Subspace Recovery From Structured Union of Subspaces

Thakshila Wimalajeewa, *Member, IEEE*, Yonina C. Eldar, *Fellow, IEEE*, and Pramod K. Varshney, *Fellow, IEEE*

Abstract—Lower dimensional signal representation schemes frequently assume that the signal of interest lies in a single vector space. In the context of the recently developed theory of compressive sensing, it is often assumed that the signal of interest is sparse in an orthonormal basis. However, in many practical applications, this requirement may be too restrictive. A generalization of the standard sparsity assumption is that the signal lies in a union of subspaces. Recovery of such signals from a small number of samples has been studied recently in several works. Here, we consider the problem of only subspace recovery in which our goal is to identify the subspace (from the union) in which the signal lies using a small number of samples, in the presence of noise. More specifically, we derive performance bounds and conditions under which reliable subspace recovery is guaranteed using maximum likelihood (ML) estimation. We begin by treating general unions and then obtain the results for the special case in which the subspaces have structure leading to block sparsity. In our analysis, we treat both general sampling operators and random sampling matrices. With general unions, we show that under certain conditions, the number of measurements required for reliable subspace recovery in the presence of noise via ML is less than that implied using the restricted isometry property, which guarantees complete signal recovery. In the special case of block sparse signals, we quantify the gain achievable over standard sparsity in subspace recovery. Our results also strengthen existing results on sparse support recovery in the presence of noise under the standard sparsity model.

Index Terms—Maximum likelihood estimation, union of linear subspaces, subspace recovery, compressive sensing, block sparsity.

I. INTRODUCTION

COMPRESSIVE sensing (CS) theory has established that a small number of measurements acquired via random projections are sufficient for signal recovery when the signal of interest is sparse in a certain basis. Consider a length-N signal $\mathbf{x}$ which can be represented in a basis $\mathbf{V}$ such that $\mathbf{x} = \mathbf{V}\mathbf{c}$. The signal $\mathbf{x}$ is said to be k-sparse in the basis $\mathbf{V}$

if $\mathbf{c}$ has only k nonzero coefficients where k is much smaller than N. It has been shown in [1]–[3] that $\mathcal{O}(k \log(N/k))$ compressive measurements are sufficient to recover $\mathbf{x}$ when the elements of the measurement matrix are random. Signal recovery may be performed via optimization or greedy based approaches. A detailed overview of CS can be found in [4].

There are a variety of applications in which complete signal recovery is not necessary. The problem of sparse support recovery (equivalently sparsity pattern recovery or finding the locations of nonzero coefficients of a sparse signal) arises in a wide variety of areas including source localization [5], [6], sparse approximation [7], subset selection in linear regression [8], [9], estimation of frequency band locations in cognitive radio networks [10]–[12], and signal denoising [13]. In these applications, often finding the sparsity pattern of the signal is more important than approximating the signal itself. Further, in the CS framework, once the sparse support is identified, the signal can be estimated using standard techniques.

For the problem of complete sparse signal recovery, there is a significant amount of work in the literature that focuses on deriving recovery guarantees and stability with respect to various l_q norms of the reconstruction error. However, as pointed out in [14], recovery guarantees derived for sparse signals do not always imply exact recovery of the sparse support. Although a signal estimate can be close to the original sparse signal, the estimated signal may have a different support [14]. For example, least-absolute shrinkage and selection operator (Lasso) has been shown to be information theoretically optimal in certain regimes of the signal-to-noise ratio (SNR) for sparse support recovery, while in other regimes of SNR, Lasso fails with high probability in recovering the sparsity pattern [14], [15]. Thus, investigation of recovery conditions for sparse support at any given SNR is an important problem. Performance limits on reliable recovery of the sparsity pattern have been derived by several authors in recent work exploiting information theoretic tools [14], [16]–[22]. Most of these works focused on the standard sparsity model.

There are practical scenarios where structured properties of the signal are available. Reduced dimensional signal processing for several signal models which go beyond simple sparsity has been treated in recent literature [23]–[28]. One general model that can describe many structured problems is that of a union of subspaces. In this setting, the signal is known to lie in one out of a possible set of subspaces but the specific subspace chosen is unknown. Examples include wideband spectrum sensing [11], time delay estimation with overlapping

echoes [24], [29], [30], and signals having finite rate innovation [31], [32]. Conditions under which stable sampling and recovery is possible in a general union of subspaces model are derived in [23]–[26]. In [33]–[35], the authors discuss conditions under which group Lasso is capable of recovering block sparse signals with and/or without overlaps. However, the problem of recovering the subspace in which the signal lies without completely recovering the signal (or the problem of subspace recovery) has not been treated in this more general setting.

In this paper, our goal is to investigate the problem of subspace recovery in the union of subspaces model with a given sampling operator. We consider subspace recovery based on the optimal ML decoding scheme in the presence of noise. While ML is computationally intractable as the signal dimension increases, the analysis provides a benchmark for the optimal performance that is achievable with any practical algorithm. We depict performance in terms of probability of error of the ML decoder when sampling is performed via an arbitrary linear sampling operator. Based on an upper bound on the probability of error, we derive the minimum number of samples required for asymptotically reliable recovery of subspaces in terms of an SNR measure, the dimension of each subspace in the union and a term which quantifies the dependence or overlap among the subspaces. In the special case where sampling is performed via random projections and the subspaces in the union have a specific structure such that each subspace is a sum of some other k_0 subspaces, we obtain a more explicit expression for the minimum number of measurements. This number depends on the number of underlying subspaces, the dimension of each subspace, and the minimum nonzero block SNR (defined in Section IV.B). The conventional sparsity model is a special case of this structure.

The asymptotic probability of error of the ML decoder for sparse support recovery in the presence of noise for the standard sparsity model was first investigated in [14] followed by several other authors [16], [17], [22]. In [14], sufficient conditions were derived on the number of noisy compressive measurements needed to achieve a vanishing probability of error asymptotically for sparsity pattern recovery while necessary conditions were considered in [17]. The analyses in both [14], [17] are based on the assumption that the sampling operator is random. Here, we follow a similar path assuming the union of subspaces model. However, there are some key differences between our derivations and that in [14]. First, we treat arbitrary (not necessarily random) sampling operators and assume a general union of subspaces model as opposed to the standard sparsity model. Further, the results in [14] were derived based on weak bounds on the probability of error, thus there is a gap between those results and the number of measurements required for the exact probability of error to vanish asymptotically at finite SNR. We consider tighter bounds on the probability of error leading to tighter results.

The rest of the paper is organized as follows. In Section II, the problem of subspace recovery from a union of subspace model is introduced. In Section III, performance limits with ML decoder for subspace recovery in terms of the probability of error are derived with a given linear sampling operator considering a general union of subspaces model. Conditions under which asymptotically reliable subspace recovery in the presence of noise is guaranteed are obtained based on the derived upper bound. The results are extended in Section IV to the setting where structured properties of the subspaces in the union are available. We also derive sufficient conditions for subspace recovery when sampling is performed via random projections. In Section V, we compare our results with some existing results in the literature. Practical algorithms to recover subspaces in the union of subspace model and numerical results to validate the theoretical claims are presented in Section VI.

Throughout the paper, we use the following notation. Arbitrary vectors in a Hilbert space $\mathcal{H}$, are denoted by lower case letters, e.g., x. Calligraphic letters, e.g., $\mathcal{S}$, are used to represent subspaces in $\mathcal{H}$. Vectors in $\mathbb{R}^N$ are written in boldface lower case letters, e.g. $\mathbf{x}$. Scalars (in $\mathbb{R}$) are also denoted by lower case letters, e.g., x, when there is no confusion. Matrices are written in boldface upper case letters, e.g., $\mathbf{A}$. Linear operators and a set of basis vectors for a given subspace $\mathcal{S}$ are denoted by upper case letters, e.g., A. By $\mathbf{0}$, we denote a vector with appropriate dimension in which all elements are zeros, and $\mathbf{I}_k$ is the identity matrix of size k. The conjugate transpose of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}^*$. Finally, $||.||_2$ denotes the l_2 norm and $|.|$ is used for both the cardinality (of a set) and the absolute value (of a scalar). The notation $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ means that the random vector $\mathbf{x}$ is distributed as multivariate Gaussian with mean $\boldsymbol{\mu}$ and the covariance matrix Σ; $x \sim \mathcal{X}_m^2(\lambda)$ denotes that the random variable x is distributed as Chi squared with m degrees of freedom and non centrality parameter λ. (The central Chi squared distribution is denoted by $\mathcal{X}_m^2$). Special functions used in the paper are: Gaussian Q-function:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt \qquad (1)$$

Gamma function:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \qquad (2)$$

and modified Bessel function with real arguments:

$$K_v(x) = \int_0^\infty e^{-x\cosh t} \cosh(vt) dt. \qquad (3)$$

II. PROBLEM FORMULATION

A. Union of Subspaces

As discussed in [23]–[26], there are many practical scenarios where the signals of interest lie in a union of subspaces.

Definition 1 (Union of Subspaces): A signal $x \in \mathcal{H}$ lies in a union of subspaces if $x \in \mathcal{X}$ where $\mathcal{X}$ is defined as

$$\mathcal{X} = \bigcup_i \mathcal{S}_i \qquad (4)$$

and $\mathcal{S}_i$'s are subspaces of $\mathcal{H}$. A signal $x \in \mathcal{X}$ if and only if there exists i_0 such that $x \in \mathcal{S}_{i_0}$.

Throughout the paper, we assume that the subspaces $\mathcal{S}_i$'s are finite dimensional and that the number of subspaces is finite.

Let $V_i = \{v_{im}\}_{m=0}^{k-1}$ be a basis for $\mathcal{S}_i$, and let k be its dimension (it is noted that while we assume all subspaces to have the same dimension, the analysis can be easily extended to the case where different subspaces have different dimensions). Then each $x \in \mathcal{S}_i$ can be expressed in terms of a basis expansion

$$x = \sum_{m=0}^{k-1} c_i(m)v_{im}$$

where $c_i(m)$ for $m = 0, 1, \cdots k - 1$ are the coefficients corresponding to the basis V_i. We assume that the subspaces are distinct (i.e. there are no subspaces such that $\mathcal{S}_i \subseteq \mathcal{S}_j$ for $i \neq j$ in the union (4)). We denote by $T < \infty$ the number of subspaces in the union $\mathcal{X}$.

B. Structured Union of Subspaces and Block Sparsity

There are certain scenarios in which the signals can be assumed to lie in more structured union of subspaces as considered in [25], [28], and [36]. Suppose that each subspace in the union (4) can be represented as a sum of k_0 (out of L) disjoint subspaces [25], [36]. More specifically,

$$\mathcal{S}_i = \bigoplus_{j \in \Sigma_{k_0}} \mathcal{V}_j \qquad (5)$$

where $\{\mathcal{V}_j\}_{j=0}^{L-1}$ are disjoint subspaces, and Σ_{k_0} contains k_0 indices from $\{0, 1, \cdots, L - 1\}$. Let $d_j = \dim(\mathcal{V}_j)$ and $N = \sum_{j=0}^{L-1} d_j$. Then there are $T = \binom{L}{k_0}$ subspaces in the union. Under this formulation, the dimension of each subspace in (4) is

$$k = \sum_{j \in \Sigma_{k_0}} d_j.$$

In the special case where $d_j = d$ for all j, $k = k_0 d$.

Taking V_j as a basis for $\mathcal{V}_j$, a signal in the union can be written as

$$x = \sum_{j \in \Sigma_{k_0}} V_j \mathbf{c}_j \qquad (6)$$

where $\mathbf{c}_j = [c_j(0), \cdots, c_j(d_j - 1)]^T \in \mathbb{R}^{d_j}$ is a $d_j \times 1$ coefficient vector corresponding to the basis V_j. Note that we use the same notation V_j to denote a basis of the subspace $\mathcal{S}_j$ in (4) for $j = 0, 1, \cdots, T - 1$ (when discussing the general union of subspaces model) and also to denote a basis of the subspace $\mathcal{V}_j$ in (5) for $j = 0, 1, \cdots, L - 1$ (when discussing the structured union of subspace model). Let V be a matrix constructed by concatenating V_i's column wise, such that $V = [V_0|V_1|\cdots|V_{L-1}]$ and $\mathbf{c}$ be a $N \times 1$ vector with $\mathbf{c} = [\mathbf{c}_0^T|\cdots|\mathbf{c}_{L-1}^T]^T$. As defined in [25], the vector $\mathbf{c} \in \mathbb{R}^N$ is called block k-sparse over $\mathcal{I} = \{d_0, d_1, \cdots, d_{L-1}\}$ if all the elements in $\mathbf{c}_i$ are zeros for all but k_0 indices where $N = \sum_{j=0}^{L-1} d_j$. In this paper, we assume $d_j = d$ for all j so that $N = Ld$.

The standard sparsity model used in the CS literature is a special case of this structured union of subspaces model. In the standard CS framework, $x = \mathbf{x}$ is a length-N signal vector which is k-sparse in an N-dimensional orthonormal basis $\mathbf{V}$ so that $\mathbf{x}$ can be represented as $\mathbf{x} = \mathbf{V}\mathbf{c}$ with $\mathbf{c}$ having only $k \ll N$ significant coefficients. This fits our framework when $d = 1$ and V_j is chosen as the jth column vector of the orthonormal basis $\mathbf{V}$ for $j = 0, 1, \cdots, N - 1$. In this case, $L = N$ and there are $T = \binom{N}{k}$ subspaces in the union.

C. Observation Model: Linear Sampling

Consider a sampling operator defined by a bounded linear mapping A of a signal x that lies in an ambient Hilbert space $\mathcal{H}$. Let A be specified by a set of unique sampling vectors $\{a_m\}_{m=0}^{M-1}$. With these notations, the noisy samples are given by,

$$\mathbf{y} = Ax + \mathbf{w}, \qquad (7)$$

where $\mathbf{y}$ is the $M \times 1$ measurement vector, and the mth element of the vector Ax is given by $(Ax)_m = \langle a_m, x \rangle$ for $m = 0, 1, \cdots, M - 1$ where $\langle ., . \rangle$ denotes the inner product. The noise vector $\mathbf{w}$ is assumed to be Gaussian with mean $\mathbf{0}$ and covariance matrix $\sigma_w^2 \mathbf{I}_M$.

When $x \in \mathcal{S}_i$ for some i in the model (4), the vector of samples can be equivalently represented in the form of a matrix vector multiplication:

$$\mathbf{y} = \mathbf{B}_i \mathbf{c}_i + \mathbf{w}, \qquad (8)$$

where

$$\mathbf{B}_i = A V_i$$
$$= \begin{pmatrix} \langle a_0, v_{i0} \rangle & \langle a_0, v_{i1} \rangle & \cdots & \langle a_0, v_{i(k-1)} \rangle \\ \langle a_1, v_{i0} \rangle & \langle a_1, v_{i1} \rangle & \cdots & \langle a_1, v_{i(k-1)} \rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle a_{M-1}, v_{i0} \rangle & \langle a_{M-1}, v_{i1} \rangle & \cdots & \langle a_{M-1}, v_{i(k-1)} \rangle \end{pmatrix}$$

and $\mathbf{c}_i = [c_i(0) \ c_i(1) \ \cdots c_i(k - 1)]^T$ is the coefficient vector with respect to the basis V_i. Further, let $\mathbf{b}_{im}$ denote the mth column vector of the matrix $\mathbf{B}_i$ for $m = 0, 1, \cdots, k - 1$ and $i = 0, 1, \cdots, T - 1$. We assume that the linear sampling operator A is a one-to-one mapping between $\mathcal{X}$ and $A\mathcal{X}$. Since $\{v_{i0}, \cdots, v_{i(k-1)}\}$ is a set of linearly independent basis vectors, $\{\mathbf{b}_{i0}, \cdots, \mathbf{b}_{i(k-1)}\}$ are also linearly independent for each $i = 0, 1, \cdots, T - 1$. It is worth noting that, while this one-to-one condition ensures uniqueness, stronger conditions are required to recover x in a stable manner as discussed in [24].

D. Subspace Recovery From the Union of Subspaces Model

As discussed in the Introduction, there are applications where it is sufficient to recover the subspace in which the signal of interest lies from the union of subspaces model (4) instead of complete signal recovery. Moreover, if there is a procedure to correctly identify the subspace with vanishing probability of error, then the signal x can be reconstructed with a small l_2-norm error using standard techniques. However, if an algorithm developed for complete signal recovery is used for subspace recovery, then it may not give an equivalent performance guarantee. This is because, even if such an estimate of the signal may be close to the true signal with respect to the considered performance metric (e.g., l_2-norm), the subspace in which the estimated signal lies may be different from the true subspace. This can happen especially when the SNR is not

sufficiently large. Our focus in this paper is the investigation of subspace recovery.

The problem of subspace recovery is to identify the subspace in which the signal x lies. The estimated subspace, $\hat{\mathcal{S}}$, via any recovery scheme can be expressed as

$$\hat{\mathcal{S}} = \zeta(\mathbf{y}) \tag{9}$$

where $\zeta(\cdot)$ is a mapping from the observation vector $\mathbf{y}$ to an estimated subspace $\hat{\mathcal{S}} \in \{\mathcal{S}_0, \cdots, \mathcal{S}_{T-1}\}$. The performance metric used to evaluate the quality of the estimate (9) is taken as the average probability of error:

$$P_e = \sum_{\mathcal{S}} Pr(\zeta(\mathbf{y}) \neq \mathcal{S}|\mathcal{S}) Pr(\mathcal{S}).$$

We say that the mapping $\zeta(\mathbf{y})$ is capable of providing *asymptotically reliable* subspace recovery if $P_e \rightarrow 0$ as $M \rightarrow \infty$. In this paper, we consider subspace recovery via ML estimation. Our goal is to address the following issues.

- Performance of the ML estimation scheme in terms of probability of error in recovering the subspaces from the union of subspaces model (4) in the presence of noise. We are also interested in conditions under which asymptotically reliable subspace recovery is guaranteed with a given sampling operator.
- How much gain in terms of the number of samples required for subspace recovery can be achieved if further information on structure is available for the subspaces in (4) compared to the case of no additional structure.
- Illustration of the performance gap between ML estimation and computationally tractable algorithms for subspace recovery at finite SNR.

The main results of the paper can be summarized as follows. With the general union of subspaces model as defined in (4), and for a given sampling operator, the minimum number of samples required for asymptotically reliable subspace recovery in the presence of noise is

$$M > k + \frac{\eta_3}{f(SNR)} \log(\bar{T}_0) \tag{10}$$

where k is the dimension of each subspace, $f(SNR)$ is a measure of the minimum SNR of the sampled signal projected onto the null space of any subspace in the union, $\bar{T}_0$ is a measure of the number of subspaces in the union with maximum dependence where $\bar{T}_0 \leq T$ and η_3 is a constant (formal definitions of all these terms are given in Section III). For the special case where each subspace in the union (4) can be expressed as a sum of k_0 subspaces out of L the problem of subspace recovery reduces to that of block sparsity pattern recovery. Assuming that the sampling operator is represented by random projections, the number of samples required for asymptotically reliable block sparsity pattern recovery is given by

$$M > k + \frac{\eta_4}{BSNR_{\min}} \log(L - k_0) \tag{11}$$

where $BSNR_{\min}$ is the minimum nonzero block SNR and η_4 is a constant. When $d = 1$ and $L = N$ where N is the signal

dimension, the block sparsity model reduces to the standard sparsity model. Then, our result shows that

$$M > k + \frac{\eta_2}{CSNR_{\min}} \log(N - k) \tag{12}$$

measurements are required for reliable sparsity pattern recovery where $CSNR_{\min}(\leq \frac{BSNR_{\min}}{d})$ is the minimum component SNR of the signal. Thus, from (11) and (12), we observe that the number of measurements required for asymptotically reliable subspace recovery beyond the sparsity index (i.e., $M - k$) reduces approximately d times with a block sparsity model (so that $k = k_0 d$) compared to the standard sparsity model. A detailed comparison between our results and existing results in the literature is given in Section V.

III. SUBSPACE RECOVERY WITH GENERAL UNIONS

The problem of finding the true subspace via ML estimation becomes finding the index $\hat{i}$ such that,

$$\hat{i} = \underset{i=0,\cdots,T-1}{\arg\max} \; p(\mathbf{y}|\mathbf{B}_i).$$

When $x \in \mathcal{S}_i$ in (4) for some i, and using the observation model (8), we have $p(\mathbf{y}|\mathbf{B}_i) = \mathcal{N}(\mathbf{B}_i \mathbf{c}_i, \sigma_w^2 \mathbf{I}_M)$. The signal x is assumed to be deterministic but unknown. Thus, when $x \in \mathcal{S}_i$, the coefficient vector $\mathbf{c}_i$ with respect to a given basis $\mathbf{B}_i$ is unknown. Assuming that each $\mathbf{B}_i$ has rank k for $i = 0, \cdots, T - 1$, the ML estimate of $\mathbf{c}_i$ which maximizes $p(\mathbf{y}|\mathbf{B}_i)$ is given by $\hat{\mathbf{c}}_i = (\mathbf{B}_i^* \mathbf{B}_i)^{-1} \mathbf{B}_i^* \mathbf{y}$. This results in

$$\log(\max_{\mathbf{c}_i} p(\mathbf{y}|\mathbf{B}_i)) = \log\left(\frac{1}{(2\pi\sigma_w^2)^{M/2}}\right) - \frac{1}{2\sigma_w^2}||\mathbf{y} - \mathbf{P}_i \mathbf{y}||_2^2$$

$$= \log\left(\frac{1}{(2\pi\sigma_w^2)^{M/2}}\right) - \frac{1}{2\sigma_w^2}||\mathbf{P}_i^\perp \mathbf{y}||_2^2$$

where $\mathbf{P}_i = \mathbf{B}_i(\mathbf{B}_i^* \mathbf{B}_i)^{-1}\mathbf{B}_i^*$ is the orthogonal projector onto the span of $\{\mathbf{b}_{im}\}_{m=0}^{k-1}$ and $\mathbf{P}_i^\perp = \mathbf{I} - \mathbf{P}_i$. Thus, the estimated index of the subspace by ML estimation is

$$\hat{i} = \underset{i=0,\cdots,T-1}{\arg\min} \; ||\mathbf{P}_i^\perp \mathbf{y}||_2^2.$$

The probability of error using ML estimation is given by

$$P_e = Pr(\mathbf{B}_{estimated} \neq \mathbf{B}_{true}) = \sum_i Pr(\hat{i} \neq i|\mathbf{B}_i) Pr(\mathbf{B}_i)$$

$$\leq \sum_i \sum_{j \neq i} Pr(\hat{i} = i|\mathbf{B} = \mathbf{B}_j) Pr(\mathbf{B} = \mathbf{B}_j) \tag{13}$$

where $Pr(\hat{i} = i|\mathbf{B} = \mathbf{B}_j)$ is the probability of selecting $\mathcal{S}_i$ when the true subspace is $\mathcal{S}_j$. Since ML estimation chooses the subspace $\mathcal{S}_i$ over $\mathcal{S}_j$ when $||\mathbf{P}_i^\perp \mathbf{y}||_2^2 - ||\mathbf{P}_j^\perp \mathbf{y}||_2^2 < 0$, $Pr(\hat{i} = i|\mathbf{B} = \mathbf{B}_j)$ for $i \neq j$ is given by

$$Pr(\hat{i} = i|\mathbf{B} = \mathbf{B}_j) = Pr(||\mathbf{P}_i^\perp \mathbf{y}||_2^2 - ||\mathbf{P}_j^\perp \mathbf{y}||_2^2 < 0).$$

Let $\Delta_{ij}(\mathbf{y}) = ||\mathbf{P}_i^\perp \mathbf{y}||_2^2 - ||\mathbf{P}_j^\perp \mathbf{y}||_2^2$ for $i \neq j$. When the true subspace is $\mathcal{S}_j$ so that $Ax = \mathbf{B}_j \mathbf{c}_j$, we have $||\mathbf{P}_j^\perp \mathbf{y}||_2^2 = ||\mathbf{P}_j^\perp \mathbf{w}||_2^2$ and

$$\mathbf{P}_i^\perp \mathbf{y} = \mathbf{P}_i^\perp Ax + \mathbf{P}_i^\perp \mathbf{w}$$
$$= \mathbf{P}_i^\perp \mathbf{B}_j \mathbf{c}_j + \mathbf{P}_i^\perp \mathbf{w} = \mathbf{P}_i^\perp \mathbf{B}_{j\backslash i} \mathbf{c}_{j\backslash i} + \mathbf{P}_i^\perp \mathbf{w}$$

where

$$\mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i} = \sum_{\mathbf{b}_{jm} \notin \mathcal{R}(\mathbf{B}_i)} \mathbf{b}_{jm} c_j(m)$$

and $\mathcal{R}(\mathbf{A})$ denotes the range space of the matrix $\mathbf{A}$. More specifically, the $M \times l$ matrix $\mathbf{B}_{j\backslash i}$ contains the columns of $\mathbf{B}_j$ which are not in the range space of the matrix $\mathbf{B}_i$ and l is the number of columns in $\mathbf{B}_{j\backslash i}$. The $l \times 1$ vector $\mathbf{c}_{j\backslash i}$ contains the elements of $\mathbf{c}_j$ corresponding to the column vectors in $\mathbf{B}_{j\backslash i}$.

We conclude that, the decision statistic for selecting $\mathcal{S}_i$ over $\mathcal{S}_j$ is given by

$$\Delta_{ij}(\mathbf{y}) = ||\mathbf{P}_i^\perp(\mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i} + \mathbf{w})||_2^2 - ||\mathbf{P}_j^\perp \mathbf{w}||_2^2$$

and

$$Pr(\Delta_{ij}(\mathbf{y}) < 0) = Pr\left(\frac{||\mathbf{P}_i^\perp(\mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i} + \mathbf{w})||_2^2}{||\mathbf{P}_j^\perp \mathbf{w}||_2^2} < 1\right).$$

When $\mathbf{B}_j$ is given, the random variable $g_1 = ||\mathbf{P}_i^\perp(\mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i} + \mathbf{w})||_2^2/\sigma_w^2$ is a non-central Chi squared random variable with $M - k$ degrees of freedom and non-centrality parameter $||\mathbf{P}_i^\perp(\mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i})||_2^2/\sigma_w^2$. The random variable $g_2 = ||\mathbf{P}_j^\perp \mathbf{w}||_2^2/\sigma_w^2$ is a (central) Chi-squared random variable with $M - k$ degrees of freedom. The two random variables g_1 and g_2 are, in general, correlated and the computation of the exact value of $Pr(\Delta_{ij}(\mathbf{y}) < 0)$ is difficult. Below we derive an upper bound on the quantity $Pr(\Delta_{ij}(\mathbf{y}) < 0)$ following techniques similar to those proposed in [14].

A. Upper Bound on $Pr(\Delta_{ij}(\mathbf{y}) < 0)$

For clarity, we introduce the following notation. Let $\mathcal{W}_{j\backslash i}$ be the set consisting of column indices of $\mathbf{B}_j$ such that $\mathbf{b}_{jm} \notin \mathcal{R}(\mathbf{B}_i)$ for $m = 0, 1, \cdots k - 1$ and $i \neq j$ ($i, j = 0, 1, \cdots, T - 1$). We then have that $|\mathcal{W}_{j\backslash i}| = l$, where l can take values from $1, 2, \cdots, k$.

Lemma 1: Consider a sampling operator A. For any given signal $x \in \mathcal{S}_j$ so that $Ax = \mathbf{B}_j\mathbf{c}_j$, the probability of error in selecting the subspace $\mathcal{S}_i$ over $\mathcal{S}_j$, is upper bounded by

$$Pr(\Delta_{ij}(\mathbf{y}) < 0) \leq Q\left(\frac{1}{2}(1 - 2\eta_0)\sqrt{\lambda_{j\backslash i}}\right) + \Psi(l, \lambda_{j\backslash i})$$

(14)

where

$$\lambda_{j\backslash i} = \frac{1}{\sigma_w^2}||\mathbf{P}_i^\perp \mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i}||_2^2, \quad (15)$$

$$\Psi(l, \lambda_{j\backslash i}) = \frac{\sqrt{2}}{2^l \Gamma(l/2)}(\eta_0 \lambda_{j\backslash i})^{l/2 - 1/2} K_{l/2 - 1/2}\left(\frac{\eta_0 \lambda_{j\backslash i}}{2}\right),$$

$Q(x)$ *is the Gaussian Q function (1), $\Gamma(x)$ is the Gamma function (2), $K_\nu(x)$ is the modified Bessel function (3), and $0 < \eta_0 < \frac{1}{2}$.*

Proof: See Appendix A. ∎

Theorem 1: Assuming that the true subspace is chosen uniformly at random from T subspaces in the union (4), the

average probability of error of ML estimation for subspace recovery is upper bounded by

$$P_e \leq \frac{1}{T}\sum_{i=0}^{T-1} \sum_{j=0}^{T-1} Q\left(\frac{1}{2}(1 - 2\eta_0)\sqrt{\lambda_{j\backslash i}}\right) + \Psi(l, \lambda_{j\backslash i})$$

(16)

where $\lambda_{j\backslash i}$, η_0, Q, Ψ are as defined in Lemma 1.

Proof: The proof follows from Lemma 1 and (13). ∎

In general, the subspaces $\mathcal{S}_i$ and $\mathcal{S}_j$ can overlap; i.e. there can be elements in $\mathcal{S}_j$ which are also in $\mathcal{S}_i$. However, one subspace can not lie in another subspace entirely; i.e. all subspaces $\mathcal{S}_i$ are distinct for $i = 0, 1, \cdots, T - 1$. As defined before, $\mathcal{W}_{j\backslash i}$ contains the column indices of $\mathbf{B}_j$ which are not in $\mathcal{R}(\mathbf{B}_i)$ and $|\mathcal{W}_{j\backslash i}| = l$ for any $i \neq j$ where l takes values from $1, 2, \cdots, k$. As l increases, the overlap of the two subspaces decreases resulting in more separable subspaces. In the special case where $\mathcal{S}_j$ and $\mathcal{S}_i$ do not intersect at all, we have $l = k$. Thus, l can be considered as a measure of overlap between any two subspaces $\mathcal{S}_j$ and $\mathcal{S}_i$ for $i \neq j$ in the union (4). For given l, the probability $Pr(\Delta_{ij}(\mathbf{y}) < 0)$ in (14) monotonically decreases as a function of $\lambda_{j\backslash i}$, defined in Lemma 1. This implies that when $\lambda_{j\backslash i}$ is large, the probability of selecting $\mathcal{S}_i$ given that the true subspace is $\mathcal{S}_j$ decreases. In other words $\lambda_{j\backslash i}$, characterizes the error in selecting the subspace $\mathcal{S}_i$ over $\mathcal{S}_j$ for $i \neq j$ when the true subspace is $\mathcal{S}_j$. It is, therefore, of interest to further investigate the quantity $\lambda_{j\backslash i}$.

B. Evaluation of $\lambda_{j\backslash i}$

The value of $\lambda_{j\backslash i}$ given by (15) depends on $||\mathbf{P}_i^\perp \mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i}||_2^2$ ($= ||\mathbf{P}_i^\perp \mathbf{B}_j\mathbf{c}_j||_2^2 = ||\mathbf{P}_i^\perp Ax||_2^2$). When the true subspace is assumed to be $\mathcal{S}_j$, this quantity denotes the energy of the sampled signal Ax projected onto the null space of $\mathbf{B}_i$; i.e., the energy of the sampled signal which is unaccounted for by $\mathcal{S}_i$ for $i \neq j$. Therefore, when $||\mathbf{P}_i^\perp \mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i}||_2^2$ is large, the probability that the subspace $\mathcal{S}_i$ is selected as the true subspace becomes small. Further, if $\mathcal{S}_j \subseteq \mathcal{S}_i$ for any $\mathcal{S}_i$, we have $||\mathbf{P}_i^\perp \mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i}||_2^2 = 0$. However, this cannot happen based on our assumption that there is no subspace in the union which completely overlaps another. Thus, $\lambda_{j\backslash i} > 0$.

Let the eigendecomposition of $\mathbf{P}_i^\perp$ be $\mathbf{P}_i^\perp = \mathbf{Q}_i \mathbf{\Lambda}_i \mathbf{Q}_i^T$ where $\mathbf{Q}_i$ is a unitary matrix consisting of eigenvectors of $\mathbf{P}_i^\perp$ and $\mathbf{\Lambda}_i$ is a diagonal matrix in which the diagonal elements represent eigenvalues of $\mathbf{P}_i^\perp$ which are $M - k$ ones and k zeros. Then, for given l,

$$\lambda_{j\backslash i} = \sum_{m \in \mathcal{Q}_i} \alpha_{m,i}^2(l) \geq (M - k)\alpha_{\min,l}^2 \quad (17)$$

where $\alpha_{m,i}(l) = \frac{1}{\sigma_w}\langle \mathbf{q}_{m,i}, \mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i}\rangle$ for given l, $\mathbf{q}_{m,i}$ is the mth eigenvector of $\mathbf{P}_i^\perp$, $\mathcal{Q}_i$ is the set containing indices corresponding to nonzero eigenvalues where $|\mathcal{Q}_i| = M - k$ and $\alpha_{\min,l} = \min_{i; i \neq j} |\alpha_{m,i}(l)|$.

Note that $(M - k)\alpha_{\min,l}^2$ is a measure of the minimum SNR of the sampled signal, Ax, projected onto the null space of any subspace $\mathcal{S}_i$ for $i \neq j$, $i = 0, 1, \cdots, T - 1$ such that

$|\mathcal{W}_{j\setminus i}| = l$ given that the true subspace in which the signal lies is $\mathcal{S}_j$.

For a given subspace $\mathcal{S}_j$, define $T_j(l)$ to be the number of subspaces $\mathcal{S}_i$ such that $|\mathcal{W}_{j\setminus i}| = l$. With these notations, the probability of error in (16) can be further upper bounded by,

$$P_e \leq \frac{1}{T} \sum_{j=0}^{T-1} \sum_{l=1}^{k} T_j(l) \left(Q\left(\frac{1}{2}(1-2\eta_0)\sqrt{(M-k)\alpha_{\min,l}^2}\right) \right. $$
$$\left. + \Psi\left(l, (M-k)\alpha_{\min,l}^2\right) \right) \quad (18)$$

where

$$\Psi\left(l, (M-k)\alpha_{\min,l}^2\right)$$
$$= \frac{\sqrt{2}}{2^l \Gamma\left(\frac{l}{2}\right)} (\eta_0(M-k)\alpha_{\min,l}^2)^{l/2-1/2} K_{l/2-1/2}$$
$$\times \left(\eta_0(M-k)\frac{\alpha_{\min,l}^2}{2}\right).$$

To obtain (18) we used the facts that $Q(x)$ is monotonically non increasing in x, and $\Psi(s, x)$ is monotonically non increasing in x for given s when $x > 0$.

The quantity $T_j(l)$ is a measure of the overlap between $\mathcal{S}_j$ and any subspace $\mathcal{S}_i$ for $i \neq j$. To compute $T_j(l)$ explicitly, the specific structures of the subspaces should be known. For example, in the standard sparsity model used in CS in which the union in (4) consists of $T = \binom{N}{k}$ subspaces from an orthonormal basis $\mathbf{V}$ of dimension N, there are $\binom{k}{l}\binom{N-k}{l}$ number of sets such that $|\mathcal{W}_{j\setminus i}| = l$, thus $T_j(l) = \binom{k}{l}\binom{N-k}{l}$. In that particular case, $T_j(l)$ is the same for all $j = 0, 1, \cdots, T-1$. To further upper bound (18), we let

$$T_0(l) = \max_{j=0,1,\cdots,T-1} T_j(l). \quad (19)$$

Then,

$$P_e \leq \sum_{l=1}^{k} T_0(l) \left(Q\left(\frac{1}{2}(1-2\eta_0)\sqrt{(M-k)\alpha_{\min,l}^2}\right) \right.$$
$$\left. + \Psi\left(l, (M-k)\alpha_{\min,l}^2\right) \right). \quad (20)$$

Theorem 2: Let $\alpha_{\min,l}^2$ and $T_0(l)$ be as defined in (17) and (19), respectively. Suppose that sampling is performed via a sampling operator A. Then P_e in (20) vanishes asymptotically (i.e., $\lim_{(M-k)\to\infty} P_e \to 0$) if the following condition is satisfied:

$$M > k + \max\{M_1, M_2\}$$

where

$$M_1 = \max_{l=1,\cdots,k}\{f_1(l)\}$$

$$M_2 = \max_{l=1,\cdots,k}\{f_2(l)\}$$

with

$$f_1(l) = \frac{8}{(1-2\eta_0)^2\alpha_{\min,l}^2}\{\log(T_0(l)) + \log(1/2)\},$$

$$f_2(l) = \frac{2(k/2+r_0-1)}{r_0\eta_0\alpha_{\min,l}^2}\left\{\log(T_0(l)) + \log\left(\frac{2b_0}{\sqrt{\pi}}\right)\right\}$$

$0 < \eta_0 < 1/2$, $b_0 = \frac{\sqrt{2\pi}}{4}$ and $r_0 > 0$.

Proof: See Appendix B. ∎

Let $l_i \in \{1, \cdots, k\}$ be the value of l which maximizes $f_i(l)$ as defined in Theorem 2 for $i = 1, 2$. For M_2, it can be verified that we can find constants η_0 and r_0 in the defined regimes such that $\frac{8}{(1-2\eta_0)^2} > \frac{2(k/2+r_0-1)}{r_0\eta_0}$ if k is fairly small. The dominant factor of M_1 and M_2 can be written in the form of $\frac{\eta_3}{\bar{\alpha}_{\min}^2}\log(\bar{T}_0)$ where $\bar{\alpha}_{\min}^2$ and $\bar{T}_0$ are the corresponding values of $\alpha_{\min,l}^2$ and $T_0(l)$ when $l = l_0$ for $l_0 \in \{l_1, l_2\}$ and η_3 is an appropriate constant. Since, most of the scenarios we are interested in are for the case where k is sufficiently small, the minimum number of samples required for reliable subspace recovery becomes

$$M \geq k + \frac{\eta_3}{\bar{\alpha}_{\min}^2}\log(\bar{T}_0).$$

It is further noted that $T_0(l) \leq T$ for all l and thus $\bar{T}_0 \leq T$ where T is the total number of subspaces in the union (4).

C. Random Sampling

Next, we consider the special case where the sampling operator is a $M \times N$ matrix in which the elements are realizations of a random variable (e.g. Gaussian). In this case, $\mathbf{B}_i = \mathbf{A}\mathbf{V}_i$ in (8) where $\mathbf{A}$ is the random sampling matrix and $\mathbf{V}_i = [\mathbf{v}_{i0}|\cdots|\mathbf{v}_{i(k-1)}]$ is the $N \times k$ matrix in which columns consist of the basis vectors of the subspace $\mathcal{S}_i$ for $i = 0, 1, \cdots, T-1$. The only term which depends on the sampling operator in the expression for the upper bound on the probability of error in (16) is $\lambda_{j\setminus i}$. When the sampling operator is a random matrix, $\lambda_{j\setminus i}$ can be evaluated as follows.

Proposition 1: Suppose that the sampling matrix $\mathbf{A}$ consists of elements drawn from a Gaussian ensemble with mean zero and variance 1. When $M - k$ is sufficiently large, we may approximate $\lambda_{j\setminus i}$ as

$$\lambda_{j\setminus i} \to \frac{1}{\sigma_w^2}(M-k)||\sum_{m\in\mathcal{W}_{j\setminus i}}\mathbf{v}_{jm}c_j(m)||_2^2$$

where $\mathcal{W}_{j\setminus i}$ ($l = |\mathcal{W}_{j\setminus i}|$) denotes the set consisting of indices of basis vectors in $\mathcal{S}_j$ which are not in $\mathcal{S}_i$.

Proof: See Appendix C. ∎

The quantity $\sum_{m\in\mathcal{W}_{j\setminus i}}\mathbf{v}_{jm}c_j(m)$ is the portion of the original signal $\mathbf{x}$ that is unaccounted for by the subspace $\mathcal{S}_i$ when the true subspace is $\mathcal{S}_j$ for $j \neq i$. Let

$$\tilde{\alpha}_{\min,l}^2 = \frac{1}{\sigma_w^2}\min_{i,j,j\neq i}||\sum_{m\in\mathcal{W}_{j\setminus i}}\mathbf{v}_{jm}c_j(m)||_2^2$$

be the minimum (over $i, j = 0, 1, \cdots, T-1$) SNR of the original signal $\mathbf{x}$ which is unaccounted for by the subspace $\mathcal{S}_i$ when the true subspace is $\mathcal{S}_j$ such that $|\mathcal{W}_{j\setminus}| = l$

for $j \neq i$. Then, with random sampling, the upper bound on the probability of error of ML estimation in (16) reduces to (20) after replacing $\alpha^2_{\min,l}$ in (20) by $\tilde{\alpha}^2_{\min,l}$. It is worth mentioning that $\alpha^2_{\min,l}$ in (20) is a measure of SNR after sampling while $\tilde{\alpha}^2_{\min,l}$ is a measure of SNR before sampling the signal.

IV. SUBSPACE RECOVERY FROM STRUCTURED UNION OF SUBSPACES

In this section, we simplify the results obtained in Section III when the subspaces in the union (4) have structured properties leading to block sparsity.

A. Block Sparsity

With the block sparsity model as discussed in Subsection II.B, the observation vector $\mathbf{y}$ can be written in the form of

$$\mathbf{y} = AV\mathbf{c} + \mathbf{w} = B\mathbf{c} + \mathbf{w} \tag{21}$$

where $\mathbf{B} = AV$ is a $M \times N$ matrix, $V = [V_0|V_1|\cdots|V_{L-1}]$ is as defined in Subsection II.B and $\mathbf{c}$ has L blocks (of size d each) in which all but k_0 blocks are zeros; i.e., $\mathbf{c}$ is a block k_0-sparse vector. Letting $\mathbf{B}[i] = AV_i$ be a $M \times d$ matrix, we can represent $\mathbf{B}$ as a concatenation of column blocks $\mathbf{B}[i]$ for $i = 0, 1, \cdots, L - 1$. With this specific structure, the subspace recovery problem reduces to finding the indices of blocks in $\mathbf{c}$ such that the elements inside that block are nonzero, i.e., the problem of finding the block sparsity pattern. In addition to the structured union of subspaces model considered here in which the block sparsity pattern is observed, there are other instances where block sparsity arises such as in multiband signals [37], and in measurements of gene expression levels [27], [38].

Define the support set of the block sparse signal $\mathbf{c}$ as

$$\mathcal{U} := \{i \in \{0, 1, \cdots, L - 1\}|\mathbf{c}_i \neq 0\}$$

which consists of the indices of the subspaces in the sum in (6) or the indices of the nonzero blocks in $\mathbf{c}$. With the above formulation, there are $T = \binom{L}{k_0}$ such support sets and the jth support set is denoted by $\mathcal{U}_j$ for $j = 0, 1, \cdots, T - 1$.

Given that the true block support set is $\mathcal{U}_j$, the measurement vector in (21) can be written as,

$$\mathbf{y} = \bar{\mathbf{B}}_j \bar{\mathbf{c}}_j + \mathbf{w}$$

where $\bar{\mathbf{B}}_j = A\bar{V}_j$, $\bar{V}_j = [V_{u^0_j}|\cdots|V_{u^{k_0-1}_j}]$ and u^m_j denotes the mth index in the set $\mathcal{U}_j$ for $m = 0, 1, \cdots, k_0 - 1$. Similar interpretation holds for the vector $\bar{\mathbf{c}}_j$. To compute the minimum number of samples required for asymptotically reliable subspace recovery based on ML estimation, we follow a similar approach as in Theorem 2 with appropriate notation changes. In this case, we can explicitly find $T_0(l)$: for given l, there are $\binom{k_0}{l}\binom{L-k_0}{l}$ number of sets such that $|\mathcal{U}_{j\setminus i}| = l$ for any given $\mathcal{U}_j$. Thus $T_j(l) = T_0(l) = \binom{k_0}{l}\binom{L-k_0}{l}$. In the next section, we extend the analysis to the case where the sampling operator is represented by random projections.

B. Sampling via Random Projections

We assume that the signal of interest $\mathbf{x}$ is a $N \times 1$ vector and the sampling operator is a $M \times N$ matrix with random elements. Further, assume that the $N \times N$ basis matrix V defined in Section II.B is orthonormal. Then,

$$\mathbf{y} = B\mathbf{c} + \mathbf{w} \tag{22}$$

where $\mathbf{B} = AV$, V is a $N \times N$ orthonormal matrix, $\mathbf{c}$ is a block sparse signal with k_0 nonzero blocks each of length d and elements in $\mathbf{A}$ are drawn from a random ensemble.

Compared to the analysis in Subsection III.C with the block sparsity model, we can further simplify the expression obtained for $\lambda_{j\setminus i}$ in Proposition 1. We define the minimum nonzero block SNR as follows:

Definition 2: The minimum nonzero block SNR is defined as $\mathrm{BSNR}_{\min} = \frac{1}{\sigma_w^2} \min_{m \in \mathcal{U}} ||\mathbf{c}_m||_2^2$ where $\mathcal{U}$ is the set containing the indices corresponding to nonzero blocks.

Proposition 2: Let $BSNR_{\min}$ be the minimum nonzero block SNR of a block sparse signal. When the matrix $\mathbf{A}$ consists of elements drawn from a Gaussian ensemble with mean zero and variance 1, for any $\mathcal{U}_j$ and $\mathcal{U}_i$ with $l = |\mathcal{U}_{j\setminus i}|$ we have

$$\lambda_{j\setminus i} = \frac{1}{\sigma_w^2}(M - k_0 d) \sum_{m=0}^{l-1} ||\mathbf{V}_{u^m_{j\setminus i}} \mathbf{c}_{u^m_{j\setminus i}}||_2^2$$

$$\geq (M - k_0 d)l\mathrm{BSNR}_{\min}$$

where $u^m_{j\setminus i}$ denotes the mth index of the set $\mathcal{U}_{j\setminus i}$ which contains the indices of the subspaces in $\mathcal{U}_j$ which are not in $\mathcal{U}_i$.

Proof: The proof follows from Proposition 1 and the following results:

$$||\sum_{m=0}^{l-1} \mathbf{V}_{u^m_{j\setminus i}} \mathbf{c}_{u^m_{j\setminus i}}||_2^2 = \langle \sum_{m=0}^{l-1} \mathbf{V}_{u^m_{j\setminus i}} \mathbf{c}_{u^m_{j\setminus i}}, \sum_{m=0}^{l-1} \mathbf{V}_{u^m_{j\setminus i}} \mathbf{c}_{u^m_{j\setminus i}} \rangle$$

$$= \sum_{m=0}^{l-1} \langle \mathbf{V}_{u^m_{j\setminus i}} \mathbf{c}_{u^m_{j\setminus i}}, \mathbf{V}_{u^m_{j\setminus i}} \mathbf{c}_{u^m_{j\setminus i}} \rangle$$

$$+ \sum_{m \neq t} \langle \mathbf{V}_{u^m_{j\setminus i}} \mathbf{c}_{u^m_{j\setminus i}}, \mathbf{V}_{u^t_{j\setminus i}} \mathbf{c}_{u^t_{j\setminus i}} \rangle$$

$$= \sum_{m=0}^{l-1} ||\mathbf{V}_{u^m_{j\setminus i}} \mathbf{c}_{u^m_{j\setminus i}}||_2^2 \tag{23}$$

where the last equality is due to the fact that the columns of $\mathbf{V}$ are orthogonal. Then (23) is lower bounded by,

$$||\sum_{m=0}^{l-1} \mathbf{V}_{u^m_{j\setminus i}} \mathbf{c}_{u^m_{j\setminus i}}||_2^2 \geq \sigma_w^2 l\mathrm{BSNR}_{\min}$$

which completes the proof. ∎

Corollary 1: When the sampling operator is a random matrix with elements drawn from a Gaussian ensemble with mean zero and the variance 1, the upper bound on the probability of error of ML estimation in (16) for block sparsity

pattern recovery reduces to

$$P_e \leq \sum_{l=1}^{k_0} \binom{k_0}{l}\binom{L-k_0}{l}$$
$$\times \left(Q\left(\frac{1}{2}(1-2\eta_0)\sqrt{(M-k)l\mathrm{BSNR}_{\min}}\right) \right.$$
$$\left. + \Psi\left(l, \mathrm{BSNR}_{\min}\right)\right) \tag{24}$$

where $k_0 = k/d$,

$$\Psi\left(l, \mathrm{BSNR}_{\min}\right) = \frac{\sqrt{2}}{2^l \Gamma(l/2)}(\eta_0(M-k_0d)l\mathrm{BSNR}_{\min})^{l/2-1/2}$$
$$K_{l/2-1/2}(\eta_0(M-k_0d)l\mathrm{BSNR}_{\min}/2) \tag{25}$$

and $0 < \eta_0 < 1/2$.

Next, we investigate sufficient conditions which state how the number of samples M scales with the other parameters $(L, k_0, d, \mathrm{BSNR}_{\min})$ to ensure that the probability of error in (24) vanishes asymptotically with the block sparse model (22).

Lemma 2: When $(M - k)\mathrm{BSNR}_{\min} \to \infty$, the probability of error of ML estimation (24) in recovering the block sparsity pattern vanishes asymptotically if the following conditions are satisfied:

$$M > k + \max\{\bar{M}_1, \bar{M}_2\} \tag{26}$$

where

$$\bar{M}_1 = \frac{16}{BSNR_{\min}(1-2\eta_0)^2}\left(\log(L-k_0) + \log\left(\frac{e}{\sqrt{2}}\right)\right) \tag{27}$$

$$\bar{M}_2 = \frac{4(k_0/2 + r_0 - 1)}{\eta_0 r_0 \mathrm{BSNR}_{\min}}\left\{\log(L-k_0) + \frac{1}{2}\log\left(\frac{2b_0 e^2}{\sqrt{\pi}}\right)\right\} \tag{28}$$

and $0 < \eta_0 < \frac{1}{2}$, $r_0 > 0$ and $b_0 = \frac{\sqrt{2\pi}}{4}$ are constants.

Proof: Proof follows from Theorem 2 and using the relations, that $\binom{k_0}{l} \leq \binom{L-k_0}{l}$ for $k_0 \leq L/2$, and $\log\left(\binom{L-k_0}{l}\right) \leq l\log\left(\frac{e(L-k_0)}{l}\right)$. ∎

From Lemma 2, we can write the minimum number of random samples required for reliable block sparsity pattern recovery asymptotically in the form of $\mathcal{O}(k + \frac{\eta_4}{BSNR_{\min}}\log(L-k_0))$ for some constant η_4 in the case where k_0 is sufficiently small.

C. Revisiting the Standard Sparsity Model

In the standard sparsity model considered widely in the CS literature, the subspaces in the union (4) are assumed to be k-dimensional subspaces of an orthonormal basis. Define the minimum component SNR,

$$\mathrm{CSNR}_{\min} = \min_{m \in \mathcal{U}, i=0,\cdots,d-1} \frac{||\mathbf{c}_m(i)||_2^2}{\sigma_w^2} \tag{29}$$

so that $BSNR_{\min} \geq dCSNR_{\min}$. Then, when the sampling is performed via a random matrix, the probability of error of ML estimation with the standard sparsity model is upper bounded as in (30), as shown at the bottom of this page, where $\Psi(l, \mathrm{CSNR}_{\min})$ is as defined in Corollary 1. With these notations, the probability of error of ML estimation with block sparsity model (24) can be rewritten as in (31), as shown at the bottom of this page, where $L = N/d$ and $k_0 = k/d$ as defined previously. By obtaining the conditions under which P_e in (30) and (31) vanishes asymptotically, it can be shown that the dominant part of the required number of random samples for reliable subspace recovery asymptotically in the presence of noise can be expressed in the form of $\mathcal{O}(k + \frac{1}{d}\frac{\hat{\eta}_1}{CSNR_{\min}}\log(L-k_0))$ with block sparsity model and $\mathcal{O}(k + \frac{\hat{\eta}_2}{CSNR_{\min}}(\log(N-k)))$ with the standard sparsity model where $\hat{\eta}_1$ and $\hat{\eta}_2$ are positive constants. Thus, when the signal $\mathbf{x}$ exhibits block sparsity pattern with $k = k_0 d$ where k is the total number of non zero coefficients of the sparse signal, k_0 is the number of blocks and d is the block size, the required number of random samples beyond k (i.e. in terms of $M-k$) is reduced by approximately a factor of d compared to that with the standard sparsity model. Note that the above analysis is for the worst case, i.e. the upper bounds on the probability of error are obtained considering the minimum block/component SNR. The actual number of measurements required for reliable subspace recovery can be less than that predicted in Lemma 2.

V. COMPARISON WITH EXISTING RESULTS

A. Support Recovery in the Standard Sparsity Model

The most related existing work on sufficient conditions for ML estimation to succeed in the presence of noise with the standard sparsity model is presented in [14]. There, taking the canonical basis as the sparsifying basis, the results are derived based on the following bound on the probability of error:

$$P_e \leq \sum_{l=1}^{k} \binom{k}{l}\binom{N-k}{l}4\exp\left\{-\frac{(M-k)lCSNR_{\min}}{64(lCSNR_{\min}+8)}\right\}. \tag{32}$$

When $CSNR_{\min} \to \infty$, it can be easily seen that this upper bound is bounded away from zero. Based on (32), it was shown in [14] that

$$M > k + (\eta_1 + 2048)\max\left\{\tilde{M}_1, \tilde{M}_2\right\} \tag{33}$$

$$P_e \leq \sum_{l=1}^{k}\binom{k}{l}\binom{N-k}{l}\left(Q\left(\frac{1}{2}(1-2\eta_0)\sqrt{(M-k)l\mathrm{CSNR}_{\min}}\right) + \Psi\left(l, \mathrm{CSNR}_{\min}\right)\right) \tag{30}$$

$$P_e \leq \sum_{l=1}^{k_0}\binom{k_0}{l}\binom{L-k_0}{l}\left(Q\left(\frac{1}{2}(1-2\eta_0)\sqrt{(M-k)ld\mathrm{CSNR}_{\min}}\right) + \Psi\left(l, d\mathrm{CSNR}_{\min}\right)\right) \tag{31}$$

measurements are required for asymptotically reliable sparsity pattern recovery where

$$\tilde{M}_1 = \log \binom{N-k}{k}, \quad \tilde{M}_2 = \frac{\log(N-k)}{CSNR_{\min}}$$

and η_1 is a constant (which is different from the one used earlier in the paper). When the minimum component SNR, $CSNR_{\min} \to \infty$, ML estimation requires $k + (\eta_1 + 2048)k \log((N-k)/k)$ measurements for asymptotically reliable recovery, which is much larger than k. However, as shown in [17] and [39], when the measurement noise power is negligible (or in the no noise case), the exhaustive search decoder is capable of recovering the sparsity pattern with $M = k + 1$ measurements with high probability. Thus, the limits predicted by existing results show a gap with what is actually required.

When $d = 1$, **V** is the standard canonical basis, and **A** is a random Gaussian matrix, the structured union of subspaces model considered in Section IV (specifically the equation (22)) is the same as the model considered in [14]. Our results show that when $CSNR_{\min} \to \infty$, the upper bound on the probability of error in (30) vanishes with the standard sparsity model when $M > k$. More specifically, when $CSNR_{\min} \to \infty$, our results imply that $\mathcal{O}(k)$ measurements are sufficient for asymptotically reliable sparsity pattern recovery. Further, at finite $CSNR_{\min}$, when $\tilde{M}_2$ dominates $\tilde{M}_1$ in (33) the lower bound in [14] has the same scaling with respect to L, k, d and $CSNR_{\min}$ to that is obtained in this paper with the standard sparsity model.

B. Signal Recovery With Union of Subspaces

The problem of stable recovery of signals that lie in a union of subspaces model is addressed in [23]–[26], and [28]. In these works, the main focuss is to derive sufficient conditions that ensure reliable recovery of the complete signals. As an example, the following result is shown in [26].

Theorem 3: For any given $t > 0$, if

$$M > \frac{2}{c\delta} \left(\log(2T) + k \log\left(\frac{12}{\delta}\right) + t \right) \tag{34}$$

then, A in (7) satisfies the restricted isometry property (RIP) with the restricted isometry constant δ (for formal definition of RIP readers may refer to [26]).

In [27], the authors derived sufficient conditions for complete signal recovery in the block sparsity model. When the samples are acquired via a random matrix (elements in **A** are Gaussian) with the notations used in Section IVB, the minimum number of samples required for the sampling matrix to satisfy block RIP with high probability is given by (from Theorem 3 and [27])

$$M \geq \frac{36}{7\delta} \left(\log\left(2\binom{L}{k_0}\right) + k \log\left(\frac{12}{\delta}\right) + t \right) \tag{35}$$

for some $t > 0$ where $0 < \delta < 1$ is the restricted isometry constant. This is roughly on the order of $\tilde{\eta}_1 k + \tilde{\eta}_2 k_0 \log(L/k_0)$ for some positive constants $\tilde{\eta}_1$ and $\tilde{\eta}_2$. Thus, block sparse signals

---

Algorithm 1 Block-OMP (B-OMP) for Block Sparsity Pattern Recovery

---

Input: **y**, **B**, k_0

1) Initialize $t = 1$, $\hat{\mathcal{U}}(0) = \emptyset$, residual vector $\mathbf{r}_0 = \mathbf{y}$
2) Find $\lambda(t)$ such that $\lambda(t) = \underset{i=0,\cdots,L-1}{\arg \max} \ ||\mathbf{B}[i]^*\mathbf{r}_{t-1}||_2$
3) Set $\hat{\mathcal{U}}(t) = \hat{\mathcal{U}}(t-1) \cup \{\lambda(t)\}$
4) Compute the projection operator

$$\mathbf{P}(t) = \mathbf{B}(\hat{\mathcal{U}}(t)) \left(\mathbf{B}(\hat{\mathcal{U}}(t))^T \mathbf{B}(\hat{\mathcal{U}}(t)) \right)^{-1} \mathbf{B}(\hat{\mathcal{U}}(t))^T$$

Update the residual vector: $\mathbf{r}_t = (\mathbf{I} - \mathbf{P}(t))\mathbf{y}$ (note: $\mathbf{B}(\hat{\mathcal{U}}(t))$ denotes the submatrix of **B** in which columns are taken from **B** corresponding to the indices in $\hat{\mathcal{U}}(t)$)

5) Increment $t = t+1$ and go to step 2 if $t \leq k_0$, otherwise, stop and set $\hat{\mathcal{U}} = \hat{\mathcal{U}}(t-1)$

---

can be reliably recovered using computationally tractable algorithms with $\tilde{\eta}_1 k + \tilde{\eta}_2 k_0 \log(L/k_0)$ measurements when there is no noise. In the presence of noise, the BP based algorithm developed in [25] is shown to be robust so that the norm of the recovery error is bounded by the noise level.

As shown in Section IV.B, it requires roughly $k + (\eta_4/BSNR_{\min}) \log(L - k_0)$ measurements (when k_0 is fairly small) for reliable block sparsity pattern recovery with ML estimation. Here, the second term is significant at finite $BSNR_{\min}$ vanishes when $BSNR_{\min} \to \infty$. At finite $BSNR_{\min}$, when k_0 is sublinear w.r.t. L, it can be shown that $k_0 \log(L/k_0) >> \log(L - k_0)$. Thus, in that region of k_0, the relevant scaling obtained in (35) is larger than what is required by ML estimation at finite $BSNR_{\min}$. The exact difference between them depends on the value of $BSNR_{\min}$ and the relevant constants.

In [34], the authors considered the problem of recovering a block sparse signal using an overlapping group Lasso algorithm when the measurement matrix is Gaussian. The authors have shown that with $\mathcal{O}(k + (\sqrt{2\log(L - k_0)} + \sqrt{\tilde{d}})^2 \tilde{d})$ measurements, exact signal recovery with block sparsity is possible where $\tilde{d}$ is the largest block size. It can be seen that this scaling (with equal block sizes where $\tilde{d} = d$) is larger than that is obtained in this paper, $\mathcal{O}(k + \frac{\eta_4}{BSNR_{\min}} \log(L - k_0))$. This observation is intuitive since the ML method requires less measurements than suboptimal methods like overlapping Lasso.

VI. NUMERICAL RESULTS

Several computationally tractable algorithms for sparsity pattern recovery with standard sparsity have been derived and discussed quite extensively in the literature. Extensions of such algorithms for model based or structured CS have also been considered. For example, extensions of CoSamp and iterative hard thresholding algorithms for model based CS were developed in [23]. Extensions of OMP for block sparsity pattern recovery (BOMP) were considered in [27] and [40] while [25], [41], [42] considered the Group Lasso algorithm for block sparse signal recovery.
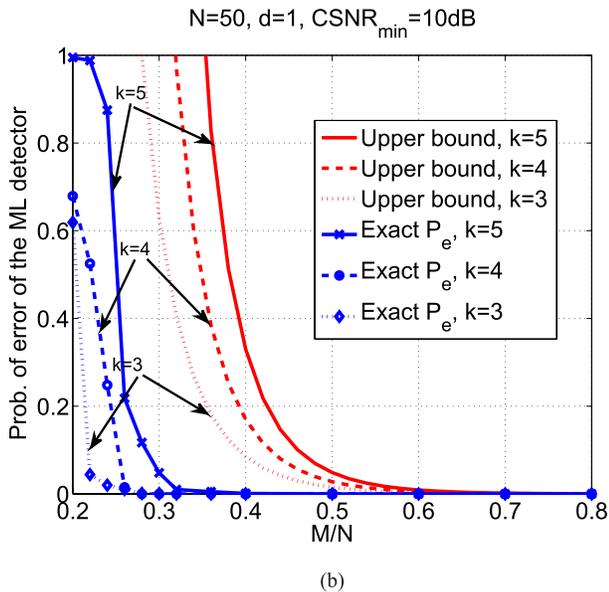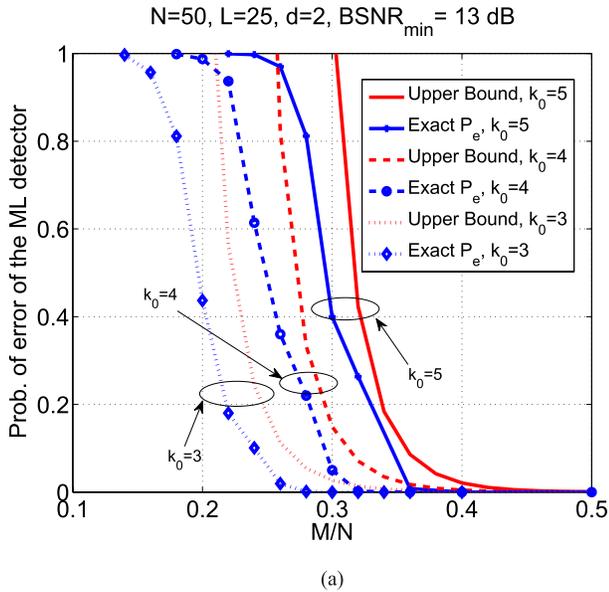
Fig. 1. Exact probability of error and the derived upper bound on the probability of error of the ML recovery for block sparsity pattern recovery. (a) $N = 50$, $L = 25$, $d = 2$, $BSNR_{\min} = 13dB$. (b) $N = L = 50$, $d = 1$, $CSNR_{\min} = 10dB$.

Fig. 2. Performance of ML estimation and the B-OMP algorithm for block sparsity pattern recovery; $L = 25$, $k_0 = 5$, $d = 2$, and thus $k = 10$, $N = 50$.

and variance 1. Thus, $\mathbf{V}$ is the standard canonical basis. In Fig. 1(a), the exact probability of error of ML estimation (obtained via simulation) and the upper bound on the probability of error derived in (24) vs M/N are shown. In the block sparsity model, we let $N = 50$, $d = 2$, $L = 25$, $BSNR_{\min} = 13dB$. The three different curves correspond to $k_0 = 3, 4, 5$. In Fig. 1 (b), we let $d = 1$ (i.e. the standard sparsity model) so that the upper bound on the probability of error reduces to (30). We also let $CSNR_{\min} = 10dB$. The different curves correspond to different values of k. To compute the exact probability of error 10^5 Monte Carlo runs were used. In the upper bounds (24) and (30), we let $\eta_0 = 1/4$. It can be seen from Fig. 1(a) and 1(b) that the derived upper bound on the probability of error is fairly tight especially as M/N increases. The tightness is more significant in Fig. 1(a). It should be noted that for $d = 2$, we have $k = k_0d$, thus the total number of non zero coefficients is larger in Fig. 1(a) than that with $d = 1$ in Fig. 1(b). Thus, the derived upper bound becomes tighter as k increases. It is also worth mentioning that the derived upper bound on the probability of error in [14] with the standard sparsity model (as in (32)) is bounded away from 1 for the selected parameter values mentioned above.

In Fig. 2, the performance of the block sparsity pattern recovery with ML and B-OMP algorithms is shown when $BSNR_{\min}$ varies. In Fig. 2, we let $k_0 = 5$, $L = 25$, $d = 2$ and $N = 50$. For B-OMP, 10^4 runs are performed for a given matrix and averaged over 100 runs. The ratio between the minimum and maximum block SNR in both cases considered is set at 1.825. From Fig. 2 it can be seen that the derived upper bound on the probability of error of ML estimation is fairly close to the exact probability error obtained via Monte Carlo simulations, especially as $BSNR_{\min}$ increases. Further, for a given finite $BSNR_{\min}$, there seems to be a considerable performance gap between B-OMP and ML estimation. That is the price to pay for the computational complexity of ML estimation vs the computationally efficient B-OMP algorithm.

Our goal in this section is to validate the tightness of the derived upper bounds on the probability of error of ML estimation and provide numerical results to illustrate the performance gap when employing practical algorithms for subspace recovery. Simulating ML algorithm is difficult due to its high computational complexity in high dimensions. Nevertheless, we show the performance for reasonably sized signal dimensions and samples just to demonstrate the tightness of the probability of error bound. The performance of the ML algorithm is compared to block-OMP as proposed in [27] which is provided in Algorithm 1 where the set $\hat{\mathcal{U}}$ contains the estimated indices of the nonzero blocks of a block sparse signal.

Results in Figures 1 and 2, assume that $\mathbf{x}$ is block sparse and the sampling operator a random matrix in which elements are drawn from a Gaussian ensemble with mean zero
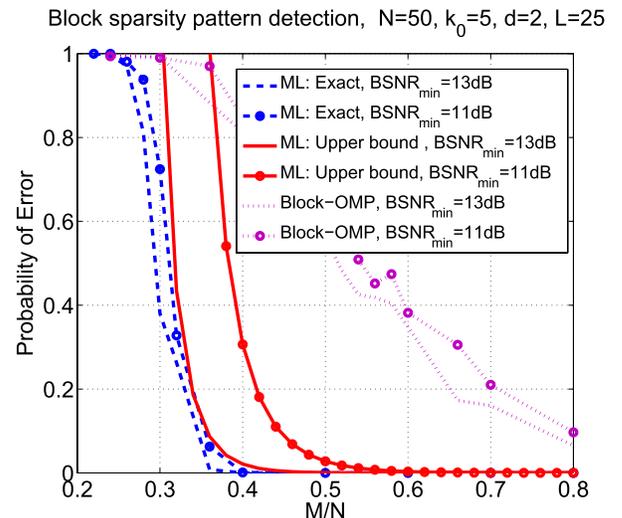
VII. CONCLUSION

In this paper, we investigated the problem of subspace recovery based on reduced dimensional samples when the signal of interest lies in a union of subspaces. With a given sampling operator, we considered the performance of ML estimation for subspace recovery in the presence of noise in terms of the probability of error. We further obtained conditions under which asymptotically reliable subspace recovery is guaranteed.

We extended the analysis to a special case of union of subspaces model which reduces to block sparsity. When the samples are obtained via a random matrix, sufficient conditions required for asymptotically reliable block sparsity pattern recovery with ML estimation were derived. Performance gain in terms of the minimum number of samples required for asymptotically reliable subspace recovery with the block sparse model was quantified compared to that with the standard sparsity model. Our results further strengthen the existing results for sparsity pattern recovery with the standard sparsity model used in CS as they are derived based on a tighter bound on the probability of error. We also discussed and illustrated numerically the performance gap between ML estimation and computationally tractable algorithms (e.g. B-OMP) used for subspace recovery with the structured union of subspaces model.

APPENDIX A
PROOF OF LEMMA 1

To prove Lemma 1, we use a similar argument to that considered in [14] with certain differences as noted in the following. As shown in [14], we may write,

$$\Delta_{ij}(\mathbf{y}) = ||\mathbf{P}_i^\perp \mathbf{y}||_2^2 - ||\mathbf{P}_i^\perp \mathbf{w}||_2^2 + ||\mathbf{P}_i^\perp \mathbf{w}||_2^2 - ||\mathbf{P}_j^\perp \mathbf{y}||_2^2.$$

For any given $\delta > 0$, define the events

$$h_1(\delta) = \left\{ |\frac{||\mathbf{P}_j^\perp \mathbf{y}||_2^2 - ||\mathbf{P}_i^\perp \mathbf{w}||_2^2}{\sigma_w^2}| \geq \delta \right\} \qquad (36)$$

and

$$h_2(\delta) = \left\{ \frac{||\mathbf{P}_i^\perp \mathbf{y}||_2^2 - ||\mathbf{P}_i^\perp \mathbf{w}||_2^2}{\sigma_w^2} \leq 2\delta \right\}. \qquad (37)$$

Then $Pr(\Delta_{ij}(\mathbf{y}) < 0)$ implies that at least one event in (36) and (37) is true. Based on the union bound, we can write

$$Pr(\Delta_{ij}(\mathbf{y}) < 0) \leq Pr(h_1(\delta)) + Pr(h_2(\delta)).$$

With the standard sparsity model and assuming that the sampling is performed via random projections, upper bounds on the probabilities $Pr(h_1(\delta))$ and $Pr(h_2(\delta))$ are derived in [14]. In contrast, in the following, we derive an exact value for $Pr(h_2(\delta))$ and a tighter bound for $Pr(h_1(\delta))$. Thus, even for the standard sparsity model, the results presented in this paper tighten those derived in [14].

We first evaluate $Pr(h_1(\delta))$. Let $\Delta_{ij}^1(\mathbf{y}) = \frac{1}{\sigma_w^2}(||\mathbf{P}_j^\perp \mathbf{y}||_2^2 - ||\mathbf{P}_i^\perp \mathbf{w}||_2^2)$. Assuming the true subspace is $\mathcal{S}_j$, $\Delta_{ij}^1(\mathbf{y})$ reduces to $\Delta_{ij}^1(\mathbf{y}) = \frac{1}{\sigma_w^2}(||\mathbf{P}_j^\perp \mathbf{w}||_2^2 - ||\mathbf{P}_i^\perp \mathbf{w}||_2^2)$. As shown in [14], the random variable $\Delta_{ij}^1(\mathbf{y})$ can be represented as $\Delta_{ij}^1(\mathbf{y}) = x_1 - x_2$

where x_1 and x_2 are independent and $x_1, x_2 \sim \mathcal{X}_l^2$ where l is the cardinality of the set $\mathcal{W}_{j \setminus i}$. With these notations, we can write

$$Pr(h_1(\delta)) = Pr(|x_1 - x_2| \geq \delta)$$
$$= Pr((x_1 - x_2) \geq \delta) + Pr((x_1 - x_2) < -\delta).$$

The pdf of the random variable $w = x_1 - x_2$ is symmetric around zero and thus

$$Pr(h_1(\delta)) = 2Pr((x_1 - x_2) \geq \delta).$$

Proposition 3: When $x_1 \sim \mathcal{X}_l^2$ and $x_2 \sim \mathcal{X}_l^2$, the random variable $w = x_1 - x_2$ has the following pdf:

$$f_w(w) = \begin{cases} f_w^+(w) = \frac{w^{\frac{l}{2}-\frac{1}{2}}}{\sqrt{\pi}2^l \Gamma(l/2)} K_{1/2-l/2}\left(\frac{w}{2}\right); & \text{if } w \geq 0 \\ f_w^-(w) = \frac{(-w)^{\frac{l}{2}-\frac{1}{2}}}{\sqrt{\pi}2^l \Gamma(l/2)} K_{1/2-l/2}\left(\frac{-w}{2}\right); & \text{if } w < 0 \end{cases}$$
$$(38)$$

where $K_v(x)$ is the modified Bessel function.

Proof: Since x_1 and x_2 are independent, the pdf of $w = x_1 - x_2$ is given by [43]

$$f_w(w) = \begin{cases} \int_0^\infty f_{x_1}(w + x_2) f_{x_2}(x_2) dx_2; & \text{if } w \geq 0 \\ \int_{-w}^\infty f_{x_1}(w + x_2) f_{x_2}(x_2) dx_2; & \text{if } w < 0 \end{cases}$$

First consider the case where $w > 0$. Then

$$f_w^+(w) = \int_0^\infty \frac{(w + x_2)^{l/2-1} e^{-(w+x_2)/2}}{2^{l/2}\Gamma(l/2)} \frac{x_2^{l/2-1} e^{-x_2/2}}{2^{l/2}\Gamma(l/2)} dx_2$$
$$= \frac{e^{-w/2}}{2^l(\Gamma(l/2))^2} \int_0^\infty x_2^{l/2-1}(w + x_2)^{l/2-1} e^{-x_2} dx_2$$
$$= \frac{e^{-w/2}}{2^l(\Gamma(l/2))^2} \frac{1}{\sqrt{\pi}} w^{l/2-1/2} e^{w/2} \Gamma(l/2) K_{1/2-l/2}(w/2)$$
$$= \frac{w^{l/2-1/2} K_{1/2-l/2}(w/2)}{\sqrt{\pi}2^l \Gamma(l/2)}$$

where $K_v(x)$ is the modified Bessel function and the third equality is obtained using the integral result

$$\int_0^\infty x^{v-1}(x + \beta)^{v-1} e^{-\mu x} dx$$
$$= \frac{1}{\sqrt{\pi}} \left(\frac{\beta}{\mu}\right)^{v-1/2} e^{\beta\mu/2} \Gamma(v) K_{1/2-v}\left(\frac{\beta\mu}{2}\right) \qquad (39)$$

for $\mu, v > 0$ in [44, p. 348].

When $w < 0$, we have,

$$f_w^-(w) = \frac{e^{-w/2}}{2^l(\Gamma(l/2))^2} \int_{-w}^\infty x_2^{l/2-1}(w + x_2)^{l/2-1} e^{-x_2} dx_2. \qquad (40)$$

Letting $z = -w$ where $z > 0$, (40) can be rewritten as,

$$f_w^-(w) = \frac{e^{z/2}}{2^l(\Gamma(l/2))^2} \int_z^\infty x_2^{l/2-1}(x_2 - z)^{l/2-1} e^{-x_2} dx_2. \qquad (41)$$

Using the integral result,

$$\int_u^\infty x^{v-1}(x - u)^{v-1} e^{-\mu x} dx$$
$$= \frac{1}{\sqrt{\pi}} \left(\frac{u}{\mu}\right)^{v-1/2} e^{-\mu u/2} \Gamma(v) K_{v-1/2}\left(\frac{\mu u}{2}\right) \qquad (42)$$

in [44, p. 347] and the relation $K_\nu(x) = K_{-\nu}(x)$, we get $f_w^-(w)$ as in (38), completing the proof. ■

Proposition 4: For $\delta > 0$, the probability $Pr(w > \delta)$ is given by

$$Pr(w > \delta) \leq \frac{\sqrt{2}}{2^{l+1}\Gamma(l/2)}\delta^{l/2-1/2}K_{l/2-1/2}(\delta/2)$$

where $K_\nu(x)$ is the modified Bessel function, and $\Gamma(.)$ is the Gamma function.

Proof: Based on (38), we have

$$Pr(w > \delta) = \int_\delta^\infty f_w^+(w)dw$$
$$= \int_\delta^\infty \frac{w^{l/2-1/2}K_{1/2-l/2}(w/2)}{\sqrt{\pi}2^l\Gamma(l/2)}dw. \quad (43)$$

Using the equivalent integral representation [44, p. 917]

$$K_\nu(az) = \frac{z^\nu}{2}\int_0^\infty e^{-\frac{a}{2}\left(t+\frac{z^2}{t}\right)}t^{-\nu-1}dt$$

we can write the integral in (43) as,

$$Pr(w > \delta)$$
$$= \frac{1}{\sqrt{\pi}2^{l+1}\Gamma(l/2)}\int_\delta^\infty\int_0^\infty e^{-\frac{1}{4}\left(t+\frac{w^2}{t}\right)}t^{l/2-3/2}dtdw.$$
$$(44)$$

Since $\int_\delta^\infty e^{-\frac{w^2}{4t}}dw = \sqrt{2\pi}Q\left(\frac{\delta}{\sqrt{2t}}\right)$, (44) reduces to,

$$Pr(w > \delta) = \frac{\sqrt{2}}{2^{l+1}\Gamma(l/2)}\int_0^\infty e^{-t/4}t^{l/2-3/2}Q\left(\frac{\delta}{\sqrt{2t}}\right)dt$$
$$\leq \frac{\sqrt{2}}{2^{l+2}\Gamma(l/2)}\int_0^\infty t^{l/2-3/2}e^{-t/4-\frac{\delta^2}{4t}}dt \quad (45)$$
$$= \frac{\sqrt{2}}{2^{l+1}\Gamma(l/2)}\delta^{l/2-1/2}K_{l/2-1/2}(\delta/2) \quad (46)$$

where we used the inequality $Q(x) \leq \frac{1}{2}e^{-\frac{x^2}{2}}$ for $x > 0$, and the relation, $\int_0^\infty x^{\nu-1}e^{-\beta/x-\gamma x}dx = 2\left(\frac{\beta}{\gamma}\right)^{\nu/2}K_\nu(2\sqrt{\beta\gamma})$ for $\beta > 0$ and $\gamma > 0$ [44, p. 368] to obtain (45) and (46), respectively, which completes the proof. ■

We have

$$Pr(h_1(\delta)) = \frac{\sqrt{2}}{2^l\Gamma(l/2)}\delta^{l/2-1/2}K_{l/2-1/2}(\delta/2).$$

Next we compute the quantity $Pr(h_2(\delta))$. Let $\Delta_{ij}^2(\mathbf{y}) = \frac{1}{\sigma_w^2}(||\mathbf{P}_i^\perp\mathbf{y}||_2^2 - ||\mathbf{P}_i^\perp\mathbf{w}||_2^2)$. Then

$$\Delta_{ij}^2(\mathbf{y}) = \frac{1}{\sigma_w^2}(||\mathbf{P}_i^\perp\mathbf{B}_{j\setminus i}\mathbf{c}_{j\setminus i}||_2^2 + 2\mathbf{w}^T\mathbf{P}_i^\perp\mathbf{B}_{j\setminus i}\mathbf{c}_{j\setminus i}).$$

Since $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma_w^2\mathbf{I}_M)$, $\Delta_{ij}^2(\mathbf{y})$ is a Gaussian random variable with pdf,

$$\Delta_{ij}^2(\mathbf{y}) \sim \mathcal{N}\left(\frac{1}{\sigma_w^2}||\mathbf{P}_i^\perp\mathbf{B}_{j\setminus i}\mathbf{c}_{j\setminus i}||_2^2, \frac{4}{\sigma_w^2}||\mathbf{P}_i^\perp\mathbf{B}_{j\setminus i}\mathbf{c}_{j\setminus i}||_2^2\right).$$

Thus,

$$Pr(h_2(\delta)) = Pr\left(\Delta_{ij}^2(\mathbf{y}) \leq 2\delta\right)$$
$$= 1 - Q\left(\frac{2\delta - \frac{1}{\sigma_w^2}||\mathbf{P}_i^\perp\mathbf{B}_{j\setminus i}\mathbf{c}_{j\setminus i}||_2^2}{\frac{2}{\sigma_w}||\mathbf{P}_i^\perp\mathbf{B}_{j\setminus i}\mathbf{c}_{j\setminus i}||_2}\right)$$
$$= 1 - Q\left(\frac{2\delta - \lambda_{j\setminus i}}{2\sqrt{\lambda_{j\setminus i}}}\right).$$

Since it is desired to control δ such that $Pr(h_2(\delta)) \leq 1/2$, we select $\delta^* = \eta_0\lambda_{j\setminus i}$ where $\eta_0 < \frac{1}{2}$. With this choice $Pr(h_2(\delta))$ reduces to

$$Pr(h_2(\delta)) = Q\left(\frac{1}{2}\sqrt{\lambda_{j\setminus i}}(1 - 2\eta_0)\right)$$

where we used the relation $1 - Q(-x) = Q(x)$ for $x > 0$, while $Pr(h_1(\delta))$ reduces to,

$$Pr(h_1(\delta)) = \frac{\sqrt{2}}{2^l\Gamma(l/2)}(\eta_0\lambda_{j\setminus i})^{l/2-1/2}K_{l/2-1/2}(\eta_0\lambda_{j\setminus i}/2).$$
$$(47)$$

APPENDIX B
PROOF OF THEOREM 2

To obtain conditions under which the probability of error bound in (20) asymptotically vanishes, we rely on the following corollary.

Corollary 2: Let $T_0(l)$ and $\alpha_{\min,l}^2$ be as defined in Subsection III.B. The probability of error of ML estimation in (20) is further upper bounded by

$$P_e \leq \sum_{l=1}^k T_0(l)\left(\frac{1}{2}e^{-\frac{1}{8}(1-2\eta_0)^2(M-k)\alpha_{\min,l}^2} + \phi_l\right) \quad (48)$$

where

$$\phi_l = \frac{\sqrt{2\pi}}{4\Gamma(l/2)}\left(\frac{1}{4}\eta_0(M-k)\alpha_{\min,l}^2\right)^{l/2-1}e^{-\frac{1}{2}\eta_0(M-k)\alpha_{\min,l}^2}$$
$$(49)$$

when $(M-k)\alpha_{\min,l}^2 >> (l/2 - 1/2)$ for all $l = 1, 2, \cdots, k$ and $0 < \eta_0 < 1/2$.

Proof: Using the Chernoff bound for the Q function $Q(x) \leq \frac{1}{2}e^{-\frac{x^2}{2}}$, we have

$$Q\left(\frac{1}{2}(1 - 2\eta_0)\sqrt{(M-k)\alpha_{\min,l}^2}\right)$$
$$\leq \frac{1}{2}e^{-\frac{1}{8}(1-2\eta_0)^2(M-k)\alpha_{\min,l}^2}$$

for $\eta_0 < \frac{1}{2}$.

To obtain (49) we used the relation $K_\nu(z) \approx \sqrt{\frac{\pi}{2z}}e^{-z}$ when $\nu << z$, completing the proof. ■

It is further noted that when k is fairly small and $\alpha_{\min,l}^2$ is sufficiently large, the condition required for (49) is often satisfied. We consider the conditions under which each term in (48) goes to 0 asymptotically, equivalently, the logarithm of

each term $\to -\infty$. Consider the first term in the summation in (48) for which the logarithm gives,

$$\log T_0(l) + \log(1/2) - \frac{1}{8}(1 - 2\eta_0)^2(M - k)\alpha_{\min,l}^2$$

$$\leq \max_l \left\{ \log(T_0(l)) + \log(1/2) \right.$$
$$\left. - \frac{1}{8}(1 - 2\eta_0)^2(M - k)\ \alpha_{\min,l}^2 \right\} \to -\infty$$

as $(M - k) \to \infty$ when $M > k + M_1$ where

$$M_1 = \max_{l=1,\cdots,k} \left\{ \frac{8}{(1 - 2\eta_0)^2\alpha_{\min,l}^2} \{\log(T_0(l)) + \log(1/2)\} \right\}.$$

To treat the second term in (48), let

$$\Pi_1 = \log T_0(l) + \log\left(\frac{b_0}{\Gamma(l/2)}\right)$$
$$+ (l/2 - 1)\log\left(\frac{1}{4}\eta_0(M - k)\alpha_{\min,l}^2\right)$$
$$- \frac{1}{2}\eta_0(M - k)\alpha_{\min,l}^2 \qquad (50)$$

where $b_0 = \frac{\sqrt{2\pi}}{4}$. When $\frac{1}{4}\eta_0(M - k)\alpha_{\min,l}^2$ is sufficiently large, we can find $0 < q_0 < \frac{1}{(k/2-1)}$ such that $\log\left(\frac{1}{4}\eta_0(M - k)\alpha_{\min,l}^2\right) < q_0\frac{1}{2}\eta_0(M - k)\alpha_{\min,l}^2$. Then (50) is upper bounded by

$$\Pi_1 \leq \max_{l=1,\cdots,k} \left\{ \log(T_0(l)) + \log\left(\frac{b_0}{\Gamma(3/2)}\right) \right.$$
$$\left. - \left(\frac{1}{2}\eta_0(M - k)\alpha_{\min,l}^2\right)(1 - q_0(k/2 - 1)) \right\}$$
$$= \Pi_2 \qquad (51)$$

where $0 < q_0 < \frac{1}{(k/2-1)}$. We can write q_0 in the form of $q_0 = \frac{1}{2(k/2+r_0-1)}$ for some $r_0 > 0$. Thus, (51) can be rewritten as

$$\Pi_2$$
$$= \max_{l=1,\cdots,k} \left\{ \log(T_0(l)) + \log\left(\frac{2b_0}{\sqrt{\pi}}\right) \right.$$
$$\left. - \left(\frac{1}{2}\eta_0(M - k)\alpha_{\min,l}^2\right)\frac{r_0}{r_0 + k/2 - 1} \right\} \to -\infty$$

as $(M - k) \to \infty$ when $M > k + M_2$ where

$$M_2 = \max_{l=1,\cdots,k} \left\{ \frac{2(k/2 + r_0 - 1)}{r_0\eta_0\alpha_{\min,l}^2} \left\{ \log(T_0(l)) + \log\left(\frac{2b_0}{\sqrt{\pi}}\right) \right\} \right\},$$

$0 < \eta_0 < 1/2$, $b_0 = \frac{\sqrt{2\pi}}{4}$, and $r_0 > 0$.

APPENDIX C
PROOF OF PROPOSITION 1

We rewrite $\lambda_{j\backslash i} = \frac{1}{\sigma_w^2}||\mathbf{P}_i^\perp\mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i}||_2^2$. The tth element of the vector $\mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i}$ can be written as $\langle \mathbf{a}_t, \sum_{m\in\mathcal{W}_{j\backslash i}} \mathbf{v}_{jm}c_j(m) \rangle$ where $\mathbf{a}_t$ is the tth row vector of $\mathbf{A}$ for $t = 0, 1, \cdots, M - 1$. Assuming that the elements of $\mathbf{A}$ are independent Gaussian with mean zero and variance 1, it can be easily seen that

$\langle \mathbf{a}_t, \sum_{m\in\mathcal{W}_{j\backslash i}} \mathbf{v}_{jm}c_j(m) \rangle$ is a realization of a Gaussian random variable with mean zero and variance $||\sum_{m\in\mathcal{W}_{j\backslash i}} \mathbf{v}_{jm}c_j(m)||_2^2$. Further, the elements of $\mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i}$ are independent of each other since $\mathbf{a}_t$'s are independent for $t = 0, 1, \cdots, M - 1$. Thus, the random vector $\mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i} \sim \mathcal{N}(\mathbf{0}, ||\sum_{m\in\mathcal{W}_{j\backslash i}} \mathbf{v}_{jm}c_j(m)||_2^2\mathbf{I}_M)$.

Consider again the transformation $\mathbf{Q}_i^T\mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i}$ where $\mathbf{Q}_i$ is the unitary matrix with eigenvectors of $\mathbf{P}_i^\perp$. Since the elements in $\mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i}$ are independent and identically distributed (iid), the unitary transformation does not change the distribution of $\mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i}$. Then $||\mathbf{P}_i^\perp\mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i}||_2^2 = ||\Lambda_i\mathbf{Q}_i^T\mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i}||_2^2$ is a sum of $M - k$ iid random variables. Thus when $(M - k)$ is sufficiently large, invoking the law of large numbers, we may approximate $||\mathbf{P}_i^\perp\mathbf{B}_{j\backslash i}\mathbf{c}_{j\backslash i}||_2^2 \to (M - k)||\sum_{m\in\mathcal{W}_{j\backslash i}} \mathbf{v}_{jm}c_j(m)||_2^2$

which completes the proof.

REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

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

[4] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[5] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[6] V. Cevher, P. Indyk, C. Hegde, and R. G. Baraniuk, "Recovery of clustered sparse signals from compressive measurements," in *Proc. Int. Conf. Sampling Theory Appl. (SAMPTA)*, Marseille, France, May 2009, pp. 18–22.

[7] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.

[8] A. J. Miller, *Subset Selection in Regression*. New York, NY, USA: Chapman & Hall, 1990.

[9] E. G. Larsson and Y. Selén, "Linear regression with a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 451–460, Feb. 2007.

[10] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, USA, Apr. 2007, pp. IV-1357–IV-1360.

[11] M. Mishali and Y. C. Eldar, "Wideband spectrum sensing at sub-Nyquist rates [applications corner]," *IEEE Signal Process. Mag.*, vol. 28, no. 4, pp. 102–135, Jul. 2011.

[12] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to digital at sub-Nyquist rates," *IET Circuits, Devices, Syst.*, vol. 5, no. 1, pp. 8–20, Jan. 2011.

[13] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[14] M. J. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5728–5741, Dec. 2009.

[15] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso)," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2183–2202, May 2009.

[16] W. Wang, M. J. Wainwright, and K. Ramchandran, "Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2967–2979, Jun. 2010.

[17] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5758–5772, Dec. 2009.

[18] M. Akcakaya and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 492–504, Jan. 2010.

[19] G. Reeves and M. Gastpar, "Sampling bounds for sparse support recovery in the presence of noise," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Toronto, ON, USA, Jul. 2008, pp. 2187–2191.

[20] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1383–1399, Mar. 2010.

[21] V. K. Goyal, A. K. Fletcher, and S. Rangan, "Compressive sampling and lossy compression," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 48–56, Mar. 2008.

[22] T. Wimalajeewa and P. K. Varshney, "Performance bounds for sparsity pattern recovery with quantized noisy random projections," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 1, pp. 43–57, Feb. 2012.

[23] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[24] Y. M. Lu and M. N. Do, "A theory for sampling signals from a union of subspaces," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2334–2345, Jun. 2008.

[25] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[26] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1872–1882, Apr. 2009.

[27] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[28] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.

[29] A. M. Bruckstein, T.-J. Shan, and T. Kailath, "The resolution of overlapping echos," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 6, pp. 1357–1367, Dec. 1985.

[30] K. Gedalyahu and Y. C. Eldar, "Time-delay estimation from low-rate samples: A union of subspaces approach," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3017–3031, Jun. 2010.

[31] Z. Ben-Haim, T. Michaeli, and Y. C. Eldar, "Performance bounds and design criteria for estimating finite rate of innovation signals," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 4993–5015, Aug. 2012.

[32] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets strang–fix," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1741–1757, May 2007.

[33] D. Percival, "Theoretical properties of the overlapping groups Lasso," *Electron. J. Statist.*, vol. 6, pp. 269–288, 2012.

[34] N. S. Rao, B. Recht, and R. D. Nowak, "Universal measurement bounds for structured sparse signal recovery," in *Proc. 15th Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2012, pp. 942–950.

[35] F. R. Bach, "Consistency of the group Lasso and multiple kernel learning," *J. Mach. Learn. Res.*, vol. 9, pp. 1179–1225, Jun. 2008.

[36] Z. Ben-Haim and Y. C. Eldar, "Near-oracle performance of greedy block-sparse estimation techniques from noisy measurements," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 1032–1047, Sep. 2011.

[37] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.

[38] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 3, pp. 275–285, Jun. 2008.

[39] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressive sensing," Dept. Elect. Comput. Eng., Rice Univ., Houston, TX, USA, Tech. Rep. TREE-0612, Nov. 2006.

[40] J. Fang and H. Li, "Block-sparsity pattern recovery from noisy observations," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2012, pp. 3321–3324.

[41] X. Lv, G. Bi, and C. Wan, "The group Lasso for stable recovery of block-sparse signal representations," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1371–1382, Apr. 2011.

[42] J. Friedman, T. Hastie, and R. Tibshirani. (2010). "A note on the group Lasso and a sparse group Lasso." [Online]. Available: http://arxiv.org/abs/1001.0736

[43] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York, NY, USA: McGraw-Hill, 2002.

[44] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*. New York, NY, USA: Academic, 2007.

Thakshila Wimalajeewa (S'07–M'10) received the B.Sc. degree in electronic and telecommunication engineering with First Class Honors from the University of Moratuwa, Sri Lanka, in 2004 and the M.S., and Ph.D. degrees in electrical and computer engineering from the University of New Mexico, Albuquerque, NM in 2007 and 2009, respectively.

She spent 2010-2012 as a postdoctoral research associate in the Department of Electrical Engineering and Computer Science, Syracuse University (SU), Syracuse, NY. She currently holds a research faculty position at SU. Her research interests lie in the broad areas of communication theory, signal processing and information theory. Her current research focuses on compressive sensing, low dimensional signal processing for communication systems, and resource optimization and decentralized processing in sensor networks.

Yonina C. Eldar (S'98–M'02–SM'07–F'12) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering both from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002. From January 2002 to July 2002, she was a Postdoctoral Fellow at the Digital Signal Processing Group at MIT. She is currently a Professor in the Department of Electrical Engineering at the Technion - Israel Institute of Technology, Haifa and holds the The Edwards Chair in Engineering. She is also a Research Affiliate with the Research Laboratory of Electronics at MIT and was a Visiting Professor at Stanford University, Stanford, CA. Her research interests are in the broad areas of statistical signal processing, sampling theory and compressed sensing, optimization methods, and their applications to biology and optics.

Dr. Eldar was a Horev Fellow of the Leaders in Science and Technology program at the Technion from 2002-2005 and an Alon Fellow. In 2004, she was awarded the Wolf Foundation Krill Prize for Excellence in Scientific Research, in 2005 the Andre and Bella Meyer Lectureship, in 2007 the Henry Taub Prize for Excellence in Research, in 2008 the Hershel Rich Innovation Award, the Award for Women with Distinguished Contributions, the Muriel & David Jacknow Award for Excellence in Teaching, and the Technion Outstanding Lecture Award, in 2009 the Technions Award for Excellence in Teaching, in 2010 the Michael Bruno Memorial Award from the Rothschild Foundation, and in 2011 the Weizmann Prize for Exact Sciences. In 2012 she was elected to the Young Israel Academy of Science and to the Israel Committee for Higher Education. In 2013 she received the Technions Award for Excellence in Teaching, the Hershel Rich Innovation Award, and the IEEE Signal Processing Technical Achievement Award, and in 2014 the IEEE/AESS Fred Nathanson Memorial Radar Award. She received several best paper awards together with her research students and colleagues. She is the Editor in Chief of *Foundations and Trends in Signal Processing* and a member of the IEEE Sensor Array and Multichannel Technical Committee. In the past, she was a Signal Processing Society Distinguished Lecturer, member of the IEEE Signal Processing Theory and Methods and Bio Imaging Signal Processing technical committees, and served as an associate editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the *EURASIP Journal of Signal Processing*, the *SIAM Journal on Matrix Analysis and Applications*, and the *SIAM Journal on Imaging Sciences*.

Pramod K. Varshney (S'72–M'77–SM'82–F'97) was born in Allahabad, India, on July 1, 1952. He received the B.S. degree in electrical engineering and computer science (with highest honors), and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana-Champaign in 1972, 1974, and 1976 respectively.

During 1972-76, he held teaching and research assistantships at the University of Illinois. Since 1976 he has been with Syracuse University, Syracuse, NY where he is currently a Distinguished Professor of Electrical Engineering and Computer Science and the Director of CASE: Center for Advanced Systems and Engineering. He is also an Adjunct Professor of Radiology at Upstate Medical University in Syracuse, NY. His current research interests are in distributed sensor networks and data fusion, detection and estimation theory, wireless communications, image processing, radar signal processing and cybersecurity. He has published extensively. He is the author of *Distributed Detection and Data Fusion*, published by Springer-Verlag in 1997.

While at the University of Illinois, Dr. Varshney was a James Scholar, a Bronze Tablet Senior, and a Fellow. He is a member of Tau Beta Pi and is the recipient of the 1981 ASEE Dow Outstanding Young Faculty Award. He was elected to the grade of Fellow of the IEEE in 1997 for his contributions in the area of distributed detection and data fusion. In 2000, he received the Third Millennium Medal from the IEEE and Chancellor's Citation for exceptional academic achievement at Syracuse University. He is on the editorial boards of *Journal on Advances in Information Fusion* and *IEEE Signal Processing Magazine*. He was the President of International Society of Information Fusion during 2001. He is the recipient of the IEEE 2012 Judith A. Resnik Award and Doctor of Engineering honoris causa from Drexel University in 2014.