Performance Analysis for Sparse Support Recovery
Gongguo Tang, Student Member, IEEE, and Arye Nehorai, Fellow, IEEE

Abstract—The performance of estimating the common support for jointly sparse signals based on their projections onto lower-dimensional space is analyzed. Support recovery is formulated as a multiple-hypothesis testing problem. Both upper and lower bounds on the probability of error are derived for general measurement matrices, by using the Chernoff bound and Fano’s inequality, respectively. The upper bound shows that the performance is determined by a quantity measuring the measurement matrix incoherence, while the lower bound reveals the importance of the total measurement gain. The lower bound is applied to derive the minimal number of samples needed for accurate direction-of-arrival (DOA) estimation for a sparse representation based algorithm. When applied to Gaussian measurement ensembles, these bounds give necessary and sufficient conditions for a vanishing probability of error for majority realizations of the measurement matrix. Our results offer surprising insights into sparse signal recovery. For example, for the support recovery case, the Chernoff bound is well known in Compressive Sensing with the Gaussian measurement matrix is generally not sufficient unless the noise level is low. Our study provides an alternative performance measure, one that is natural and important in practice, for signal recovery in Compressive Sensing and other application areas exploiting signal sparsity.

Index Terms—Chernoff bound, compressive sensing, Fano’s inequality, jointly sparse signals, multiple hypothesis testing, probability of error, support recovery.

I. INTRODUCTION

SUPPORy recovery for jointly sparse signals concerns accurately estimating the nonzero component locations shared by a set of sparse signals based on a limited number of noisy linear observations. More specifically, suppose that \( \{ x(t) \in \mathbb{F}^N, t = 1, 2, \ldots, T \} \), \( \mathbb{F} = \mathbb{R} \) or \( \mathbb{C} \), is a sequence of jointly sparse signals (possibly under a sparsity-inducing basis \( \Phi \) instead of the canonical domain) with a common support \( S \), which is the index set indicating the nonvanishing signal coordinates. This model is the same as the joint sparsity model 2 (JSM-2) in [1]. The observation model is linear

\[
g(y(t)) = Ax(t) + w(t) \quad t = 1, 2, \ldots, T.
\]

In (1), \( A \in \mathbb{F}^{M \times N} \) is the measurement matrix, \( y(t) \in \mathbb{F}^M \) the noisy data vector, and \( w(t) \in \mathbb{F}^M \) an additive noise. In most cases, the sparsity level \( K \triangleq |S| \) and the number of observations \( M \) are far less than \( N \), the dimension of the ambient space. This problem arises naturally in several signal processing areas such as Compressive Sensing [2]–[6], source localization [7]–[10], sparse approximation and signal denoising [11].

Compressive Sensing [2]–[4], a recently developed field exploiting the sparsity property of most natural signals, shows great promise to reduce signal sampling rate. In the classical setting of Compressive Sensing, only one snapshot is considered; i.e., \( T = 1 \) in (1). The goal is to recover a long vector \( x := x(1) \) with a small fraction of nonzero coordinates from the much shorter observation vector \( y := y(1) \). Since most natural signals are compressible under some basis and are well approximated by their \( K \)-sparse representations [12], this scheme, if properly justified, will reduce the necessary sampling rate beyond the limit set by Nyquist and Shannon [5], [6]. Surprisingly, for exact \( K \)-sparse signals, if \( M = O(K \log(N)) \ll N \) and the measurement matrix is generated randomly from, for example, a Gaussian distribution, we can recover \( x \) exactly in the noise-free setting by solving a linear programming task. Besides, various methods have been designed for the noisy case [13]–[17]. Along with these algorithms, rigorous theoretic analysis is provided to guarantee their effectiveness in terms of, for example, various \( l_p \)-norms of the estimation error for \( x \) [13]–[17]. However, these results offer no guarantee that we can recover the support of a sparse signal correctly.

The accurate recovery of signal support is crucial to Compressive Sensing both in theory and in practice. Since for signal recovery it is necessary to have \( K \leq M \), signal component values can be computed by solving a least squares problem once its support is obtained. Therefore, support recovery is a stronger theoretic criterion than various \( l_p \)-norms. In practice, the success of Compressive Sensing in a variety of applications relies on its ability for correct support recovery because the nonzero component indices usually have significant physical meanings. The support of temporally or spatially sparse signals reveals the timing or location for important events such as anomalies. The indices for nonzero coordinates in the Fourier domain indicate the harmonics existing in a signal [18], which is critical for tasks such as spectrum sensing for cognitive radios [19]. In compressed DNA microarrays for bio-sensing, the existence of certain target agents in the tested solution is reflected by the locations of nonvanishing coordinates, while the magnitudes are determined by their concentrations [20]–[23]. For compressive radar imaging, the sparsity constraints are usually imposed on the discretized time-frequency domain. The distance and velocity of an object have a direct correspondence to its coordinate in the time-frequency domain. The magnitude determined by coefficients of reflection is of less physical significance[24]–[26]. In sparse linear regression [27], the recovered parameter support corresponds to the few factors that explain

Manuscript received June 05, 2009; revised November 04, 2009. Current version published March 10, 2010. This work was supported by the Department of Defense under the Air Force Office of Scientific Research MURI Grant FA9550-05-1-0443, and ONR Grant N000140810849. The material in this paper was presented in part at the 47th Annual Allerton Conference on Communication, Control, and Computing, Illinois, USA, September 2009.

The authors are with the Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130 USA (e-mail: gtt2@wustl.edu; nehorai@wustl.edu).

Communicated by J. Romberg, Associate Editor for Signal Processing.

Digital Object Identifier 10.1109/TIT.2009.2039039
the data. In all these applications, the supports are physically more significant than the component values.

Our study of sparse support recovery is also motivated by the recent reformulation of the source localization problem as one of sparse spectrum estimation. In [7], the authors transform the process of source localization using sensory arrays into the task of estimating the spectrum of a sparse signal by discretizing the parameter manifold. This method exhibits super-resolution in the estimation of direction of arrival (DOA) compared with traditional techniques such as beamforming [28], Capon [29], and MUSIC [30], [31]. Since the basic model employed in [7] applies to several other important problems in signal processing (see [32] and references therein), the principle is readily applicable to those cases. This idea is later generalized and extended to other source localization settings in [8]–[10]. For source localization, the support of the sparse signal reveals the DOA of sources. Therefore, the recovery algorithm’s ability of exact support recovery is key to the effectiveness of the method. We also note that usually multiple temporal snapshots are collected, which results in a jointly sparse signal sets as in (1). In addition, since $M$ is the number of sensors while $T$ is the number of temporal samples, it is far more expensive to increase $M$ than $T$. The same comments apply to several other examples in the Compressive Sensing applications discussed in the previous paragraph, especially the compressed DNA microarrays, spectrum sensing for cognitive radios, and Compressive Sensing radar imaging.

The signal recovery problem with joint sparsity constraint [33]–[36], also termed the multiple measurement vector (MMV) problem [37]–[41], has been considered in a line of previous works. Several algorithms, among them Simultaneous Orthogonal Matching Pursuit (SOMP) [34], [37], [40]; convex relaxation [41]; $l_1$-minimization [38], [39]; and M-FOCUSS [37], are proposed and analyzed, either numerically or theoretically. These algorithms are multiple-dimension extensions of their one-dimension counterparts. Most performance measures of the algorithms are concerned with bounds on various norms of the difference between the true signals and their estimates or the closely related variants. The performance bounds usually involve the mutual coherence between the measurement matrix $A$ and the basis matrix $\Phi$ under which the measured signals $x(t)$ have a jointly sparse representation. However, with joint sparsity constraints, a natural measure of performance would be the model (1)’s potential for correctly identifying the true common support, and, hence, the algorithm’s ability to achieve this potential. As part of their research, Chen and Huo derived, in a noiseless setting, sufficient conditions on the uniqueness of solutions to (1) under $l_0$ and $l_1$ minimization. In [37], Cotter et al. numerically compared the probabilities of correctly identifying the common support by basic matching pursuit, orthogonal matching pursuit, FOCUSS, and regularized FOCUSS in the multiple-measurement setting with a range of SNRs and different numbers of snapshots.

The availability of multiple temporal samples offers several advantages to the single-sample case. As suggested by the upper bound (26) on the probability of error, increasing the number of temporal samples drives the probability of error to zero exponentially fast as long as certain condition on the inconsistency property of the measurement matrix is satisfied. The probability of error is driven to zero by scaling the SNR according to the signal dimension in [42], which is not very natural compared with increasing the samples, however. Our results also show that under some conditions increasing temporal samples is usually equivalent to increasing the number of observations for a single snapshot. The later is generally much more expensive in practice. In addition, when there is considerable noise and the columns of the measurement matrix are normalized to one, it is necessary to have multiple temporal samples for accurate support recovery as discussed in Sections IV and V-B.

Our work has several major differences compared to related work [42] and [43], which also analyze the performance bounds on the probability of error for support recovery using information theoretic tools. The first difference is in the way the problem is modeled: In [42] and [43], the sparse signal is deterministic with known smallest absolute value of the nonzero components while we consider a random signal model. This leads to the second difference: We define the probability of error over the signal and noise distributions with the measurement matrix fixed; In [42] and [43], the probability of error is taken over the noise, the Gaussian measurement matrix and the signal support. Most of the conclusions in this paper apply to general measurement matrices and we only restrict ourselves to the Gaussian measurement matrix in Section V. Therefore, although we use a similar set of theoretical tools, the exact details of applying them are quiet different. In addition, we consider a multiple measurement model while only one temporal sample is available in [42], [43]. In particular, to get a vanishing probability of error, Aeron et al. [42] require to scale the SNR according to the signal dimension, which has a similar effect to having multiple temporal measurements in our paper. Although the first two differences make it difficult to compare corresponding results in these two papers, we will make some heuristic comments in Section V.

The contribution of our work is threefold. First, we introduce a hypothesis-testing framework to study the performance for multiple support recovery. We employ well-known tools in statistics and information theory such as the Chernoff bound and Fano’s inequality to derive both upper and lower bounds on the probability of error. The upper bound we derive is for the optimal decision rule, in contrast to performance analysis for specific suboptimal reconstruction algorithms [13]–[17]. Hence, the bound can be viewed as a measure of the measurement system’s ability to correctly identify the true support. Our bounds isolate important quantities that are crucial for system performance. Since our analysis is based on measurement matrices with as few assumptions as possible, the results can be used as a guidance in system design. Second, we apply these performance bounds to other more specific situations and derive necessary and sufficient conditions in terms of the system parameters to guarantee a vanishing probability of error. In particular, we study necessary conditions for accurate source localization by the mechanism proposed in [7]. By restricting our attention to Gaussian measurement matrices, we derive a result parallel to those for classical Compressive Sensing [2], [3], namely, the number of measurements that are sufficient for signal reconstruction. Even if we adopt the probability of error as the performance criterion, we get the same bound on
from the context. Gaussian distribution for a random vector in field $F$ with mean $\mu$ and covariance matrix $\Sigma$ is represented by $F_N(\mu, \Sigma)$. Matrix variate Gaussian distribution [45] for $Y \in F_{M \times T}^N$ with mean $\Theta \in F_{N \times T}$ and covariance matrix $\Sigma \otimes \Psi$, where $\Sigma \in F_{M \times M}$ and $\Psi \in F_{T \times T}$, is denoted by $F_{N,M,T}(\Theta, \Sigma \otimes \Psi)$.

Suppose $\{f_n\}_{n=1}^{\infty}, \{g_n\}_{n=1}^{\infty}$ are two positive sequences, $f_n = o(g_n)$ means that $\lim_{n \to \infty} \frac{f_n}{g_n} = 0$. An alternative notation in this case is $g_n \gg f_n$. We use $f_n = O(g_n)$ to denote that there exists an $N \in \mathbb{N}$ and $C > 0$ independent of $N$ such that $f_n \leq C g_n$ for $n \geq N$. Similarly, $f_n = \Omega(g_n)$ means $f_n \geq C g_n$ for $n \geq N$. These simple but expedient notations introduced by G. H. Hardy greatly simplify derivations [46].

### B. Models

Next, we introduce our mathematical model. Suppose $x(t) \in F^N$, $t = 1, \ldots, T$ are jointly sparse signals with common support; that is, only a few components of $x(t)$ are nonzero and the indices corresponding to these nonzero components are the same for all $t = 1, \ldots, T$. The common support $S = \text{supp}(x(t))$ has known size $K = |S|$. We assume that the vectors $x^S(t), t = 1, \ldots, T$ formed by the nonzero components of $x(t)$ follow i.i.d. $F_N(0, 1_K)$. The measurement model is as follows:

$$y(t) = Ax(t) + w(t), \quad t = 1, 2, \ldots, T$$

where $A$ is the measurement matrix and $y(t) \in F^M$ are the measurements. The additive noise $w(t) \in F^N$ is assumed to follow i.i.d. $F_N(0, \sigma^2 I_M)$. Note that assuming unit variance for signals loses no generality since only the ratio of signal variance to noise variance appears in all subsequent analyses. In this sense, we view $1/\sigma^2$ as the signal-to-noise ratio (SNR).

Let $X = [x(1) \quad x(2) \quad \cdots \quad x(T)]$ and $Y, W$ be defined in a similar manner. Then we write the model in the more compact matrix form

$$Y = AX + W.$$  

### 2. Notations, Models, and Preliminaries

#### A. Notations

We first introduce some notations used throughout this paper. Suppose $x \in F^N$ is a column vector. We denote by $S = \text{supp}(x) \subseteq \{1, \ldots, N\}$ the support of $x$, which is defined as the set of indices corresponding to the nonzero components of $x$. For a matrix $X$, $S = \text{supp}(X)$ denotes the index set of nonzero rows of $X$. Here the underlying field $F$ can be assumed as $\mathbb{R}$ or $\mathbb{C}$. We consider both real and complex cases simultaneously. For this purpose, we denote a constant $\kappa = 1/2$ or $1$ for the real or complex case, respectively.

Suppose $S$ is an index set. We denote by $|S|$ the number of elements in $S$. For any vector column $x \in F^N$, $x_S$ is the vector in $F^{|S|}$ formed by the components of $x$ indicated by the index set $S$; for any matrix $B$, $B_S$ denotes the submatrix formed by picking the rows of $B$ corresponding to indices in $S$, while $B_S$ is the submatrix with columns from $B$ indicated by $S$. If $I$ and $J$ are two index sets, then $B_{IJ} = (B^T)^I J$, the submatrix of $B$ with rows indicated by $I$ and columns indicated by $J$.

Transpose of a vector or matrix is denoted by $\top$ while conjugate transpose by $\dagger$. $A \otimes B$ represents the Kronecker product of two matrices. For a vector $v$, $\text{diag}(v)$ is the diagonal matrix with the elements of $v$ in the diagonal. The identity matrix of dimension $M$ is $I_M$. The trace of matrix $A$ is given by $\text{tr}(A)$, the determinant by $|A|$, and the rank by $\text{rank}(A)$. Though the notation for determinant is inconsistent with that for cardinality of an index set, the exact meaning can always be understood from the context.

Bold symbols are reserved for random vectors and matrices. We use $P$ to denote the probability of an event and $E$ the expectation. The underlying probability space can be inferred as $[2], [3]$. However, our result suggests that generally it is impossible to obtain the true support accurately with only one snapshot when there is considerable noise. We also obtain a necessary condition showing that the $\log \frac{N}{K}$ term cannot be dropped in Compressive Sensing. Last but not least, in the course of studying the performance bounds we explore the eigenvalue structure of a fundamental matrix in support recovery hypothesis testing for both general measurement matrices and the Gaussian measurement ensemble. These results are of independent interest. The paper is organized as follows. In Section II, we introduce the mathematical model and briefly review the fundamental ideas in hypothesis testing. Section III is devoted to the derivation of upper bounds on the probability of error for general measurement matrices. We first derive an upper bound on the probability of error for the binary support recovery problem by employing the well-known Chernoff bound in detection theory [44] and extend it to multiple support recovery. We also study the effect of noise on system performance. In Section IV, an information theoretic lower bound is given by using the Fano’s inequality, and a necessary condition is shown for the DOA problem considered in [7]. We focus on the Gaussian ensemble in Section V. Necessary and sufficient conditions on system parameters for accurate support recovery are given and their implications discussed. The paper is concluded in Section VI.
The results we obtain for binary support recovery (4) offer insight into our second problem: the multiple support recovery. In the multiple support recovery problem we choose one among \( \binom{N}{K} \) distinct candidate supports of \( X \), which is a multiple-hypothesis testing problem

\[
\begin{align*}
H_0 : & \quad \text{supp}(X) = S_0 \\
H_1 : & \quad \text{supp}(X) = S_1 \\
& \quad \vdots \\
H_{L-1} : & \quad \text{supp}(X) = S_{L-1}.
\end{align*}
\]

(5)

C. Preliminaries for Hypothesis Testing

We now briefly introduce the fundamentals of hypothesis testing. The following discussion is based mainly on [44]. In a simple binary hypothesis test, the goal is to determine which of two candidate distributions is the true one that generates the data matrix (or vector) \( \mathbf{Y} \)

\[
\begin{align*}
H_0 : & \quad \mathbf{Y} \sim p(\mathbf{Y}|H_0) \\
H_1 : & \quad \mathbf{Y} \sim p(\mathbf{Y}|H_1).
\end{align*}
\]

(6)

There are two types of errors when one makes a choice based on the observed data \( \mathbf{Y} \). A false alarm corresponds to choosing \( H_1 \) when \( H_0 \) is true, while a miss happens by choosing \( H_0 \) when \( H_1 \) is true. The probabilities of these two types of errors are called the probability of a false alarm and the probability of a miss, which are denoted by

\[
\begin{align*}
P_F & = \mathbb{P} (\text{Choose } H_1|H_0) \\
P_M & = \mathbb{P} (\text{Choose } H_0|H_1).
\end{align*}
\]

(7)

(8)

respectively. Depending on whether one knows the prior probabilities \( \mathbb{P}(H_0) \) and \( \mathbb{P}(H_1) \) and assigns losses to errors, different criteria can be employed to derive the optimal decision rule. In this paper we adopt the probability of error with equal prior probabilities of \( H_0 \) and \( H_1 \) as the decision criterion; that is, we try to find the optimal decision rule by minimizing

\[
P_{err} = P_F \mathbb{P}(H_0) + P_M \mathbb{P}(H_1) = \frac{1}{2} P_F + \frac{1}{2} P_M.
\]

(9)

\[\ell(\mathbf{Y}) = \log \frac{p(\mathbf{Y}|H_1)}{p(\mathbf{Y}|H_0)} \geq 0 \quad \text{if } H_0 \text{ is true.} \]

(10)

where \( \log(\cdot) \) is the natural logarithm function. The probability of error associated with the optimal decision rule, namely, the likelihood ratio test (10), is a measure of the best performance a system can achieve. In many cases of interest, the simple binary hypothesis testing problem (6) is derived from a signal-generation system. For example, in a digital communication system, hypotheses \( H_0 \) and \( H_1 \) correspond to the transmitter sending digit 0 and 1, respectively, and the distributions of the observed data under the hypotheses are determined by the modulation method of the system. Therefore, the minimal probability of error achieved by the likelihood ratio test is a measure of the performance of the modulation method. For the problem addressed in this paper, the minimal probability of error reflects the measurement matrix’s ability to distinguish different signal supports.

The Chernoff bound [44] is a well-known tight upper bound on the probability of error. In many cases, the optimum test can be derived and implemented efficiently but an exact performance calculation is impossible. Even if such an expression can be derived, it is too complicated to be of practical use. For this reason, sometimes a simple bound turns out to be more useful in many problems of practical importance. The Chernoff bound, based on the moment generating function of the test statistic \( \ell(\mathbf{Y}) \) (10), provides an easy way to compute such a bound.

Define \( \mu(s) \) as the logarithm of the moment generating function of \( \ell(\mathbf{Y}) \)

\[
\mu(s) = \log \int_\infty^{\infty} \exp(\mathbf{Y}) p(\mathbf{Y}|H_0) d\mathbf{Y} = \log \int_\infty^{\infty} [p(\mathbf{Y}|H_1)]^s [p(\mathbf{Y}|H_0)]^{1-s} d\mathbf{Y}.
\]

(11)

Then the Chernoff bound states that

\[
P_F \leq \exp[\mu(s_m)] \leq \exp[\mu(s)]
\]

(12)

\[
P_M \leq \exp[\mu(s_m)] \leq \exp[\mu(s)]
\]

(13)

\[
P_{err} \leq \frac{1}{2} \exp[\mu(s_m)] \leq \frac{1}{2} \exp[\mu(s)]
\]

(14)

where \( 0 \leq s \leq 1 \) and \( s_m = \arg \min_{0 \leq s \leq 1} \mu(s) \). Note that a refined argument gives the constant \( 1/2 \) in (14) instead of 1 as obtained by direct application of (12) and (13) [44]. We use these bounds to study the performance of the support recovery problem.

We next extend to multiple-hypothesis testing the key elements of the binary hypothesis testing. The goal in a simple multiple-hypothesis testing problem is to make a choice among \( L \) distributions based on the observations

\[
\begin{align*}
H_0 : & \quad \mathbf{Y} \sim p(\mathbf{Y}|H_0) \\
H_1 : & \quad \mathbf{Y} \sim p(\mathbf{Y}|H_1) \\
& \quad \vdots \\
H_{L-1} : & \quad \mathbf{Y} \sim p(\mathbf{Y}|H_{L-1}).
\end{align*}
\]

(15)

Using the total probability of error as a decision criterion and assuming equal prior probabilities for all hypotheses, we obtain the optimal decision rule given by

\[
\hat{H} = \arg \max_{0 \leq i \leq L-1} p(\mathbf{Y}|H_i).
\]

(16)

Application of the union bound and the Chernoff bound (14) shows that the total probability of error is bounded as follows:

\[
P_{err} = \sum_{i=0}^{L-1} \mathbb{P}(\hat{H} \neq H_i|H_i) \mathbb{P}(H_i) \leq \frac{1}{2L} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \exp[\mu(s; H_i, H_j)], 0 \leq s \leq 1
\]

(17)

where \( \exp[\mu(s; H_i, H_j)] \) is the moment-generating function in the binary hypothesis testing problem for \( H_i \) and \( H_j \). Hence, we
obtain an upper bound for multiple-hypothesis testing from that for binary hypothesis testing.

III. UPPER BOUND ON PROBABILITY OF ERROR FOR NONDEGENERATE MEASUREMENT MATRICES

In this section, we apply the general theory for hypothesis testing, the Chernoff bound on the probability of error in particular, to the support recovery problems (4) and (5). We first study binary support recovery, which lays the foundation for the general support recovery problem.

A. Binary Support Recovery

Under model (3) and the assumptions pertaining to it, observations \( Y \) follow a matrix variate Gaussian distribution \([45]\) when the true support is \( S \)

\[
Y|S \sim F_{N_M,T}(0, \Sigma_S \otimes I_T)
\]

with the probability density function (pdf) given by

\[
p(Y|S) = \frac{1}{(\pi/\kappa)^{kM/2+kT/2}} \exp\left[-\kappa \text{tr} \left(Y^\dagger (\Sigma_S^{-1} - \Sigma_0^{-1}) Y\right)\right]
\]

where \( \Sigma_S = \Sigma_0 + \epsilon^2 \mathbf{I}_M \) is the common covariance matrix for each column of \( Y \). The binary support recovery problem (4) is equivalent to a linear Gaussian binary hypothesis testing problem

\[
\begin{align*}
H_0 &: Y \sim F_{N_M,T}(0, \Sigma_S \otimes I_T), \\
H_1 &: Y \sim F_{N_M,T}(0, \Sigma_S \otimes I_T).
\end{align*}
\]

From now on, for notation simplicity we will denote \( \Sigma_{S_1} \) by \( \Sigma_1 \). The optimal decision rule with minimal probability of error given by the likelihood ratio test \( \ell(Y) \) (10) reduces to

\[
\ell(Y) = -\kappa \text{tr} \left(Y^\dagger (\Sigma_1^{-1} - \Sigma_0^{-1}) Y\right) - \kappa T \log \frac{\|\Sigma_1\|}{\|\Sigma_0\|} \geq 0.
\]

To analyze the performance of the likelihood ratio test (21), we first compute the log-moment-generating function of \( \ell(Y) \) according to (11)

\[
\mu(s) = \log \left[ \kappa \text{tr} \left( \frac{\Sigma_1}{\Sigma_0} \left( 1 - s \right) \Sigma_0^{-1} \right) \right] = \kappa T \log \frac{\|\Sigma_1\|}{\|\Sigma_0\|} \geq 0,
\]

where \( \Sigma_1 \) and the columns of \( A \) are not highly correlated, for example in the case of \( A \) with i.i.d. elements, \( s_{st} \approx \frac{1}{2} \). We then take \( s = \frac{1}{2} \) in the Chernoff bounds (12), (13), and (14). Whereas the bounds obtained in this way may not be the absolute best ones, they are still valid.

As positive definite Hermitian matrices, \( H \) and \( H^{-1} \) can be simultaneously diagonalized by a unitary transformation. Suppose that the eigenvalues of \( H \) are \( \lambda_1 \geq \cdots \geq \lambda_k \), \( \sigma_1 \geq \cdots \geq \sigma_k \) and

\[
D = \text{diag}(\lambda_1, \ldots, \lambda_k, 1, \ldots, 1, \sigma_1, \ldots, \sigma_k).
\]

Then it is easy to show that

\[
\mu(1/2) = -\kappa T \log \left( \frac{D^{1/2} + D^{-1/2}}{2} \right)
\]

\[
= -\kappa T \left[ \log \left( \frac{\sqrt{\lambda_j} + 1/\sqrt{\lambda_j}}{2} \right) \right]_{j=1}^{k_0} + \left[ \log \left( \sqrt{\sigma_j} + 1/\sqrt{\sigma_j} \right) \right]_{j=k_0+1}^{k_1}.
\]

Therefore, it is necessary to count the numbers of eigenvalues of \( H \) that are greater than 1, equal to 1 and less than 1, i.e., the values of \( k_0 \) and \( k_1 \) for general nondegenerate measurement matrix \( A \). We have the following theorem on the eigenvalue structure of \( H \).

Proposition 1: For any nondegenerate measurement matrix \( A \), let \( H = \sum_1^{1/2} \Sigma_1^{-1} \Sigma_0^{-1/2}, k_0 = |S_0 \cap S_1|, k_1 = |S_0 \setminus S_1|, k_2 = |S_1 \setminus S_0| = |S_1| - k_1 \) and assume \( M \geq k_0 + k_1 \); then \( k_0 \) eigenvalues of matrix \( H \) are greater than 1, \( k_1 \) less than 1, and \( M - (k_0 + k_1) \) equal to 1.

Proof: See Appendix A.

For binary support recovery (4) with \( |S_0| = |S_1| = K \), we have \( k_0 \leq k_1 \leq k_0 \). The subscripts \( i \) and \( d \) in \( k_0 \) and \( k_1 \) are short for “intersection” and “difference,” respectively. Employing the Chernoff bounds (14) and Proposition 1, we have the following proposition.

Proposition 2: If \( M \geq 2k_0d \), the probability of error for the binary support recovery problem (4) is bounded by

\[
P_{\text{err}} \leq \frac{1}{2} \left( \sum_{j=1}^{k_0} \lambda_j \right)^{-k_0d/4}.
\]

where \( \lambda_j \) is the geometric mean of the eigenvalues of \( H = \sum_1^{1/2} \Sigma_1^{-1} \Sigma_0^{-1/2} \) that are greater than one.

Proof: According to (14) and (24), we have

\[
P_{\text{err}} \leq \frac{1}{2} \exp \left[ -\kappa T \left( \frac{1}{2} \right) \right]
\]

\[
\leq \frac{1}{2} \left( \prod_{j=1}^{k_0} \left( \sqrt{\lambda_j} \right) \prod_{j=1}^{k_1} \left( 1/\sqrt{\lambda_j} \right) \right)^{-k_0d/4}.
\]
Define \( \bar{\lambda}_{S_i, S_j} \) as the geometric mean of the eigenvalues of \( H = \sum_j S_j^{-1/2} \sum_{i,j} \alpha_{ij} S_i^{-1/2} \) that are greater than one. Then obviously we have \( \bar{\lambda}_{S_0, S_1} = \left( \prod_{j=1}^{k_1} \lambda_j^{1/k_1} \right) \). Since \( H^{-1} \) and \( \sum_j S_j^{-1/2} \sum_{i,j} \alpha_{ij} S_i^{-1/2} \) have the same set of eigenvalues, \( 1/\sigma_j, j = 1, \ldots, k_1 \) are the eigenvalues of \( \sum_j S_j^{-1/2} \sum_{i,j} \alpha_{ij} S_i^{-1/2} \) that are greater than one. We conclude that \( \bar{\lambda}_{S_1, S_0} = \left( \prod_{j=1}^{k_1} 1/\sigma_j \right)^{1/k_1} \).

Note that \( \bar{\lambda}_{S_0, S_1} \) and \( \bar{\lambda}_{S_1, S_0} \) completely determine the measurement system (3) ‘s performance in differentiating two different signal supports. It must be larger than the constant 16 for a vanishing bound when more temporal samples are taken. Once the threshold 16 is exceeded, taking more samples will drive the probability of error to 0 exponentially fast. From numerical simulations and our results on the Gaussian measurement matrix, \( \bar{\lambda}_{S_1, S_j} \) does not vary much when \( S_0, S_j \) and \( k_1 \) change, as long as the elements in the measurement matrix \( A \) are highly uncorrelated. Therefore, quite appealing to intuition, the larger the size \( k_1 \) of the difference set between the two candidate supports, the smaller the probability of error.

**B. Multiple Support Recovery**

Now we are ready to use the union bound (17) to study the probability of error for the multiple support recovery problem (5). We assume each candidate support \( S_i \) has known cardinality \( K_i \), and we have \( L = \binom{N}{K} \) such supports. Our general approach is also applicable to cases for which we have some prior information on the structure of the signal’s sparsity pattern, for example the setup in model-based Compressive Sensing [47]. In these cases, we usually have \( L \ll \binom{N}{K} \) supports, and a careful examination on the intersection pattern of these supports will give a better bound. However, in this paper we will not address this problem and will instead focus on the full support recovery problem with \( L = \binom{N}{K} \). Defining \( \tilde{\lambda} = \min_{i \neq j} \{ \bar{\lambda}_{S_i, S_j} \} \), we have the following theorem.

**Theorem 1:** If \( M \geq 2K \) and \( \tilde{\lambda} > 4 \left[ K(N-K) \right]^{1/3} \), then the probability of error for the full support recovery problem (5) with \( |S_i| = K \) and \( L = \binom{N}{K} \) is bounded by

\[
P_{err} \leq \frac{1}{2} \frac{K(N-K)}{(\tilde{\lambda}/4)^{KT}}.
\]

**Proof:** Combining the bound in Proposition 2 and (17), we have

\[
P_{err} \leq \frac{1}{2L} \sum_i \sum_{j \neq i} \sum_{k \neq j} \left[ \bar{\lambda}_{S_i, S_j} \bar{\lambda}_{S_j, S_k} \right]^{-k_{i,j}/2} \leq \frac{1}{2L} \sum_i \sum_{j \neq i} \left( \lambda_j^{-1} \right)^{-k_{i,j}T}.
\]

Here \( k_{i,j} \) depends on the supports \( S_i \) and \( S_j \). For fixed \( S_i \), the number of supports that have a difference set with \( S_i \) with cardinality \( k_{i,j} \) is \( \binom{K}{k_{i,j}} \binom{N-K}{k_{i,j}} \). Therefore, using \( \binom{K}{k_{i,j}} \leq K^{k_{i,j}} \) and \( \binom{N-K}{k_{i,j}} \leq (N-K)^{k_{i,j}} \), and the summation formula for geometric series, we obtain

\[
P_{err} \leq \frac{1}{2L} \sum_{k_{i,j}=1}^{K} \sum_{k_{i,j}=1}^{K} \left( \binom{K}{k_{i,j}} \right) \left( \binom{N-K}{k_{i,j}} \right) \frac{1}{\lambda_j^{-k_{i,j}T}} \leq \frac{1}{2L} \sum_{k_{i,j}=1}^{K} \left[ K \left( N-K \right) \frac{1}{(\tilde{\lambda}/4)^{k_{i,j}T}} \right] \]

\[
\leq \frac{1}{2} \frac{K(N-K)}{(\tilde{\lambda}/4)^{KT}}.
\]

We make several comments here. First, \( \tilde{\lambda} \) depends only on the measurement matrix \( A \) and the variance of noise \( \sigma^2 \). Compared with the results in [43], where the bounds involve the signal, we get more insight into what quantity of the measurement matrix is important in support recovery. This information is obtained by modeling the signals \( s(t) \) as Gaussian random vectors. The quantity \( \tilde{\lambda} \) effectively characterizes system (3) ‘s ability to distinguish different supports. Clearly, \( \tilde{\lambda} \) is related to the restricted isometry property (RIP), which guarantees stable sparse signal recovery in Compressive Sensing [4–6]. We discuss the relationship between RIP and \( \tilde{\lambda} \) for the special case with \( K = 1 \) at the end of Section III-C. However, a precise relationship for the general case is yet to be discovered.

Second, we observe that increasing the number of temporal samples plays two roles simultaneously in the measurement system. For one thing, it decreases the the threshold \( 4K \left( N-K \right)^{1/3} \) that \( \tilde{\lambda} \) must exceed for the bound (26) to hold. However, since \( \ln_N 4K \left( N-K \right)^{1/3} = 4 \) for fixed \( K \) and \( N \), increasing temporal samples can reduce the threshold only to a certain limit. For another, since the bound (26) is proportional to \( e^{-T k_{i,j} \mathrm{log}(\tilde{\lambda})/4} \), the probability of error turns to 0 exponentially fast as \( T \) increases, as long as \( \tilde{\lambda} > 4 \left( K(N-K) \right)^{1/3} \) is satisfied.

In addition, the final bound (26) is of the same order as the probability of error when \( k_{i,j} = 1 \). The probability of error \( P_{err} \) is dominated by the probability of error in cases for which the estimated support differs by only one index from the true support, which are the most difficult cases for the decision rule to make a choice. However, in practice we can imagine that these cases induce the least loss. Therefore, if we assign weights/costs to the errors based on \( k_{i,j} \), then the weighted probability of error or average cost would be much lower. For example, we can choose the costs to exponentially decrease when \( k_{i,j} \) increases. Another possible choice of cost function is to assume zero cost when \( k_{i,j} \) is below a certain critical number. Our results can be easily extended to these scenarios.

Finally, note that our bound (26) applies to any nondegenerate matrix. In Section V, we apply the bound to Gaussian measurement matrices. The additional structure allows us to derive more profound results on the behavior of the bound.

**C. The Effect of Noise**

In this subsection, we explore how the noise variance affects the probability of error, which is equivalent to analyzing the behavior of \( \bar{\lambda}_{S_i, S_j} \) and \( \tilde{\lambda} \) as indicated in (25) and (26).
We now derive bounds on the eigenvalues of \( H \). The lower bound is expressed in terms of the QR decomposition of a submatrix of the measurement matrix with the noise variance \( \sigma^2 \) isolated.

**Proposition 3:** For any nondegenerate measurement matrix \( A \), let \( H = \sum_{i=0}^{1/2} \sum_{j=0}^{1/2} \sum_{k=0}^{1/2} \sum_{l=0}^{1/2} k_i \), where \( k_i = A_{S_i} A_{S_i}^\dagger + \sigma^2 I_M \). We have the following:

1) if \( M \gg k_0 + k_1 \), then the sorted eigenvalues of \( H \) that are greater than 1 are lower bounded by the corresponding eigenvalues of \( I_{k_0} + \frac{1}{\sigma^2} R_{33} R_{33}^\dagger \), where \( R_{33} \) is the \( k_0 \times k_1 \) submatrix at the lower-right corner of the upper triangle matrix in the QR decomposition of \( \begin{array}{ll} [A_{S_i} A_{S_j} \cdots A_{S_l}] \end{array} \).

2) the eigenvalues of \( H \) are upper bounded by the corresponding eigenvalues of \( I_M + \frac{1}{\sigma^2} A_{S_i} A_{S_i}^\dagger A_{S_i} \); in particular, the sorted eigenvalues of \( H \) that are greater than 1 are upper bounded by the corresponding ones of \( I_{k_0} + \frac{1}{\sigma^2} A_{S_0} A_{S_0}^\dagger A_{S_0} \).

**Proof:** See Appendix B.

The importance of this proposition is twofold. First, by isolating the noise variance from the expression of matrix \( H \), this theorem clearly shows that when noise variance decreases to zero, the relatively large eigenvalues of \( H \) will blow up, which results in increased performance in support recovery. Second, the bounds provide ways to analyze special measurement matrices, especially the Gaussian measurement ensemble discussed in Section V.

We have the following corollary.

**Corollary 1:** For support recovery problems (4) and (5) with support size \( K \), suppose \( M \geq 2K \); then there exist constants \( c_1, c_2 > 0 \) that depend only on the measurement matrix \( A \) such that

\[
1 + \frac{c_2}{\sigma^2} \geq \lambda \geq 1 + \frac{c_1}{\sigma^2}.
\]

From (25) and (26), we then conclude that for any temporal sample size \( T \),

\[
\lim_{\sigma^2 \to 0} P_{err} = 0
\]

and the speed of convergence is approximately \((\sigma^2)^{k_0 T}\) for the binary and multiple cases, respectively.

**Proof:** According to Proposition 3, for any fixed \( S_i, S_j \), the eigenvalues of \( H = \sum_{i=0}^{1/2} \sum_{j=0}^{1/2} \sum_{k=0}^{1/2} \sum_{l=0}^{1/2} k_i \) that are greater than 1 are lower bounded by those of \( I_{k_0} + \frac{1}{\sigma^2} R_{33} R_{33}^\dagger \), hence, we have

\[
\lambda_{S_i, S_j} \geq \left[ I_{k_0} + 1 \frac{R_{33} R_{33}^\dagger}{\sigma^2} \right]^{1/k_0} \geq \left[ I_{k_0} + 1 \frac{R_{33} R_{33}^\dagger}{\sigma^2} \right]^{1/k_0} = \prod_{l=1}^{k_0} \left( 1 + \frac{1}{\sigma^2} r_{l}^2 \right)^{1/k_0} \geq 1 + \frac{1}{\sigma^2} \left( \prod_{l=1}^{k_0} r_{l}^2 \right)^{1/k_0}
\]

where \( r_{l} \) is the \( l \)th diagonal element of \( R_{33} \). For the second inequality we have used [48, Fact 8.11.20]. Since \( A \) is nondegenerate and \( M \geq 2K \), \([A_{S_i} A_{S_i} A_{S_i} A_{S_i}] \) is of full rank and \( r_{l}^2 > 0 \), \( 0 \leq l \leq k_0 \) for all \( S_i, S_j \). Defining \( c_1 \) as the minimal value of \( \left( \prod_{l=1}^{k_0} r_{l}^2 \right)^{1/k_0} \) over all possible support pairs \( S_i, S_j \), we then have \( c_1 > 0 \) and

\[
\lambda \geq 1 + \frac{c_1}{\sigma^2}.
\]

On the other hand, the upper bound on the eigenvalues of \( H \) yields

\[
\lambda_{S_i, S_j} \leq \left[ I_{k_0} + \frac{1}{\sigma^2} A_{S_i} A_{S_i} A_{S_i} A_{S_i} \right]^{1/k_0} \leq 1 + \frac{1}{\sigma^2} \sum_{l=1}^{k_0} \sum_{m=1}^{k_0} \left( A_{S_i} A_{S_i} A_{S_i} A_{S_i} \right)^{1/k_0} = 1 + \frac{1}{\sigma^2} \sum_{l=1}^{k_0} \sum_{m=1}^{k_0} \left( A_{S_i} A_{S_i} A_{S_i} A_{S_i} \right)^{1/k_0}
\]

Therefore, we have

\[
\lambda \leq 1 + \frac{c_2}{\sigma^2}
\]

with \( c_2 = \max_{S} |x| |x| \leq k \). This fact has been observed in [4]. Our result also shows that the optimal decision rule, which is unfortunately inefficient, is robust to noise. Another extreme case is when the noise variance \( \sigma^2 \) is very large. Then from (25) and (26), we can approximate (27) and (28) by \( e^{-k_0 T} / \sigma^2 \) and \( e^{-k_0 T} / \sigma^2 \). Therefore, the convergence exponents for the bounds are proportional to the SNR in this limiting case.

The diagonal elements of \( R_{33} \), \( r_{l}^2 \)'s, have clear meanings. Since QR factorization is equivalent to the Gram-Schmidt orthogonalization procedure, \( r_{l}^2 \) is the distance of the first column of \( A_{S_i} A_{S_i} A_{S_i} A_{S_i} \) to the subspace spanned by the columns of \( A_{S_i} A_{S_i} A_{S_i} A_{S_i} \); and \( r_{l}^2 \) is the distance of the second column of \( A_{S_i} A_{S_i} A_{S_i} A_{S_i} \) to the subspace spanned by the columns of \( A_{S_i} A_{S_i} A_{S_i} A_{S_i} \), and so on. Therefore, \( \lambda_{S_i, S_j} \) is a measure of how well the columns of \( A_{S_i} A_{S_i} A_{S_i} A_{S_i} \) can be expressed by the columns of \( A_{S_i} A_{S_i} A_{S_i} A_{S_i} \) or, in another way, a measure of the coherence between the columns of \( A_{S_i} A_{S_i} A_{S_i} A_{S_i} \). Similarly, \( \lambda \) is an indicator of the incoherence of the entire matrix \( A \) of order \( K \). To relate \( \lambda \) with the coherence, we consider the case with \( K = 1 \) and \( F = R \). By restricting our attention to matrices with unit columns, the above discussion implies that a better bound is achieved if the minimal distance of all pairs of column vectors of matrix \( A \) is maximized. Finding such a matrix \( A \) is equivalent to finding a matrix with the inner product between columns as large as possible, since the distance between two unit vectors \( u \) and \( v \) is

\[
2 - 2 |\langle u, v \rangle| = |\langle u, v \rangle| \quad \text{where} \quad |\langle u, v \rangle| = u^T v
\]

is the inner product.
between $u$ and $v$. For each integer $s$, the RIP constant $\delta_s$ is defined as the smallest number such that $[4], [5]$

$$1 - \delta_s \leq \frac{\|Ax\|^2}{\|x\|^2} \leq 1 + \delta_s, \quad |\text{supp}(x)| = s. \quad (31)$$

A direct computation shows that $\delta_2$ is equal to the minimum of the absolute values of the inner products between all pairs of columns of $A$. Hence, the requirements of finding the smallest $\delta_2$ that satisfies (31) and maximizing $\lambda$ coincide when $K = 1$. For general $K$, Milenkovic et al. established a relationship between $\delta_2$ and $\delta_K$ via Geršgorin’s disc theorem [49] and discussed them as well as some coding theoretic issues in Compressive Sensing context [50].

IV. AN INFORMATION THEORETIC LOWER BOUND ON PROBABILITY OF ERROR

In this section, we derive an information theoretic lower bound on the probability of error for any decision rule in the multiple support recovery problem. The main tool is a variant of the well-known Fano’s inequality[51]. In the variant, the average probability of error in a multiple-hypothesis testing problem is bounded in terms of the Kullback-Leibler divergence[52]. Suppose that we have a random vector or matrix $Y$ with $L$ possible densities $f_0, \ldots, f_{L-1}$. Denote the average of the Kullback-Leibler divergence between any pair of densities by

$$\beta = \frac{1}{T^2} \sum_{i,j} D_{KL}(f_i \| f_j). \quad (32)$$

Then by Fano’s inequality [43], [53], the probability of error (17) for any decision rule to identify the true density is lower bounded by

$$P_{\text{err}} \geq 1 - \frac{\beta + \log 2}{\log L}. \quad (33)$$

Since in the multiple support recovery problem (5), all the distributions involved are matrix variate Gaussian distributions with zero mean and different variances, we now compute the Kullback-Leibler divergence between two matrix variate Gaussian distributions. Suppose $f_i = \mathcal{F}_{N,M,T}(0, \Sigma_i \otimes I_T)$, $f_j = \mathcal{F}_{N,M,T}(0, \Sigma_j \otimes I_T)$, the Kullback-Leibler divergence has closed form expression

$$D_{KL}(f_i \| f_j) = \mathbb{E}_{f_i} \log \frac{f_i}{f_j} = \frac{1}{2} \mathbb{E}_{f_i} \left[ -\kappa \text{tr} \left( \Sigma_i^{-1} - \Sigma_j^{-1} \right) Y \right] + \frac{1}{2} \kappa T \log \frac{\Sigma_j}{\Sigma_i}$$

$$= \frac{1}{2} \kappa T \left[ \text{tr} (H_{i,j} - I_M) + \log \frac{\Sigma_j}{\Sigma_i} \right]$$

where $H_{i,j} = \Sigma_i^{1/2} \Sigma_j^{-1} \Sigma_i^{1/2}$. Therefore, we obtain the average Kullback-Leibler divergence (32) for the multiple support recovery problem as

$$\beta = \frac{1}{T^2} \sum_{i,j} \frac{1}{2} \kappa T \left[ \text{tr} (H_{i,j}) - M + \log \frac{\Sigma_j}{\Sigma_i} \right]$$

$$= \frac{\kappa T}{2T^2} \sum_{i,j} \left[ \text{tr} (H_{i,j}) - M \right]$$

where the $\log \frac{\Sigma_j}{\Sigma_i}$ terms all cancel out and $L = \binom{K}{2}$. Invoking the second part of Proposition 3, we get

$$\text{tr} (H_{i,j}) \leq \text{tr} \left( I_M + \frac{1}{\sigma^2} A_{S_i \backslash S_j} A_{S_j \backslash S_i}^\dagger \right)$$

$$= M + \frac{1}{\sigma^2} \sum_{1 \leq m \leq M, n \in S_i \backslash S_j} |A_{mn}|^2.$$ 

Therefore, the average Kullback-Leibler divergence is bounded by

$$\beta \leq \frac{\kappa T}{2\sigma^2 T^2} \sum_{i,j} \sum_{1 \leq m \leq M, n \in S_i \backslash S_j} \frac{1}{k_{id}}.$$ 

Due to the symmetry of the right-hand side, it must be of the form $a \sum_{i,j} \sum_{1 \leq m \leq M, n \in S_i \backslash S_j} |A_{mn}|^2 = a \|A\|_F^2$, where $\| \cdot \|_F$ is the Frobenius norm. Setting all $A_{mn} = 1$ gives

$$\frac{\kappa T}{2\sigma^2 T^2} \sum_{i,j} \sum_{1 \leq m \leq M, n \in S_i \backslash S_j} \frac{1}{k_{id}} = \frac{\kappa T}{2\sigma^2 T^2} \sum_{i,j} \sum_{1 \leq m \leq M, n \in S_i \backslash S_j} \frac{1}{k_{id}} = \frac{K}{N}. \quad (34)$$

Therefore, we get $a = \frac{\kappa T (N - K)}{2\sigma^2 N^2}$ using the mean expression for hypergeometric distribution

$$\sum_{k_{id} = 1}^{K} \binom{K}{k_{id}} (N-k_{id}) = K (N-K) \quad (34)$$

Hence, we have

$$\beta \leq \frac{\kappa T (N - K)}{2\sigma^2 N^2} \|A\|_F^2.$$ 

Therefore, the probability of error is lower bounded by

$$P_{\text{err}} \geq 1 - \frac{\kappa T K (N - K)}{2\sigma^2 N^2} \log \frac{2}{\log L}.$$ 

We conclude with the following theorem.
**Theorem 2:** For multiple decision support recovery problem (5), the probability of error for any decision rule is lower bounded by

\[
P_{\text{err}} \geq 1 - \frac{\kappa T K}{2\pi^2 \log N} \frac{(1 - \frac{K}{N}) ||A||_F^2}{K} + o(1) .
\]  

(35)

Each term in bound (35) has clear meanings. The Frobenius norm of measurement matrix \( ||A||_F \) is total gain of system (2). Since the measured signal is \( K \)-sparse, only a fraction of the gain plays a role in the measurement, and its average over all possible \( K \)-sparse signals is \( \frac{K}{N} ||A||_F^2 \). While an increase in signal energy enlarges the distances between signals, a penalty term \( (1 - \frac{K}{N}) \) is introduced because we now have more signals. The term \( \log L = \log \binom{N}{K} \) is the total uncertainty or entropy of the support variable \( S \), since we impose a uniform prior on it. As long as \( K \leq \frac{N}{3} \), increasing \( K \) increases both the average gain exploited by the measurement system, and the entropy of the support variable \( S \). The overall effect, quite counterintuitively, is a decrease of the lower bound in (35). Actually, the term involving \( K \), \( \frac{K}{N} \left( \frac{1 - K}{K} \right) \log \binom{N}{K} \), is approximated by an increasing function \( \frac{k}{N} \log \binom{N}{k} \) with \( k = \frac{K}{N} \) and the binary entropy function \( H(\alpha) = -\alpha \log \alpha - (1 - \alpha) \log (1 - \alpha) \). The reason for the decrease of the bound is that the bound only involves the effective SNR without regard to any inner structure of \( A \) (e.g., the incoherence) and the effective SNR increases with \( K \). To see this, we compute the effective SNR as \( E_{\text{eff}} \left( \frac{||A||_F^2}{M} \right) = \frac{K}{N} \left( \frac{1 - K}{K} \right)^{\frac{1}{2}} \). If we scale down the effective SNR through increasing the noise energy \( \sigma^2 \) by a factor of \( K \), then the bound is strictly increasing with \( K \).

The above analysis suggests that the lower bound (35) is weak as it disregards any incoherence property of the measurement matrix \( A \). For some cases, the bound reduces to \( 2\sigma^2 \log \frac{N}{K} \) (refer to Corollary 2, Theorem 3 and 4) and is less than \( K \) when the noise level or \( K \) is relatively large. Certainly recovering the support is not possible with fewer than \( K \) measurements. The bound is loose also in the sense that when \( T \), \( ||A||_F^2 \), or the SNR \( 1/\sigma^2 \) is large enough the bound becomes negative, but then there is noise, perfect support recovery is generally impossible.

While the original Fano’s inequality

\[
H (P_{\text{err}}) + P_{\text{err}} \log (L - 1) \geq H (S|Y)
\]

is tight in some cases\[51\], the adoption of the average divergence (32) as an upper bound on the mutual information \( I (S, Y) \) between the random support \( S \) and the observation \( Y \) reduces the tightness (see the proof of \[54, (33)]). Due to the difficulty of computing \( H (S|Y) \) and \( I (S, Y) \) analytically, it is not clear whether a direct application of (36) results in a significantly better bound.

Despite its drawbacks, the bound (35) identifies the importance of the gain \( ||A||_F \) of the measurement matrix, a quantity usually ignored in, for example, Compressive Sensing. We can also draw some interesting conclusions from (35) for measurement matrices with special properties. In particular, in the following corollary, we consider measurement matrices with rows or columns normalized to one. The rows of a measurement matrix are normalized to one in sensor network scenario (SNET) where each sensor is power limited while the columns are sometimes normalized to one in Compressive Sensing (refer to [42] and references therein).

**Corollary 2:** In order to have a probability of error \( P_{\text{err}} < \varepsilon \) with \( 0 < \varepsilon < 1 \), the number of measurements must satisfy

\[
MT \geq \frac{(1 - \varepsilon) \frac{2\sigma^2 K \log \frac{N}{K}}{K}}{\frac{2\pi^2 \log \binom{N}{K}}{K}} + o(1)
\]

(37)

if the rows of \( A \) have unit norm; and

\[
T \geq \frac{(1 - \varepsilon) \frac{\frac{2\sigma^2 \log \frac{N}{K}}{K}}{2\pi^2 \log \binom{N}{K}}}{\frac{2\pi^2 \log \binom{N}{K}}{K}} + o(1)
\]

(38)

if the columns of \( A \) have unit norm.

Note that the necessary condition (37) has the same critical quantity as the sufficient condition in Compressive Sensing. The inequality in (38) is independent of \( M \). Therefore, if the columns are normalized to have unit norm, it is necessary to have multiple temporal measurements for a vanishing probability of error. Refer to Theorem 3 and 4 and discussions following them.

In the work of \[7\], each column of \( A \) is the array manifold vector function evaluated at a sample of the DOA parameter. The implication of the bound (35) for optimal design is that we should construct an array whose geometry leads to maximal \( ||A||_F^2 \). However, under the narrowband signal assumption and narrowband array assumption \[55\], the array manifold vector for isotropic sensor arrays always has norm \( \sqrt{M} \) \[56\], which means that \( ||A||_F^2 = MN \). Hence, in this case, the probability of error is always bounded by

\[
P_{\text{err}} \geq 1 - \frac{TK}{2\pi^2 \log \binom{N}{K}} \frac{MN}{2\pi^2 \log \binom{N}{K}} + o(1) .
\]

(39)

Therefore, we have the following theorem.

**Theorem 3:** Under the narrowband signal assumption and narrowband array assumption, for an isotropic sensor array in the DOA estimation scheme proposed in \[7\], in order to let the probability of error \( P_{\text{err}} < \varepsilon \) with \( 0 < \varepsilon < 1 \) for any decision rule, the number of measurements must satisfy the following:

\[
MT \geq \frac{(1 - \varepsilon) \frac{2\sigma^2 \log \frac{N}{K}}{K}}{\frac{2\pi^2 \log \binom{N}{K}}{K}} + o(1)
\]

\[
\geq (1 - \varepsilon) 2\sigma^2 \log \frac{N}{K} \frac{N}{K} + o(1) .
\]

(40)

We comment that the same lower bound applies to Fourier measurement matrix (not normalized by \( 1/\sqrt{M} \)) due to the same line of argument. We will not explicitly present this result in the current paper.

Since in radar and sonar applications the number of targets \( K \) is usually small, our result shows that the number of samples is lower bounded by \( \log N \). Note that \( N \) is the number of intervals we use to divide the whole range of DOA; hence, it is a measure
of resolution. Therefore, the number of samples only needs to increase in the logarithm of \( N \), which is very desirable. The symmetric roles played by \( M \) and \( T \) are also desirable since \( M \) is the number of sensors and is expensive to increase. As a consequence, we simply increase the number of samples to achieve a desired probability of error. In addition, unlike the upper bound of Theorem 1, we do not need to assume that \( M \geq 2K \) in Theorem 2 and 3. Actually, Malioutov \textit{et al.} made the empirical observation that \( \ell_2 \)-SVD technique can resolve \( M-1 \) sources if they are well separated \cite{7}. Theorem 3 still applies to this extreme case.

Analysis of support recovery problem with measurement matrix obtained from sampling a manifold has considerable complexity compared with the Gaussian case. For example, it presents significant challenge to estimate \( \lambda_{S_1,S_2} \) in the DOA problem except for a few special cases that we discuss in \cite{57}. As aforementioned, unlike the Gaussian case, \( \lambda_{S_1,S_2} \) for uniform linear arrays varies greatly with \( S_1 \) and \( S_2 \). Therefore, even if we can compute \( \lambda_{S_1,S_2} \), replacing it with \( \lambda \) in the upper bound of Theorem 1 would lead to a very loose bound. On the other hand, the lower bound of Theorem 2 only involves the Frobenius norm of the measurement matrix, so we apply it to the DOA problem effortlessly. However, the lower bound is weak as it does not exploit any inner structure of the measurement matrix.

Donoho \textit{et al.} considered the recovery of a “sparse” wide-band signal from narrow-band measurements \cite{58}, \cite{59}, a problem with essentially the same mathematical structure when we sample the array manifold uniformly in the wave number domain instead of the DOA domain. It was found that the spectral norm of the product of the band-limiting and time-limiting operators is crucial to stable signal recovery measured by the \( \ell_2 \) norm. In \cite{58}, Donoho and Stark bounded the spectral norm using the Frobenius norm, which leads to the well-known uncertainty principle. The authors commented that the uncertainty principle condition demands an extreme degree of sparsity for the signal. However, this condition can be relaxed if the signal support are widely scattered. In \cite{7}, Malioutov \textit{et al.} also observed from numerical simulations that the \( \ell_2 \)-SVD algorithm performs much better when the sources are well separated than when they are located close together. In particular, they observed that presence of bias is mitigated greatly when sources are far apart. Donoho and Logan \cite{59} explored the effect of the scattering of the signal support by using “the analytic principle of the large sieve.” They bounded the spectral norm for the limiting operator by the maximum Nyquist density, a quantity that measures the degree of scattering of the signal support. We expect that our results can be improved in a similar manner. The challenges include using support recovery as a performance measure, incorporating multiple measurements, as well as developing the whole theory within a probabilistic framework.

V. SUPPORT RECOVERY FOR THE GAUSSIAN MEASUREMENT ENSEMBLE

In this section, we refine our results in previous sections from general nondegenerate measurement matrices to the Gaussian ensemble. Unless otherwise specified, we always assume that the elements in a measurement matrix \( A \) are i.i.d. samples from unit variance real or complex Gaussian distributions. The Gaussian measurement ensemble is widely used and studied in Compressive Sensing \cite{2}–\cite{6}. The additional structure and the theoretical tools available enable us to derive deeper results in this case. In this section, we assume general scaling of \((N,M,K,T)\). We do not find in our results a clear distinction between the regime of sublinear sparsity and the regime of linear sparsity as the one discussed in \cite{43}.

We first show two corollaries on the eigenvalue structure for the Gaussian measurement ensemble. Then we derive sufficient and necessary conditions in terms of \( M,N,K \) and \( T \) for the system to have a vanishing probability of error.

A. Eigenvalue Structure for a Gaussian Measurement Matrix

First, we observe that a Gaussian measurement matrix is non-degenerate with probability one, since any \( p \leq M \) random vectors \( \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_p \), from \( F\mathcal{N}(0,\Sigma) \) with \( \Sigma \in \mathbb{R}^{M\times M} \) positive definite are linearly independent with probability one (refer to \cite[Theorem 3.2.1]{45}). As a consequence, we have the following.

Corollary 3: For Gaussian measurement matrix \( A \), let \( H = \mathbf{\Sigma}_0^{1/2} \mathbf{\Sigma}_1^{-1/2} \Sigma_1^{1/2} \), \( k_1 = |S_0 \cap S_1| \), \( k_0 = |S_0 \setminus S_1| = |S_1 \setminus S_0| - k_1 \), \( k_3 = |S_1 \setminus S_0| = |S_0 \setminus S_1| - k_1 \). If \( M \geq k_0 + k_1 \), then with probability one, \( k_0 \) eigenvalues of matrix \( H \) are greater than 1, \( k_3 \) less than 1, and \( M = (k_0 + k_1) \) equal to 1.

We refine Proposition 3 based on the well-known QR factorization for Gaussian matrices \cite{45}, \cite{60}.

Corollary 4: With the same notations as in Corollary 3, then with probability one, we have:

1) if \( M \geq k_0 + k_1 \), then the sorted eigenvalues of \( H \) that are greater than 1 are lower bounded by the corresponding ones of \( I_{k_0} + \frac{1}{\sigma^2} \mathbf{R}_{33} \mathbf{R}_{33}^\dagger \), where the elements of \( \mathbf{R}_{33} = (r_{mm})_{k_0 \times k_0} \) satisfy

\[ 2\pi^2 r_{mm}^2 \sim \frac{\lambda_{S_0 \setminus S_1}}{2\pi(M - k_1 - k_0 - 1 + 1)}, \quad 1 \leq m \leq k_0 \]

\[ r_{mm} \sim F\mathcal{N}(0,1), \quad 1 \leq m < n \leq k_0. \]

2) the eigenvalues of \( H \) are upper bounded by the corresponding eigenvalues of \( I_M + \frac{1}{\sigma^2} \mathbf{A}_{S_1 \setminus S_0} \mathbf{A}_{S_1 \setminus S_0}^\dagger \); in particular, the sorted eigenvalues of \( H \) that are greater than 1 are upper bounded by the corresponding ones of \( I_{k_0} + \frac{1}{\sigma^2} \mathbf{A}_{S_1 \setminus S_0}^{\dagger} \mathbf{A}_{S_1 \setminus S_0} \).

Now with the distributions on the elements of the bounding matrices, we can give sharp estimate on \( \lambda_{S_1,S_2} \). In particular, we have the following proposition.

Proposition 4: For Gaussian measurement matrix \( A \), suppose \( S_1 \) and \( S_2 \) are a pair of distinct supports with the same size \( K \). Then we have

\[ 1 + \frac{M}{\sigma^2} \geq \lambda_{S_1,S_2} \geq 1 + \frac{M - K - k_1}{\sigma^2}. \]

Proof: We copy (29), (30) on \( \lambda_{S_1,S_2} \) here

\[ 1 + \frac{1}{\sigma^2 k_1} \sum_{1 \leq m \leq M \atop m \in S_1 \setminus S_2} |A_{mm}|^2 \geq \lambda_{S_1,S_2} \geq 1 + \frac{1}{\sigma^2} \left( \prod_{m=1}^{k_1} |r_{mm}|^2 \right)^{1/k_1}. \]
The proof then reduces to the computation of two expectations, one of which is trivial
\[ E \left( \frac{1}{\sigma^2} \sum_{m \in S_0 \setminus S_1} |A_{mn}|^2 \right) = \frac{M}{\sigma^2}. \]

Next, the independence of the \( r_{nm} \)'s and the convexity of exponential functions together with Jensen’s inequality yield
\[
E \left( \frac{1}{\sigma^2} \left( \prod_{n=1}^{k_d} r_{nm}^2 \right)^{1/k_d} \right) = \frac{1}{2\kappa\sigma^2} \exp \left[ \frac{1}{k_d} \sum_{n=1}^{k_d} \log \left( 2\kappa r_{mn}^2 \right) \right] \geq \frac{1}{2\kappa\sigma^2} \exp \left[ \frac{1}{k_d} \sum_{n=1}^{k_d} E \log \left( 2\kappa r_{mn}^2 \right) \right].
\]

Since \( 2\kappa r_{nm}^2 \sim \chi^2_{2k_d(M-K-n+1)} \), the expectation of logarithm is \( E \log \left( 2\kappa r_{mn}^2 \right) = \log 2 + \psi(\kappa(M-K-n+1)) \), where \( \psi(z) = \Gamma'(z)/\Gamma(z) \) is the digamma function. Note that \( \psi(z) \) is increasing and satisfies \( \psi(z+1) \geq \log z \). Therefore, we have
\[
E \left( \frac{1}{\sigma^2} \left( \prod_{n=1}^{k_d} r_{nm}^2 \right)^{1/k_d} \right) \geq \frac{1}{2\kappa\sigma^2} \exp \left[ \log 2 + \frac{1}{k_d} \sum_{n=1}^{k_d} \psi(\kappa(M-K-n+1)) \right] \geq \frac{1}{2\kappa\sigma^2} \exp \left[ \psi(\kappa(M-K-k_d+1)) \right] \geq \frac{1}{2\kappa\sigma^2} \exp \left[ \log(\kappa(M-K-k_d)) \right] \geq \frac{M-K-k_d}{\sigma^2}. \]

The expected value of the critical quantity \( \bar{\lambda}_{s_i,s_j} \) lies between \( 1 + \frac{M-2K}{\sigma^2} \) and \( 1 + \frac{M}{\sigma^2} \), linearly proportional to \( M \). Note that in conventional Compressive Sensing, the variance of the elements of \( A \) is usually taken to be \( \frac{1}{M} \), which is equivalent to scaling the noise variance \( \sigma^2 \) to \( M\sigma^2 \) in our model. The resultant \( \bar{\lambda}_{s_i,s_j} \) is then centered between \( 1 + \frac{1-2K}{\sigma^2} \) and \( 1 + \frac{1}{\sigma^2} \).

**B. Necessary Condition**

One fundamental problem in Compressive Sensing is how many samples should the system take to guarantee a stable reconstruction. Although many sufficient conditions are available, nontrivial necessary conditions are rare. Besides, in previous works, stable reconstruction has been measured in the sense of \( l_1 \) norms between the reconstructed signal and the true signal. In this section, we derive two necessary conditions on \( M \) and \( T \) in terms of \( N \) and \( K \) in order to guarantee respectively that, first, \( E[P_{\text{err}}] \) turns to zero and, second, for majorization realizations of \( A \), the probability of error vanishes. More precisely, we have the following theorem.

**Theorem 4:** In the support recovery problem (5), for any \( \varepsilon, \delta > 0 \), a necessary condition of \( E[P_{\text{err}}] < \varepsilon \) is
\[
MT \geq \left( 1 - \varepsilon \right) \frac{2\kappa^2 \log \left( \frac{N}{K} \right)}{ \kappa K \left( 1 - \frac{K}{N} \right)} + o(1) \quad \text{(41)}
\]
\[
\geq \left( 1 - \varepsilon \right) \frac{2\sigma^2}{K} \log \frac{N}{K} + o(1) \quad \text{(42)}
\]

and a necessary condition of \( P \{ P_{\text{err}}(A) \leq \varepsilon \} \geq 1 - \delta \) is
\[
MT \geq \left( 1 - \varepsilon - \delta \right) \frac{2\kappa^2 \log \left( \frac{N}{K} \right)}{ \kappa K \left( 1 - \frac{K}{N} \right)} + o(1) \quad \text{(43)}
\]
\[
\geq \left( 1 - \varepsilon - \delta \right) \frac{2\sigma^2}{K} \log \frac{N}{K} + o(1). \quad \text{(44)}
\]

**Proof:** Equation (35) and \( E\|A\|_{l_2}^2 = \sum_{m \in d} E|A_{mn}|^2 = MN \) give
\[
E[P_{\text{err}}] \geq 1 - \frac{\kappa T N}{2\sigma^2 \log \left( \frac{N}{K} \right)} + \frac{MN}{2\sigma^2 \log \left( \frac{N}{K} \right)} + o(1). \quad \text{(45)}
\]

Hence, \( E[P_{\text{err}}] < \varepsilon \) entails (41) and (42).

Denote by \( E \) the event \( \{ A : P_{\text{err}}(A) \leq \varepsilon \} \); then \( P \{ E^c \} \leq \delta \) and we have
\[
E[P_{\text{err}}] = \int_E P_{\text{err}}(A) + \int_{E^c} P_{\text{err}}(A) \leq \varepsilon P(E) + P(E^c) \leq \varepsilon + \delta.
\]

Therefore, from the first part of the theorem, we obtain (43) and (44). \( \Box \)

We compare our results with those of [43] and [42]. As we mentioned in the introduction, the differences in problem modeling and the definition of the probability of error make a direct comparison difficult. We first note that [43, Theorem 2] is established for the restricted problem where it is known \textit{a priori} that all nonzero components in the sparse signal are equal. Because the set of signal realizations with equal nonzero components is a rare event in our signal model, it is not fitting to compare our result with the corresponding one in [43] by computing the distribution of the smallest on-support element, e.g., the expectation. Actually, the square of the smallest on-support element for the restricted problem, \( M^2(\beta) \) (or \( \beta \) in [42]), is equivalent to the signal variance in our model: both are measures of the signal energy. If we take into account the noise variance and replace \( M^2(\beta) \) (or \( \beta \)SNR in [42]) with \( 1/\sigma^2 \), the necessary conditions in these papers coincide with ours when only one temporal sample is available. Our result shows that as far as support recovery is concerned, one cannot avoid the \( \log \frac{N}{K} \) term when only given one temporal sample. Worse, for conventional Compressive Sensing with a measurement matrix generated from a Gaussian random variable with variance \( 1/M \), the necessary condition becomes
\[
T \geq \frac{2\kappa^2 \log \left( \frac{N}{K} \right)}{ \kappa K \left( 1 - \frac{K}{N} \right)} + o(1)
\]
\[
\geq \frac{2\sigma^2}{K} \log \frac{N}{K} + o(1)
\]
which is independent of $M$. Therefore, when there isconsiderable noise ($\sigma^2 > \kappa/(2 \log K$)), it is impossible to have a vanishing $EP_{\text{err}}$, no matter how large an $M$ one takes. Basically this situation arises because while taking more samples, one scales down the measurement gains $A_{\text{gain}}$, which effectively reduces the SNR and thus is not helpful in support recovery. As discussed, after Theorem 3, $\log (N/K)$ is the uncertainty of the support variable $S$, and $\log (N/K)$ actually comes from it. Therefore, it is no surprise that the number of samples is determined by this quantity and cannot be made independent of it.

C. Sufficient Condition

We derive a sufficient condition in parallel with sufficient conditions in Compressive Sensing. In Compressive Sensing, when only one temporal sample is available, $M = \Omega(K \log N)$ is enough for stable signal reconstruction for the majority of the realizations of measurement matrix $A$ from a Gaussian ensemble with variance $\frac{1}{N}$. As shown in the previous subsection, if we take the probability of error for support recovery as a performance measure, it is impossible in this case to recover the support with a vanishing probability of error unless the noise is small. Therefore, we consider a Gaussian ensemble with unit variance. We first establish a lemma to estimate the lower tail of the distribution for $\bar{\lambda}_{S_i,S_j}$. We have shown that the $E(\bar{\lambda}_{S_i,S_j})$ lies between $1 + \frac{M-2K}{\sigma^2}$ and $1 + \frac{M}{\sigma^2}$. When $\gamma$ is much less than $1 + \frac{M-2K}{\sigma^2}$, we expect that $P\{\bar{\lambda}_{S_i,S_j} \leq \gamma\}$ decays quickly. More specifically, we have the following large deviation lemma:

**Lemma 1:** Suppose that $\gamma = \frac{1}{3} \frac{M-2K}{\sigma^2}$. Then there exists constant $c > 0$ such that for $M = 2K$ sufficiently large, we have

$$P\{\lambda_{S_i,S_j} \leq \gamma\} \leq \exp[-c(M-2K)].$$

This large deviation lemma together with the union bound yield the following sufficient condition for support recovery.

**Theorem 5:** Suppose that

$$M = \Omega(K \log N)$$

and

$$\kappa T \log \frac{M}{\sigma^2} \gg \log[K(N-K)].$$

Then given any realization of measurement matrix $A$ from a Gaussian ensemble, the optimal decision rule (16) for multiple support recovery problem (5) has a vanishing $P_{\text{err}}$ with probability turning to one. In particular, if $M = \Omega(K \log N)$ and

$$T \gg \frac{\log N}{\log \log N},$$

then the probability of error turns to zero as $N$ turns to infinity.

**Proof:** Denote $\gamma = \frac{1}{3} \frac{M-2K}{\sigma^2}$. Then according to the union bound, we have

$$P\{\bar{\lambda} \leq \gamma\} = P\left\{\bigcup_{S_i \neq S_j} [\lambda_{S_i,S_j} \leq \gamma]\right\} \leq \sum_{S_i \neq S_j} P\{\lambda_{S_i,S_j} \leq \gamma\}.$$
recovery averaged over the Gaussian ensemble. The bound and its approximation are applicable only to the special case with $K = 1$ and involve complex integrals. The authors obtained a useful SNR threshold as a function of $M,N$ and $K$ through the analytical bound. Note that our bounds are valid for general $K$ and have a simple form. Besides, most of our derivation is conditioned on a realization of the Gaussian measurement ensemble. The conditioning makes more sense than averaging because in practice we usually make observations with fixed sensing matrix and varying signals and noise. The result of Theorem 5 also exhibits several interesting properties in the general case. Compared with the necessary condition (43) and (44), the asymmetry in the sufficient condition is even more desirable in most cases because of the asymmetric cost associated with sensors and temporal samples. Once the threshold $K \log \frac{N}{K}$ of $M$ is exceeded, we can achieve a desired probability of error by taking more temporal samples. If we were concerned only with the total number of samples, we would minimize $MT$ subject to the constraints (46) and (47) to achieve a given level of probability of error. However, in applications for which timing is important, one has to increase the number of sensors to reduce $P_{\text{err}}$ to a certain limit.

The sufficient condition (46), (47), and (48) is separable in the following sense. We observe from the proof that the requirement $M = \Omega(K \log \frac{N}{K})$ is used only to guarantee that the randomly generated measurement matrix is a good one in the sense that its incoherence $\lambda$ is sufficiently large, as in the case of Compressive Sensing. It is in Lemma 1 that we use the Gaussian ensemble assumption. If another deterministic construction procedure (for attempts in this direction, see [62]) or random distribution give measurement matrix with better incoherence $\lambda$, it would be possible to reduce the orders for both $M$ and $T$.

VI. CONCLUSION

In this paper, we formulated the support recovery problems for jointly sparse signals as binary and multiple-hypothesis testings. Adopting the probability of error as the performance criterion, the optimal decision rules are given by the likelihood ratio test and the maximum a posteriori probability estimator. The latter reduces to the maximum likelihood estimator when equal prior probabilities are assigned to the supports. We then employed the Chernoff bound and Fano’s inequality to derive bounds on the probability of error. We discussed the implications of these bounds at the end of Sections III-B, III-C, IV, V-B and V-C, in particular when they are applied to the DOA estimation problem considered in [7] and Compressive Sensing with a Gaussian measurement ensemble. We derived sufficient and necessary conditions for Compressive Sensing using Gaussian measurement matrices to achieve a vanishing probability of error in both the mean and large probability senses. These conditions show the necessity of considering multiple temporal samples. The symmetric and asymmetric roles played by the spatial and temporal samples and their implications in system design were discussed. For Compressive Sensing, we demonstrated that it is impossible to obtain accurate signal support with only one temporal sample if the variance for the Gaussian measurement matrix scales with $1/M$ and there is considerable noise.

This research on support recovery for jointly sparse signals is far from complete. Several questions remain to be answered. First, we notice an obvious gap between the necessary and sufficient conditions even in the simplest case with $K = 1$. Better techniques need to be introduced to refine the results. Second, as in the case for RIP, computation of the quantity $\lambda$ for an arbitrary measurement matrix is extremely difficult. Although we derive large derivation bounds on $\lambda$ and compute the expected value for $\lambda_{S\cup S'}$ for the Gaussian ensemble, its behaviors in both the general and Gaussian cases require further study. Its relationship with RIP also needs to be clarified. Finally, our lower bound derived from Fano’s inequality identifies only the effect of the total gain. The effect of the measurement matrix’s incoherence is elusive. The answers to these questions will enhance our understanding of the measurement mechanism (2).

APPENDIX A

PROOF OF PROPOSITION 1

In this proof, we focus on the case for which both $k_0 \neq 0$ and $k_1 \neq 0$. Other cases have similar and simpler proofs. The eigenvalues of $H$ satisfy $|\lambda_{1/2} - H| = 0$, which is equivalent to $|\lambda_{1/2} - \Sigma_0 - \Sigma_1| = 0$. The substitution $\lambda = \mu + 1$ defines

$g(\mu) = |(\mu + 1) \Sigma_1 - \Sigma_0| = |\mu \Sigma_1 - (\Sigma_0 - \Sigma_1)|$.

The following algebraic manipulation

$G \triangleq \Sigma_0 - \Sigma_1$

$= A_{S_0} A_{S_0}^\dagger - A_{S_1} A_{S_1}^\dagger$

$= [A_{S_0 \setminus S_1} A_{S_0 \setminus S_1}^\dagger + A_{S_1 \setminus S_0} A_{S_1 \setminus S_0}^\dagger]$

$= [A_{S_0 \setminus S_1} A_{S_0 \setminus S_1}^\dagger + A_{S_1 \setminus S_0} A_{S_1 \setminus S_0}^\dagger]$

$= A_{S_0 \setminus S_1} A_{S_0 \setminus S_1}^\dagger + A_{S_1 \setminus S_0} A_{S_1 \setminus S_0}^\dagger$

leads to

$g(\mu) = |\mu \Sigma_1 - G|$

$= |\Sigma_1^{1/2} |\mu M - \Sigma_1^{-1/2} G \Sigma_1^{-1/2} |^{1/2}$.

Therefore, to prove the theorem, it suffices to show that $\Sigma_1^{1/2} G \Sigma_1^{-1/2}$ has $k_0$ positive eigenvalues, $k_1$ negative eigenvalues and $M - (k_0 + k_1)$ zero eigenvalues or, put another way, $\Sigma_1^{1/2} G \Sigma_1^{-1/2}$ has inertia $(k_0, k_1, M - (k_0 + k_1))$. Sylvester’s law of inertia ([49, Theorem 4.5.8, p. 223]) states that the inertia of a symmetric matrix is invariant under congruence transformations. Hence, we need only to show that $G$ has inertia $(k_0, k_1, M - (k_0 + k_1))$. Clearly $G = PQ^\dagger$ with $P = [A_{S_0 \setminus S_1} A_{S_0 \setminus S_1}^\dagger]$ and $Q = [A_{S_1 \setminus S_0} A_{S_1 \setminus S_0}^\dagger]$. To find the number of zero eigenvalues of $G$, we calculate the rank of $G$. The nondegenerateness of measurement matrix $A$ implies that $\text{rank}(P) = \text{rank}(Q) = k_0 + k_1$, therefore, from rank inequality ([49, Theorem 0.4.5, p. 13]),

$\text{rank}(P) + \text{rank}(Q^\dagger) = (k_0 + k_1)$

$\leq \text{rank}(PQ^\dagger)$

$\leq \min\{\text{rank}(P), \text{rank}(Q^\dagger)\}$

we conclude that $\text{rank}(G) = k_0 + k_1$. 

Authorized licensed use limited to: WASHINGTON UNIVERSITY LIBRARIES. Downloaded on March 11,2010 at 17:22:28 EST from IEEE Xplore. Restrictions apply.
To count the number of negative eigenvalues of $G$, we use the Jacobi-Sturm rule ([63, Theorem A.1.4, p. 320]), which states that for an $M \times M$ symmetric matrix whose $j$th leading principal minor has determinant $d_j$, $j = 1, \ldots, M$, the number of nonnegative eigenvalues is equal to the number of sign changes of sequence $\{1, d_1, \ldots, d_M\}$. We consider only the first $k_0 + k_1$ leading principal minors, since higher order minors have determinant 0.

Suppose $J = \{1, \ldots, k_0 + k_1\}$ is an index set. Without loss of generality, we assume that $P^T$ is nonsingular. Applying QL factorization (one variation of QR factorization, see [64]) to matrix $P^T$, we obtain $P^T = QL$, where $O$ is an orthogonal matrix, $QQ^T = I_{k_0 + k_1}$, and $L = (l_{ij})_{(k_0 + k_1) \times (k_0 + k_1)}$ is an lower triangular matrix. The diagonal entries of $L$ are nonzero because $P^T$ is nonsingular. The partition of $L$ into

$$
L = \begin{bmatrix} L_1 & L_2 \\
0 & L_3 \end{bmatrix}
$$

with $L_1 \in \mathbb{F}^{(k_0 + k_1) \times k_0}$, $L_2 \in \mathbb{F}^{(k_0 + k_1) \times k_1}$, and $L_3 \in \mathbb{F}^{k_1 \times k_1}$ implies

$$
G^T = P^T(QT)^T = Q\begin{bmatrix} L_1^T & L_2^T \\
- & L_3 \end{bmatrix} Q^T.
$$

Again using the invariance property of inertia under congruence transformation, we focus on the leading principal minors of

$$
U \triangleq \begin{bmatrix} L_1 & L_2 \\
0 & L_3 \end{bmatrix}.
$$

Suppose $J = \{1, \ldots, j\}$. For $1 \leq j \leq k_0$, from the lower triangularity of $L$, it is clear that

$$
|\left(U^j\right)| = |(L_1^j)^T| = \prod_{i=1}^{j} |l_{ii}|^2 > 0.
$$

For $k_0 + 1 \leq j \leq k_0 + k_1$, suppose $J_0 = \{1, \ldots, k_0\}$ and $J_1 = \{1, \ldots, j - k_0\}$. We then have

$$
|U^j| = \left|(L_1)^j_{k_0}ight|^2 \left|(L_3)^j_{k_1}ight|^2 - \left|(L_2)^j_{k_1}ight|^2
= (-1)^j \prod_{i=1}^{j} |l_{ii}|^2.
$$

Therefore, the sequence $1, d_1, d_2, \ldots, d_{k_0 + k_1}$ has $k_1$ sign changes, which implies that $G^T$—hence $G$—has $k_1$ negative eigenvalues. Finally, we conclude that the theorem holds for $H$. □

APPENDIX B
PROOF OF PROPOSITION 3

We first prove the first claim. From the proof of Proposition 1, it suffices to show that the sorted positive eigenvalues of

$$
\sum_{i=1}^{2} G_{i}^2 \sum_{j=1}^{\frac{k_1+1}{2}} G_{j}^2 \text{ are greater than those of } \frac{1}{2} R_{31} R_{32}^T,
$$

where $G = A_{S_0 \setminus S_1} A_{S_1 \setminus S_1}^T - A_{S_0 \setminus S_0} A_{S_1 \setminus S_1}^T$, $S_0$ is an upper triangular matrix with partitions $Q_1 \in \mathbb{F}^{M \times k_1}$, $Q_2 \in \mathbb{F}^{M \times k_1}$, $Q_3 \in \mathbb{F}^{M \times k_0}$, $R \in \mathbb{F}^{M \times (k_1 + k_2 + k_0)}$ is an upper triangular matrix with partitions $R_{11} \in \mathbb{F}^{k_1 \times k_1}$, $R_{22} \in \mathbb{F}^{k_2 \times k_2}$, $R_{33} \in \mathbb{F}^{k_0 \times k_0}$, and other submatrices have corresponding dimensions.

First, we note that

$$
Q^T G Q
= \begin{bmatrix} R_{13} & R_{11} \\
R_{23} & 0 \\
R_{33} & 0 \\
0 & 0 \\
0 & 0 \end{bmatrix}
\begin{bmatrix} R_{13} & R_{12} & R_{13} \\
R_{12} & R_{23} & R_{23} \\
R_{13} & R_{23} & R_{33} \\
R_{11} & R_{12} & R_{13} \\
R_{11} & R_{12} & R_{13} \\
R_{11} & R_{12} & R_{13} \end{bmatrix}
\begin{bmatrix} R_{13} & R_{12} & R_{13} \\
R_{12} & R_{23} & R_{23} \\
R_{13} & R_{23} & R_{33} \\
R_{11} & R_{12} & R_{13} \\
R_{11} & R_{12} & R_{13} \\
R_{11} & R_{12} & R_{13} \end{bmatrix}
\begin{bmatrix} 0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \end{bmatrix}.
$$

Therefore, the last $M - (k_1 + k_1 + k_0)$ rows and columns of $Q^T G Q$—and hence of $(Q^T \Sigma Q)^{-1} (Q^T G Q)$—are zeros, which lead to the last $M - (k_1 + k_1 + k_0)$ zero eigenvalues of $\sum_{i=1}^{2} G_{i}^2 \sum_{j=1}^{\frac{k_1+1}{2}} G_{j}^2 \text{.}$ We then drop these rows and columns in all matrices involved in subsequent analysis. In particular, the submatrix of $Q^T \Sigma Q = Q^T (\sigma^2 I_M + A_{S_1} A_{S_1}^T) Q$ without the last $M - (k_1 + k_1 + k_0)$ rows and columns is

$$
\sigma^2 I_M + \begin{bmatrix} R_{11} & R_{12} & R_{13} \\
R_{12} & R_{22} & R_{23} \\
R_{13} & R_{23} & R_{33} \\
R_{11} & R_{12} & R_{13} \\
R_{11} & R_{12} & R_{13} \\
R_{11} & R_{12} & R_{13} \end{bmatrix}
\begin{bmatrix} 0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \end{bmatrix}.
$$

Define

$$
\begin{bmatrix} V & K^T \\
K & R_{33} R_{33}^T \end{bmatrix}
\triangleq \begin{bmatrix} R_{13} & R_{11} \\
R_{23} & 0 \\
R_{33} & 0 \\
R_{13} & R_{11} \\
R_{23} & 0 \\
R_{33} & 0 \\
R_{13} & R_{11} \\
R_{23} & 0 \\
R_{33} & 0 \\
R_{13} & R_{11} \\
R_{23} & 0 \\
R_{33} & 0 \end{bmatrix}.
$$
Due to the invariance of eigenvalues with respect to orthogonal transformations and switching to the symmetrized version, we focus on

\[
\begin{bmatrix}
 F & 0 \\
 0 & \sigma^2 I_{k_0}
\end{bmatrix}^{-\frac{1}{2}}
\begin{bmatrix}
 V & K^\dagger \\
 K & R_{33} R_{33}^T
\end{bmatrix}
\begin{bmatrix}
 F & 0 \\
 0 & \sigma^2 I_{k_0}
\end{bmatrix}^{-\frac{1}{2}}
= 
\begin{bmatrix}
 F - \frac{1}{2} V F^{-\frac{1}{2}} & F - \frac{1}{2} K^\dagger \\
 \frac{1}{2} K F^{-\frac{1}{2}} & \frac{1}{\sigma^2} R_{33} R_{33}^T
\end{bmatrix}. 
\]

Next we argue that the sorted positive eigenvalues of

\[
\begin{bmatrix}
 F - \frac{1}{2} V F^{-\frac{1}{2}} & F - \frac{1}{2} K^\dagger \\
 \frac{1}{2} K F^{-\frac{1}{2}} & \frac{1}{\sigma^2} R_{33} R_{33}^T
\end{bmatrix}
\]

are greater than the corresponding sorted eigenvalues of \(\frac{1}{\sigma^2} R_{33} R_{33}^T\).

For any \(\varepsilon > 0\), we define a matrix

\[
M_{\varepsilon,N} = 
\begin{bmatrix}
 -NI_{k_1+k_2} + k_1 & 0 \\
 0 & \frac{1}{\sigma^2} R_{33} R_{33}^T - \varepsilon I_{k_0}
\end{bmatrix}.
\]

Then we have

\[
\begin{bmatrix}
 F - \frac{1}{2} V F^{-\frac{1}{2}} & F - \frac{1}{2} K^\dagger \\
 \frac{1}{2} K F^{-\frac{1}{2}} & \frac{1}{\sigma^2} R_{33} R_{33}^T
\end{bmatrix} - M_{\varepsilon,N}
= 
\begin{bmatrix}
 F - \frac{1}{2} V F^{-\frac{1}{2}} + NI_{k_1+k_2} & F - \frac{1}{2} K^\dagger \\
 \frac{1}{2} K F^{-\frac{1}{2}} & \varepsilon I_{k_0}
\end{bmatrix}.
\]

Note that

\[
\begin{bmatrix}
 F - \frac{1}{2} V F^{-\frac{1}{2}} + NI_{k_1+k_2} & F - \frac{1}{2} K^\dagger \\
 \frac{1}{2} K F^{-\frac{1}{2}} & \varepsilon I_{k_0}
\end{bmatrix}
\]

is congruent to

\[
\begin{bmatrix}
 F - \frac{1}{2} V F^{-\frac{1}{2}} + NI_{k_1+k_2} & 0 \\
 0 & \frac{1}{\sigma^2} F - \frac{1}{2} K^\dagger K F^{-\frac{1}{2}} - \varepsilon I_{k_0}
\end{bmatrix}.
\]

Clearly \(F - \frac{1}{2} V F^{-\frac{1}{2}} + NI_{k_1+k_2} - \frac{1}{\sigma^2} F - \frac{1}{2} K^\dagger K F^{-\frac{1}{2}}\) is positive definite when \(N\) is sufficiently large. Hence, when \(N\) is large enough, we obtain

\[
\begin{bmatrix}
 F - \frac{1}{2} V F^{-\frac{1}{2}} & F - \frac{1}{2} K^\dagger \\
 \frac{1}{2} K F^{-\frac{1}{2}} & \frac{1}{\sigma^2} R_{33} R_{33}^T
\end{bmatrix} \succeq M_{\varepsilon,N}.
\]

Using [49, Corollary 4.3.3], we conclude that the eigenvalues of

\[
\begin{bmatrix}
 F - \frac{1}{2} V F^{-\frac{1}{2}} & F - \frac{1}{2} K^\dagger \\
 \frac{1}{2} K F^{-\frac{1}{2}} & \frac{1}{\sigma^2} R_{33} R_{33}^T
\end{bmatrix}
\]

are greater than those of \(M_{\varepsilon,N}\) if sorted. From Proposition 1, we know that

\[
\begin{bmatrix}
 F - \frac{1}{2} V F^{-\frac{1}{2}} + NI_{k_1+k_2} & F - \frac{1}{2} K^\dagger \\
 \frac{1}{2} K F^{-\frac{1}{2}} & \varepsilon I_{k_0}
\end{bmatrix}
\]

has exactly \(k_0\) positive eigenvalues, which are the only eigenvalues that could be greater than \(\lambda \left( \frac{1}{\sigma^2} R_{33} R_{33}^T \right) - \varepsilon\). Since \(\varepsilon\) is arbitrary, we finally conclude that the positive eigenvalues of \(\Sigma^{-1} G\) are greater than those of \(\frac{1}{\sigma^2} R_{33} R_{33}^T\) if sorted in the same way.

For the second claim, we need some notions and properties of symmetric and Hermitian matrices. For any pair of symmetric (or Hermitian) matrices \(P\) and \(Q\), \(P \preceq Q\) means that \(Q - P\) is positive definite and \(P \preceq Q\) means \(Q - P\) is nonnegative definite. Note that if \(P\) and \(Q\) are positive definite, then from [49, Corollary 7.7.4] \(P \preceq Q\) if and only if \(Q^{-1} \preceq P^{-1}\); if \(P \preceq Q\) then the eigenvalues of \(P\) and \(Q\) satisfy \(\lambda_k(Q) \leq \lambda_k(P)\), where \(\lambda_k(P)\) denotes the \(k\)th largest eigenvalue of \(P\); furthermore, \(A \preceq B\) implies that \(P A P^T \preceq P B P^T\) for any \(P\), square or rectangular. Therefore, \(\sigma^2 I_{3M} + A_{S_0 \backslash S_1} A_{S_0 \backslash S_1}^T \preceq \sigma^2 I_{3M} + A_{S_0 \backslash S_1} A_{S_0 \backslash S_1}^T = \Sigma_1\) yields

\[
\sigma_0^{1/2} \Sigma_1^{1/2} \preceq \sigma_0^{1/2} \left( \sigma^2 I_{3M} + A_{S_0 \backslash S_1} A_{S_0 \backslash S_1}^T \right)^{-1} \Sigma_0^{1/2}.
\]

Recall that from the definition of eigenvalues, the nonzero eigenvalues of \(AB\) and \(BA\) are the same for any matrices \(A\) and \(B\). Since we are interested only in the eigenvalues, a cyclic permutation in the matrix product on the previous inequality’s right-hand side gives us

\[
\left( \sigma^2 I_{3M} + A_{S_0 \backslash S_1} A_{S_0 \backslash S_1}^T \right)^{-\frac{1}{2}} \Sigma_0 \left( \sigma^2 I_{3M} + A_{S_0 \backslash S_1} A_{S_0 \backslash S_1}^T \right)^{-\frac{1}{2}} 
= I_{3M} + \left( \sigma^2 I_{3M} + A_{S_0 \backslash S_1} A_{S_0 \backslash S_1}^T \right)^{-\frac{1}{2}} A_{S_0 \backslash S_1}
\times A_{S_0 \backslash S_1}^T \left( \sigma^2 I_{3M} + A_{S_0 \backslash S_1} A_{S_0 \backslash S_1}^T \right)^{-\frac{1}{2}} 
\preceq I_{3M} + Q^{-\frac{1}{2}} A_{S_0 \backslash S_1} A_{S_0 \backslash S_1}^T Q^{-\frac{1}{2}} 
\preceq I_{3M} + P.
\]

Until now we have shown that the sorted eigenvalues of \(H\) are less than the corresponding ones of \(I_{3M} + P\). The nonzero eigenvalues of \(Q^{-\frac{1}{2}} A_{S_0 \backslash S_1} A_{S_0 \backslash S_1}^T Q^{-\frac{1}{2}}\) are the same as the nonzero eigenvalues of \(A_{S_0 \backslash S_1}^T A_{S_0 \backslash S_1}\). Using the same fact again, we conclude that the nonzero eigenvalues of \(A_{S_0 \backslash S_1}^T A_{S_0 \backslash S_1}\) are the same as the nonzero eigenvalues of \(\frac{1}{\sigma^2} A_{S_0 \backslash S_1} A_{S_0 \backslash S_1}^T\). Therefore, we obtain that

\[
\lambda_k \left( \Sigma_0^{1/2} \Sigma_1^{-1} \Sigma_0^{1/2} \right) \leq \lambda_k \left( I_{3M} + P \right) 
\leq \lambda_k \left( I_{3M} + \frac{1}{\sigma^2} A_{S_0 \backslash S_1} A_{S_0 \backslash S_1}^T \right).
\]

In particular, the eigenvalues of \(H\) that are greater than 1 are upper bounded by the corresponding ones of \(I_{3M} + \frac{1}{\sigma^2} A_{S_0 \backslash S_1} A_{S_0 \backslash S_1}^T\), if they are both sorted ascendently. Hence, we get that the eigenvalues of \(H\) that are greater than 1 are less than those of \(I_{3M} + \frac{1}{\sigma^2} A_{S_0 \backslash S_1} A_{S_0 \backslash S_1}^T\).

Therefore, the conclusion of the second part of the theorem holds. We comment here that usually it is not true that \(H \preceq I_{3M} + \frac{1}{\sigma^2} A_{S_0 \backslash S_1} A_{S_0 \backslash S_1}^T\). Only the inequality on eigenvalues holds. \(\square\)
For arbitrary fixed supports \( S_i, S_j \), we have

\[
\lambda_{S_i, S_j} \geq 1 + \frac{1}{2\kappa \sigma^2} \left( \prod_{l=1}^{k_d} 2\eta_l^2 \right)^{1/k_d}
\]

where \( 2\eta_l^2 \sim 2\beta^2(M-K-l+1) \) can be written as a sum of \( 2\kappa(M-K-l+1) \) independent squared standard Gaussian random variables and \( \eta_l \sim \beta^2(M-2K) \) is obtained by dropping \( K-l+1 \) of them. Therefore, using the union bound we obtain

\[
P\{\lambda_{S_i, S_j} \leq \gamma\} \leq \frac{1}{2\kappa \sigma^2} \min_{1 \leq l \leq k_d} \eta_l \leq \gamma \}
\]

\[
\leq \frac{1}{k_d} \sum_{1 \leq l \leq k_d} \eta_l \leq 2\kappa \sigma^2 \gamma
\]

Since \( \gamma = \frac{M-2K}{\sigma^2} \) implies that \( 2\kappa \sigma^2 \gamma = \frac{2\kappa}{3} (M-2K) < 2\kappa(M-2K) \), the mode of \( \beta^2(M-2K) \) when \( M-2K \) is sufficiently large, we have

\[
P\{\eta_l \leq 2\kappa \sigma^2 \gamma\} \leq \int_0^{2\kappa \sigma^2 \gamma} \frac{(1/2)^{1/2}(M-2K)}{\Gamma((M-2K)/2)} e^{-x^2/2} dx
\]

\[
\leq \frac{\Gamma(\kappa \sigma^2 \gamma)^{1/2}(M-2K)}{\Gamma((M-2K)/2)} e^{-\kappa \sigma^2 \gamma}.
\]

The inequality \( \log(\Gamma(z)) \geq (z - 1/2) \log z - z \) says that when \( M-2K \) is large enough,

\[
P\{\eta \leq 2\kappa \sigma^2 \gamma\} \leq \exp\left(\kappa(M-2K) \log(\kappa \sigma^2 \gamma) - \kappa \sigma^2 \gamma\right) - \left[\kappa(M-2K) - 1/2\right] \log \left[\kappa(M-2K)\right]
\]

\[
+ \exp\left(-c(M-2K)\right)
\]

where \( c < \kappa(\log 3 - 1) \). Therefore, we have

\[
P\{\lambda_{S_i, S_j} \leq \gamma\} \leq K \exp\left(-c(M-2K)\right).
\]

\[
\square
\]

ACKNOWLEDGMENT

The authors thank the anonymous referees for their careful and helpful comments.

REFERENCES


Gongguo Tang (S’09) received the B.Sc. degree in mathematics from Shandong University, China, in 2003, and the M.Sc. degree in system science from the Chinese Academy of Sciences, in 2006. Currently, he is a Ph.D. degree candidate with the Department of Electrical and Systems Engineering, Washington University, St. Louis, MO, under the guidance of Dr. Arne Nehorai. His research interests are in the area of compressive sensing, statistical signal processing, detection and estimation, and their applications.

Arne Nehorai (S’80–M’83–SM’90–F’94) received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA. From 1985 to 1995, he was a Faculty Member with the Department of Electrical Engineering, Yale University, New Haven, CT. In 1995, he joined the Department of Electrical Engineering and Computer Science, The University of Illinois at Chicago (UIC), as a Full Professor. From 2000 to 2001, he was Chair of the department’s Electrical and Computer Engineering (ECE) Division, which then became a new department. In 2001, he was named University Scholar of the University of Illinois. In 2006, he became Chair of the Department of Electrical and Systems Engineering, Washington University in St. Louis (WUSTL), St. Louis, MO. He is the inaugural holder of the Eugene and Martha Loohan Professorship and the Director of the Center for Sensor Signal and Information Processing (CSSIP) at WUSTL since 2006. Dr. Nehorai was Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2000 to 2002. From 2003 to 2005, he was Vice President (Publications) of the IEEE Signal Processing Society (SPS), Chair of the Publications Board, member of the Board of Governors, and member of the Executive Committee of this Society. From 2003 to 2006, he was the founding editor of the special columns on Leadership Reflections in the IEEE SIGNAL PROCESSING MAGAZINE. He was coreipient of the IEEE SPS 1989 Senior Award for Best Paper with P. Stoica, coauthor of the 2003 Young Author Best Paper Award, and coreipient of the 2004 Magazine Paper Award with A. Dogandzic. He was elected Distinguished Lecturer of the IEEE SPS for the term 2004 to 2005. He received the 2006 IEEE SPS Technical Achievement Award and the 2010 IEEE SPS Meritorious Service Award. He is the Principal Investigator of the Multidisciplinary University Research Initiative (MURI) project entitled Adaptive Waveform Diversity for Full Spectral Dominance. He has been a Fellow of the Royal Statistical Society since 1996.