrix

$$
\begin{aligned}
\mathbb{E}[\mathbf{x}|\mathbf{y},\mathcal{S}] &= \begin{bmatrix} \mathbb{E}[\mathbf{x}_1|\mathbf{y},\mathcal{S}] \\ \mathbb{E}[\mathbf{x}_2|\mathbf{y},\mathcal{S}] \\ \vdots \\ \mathbb{E}[\mathbf{x}_P|\mathbf{y},\mathcal{S}] \end{bmatrix} && (10)\\
&= \left(\begin{bmatrix} \mathbf{\Psi}_{\mathcal{S}_1}^{\mathsf{H}} \\ \mathbf{\Psi}_{\mathcal{S}_2}^{\mathsf{H}} \\ \vdots \\ \mathbf{\Psi}_{\mathcal{S}_P}^{\mathsf{H}} \end{bmatrix} \begin{bmatrix} \mathbf{\Psi}_{\mathcal{S}_1} & \mathbf{\Psi}_{\mathcal{S}_2} & \cdots & \mathbf{\Psi}_{\mathcal{S}_P} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{\Psi}_{\mathcal{S}_1}^{\mathsf{H}} \\ \vdots \\ \mathbf{\Psi}_{\mathcal{S}_P}^{\mathsf{H}} \end{bmatrix} \mathbf{y} && (11)\\
&\simeq \begin{bmatrix} (\mathbf{\Psi}_{\mathcal{S}_1}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{S}_1})^{-1} & 0 & \cdots & 0 \\ 0 & (\mathbf{\Psi}_{\mathcal{S}_2}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{S}_2})^{-1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & (\mathbf{\Psi}_{\mathcal{S}_P}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{S}_P}^{\mathsf{H}})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{\Psi}_{\mathcal{S}_1}^{\mathsf{H}} \\ \mathbf{\Psi}_{\mathcal{S}_2}^{\mathsf{H}} \\ \vdots \\ \mathbf{\Psi}_{\mathcal{S}_P}^{\mathsf{H}} \end{bmatrix} \mathbf{y} && (12)\\
&= \begin{bmatrix} (\mathbf{\Psi}_{\mathcal{S}_1}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{S}_1})^{-1}\mathbf{\Psi}_{\mathcal{S}_1}^{\mathsf{H}}\mathbf{y} \\ (\mathbf{\Psi}_{\mathcal{S}_2}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{S}_2})^{-1}\mathbf{\Psi}_{\mathcal{S}_2}^{\mathsf{H}}\mathbf{y} \\ \vdots \\ (\mathbf{\Psi}_{\mathcal{S}_P}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{S}_P})^{-1}\mathbf{\Psi}_{\mathcal{S}_P}^{\mathsf{H}}\mathbf{y} \end{bmatrix} && (13)\\
&= \begin{bmatrix} \mathbb{E}[\mathbf{x}_1|\mathbf{y},\mathcal{S}_1] \\ \mathbb{E}[\mathbf{x}_2|\mathbf{y},\mathcal{S}_2] \\ \vdots \\ \mathbb{E}[\mathbf{x}_P|\mathbf{y},\mathcal{S}_P] \end{bmatrix} && (14)
\end{aligned}
$$

$\mathbf{\Psi}$, we can determine the dominant regions to which the support of $\mathbf{x}$ is most likely to belong.[2]

B. *Form semi-orthogonal clusters*

Select the index with the largest correlation and form a cluster of length L around it (The cluster length L can be estimated by investigating the correlation between the columns of the sensing matrix [25]). Continue in a similar fashion for the remaining large indices untill P clusters are formed. If a particular index is already present in the clusters formed by the previous indices, it is discarded and the next dominant index is considered.

C. *Find the dominant supports and their likelihoods*

Let L be the cluster size and let P_c denote the maximum possible support size in a cluster[3]. For each of the P clusters, find the most probable support of size $|\mathcal{S}| = 1, 2, \cdots, P_c$ by calculating the likelihoods for each size (using (7)). Each cluster is processed independently due to the semi-orthogonality between clusters. The expected value of $\mathbf{x}$ given $\mathbf{y}$ can also be evaluated using (4) for the dominant support of each size.

[2] The dominant support regions could also be found by employing some convex relaxation approach and retaining the support corresponding to the dominant values.

[3] P_c is evaluated in a way similar to P (see [25] for further details).

D. *Evaluate the estimate of* $\mathbf{x}$

Once we have the dominant supports for each cluster, their likelihoods, and $\mathbb{E}[\mathbf{x}|\mathbf{y},\mathcal{S}]$, the MMSE estimate of $\mathbf{x}$ can be evaluated using equation (18).

VI. Partial DFT Matrix: An Example

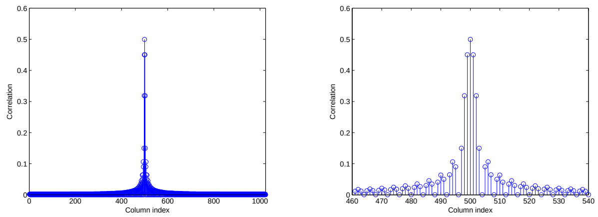
In the preceding sections, we demonstrated how semi-orthogonality allowed us to calculate the MMSE estimate of the sparse signal in a divide and conquer manner. It turns out that we can reduce the complexity even further by utilizing other structural properties of the sensing matrix. We demonstrate this by considering the case when the sensing matrix is a partial DFT matrix, i.e.

$$\mathbf{\Psi} = \mathbf{SF}$$

where $\mathbf{S}$ is an $M \times N$ selection matrix consisting of all zeros and exactly one entry equal to 1 per row and $\mathbf{F}$ denotes the $N \times N$ unitary DFT matrix given by

$$[\mathbf{F}]_{a,b} = \frac{1}{\sqrt{N}} e^{-j2\pi ab/N}$$

with $a, b \in \{0, 1, \ldots, N-1\}$. To impose the desired semi-orthogonal structure, the sensing matrix must consist of a continuous band of sensing frequencies (as opposed to the random sensing matrix required in general in CS). This form of sensing matrix is encountered in many OFDM problems [2], [3], [5], [25]. The correlation between two columns in



(a) (b)

Fig. 2: The 500^{th} column has high correlation with its neighbours only (a) Correlation with all columns (b) Correlation with neighbouring columns

a partial DFT sensing matrix can be shown to be

$$\psi_k^{\mathsf{H}}\psi_{k'} = \begin{cases} 1, & (k = k') \\ \left|\frac{\sin\left(\pi(k-k')M/N\right)}{M\sin(\pi(k-k')/N)}\right|, & (k \neq k') \end{cases} \quad (19)$$

which is a function of the difference $(k - k') \bmod N$, so it suffices to consider correlation of one column with the remaining ones. Figure 2 illustrates this correlation for $N = 1024$ and $M = 256$. It can be seen that the selected column (500^{th} in this case) has high correlation only with its immediate neighbours and is almost uncorrelated with other columns. This implies that a cluster formed around a particular column will be almost orthogonal to the clusters formed around the other farther columns.

A. Reducing the Computational Complexity

In this subscetion, we will demonstrate how the structure of the partial DFT sensing matrix enables us to reduce the computational complexity in the following two dimensions.

1) From one cluster to the other: For the case when the sensing matrix is a partial DFT matrix, we show that different clusters are very similar. In fact, one cluster can be obtained from the other by a simple modulation operation. This enables us to reduce the complexity involved in the likelihood calculations across clusters.

To this end, let $\psi_1, \ \psi_2, \ \cdots, \ \psi_L$ denote the sensing columns associated with the first cluster. Then, it is easy to show that the corresponding columns for the ith cluster of equal length that is $\triangle_i$ columns away are

$$\psi_1 \odot \psi_{\triangle_i}, \ \psi_2 \odot \psi_{\triangle_i}, \ \cdots, \ \psi_L \odot \psi_{\triangle_i}$$

where $\psi_{\triangle_i}$ is some constant vector that depends only on the distane $\triangle_i$ between the two columns.[4] Now assume that we calculate the projection matrix $\mathbf{P}_{\Omega_1}^{\perp}$ for a set of columns Ω_1 in the first cluster and let Ω_i denote the same set of columns chosen in the ith cluster (i.e. a distance $\triangle_i$ away). For the set of columns Ω_i in the ith cluster, the projection matrix in (8) can be written as

$$\begin{aligned} \mathbf{P}_{\Omega_i}^{\perp} &= \mathbf{I} - \Omega_i \left(\Omega_i^{\mathsf{H}}\Omega_i\right)^{-1}\Omega_i^{\mathsf{H}} \\ &= \mathbf{I} - (\Omega_1 \odot \mathbf{A}_{\triangle_i}) \\ &\quad \left[(\Omega_1 \odot \mathbf{A}_{\triangle_i})^{\mathsf{H}}(\Omega_1 \odot \mathbf{A}_{\triangle_i})\right]^{-1}(\Omega_1 \odot \mathbf{A}_{\triangle_i})^{\mathsf{H}} \end{aligned}$$

where $\mathbf{A}_{\triangle_i}$ is a matrix consisting of L identical columns $\psi_{\triangle_i}$, i.e. $\mathbf{A}_{\triangle_i} = \left[\psi_{\triangle_i} \ \psi_{\triangle_i} \ \cdots \ \psi_{\triangle_i}\right]$. Thus,

$$\mathbf{P}_{\Omega_i}^{\perp} = \mathbf{I} - (\Omega_1 \odot \mathbf{A}_{\triangle_i})\left(\Omega_1^{\mathsf{H}}\Omega_1\right)^{-1}(\Omega_1 \odot \mathbf{A}_{\triangle_i})^{\mathsf{H}} \quad (20)$$

Let's see how this helps to write $\|\mathbf{P}_{\Omega_i}^{\perp}\mathbf{y}\|^2$ in terms of $\mathbf{P}_{\Omega_i}^{\perp}$,

$$\begin{aligned} \|\mathbf{P}_{\Omega_i}^{\perp}\mathbf{y}\|^2 &= \mathbf{y}^{\mathsf{H}}\big[\mathbf{I} - (\Omega_1 \odot \mathbf{A}_{\triangle_i}) \\ &\quad \left(\Omega_1^{\mathsf{H}}\Omega_1\right)^{-1}(\Omega_1 \odot \mathbf{A}_{\triangle_i})^{\mathsf{H}}\big]\mathbf{y} \\ &= \|\mathbf{y}\|^2 - (\mathbf{y}^{\mathsf{H}} \odot \psi_{\triangle_i}^{\mathsf{T}}) \\ &\quad (\Omega_1\left(\Omega_1^{\mathsf{H}}\Omega_1\right)^{-1}\Omega_1^{\mathsf{H}})(\psi_{\triangle_i}^{*} \odot \mathbf{y}) \\ &= \|\mathbf{y} \odot \psi_{\triangle_i}^{*}\|^2_{\mathbf{P}_{\Omega_1}^{\perp}} \end{aligned} \quad (21)$$

In other words, to calculate the likelihood for a particular set of columns Ω_i of the ith cluster, we can re-use $\mathbf{P}_{\Omega_1}^{\perp}$ (evaluated for the same set of columns of the first cluster) and just modulate $\mathbf{y}$ by $\psi_{\triangle_i}^{*}$.

2) Within a cluster: Consider a single cluster for which we need to calculate the likelihoods for the supports of size $|\mathcal{S}| = 1, 2, ..., P_c$. We proceed to evaluate these likelihoods in an order recursive manner i.e. we utilize the computations for $|\mathcal{S}| = 1$ to evaluate the likelihood for $|\mathcal{S}| = 2$ and so on. Recall that evaluating the likelihood involves computing the following norm

$$\|\mathbf{y}\|^2_{\mathbf{P}^{\perp}} = \|\mathbf{y}\|^2 - \mathbf{y}^{\mathsf{H}}\mathbf{\Psi}\left(\mathbf{\Psi}^{\mathsf{H}}\mathbf{\Psi}\right)^{-1}\mathbf{\Psi}^{\mathsf{H}}\mathbf{y}$$

where $\mathbf{\Psi}$ corresponds to the candidate sensing matrix. We would like to compute this norm recursively. When $\mathbf{\Psi}$ consists of $|\mathcal{S}|$ columns i.e. it is of size $M \times |\mathcal{S}|$ we use the notation $\mathbf{\Psi}_{|\mathcal{S}|}$ instead and the corresponding likelihood is given by

$$\mathcal{L}_{|\mathcal{S}|} = \exp\left(-\frac{1}{\sigma_n^2}\left\{\|\mathbf{y}\|^2 - \mathbf{y}^{\mathsf{H}}\mathbf{\Psi}_{|\mathcal{S}|}\left(\mathbf{\Psi}_{|\mathcal{S}|}^{\mathsf{H}}\mathbf{\Psi}_{|\mathcal{S}|}\right)^{-1}\mathbf{\Psi}_{|\mathcal{S}|}^{\mathsf{H}}\mathbf{y}\right\}\right)$$

Now we add another column to $\mathbf{\Psi}_{|\mathcal{S}|}$ to obtain $\mathbf{\Psi}_{|\mathcal{S}|+1} = \left[\mathbf{\Psi}_{|\mathcal{S}|} \ \ \psi_i\right]$ and we calculate the corresponding likelihood $\mathcal{L}_{|\mathcal{S}|+1}$ in a recursive manner from the previous likelihood $\mathcal{L}_{|\mathcal{S}|}$. This approach to calculate the likelihood in recursion hinges upon calculating the inverse term $\mathbf{\Sigma}_{|\mathcal{S}|+1} \triangleq (\mathbf{\Psi}_{|\mathcal{S}|+1}^{\mathsf{H}}\mathbf{\Psi}_{|\mathcal{S}|+1})^{-1}$ recursively. To this end, note that using the block inversion formula, we can relate $\mathbf{\Sigma}_{|\mathcal{S}|+1}$ to $\mathbf{\Sigma}_{|\mathcal{S}|}$ as

$$\mathbf{\Sigma}_{|\mathcal{S}|+1} = \begin{bmatrix} \mathbf{\Sigma}_{|\mathcal{S}|} + \frac{1}{\xi_i}\omega_i\omega_i^{\mathsf{H}} & -\frac{1}{\xi_i}\omega_i \\ -\frac{1}{\xi_i}\omega_i & \frac{1}{\xi_i} \end{bmatrix} \quad (22)$$

where

$$\omega_i \triangleq \mathbf{\Sigma}_{|\mathcal{S}|}(\mathbf{\Psi}_{|\mathcal{S}|}^{\mathsf{H}}\psi_i) \quad (23)$$

[4] For example, if we use the last M columns of sensing matrix, $\psi_{\triangle_i} = \exp\left(-\frac{\jmath 2\pi(N-(M-k))}{N}\triangle_i\right)$ with $k = 0, 1, \cdots, M-1$.

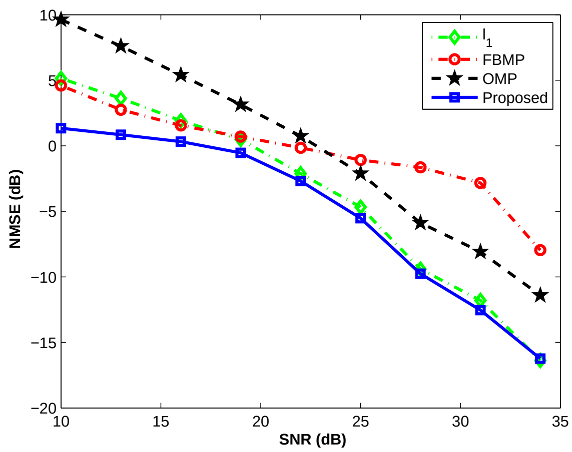


Fig. 3: NMSE vs SNR comparison for DFT matrix

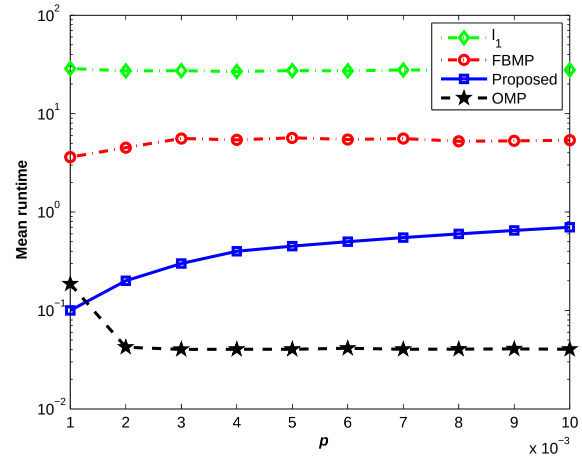


Fig. 4: Mean runtime comparison for DFT matrix

and

$$\xi_i = \|\boldsymbol{\psi}_i\|^2 + (\boldsymbol{\psi}_i^{\mathsf{H}} \boldsymbol{\Psi}_{|\mathcal{S}|}) \boldsymbol{\Sigma}_{|\mathcal{S}|} (\boldsymbol{\Psi}_{|\mathcal{S}|}^{\mathsf{H}} \boldsymbol{\psi}_i) \tag{24}$$

$$= \|\boldsymbol{\psi}_i\|^2 + \boldsymbol{\omega}_i^{\mathsf{H}} (\boldsymbol{\Psi}_{|\mathcal{S}|}^{\mathsf{H}} \boldsymbol{\psi}_i) \tag{25}$$

Following this recursion, we can express $\mathcal{L}_{|\mathcal{S}|+1}$ as shown in equation (26) where in the i^{th} step of recursion δ_i is the part which needs to be calculated.

VII. Simulation Results

In this section, the performance of the proposed OC algorithm is compared with some popular sparse reconstruction techniques including convex relaxation methods [12], OMP [18], and FBMP [23]. The parameters for these algorithms are set in such a way that they perform to the best of their abilities. For a fair comparison, we perform MMSE refinement to the output of convex relaxation methods and OMP. The parameters used in all the simulations are $N = 800$, $M = \frac{N}{4} = 200$, and $p = 10^{-2}$ (unless stated otherwise). Figure 3 compares the Normalized Mean Square Error (NMSE) of the algorithms for the case when sensing matrix is a DFT matrix. In FBMP implementation, the number of branches to explore (D) is set to 10 and it is allowed to estimate the hyper-parameters using its approximate maximum-likelihood estimator (with E set to 5)[23]. It can be seen that the proposed algorithm easily outperforms OMP and FBMP while ℓ_1 minimization method performs quite close to it (but at the cost of much higher complexity). OC easily beats FBMP and ℓ_1 minimization method in execution time as demonstrated in Figure 4.

VIII. Conclusion

In this paper, we present a fast Bayesian sparse reconstruction algorithm which we call Orthogonal Clustering (OC) that uses the sparsity, non-Gaussian (or unknown) prior, and the structure of the sensing matrix in a collective manner. Two unique aspects of the paper are 1) its ability to deal with non-Gaussian prior (or prior with unknown statistics) in a Bayesian framework and 2) the ability to use the structure of the sensing matrix for drastic reduction in complexity of sparse reconstruction algorithm. The paper focussed on the partial DFT matrix to illustrate how OC works and its superior performance compared to the popular sparse recovery algorithms present in literature (see [25] for other examples and for extenstion of OC to the case of Gaussian prior).

Appendix: Proof of Equation (16)

To prove equation (16), we start from the likelihood (7) which involves the orthogonal projector $\mathbf{P}_{\mathcal{S}}^{\perp}$ given by

$$\mathbf{P}_{\mathcal{S}}^{\perp} = \mathbf{I} - \boldsymbol{\Psi}_{\mathcal{S}} \left(\boldsymbol{\Psi}_{\mathcal{S}}^{\mathsf{H}} \boldsymbol{\Psi}_{\mathcal{S}} \right)^{-1} \boldsymbol{\Psi}_{\mathcal{S}}^{\mathsf{H}} \tag{27}$$

We assume for simplicity that $\boldsymbol{\Psi}_{\mathcal{S}}$ consists of two clusters. Thus,

$$\boldsymbol{\Psi}_{\mathcal{S}} = [\boldsymbol{\Psi}_{\mathcal{S}_1} \ \boldsymbol{\Psi}_{\mathcal{S}_2}]$$

In this case, the inverse involved in (27) is given by

$$\left(\boldsymbol{\Psi}_{\mathcal{S}}^{\mathsf{H}} \boldsymbol{\Psi}_{\mathcal{S}} \right)^{-1} = \left(\begin{bmatrix} \boldsymbol{\Psi}_{\mathcal{S}_1}^{\mathsf{H}} \\ \boldsymbol{\Psi}_{\mathcal{S}_2}^{\mathsf{H}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Psi}_{\mathcal{S}_1} & \boldsymbol{\Psi}_{\mathcal{S}_2} \end{bmatrix} \right)^{-1}$$

$$= \begin{bmatrix} \boldsymbol{\Psi}_{\mathcal{S}_1}^{\mathsf{H}} \boldsymbol{\Psi}_{\mathcal{S}_1} & \boldsymbol{\Psi}_{\mathcal{S}_1}^{\mathsf{H}} \boldsymbol{\Psi}_{\mathcal{S}_2} \\ \boldsymbol{\Psi}_{\mathcal{S}_2}^{\mathsf{H}} \boldsymbol{\Psi}_{\mathcal{S}_1} & \boldsymbol{\Psi}_{\mathcal{S}_2}^{\mathsf{H}} \boldsymbol{\Psi}_{\mathcal{S}_2} \end{bmatrix}^{-1}$$

As $\boldsymbol{\Psi}_{\mathcal{S}_1}$ and $\boldsymbol{\Psi}_{\mathcal{S}_2}$ are almost orthogonal, (28) results in a block diagonal matrix given by

$$\left(\boldsymbol{\Psi}_{\mathcal{S}}^{\mathsf{H}} \boldsymbol{\Psi}_{\mathcal{S}} \right)^{-1} \simeq \begin{bmatrix} \boldsymbol{\Psi}_{\mathcal{S}_1}^{\mathsf{H}} \boldsymbol{\Psi}_{\mathcal{S}_1} & 0 \\ 0 & \boldsymbol{\Psi}_{\mathcal{S}_2}^{\mathsf{H}} \boldsymbol{\Psi}_{\mathcal{S}_2} \end{bmatrix}^{-1}$$

$$\simeq \begin{bmatrix} (\boldsymbol{\Psi}_{\mathcal{S}_1}^{\mathsf{H}} \boldsymbol{\Psi}_{\mathcal{S}_1})^{-1} & 0 \\ 0 & (\boldsymbol{\Psi}_{\mathcal{S}_2}^{\mathsf{H}} \boldsymbol{\Psi}_{\mathcal{S}_2})^{-1} \end{bmatrix}$$

$$
\begin{aligned}
\mathcal{L}_{|\mathcal{S}|+1} &= \exp\left(-\frac{1}{\sigma_n^2}\left[\|\mathbf{y}\|^2 - \mathbf{y}^{\mathsf{H}}\mathbf{\Psi}_{|\mathcal{S}|+1}\mathbf{\Sigma}_{|\mathcal{S}|+1}\mathbf{\Psi}_{|\mathcal{S}|+1}^{\mathsf{H}}\mathbf{y}\right]\right)\\
&= \exp\left(-\frac{1}{\sigma_n^2}\left[\|\mathbf{y}\|^2 - \mathbf{y}^{\mathsf{H}}\mathbf{\Psi}_{|\mathcal{S}|}\mathbf{\Sigma}_{|\mathcal{S}|}\mathbf{\Psi}_{|\mathcal{S}|}^{\mathsf{H}}\mathbf{y}\right]\right)\\
&\quad \exp\left(-\frac{1}{\sigma_n^2}\left[\frac{1}{\xi_i}|(\mathbf{y}^{\mathsf{H}}\mathbf{\Psi}_{|\mathcal{S}|})\boldsymbol{\omega}_i|^2 - \frac{2}{\xi_i}\mathrm{Re}\{(\mathbf{y}^{\mathsf{H}}\boldsymbol{\psi}_i)\boldsymbol{\omega}_i^{\mathsf{H}}(\mathbf{\Psi}_{|\mathcal{S}|}^{\mathsf{H}}\mathbf{y})\} + \frac{1}{\xi_i}|\mathbf{y}^{\mathsf{H}}\boldsymbol{\psi}_i|^2\right]\right)\\
&= \mathcal{L}_{|\mathcal{S}|}\underbrace{\exp\left(-\frac{1}{\sigma_n^2}\left[\frac{1}{\xi_i}|(\mathbf{y}^{\mathsf{H}}\mathbf{\Psi}_{|\mathcal{S}|})\boldsymbol{\omega}_i|^2 - \frac{2}{\xi_i}\mathrm{Re}\{(\mathbf{y}^{\mathsf{H}}\boldsymbol{\psi}_i)\boldsymbol{\omega}_i^{\mathsf{H}}(\mathbf{\Psi}_{|\mathcal{S}|}^{\mathsf{H}}\mathbf{y})\} + \frac{1}{\xi_i}|\mathbf{y}^{\mathsf{H}}\boldsymbol{\psi}_i|^2\right]\right)}_{\delta_i} \qquad (26)
\end{aligned}
$$

Substituting this in (27) yields

$$
\begin{aligned}
\mathbf{P}_{\mathcal{S}}^{\perp} &= \mathbf{I} - \mathbf{\Psi}_{\mathcal{S}_1}(\mathbf{\Psi}_{\mathcal{S}_1}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{S}_1})^{-1}\mathbf{\Psi}_{\mathcal{S}_1}^{\mathsf{H}} - \mathbf{\Psi}_{\mathcal{S}_2}(\mathbf{\Psi}_{\mathcal{S}_2}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{S}_2})^{-1}\mathbf{\Psi}_{\mathcal{S}_2}^{\mathsf{H}}\\
&= -\mathbf{I} + \left(\mathbf{I} - \mathbf{\Psi}_{\mathcal{S}_1}(\mathbf{\Psi}_{\mathcal{S}_1}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{S}_1})^{-1}\mathbf{\Psi}_{\mathcal{S}_1}^{\mathsf{H}}\right)\\
&\quad + \left(\mathbf{I} - \mathbf{\Psi}_{\mathcal{S}_2}(\mathbf{\Psi}_{\mathcal{S}_2}^{\mathsf{H}}\mathbf{\Psi}_{\mathcal{S}_2})^{-1}\mathbf{\Psi}_{\mathcal{S}_2}^{\mathsf{H}}\right)\\
&= -\mathbf{I} + \mathbf{P}_{\mathcal{S}_1}^{\perp} + \mathbf{P}_{\mathcal{S}_2}^{\perp}
\end{aligned}
$$

Incorporating this result in the MAP-metric (7) yields

$$
\begin{aligned}
p(\mathbf{y}|\mathcal{S}) &= \exp\left(\frac{1}{\sigma_n^2}\|\mathbf{y}\|^2\right)\exp\left(-\frac{1}{\sigma_n^2}\|\mathbf{P}_{\mathcal{S}_1}^{\perp}\mathbf{y}\|^2\right)\\
&\quad \exp\left(-\frac{1}{\sigma_n^2}\|\mathbf{P}_{\mathcal{S}_2}^{\perp}\mathbf{y}\|^2\right)\\
&= \exp\left(\frac{1}{\sigma_n^2}\|\mathbf{y}\|^2\right)p(\mathbf{y}|\mathcal{S}_1)p(\mathbf{y}|\mathcal{S}_2) \qquad (28)
\end{aligned}
$$

which can be written in general form for P clusters as

$$
p(\mathbf{y}|\mathcal{S}) = \exp\left(\frac{P-1}{\sigma_n^2}\|\mathbf{y}\|^2\right)\prod_{i=1}^{P} p(\mathbf{y}|\mathcal{S}_i) \qquad (29)
$$

References

[1] E. J. Candes and M. B. Wakin, "An Introduction To Compressive Sampling," IEEE Signal Processing Magazine, vol.25, no.2, pp.21-30, March 2008

[2] T. Y. Al-Naffouri, E. B. Al-Safadi, and M. E. Eltayeb, "OFDM Peak-to-Average Power Ratio Reduction Method," *US patent pending.*

[3] E. B. Al-Safadi and T. Y. Al-Naffouri, "On Reducing the Complexity of Tone-Reservation Based PAPR Reduction Techniques by Compressive Sensing," *IEEE Global Telecommunications Conference*, pp.1-6, Nov. 30 2009-Dec. 4 2009.

[4] G. Caire, T. Y. Al-Naffouri, and A. Narayanan, "Impulse noise cancellation in OFDM: an application of compressed sensing," *IEEE Int. Symp. on Inform. Theory*, Canada, Jul. 2008.

[5] T. Y. Al-Naffouri, A. A. Quadeer, and G. Caire, "Impulse noise estimation and cancellation in DSL using Orthogonal Clustering," *IEEE Int. Symp. Inform. Theory*, Russia, 2011.

[6] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182-1195, 2007.

[7] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz Compressed Sensing Matrices With Applications to Sparse Channel Estimation," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, Nov. 2010.

[8] J. L. Paredes, G. R. Arce, and Z. Wang, "Ultra-Wideband Compressed Sensing: Channel Estimation," *IEEE Jrnl. of Selected Topics in Signal Proc.*, vol. 1, no. 3, pp. 383-395, Oct. 2007.

[9] S. F. Ahmed, T. Y. Al-Naffouri and A. H. Muqaibel, "Low-Complexity MAP Based Channel Support Estimation for Impulse Radio Ultra-Wideband (IR-UWB) Communications," *2011 IEEE International Conference on Ultra-Wideband*, vol. 1, pp. 133-137, Sep. 2011.

[10] M. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Proc.*, vol. 57, no. 6, 2007.

[11] E. Candes, J. Romberg and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[12] E. J. Candes and P. Randall. "Highly robust error correction by convex programming," *submitted for publication.*

[13] E. Candes and T. Tao, "The Dantzig selector: statistical estimation when p is much larger than n," *submitted for publication.*

[14] J. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. on Inform. Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.

[15] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. on Inform. Theory*, vol. 51, no. 12, pp. 4203-4215, Dec. 2005.

[16] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. on Inform. Theory*, vol. 52, no. 12, pp. 5406-5425, Dec. 2006.

[17] I. E. Nesterov, A. Nemirovskii, and Y. Nesterov, *Interior-Point Polynomial Algorithms in Convex Programming.* SIAM, 1994.

[18] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. on Inform. Theory*, vol. 53 no. 12, pp. 4655-4666, Dec. 2007.

[19] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of Computational Mathematics, Springer*, 2009.

[20] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comp. Harmonic Analysis*, 2008.

[21] T. Y. Al-Naffouri, A. A. Quadeer, H. Hmida, and F. F. Al-Shaalan, "Impulsive Noise Estimation and Cancellation in DSL Using Compressive Sensing," *IEEE Int. Symp. Circuits and Systems*, Brazil, May 2011.

[22] A. Montanari, B. Prabhakar, and D. Tse, "Belief propagation based multi-user detection," *Allerton Conf. on Communications, Control and Computing*, USA, Sep. 2005.

[23] P. Schniter, L. C. Potter, and J. Ziniel, "Fast Bayesian Matching Pursuit," *Inform. Theory and Appl. Workshop*, pp. 326-333, 2008.

[24] J. Wen, Z. Chen, Y. Han, J. D. Villasenor, and S. Yang, "A compressive sensing image compression algorithm using quantized DCT and noiselet information," *IEEE Int. Conf. on Acoust. Speech and Signal Proc.*, pp. 1294-1297, USA, Mar. 2010.

[25] T. Y. Al-Naffouri and A. A. Quadeer, "Structure based Bayesian sparse signal reconstruction," *submitted to IEEE Trans. Signal Proc.*.