# Pushing the Limits of Sparse Support Recovery Using Correlation Information

Piya Pal, *Student Member, IEEE*, and P. P. Vaidyanathan, *Fellow, IEEE*

***Abstract*—A new framework for the problem of sparse support recovery is proposed, which exploits statistical information about the unknown sparse signal in the form of its correlation. A key contribution of this paper is to show that if existing algorithms can recover sparse support of size $s$, then using such correlation information, the guaranteed size of recoverable support can be increased to $O(s^2)$, although the sparse signal itself may not be recoverable. This is proved to be possible by (a) formulating the sparse support recovery problem in terms of the covariance matrix of the measurements, and (b) designing a suitable measurement/sampling matrix which inherently exploits the correlation priors. The so-called Khatri-Rao product of the measurement matrix is shown to play an important role in deciding the level of recoverable sparsity. A systematic analysis of the proposed framework is also presented for the cases when the covariance matrix is only approximately known, by estimating it from finite number of measurements, obtained from the Multiple Measurement Vector (MMV) model. In this case, the use of LASSO on the estimated covariance matrix is proposed for recovering the support. However, the recovery may not be exact and hence a probabilistic guarantee is developed both for sources with arbitrary distribution as well as for Gaussian sources. In the latter case, it is shown that such recovery can happen with overwhelming probability as the number of available measurement vectors increases.**

***Index Terms*—Compressed sensing, correlation, joint support recovery, Khatri-Rao product, Kruskal rank, $l_1$ minimization, LASSO, multiple measurement vector, sparse.**

## I. Introduction

THE problem of underdetermined estimation has received great attention in recent times particularly due to its connection with the theory of sparse sampling and recovery. Fundamentally, it seeks to solve for $\mathbf{x} \in \mathbb{C}^{N\times 1}$ from an underdetermined system of linear equations: $\mathbf{y} = \mathbf{A}\mathbf{x}$ where $\mathbf{A} \in \mathbb{C}^{M\times N}$ is a fat matrix with $N \gg M$. Here $\mathbf{y}$ represents a measurement vector, the matrix $\mathbf{A}$ is known as the measurement matrix and the unknown vector $\mathbf{x}$ is assumed to be sparse, i.e., it has very few non-zero elements. This is a central problem in the field of compressed sensing [3], [9], and attempts at solving this problem have led to a series of breakthrough results [4], [7], [11], [39], [54] over the past decade.

In this paper, we investigate if it is possible to recover sparsity levels that can be orders of magnitude higher than that achieved by existing methods. In many applications, the support of sparse vectors is of importance rather than the vector $\mathbf{x}$ itself. They include spectral estimation [32], [36], imaging [26], [46], [50], direction finding in array processing [25], [27], [34], [51] etc. We introduce a statistical model for the unknown signal and incorporate *priors* in the form of additional information about the nature of correlation between the non-zero components of the unknown signal $\mathbf{x}$. We will propose a new approach to "sparse support recovery" based on the covariance matrix of the observations. Our formulation specifically makes use of the following prior: *The non-zero elements of the sparse vector* $\mathbf{x}$ *are assumed to be statistically uncorrelated.* This was a fundamental assumption in our earlier work on nested [30] and coprime arrays [42], in which we proposed that it is possible to identify $O(M^2)$ sources using $M$ physical sensors. In this paper, we will establish a similar implication of this assumption with respect to the size of the recoverable sparse support.

The central contribution of this paper is to demonstrate the following. If existing algorithms for sparse support recovery can identify support of size $s$, then our formulation exploiting the correlation priors can increase the guaranteed size to $O(s^2)$. Under suitable conditions, this can imply recovering support of size as large as $O(M^2)$. Measurement matrix design will be shown to play a crucial role in achieving this, and we will develop new conditions on the matrix for such recovery to be successful. This will in turn lead to the development of new sampling schemes and justify the need for the use of nested and coprime sampling. Another noteworthy point is that *we distinguish the recovery of the sparse vector* $\mathbf{x}$ *from that of its support*. Using our correlation-aware framework, we are able to guarantee the recovery of supports of size as large as $O(M^2)$ but in such a case, we cannot guarantee the recovery of $\mathbf{x}$ itself. This is because our approach to support recovery is based on the covariance matrix of the measurements (exact or estimate), instead of the raw measurements themselves. This observation is also consistent with existing guarantees on recovery of sparse $\mathbf{x}$ that imply that recovery of $\mathbf{x}$ itself is not possible when the sparsity exceeds $M$. This is further discussed in Section II-C2.

In practice however, we can only estimate the covariance matrix from multiple measurements which share a common sparse support. Our formulation in this paper is largely based on the Multiple Measurement Vector (MMV) model to obtain an estimate of the covariance matrix. Joint sparse support recovery

Manuscript received August 01, 2013; revised February 09, 2014, June 21, 2014, and August 27, 2014; accepted September 23, 2014. Date of publication December 22, 2014; date of current version January 08, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. John McAllister. This work was supported in part by the ONR under Grant N00014-11-1-0676, by the University of Maryland, College Park, and by the California Institute of Technology .

P. Pal was with California Institute of Technology, Pasadena, CA 91125 USA, she is now with the University of Maryland, College Park, MD 20742 USA (e-mail: ppal@ece.umd.edu).

P. P. Vaidyanathan is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: ppvnath@systems.caltech.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2014.2385033

from MMV is an important problem that seeks to recover the common support shared by the set of $L$ unknown sparse vectors $\mathbf{x}_l$ from the observation model: $\mathbf{Y}_L = \mathbf{A}\underbrace{[\mathbf{x}_1 \quad \mathbf{x}_2 \cdots \mathbf{x}_L]}_{\mathbf{X}_L}$. As we will demonstrate, the effect of exploiting only a finite number of measurements on the estimated covariance matrix leads to an additive noise-like term. We propose the use of LASSO [71] for support recovery from such a model derived from the covariance matrix and analyze its performance. The performance of LASSO for sparse models with additive noise has been previously analyzed [43], [60], [75]. In this paper, we analyze LASSO for its capability to recover the underlying sparse support, as opposed to developing stability of the recovered vector in terms of $l_2$ norm of the error. The main difference between these algorithms and our use of LASSO is that we apply LASSO on the covariance matrix of the measurements as opposed to the measured vectors themselves. While [60] considers a bounded (non-random) noise, we have a statistical model for our noise term, which is derived from the finite sample estimated covariance matrix. The analysis in [43] considers a random noise model and derives bounds on the probability of support recovery, both for deterministic and random ensembles of measurement matrices. But the signal itself is considered non-random. In contrast, our formulation makes use of the second order moments of the distribution of the random signal model. In our analysis, *the measurement matrix is deterministic and the probability is measured over the distribution of the unknown signals*. Consequently, our recovery guarantees are fundamentally different from those developed in [43]. We prove that without any specific assumption on the distribution of the sparse signal, the probability of failure decreases only as $\frac{1}{L}$. However, for Gaussian distributed signals, the probability of failure decays exponentially fast (i.e., as $\alpha^{-L}$ for $\alpha > 1$). Hence in this case, we can claim successful recovery with overwhelming probability with respect to $L$. However, these probabilistic guarantees for estimated covariance matrix are based on coherence of the measurement matrix $\mathbf{A}$ and the possibility of recovering supports of size $O(M^2)$ is not theoretically established in this case. However, the simulations indicate that in practice, supports of size much larger than $M$ can be recovered.

### A. Paper Outline

The rest of the paper is organized as follows. In Section II-A, we review existing guarantees on the size of the recoverable support for SMV and MMV models. In Section II-C, we summarize the main new contributions of our paper. In Section III, several key properties of the Khatri-Rao product are developed, which play important roles for establishing conditions for uniqueness of support recovery. These properties lead to new strategies for deterministic sampling and naturally generalize the concepts of nested and coprime sampling. In Section III-E, we formulate a convex relaxation for the correlation-aware support recovery problem and develop new theoretical conditions for its performance. We demonstrate that the correlation-aware framework naturally leads to a positive solution and this can be crucially exploited to guarantee the recovery of supports of size as large as $O(M^2)$. In Section V, we formulate the problem of support recovery from an estimated covariance matrix using the MMV model. We show that the finite number of measurements has a similar effect as an additive noise-like term on the model derived from the covariance matrix. We thereby propose the use of LASSO on the model derived from the estimated covariance matrix in Section V. We first analyze its performance regardless of the specific distribution of the unknown signal. We then compute bounds on the probability of successful support recovery when the unknown sparse signal follows a multivariate Gaussian distribution. In this case, it becomes possible to recover the support with overwhelming probability as the number of measurement vectors increases. In Section VII, we provide numerical experiments to validate the proposed theory. Some of the results appearing in this paper have been presented at recent conferences [31], [67]–[69].

### Notation

Matrices are denoted by uppercase letters in boldface (e.g., $\mathbf{A}$). Vectors are denoted by lowercase letters in boldface (e.g., $\mathbf{a}$). Superscript $^H$ denotes transpose conjugate, whereas superscript $^*$ denotes conjugation without transpose. The notation $[\mathbf{A}]_{i,j}$ denotes the $(i, j)$th element of matrix $\mathbf{A}$. For a set $S$, $|S|$ denotes its cardinality. Similarly, given two sets $S_1$ and $S_2$, $S_1 \setminus S_2$ denotes the difference of the sets, $S_1 \cap S_2$ denotes their intersection and $S_1 \cup S_2$ denotes their union. Given a set of integers $S \subset \{1, 2, \ldots, N\}$ and a vector $\mathbf{v} \in \mathbb{C}^{N \times 1}$, the vector $[\mathbf{v}]_S \in \mathbb{C}^{|S|} \times 1$ consists of elements of $\mathbf{v}$ indexed by $S$. Similarly given a matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$, the matrix $\mathbf{A}_S \in \mathbb{C}^{M \times |S|}$ consists of the columns of $\mathbf{A}$ indexed by $S$. The notation $\text{vec}(\mathbf{A})$ denotes vectorization of a matrix $\mathbf{A}$. The symbol $\odot$ denotes the Khatri-Rao product [23] between two matrices of appropriate size and the symbol $\otimes$ is used to denote the left Kronecker product. For two matrices $\mathbf{A}$ and $\mathbf{B}$ of same dimensions, $\mathbf{A} \circ \mathbf{B}$ denotes the Hadamard product [20].

## II. Background and Problem Formulation

In this section, we present a new formulation of the support recovery problem using the covariance matrix of the data, after reviewing relevant results and distinguishing them from our proposed approach.

### A. Review of Conditions for Unique Recovery

A central problem in the field of Compressive Sensing and Sparse Reconstruction is the recovery of the sparsest solution to an underdetermined system of linear equations, also known as the Single Measurement Vector (SMV) model. Such a model is given by

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 \qquad \text{(SMV)} \tag{1}$$

where $\mathbf{A} \in \mathbb{C}^{M \times N}$ is a fat matrix with $N \gg M$ and $\mathbf{x}_0$ is a sparse vector with $|\text{Supp}(\mathbf{x}_0)| \ll N$. This problem is cast as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{x} \quad (P0)_{SMV} \tag{2}$$

The sole property of $\mathbf{A}$ that decides when such recovery is possible, is the Kruskal Rank [24] of $\mathbf{A}$, denoted as $\text{Krank}(\mathbf{A})$, as defined below:

*Definition 1:* (**Kruskal Rank:**) The Kruskal Rank of $\mathbf{A}$, denoted as $\text{Krank}(\mathbf{A})$ is defined as the largest integer $r$ such that *every* $r$ columns of $\mathbf{A}$ are linearly independent. ♦

The Kruskal rank of a matrix is related very closely to its so-called "spark" [11], which denotes the smallest integer $i$ such that there exists a set of $i$ linearly dependent columns in the matrix. Denoting $S_0$ as the support of $\mathbf{x}_0$, the following condition provides an upper limit on $\mid S_0 \mid$ (also known as the *sparsity level*) that ensures unique recovery of $\mathbf{x}_0$ [11]:

*Theorem 1:* Consider the (SMV) model $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ with $S_0$ denoting the support of $\mathbf{x_0}$. Solving $(P0)_{SMV}$ will uniquely recover $\mathbf{x}_0$ if and only if $\mid S_0 \mid \leq \frac{\mathrm{Krank}(\mathbf{A})}{2}$. ♦

This guarantee on the maximum level of sparsity that can be recovered is fundamental in the sense that it needs to be satisfied for *any* recovery algorithm (convex or otherwise) to be successful.

Since $\mathrm{Krank}(\mathbf{A}) \leq \mathrm{rank}(\mathbf{A})$, it can be easily shown that the maximum level of recoverable sparsity from the SMV model satisfies

$$\mid S_0 \mid \leq \frac{M}{2} \tag{3}$$

Next, consider the MMV model (16). Here a set of $L > 1$ unknown sparse vectors $\mathbf{x}_l \in \mathbb{C}^{N\times 1}, l = 1, 2, \dots L$ share a common support $S_0$, i.e., $\mathrm{Supp}(\mathbf{x}_l) = S_0, \forall l$. Let $\mathbf{r}_{\mathbf{X}_L} \in \mathbb{R}^{N\times 1}$ be defined as $[\mathbf{r}_{\mathbf{X}_L}]_n = \left(\sum_{l=1}^{L}([\mathbf{X}_L]_{n,l})^q\right)^{\frac{1}{q}}$, for some $q \geq 1$. The corresponding problem for recovering the sparsest set of vectors now becomes

$$\min_{\mathbf{X}\in\mathbb{C}^{N\times L}} \|\mathbf{r}_{\mathbf{X}_L}\|_0 \quad \text{subject to } \mathbf{Y}_L = \mathbf{A}\mathbf{X}_L \quad (P0)_{MMV}$$

The recovery guarantee (independent of any specific algorithm used) for $\mid S_0 \mid$ now obeys [57], [59]:

*Theorem 2:* Consider the MMV model (16) where $S_0$ denotes the common support of the $L$ vectors. Solving $(P0)_{MMV}$ will recover the support $S_0$ if and only if

$$\mid S_0 \mid < \frac{\mathrm{Krank}(\mathbf{A}) + \mathrm{rank}(\mathbf{Y}_L)}{2} \tag{4}$$

♦

Since $\mathrm{rank}(\mathbf{Y}_L) \leq M$ and $\mathrm{Krank}(\mathbf{A}) \leq M$, (4) implies $\mid S_0 \mid < M$. It is to be noted that the maximum recoverable sparsity is independent of $L$ as long as $L > M$. From Theorem 2, it is clear that with no further assumption on $\mathbf{X}_L$, *the maximum sparsity level that can be recovered from the MMV model is always* $O(M)$.

### B. Relation to Past Work

Our paper develops explicit conditions for recovering sparsity levels that can be much larger than the size of the measurements, by exploiting correlation information. For instance, in [38], the use of cumulants was proposed to identify more sources than the number of physical sensors. It is well known in the radio astronomical community that correlation information can be exploited to improve imaging and in this regard sparse reconstruction techniques have also been proposed [53]. However, our work develops a rigorous theoretical framework for a generic sparse observation model and explicit guarantees are established for support recovery, under both ideal and non-ideal conditions. As such, the results developed in this paper stand in contrast to existing approaches to sparse reconstruction, such as basis pursuit[4], [6], greedy algorithms [29], [39], [41], or non convex norm based formulations [17], [18], [21]. In few cases, the sparse signal recovery problem has been cast in an estimation/detection setting [8]. More generally, a family of recovery techniques which explicitly compute a MAP-like estimate of the sparse unknown vector, has been unified under the umbrella of Bayesian Compressive Sensing [22]. Sparse Bayesian Learning (SBL) [37], [47]–[49] provides an alternative to the usual $l_1$ minimization approach to recovery of such sparse vectors by computing the MAP estimate of the sparse unknown vector. None of these methods however exploits the assumption that the non-zero elements of the unknown signal are uncorrelated, although the assumption is implicit in the signal model. For instance, the SBL framework [47], [49] models the unknown vector as a Gaussian random vector with diagonal covariance matrix, but it does not investigate what advantage (other than the algebraic simplifications of some of the results) this structure might offer in terms of size of the sparse support. The work of [25] uses a hybrid of subspace based methods like MUSIC and $l_1$ norm minimization algorithms for joint support recovery. Although the computations are based on the covariance matrix of the data, it does not exploit the uncorrelated structure of the unknown signal. The theoretical guarantees developed in [25] are for recovering sparse supports of size $O(M)$. More recently, the work in [51] represents the sparse recovery problem using the covariance matrix in a form that closely matches our own formulation. However, their focus is on developing a sparse spectrum fitting algorithm for DOA estimation, establishing consistency results, and finding a suitable regularization parameter. Our focus in this paper is to fundamentally establish conditions under which it becomes possible to recover larger supports from the covariance based model. Unlike [51], we point out explicit sampling strategies that help to guarantee such recovery. As we show in Section III, the so-called Khatri-Rao product [23] plays a pivotal role in many of our results. This product is closely related to the Kronecker product of matrices. The use of Kronecker product in compressive sensing has been studied in [16]. However, in spite of its apparent similarity with the Kronecker product, the Khatri-Rao product has very different properties and the results in [16] are not applicable to our case.

For the MMV model, a family of algorithms aims to exploit structure in the unknown sparse signals $\{\mathbf{x}_l\}_{l=1}^{L}$, such as rank [25], [59], correlation [49] etc. There is also a body of recent work that exploits correlation (for the tasks of spectrum sensing and DOA estimation) of the data to reduce sampling rates [55], [56], [70], [74]. However, their reconstruction techniques are not based on a sparse formulation. In this paper we consider a specific prior on the correlation of the received signal. In particular, we assume a diagonal covariance matrix for the unknown signal. Although such a diagonal structured covariance matrix has been implicitly considered in [55], [74], it has not been discussed how such a structure can improve the size of the recoverable sparse support. Hence, in compressive sensing literature, no past work so far exploits the specific prior that we consider in this paper. While the work in [49] considers the case where the unknown vectors have temporal correlation and thereby proposes algorithms to learn the correlation by Sparse Bayesian Learning (SBL) techniques, we actually consider the case when the non-zero entries of the unknown vectors have *no correlation* among themselves. The work in [49], however, does not specifically exploit the lack of correlation among the non-zero elements, although it implicitly assumes a block diagonal covariance matrix for the unknown sparse vectors. While the approach in [49] is based on SBL, we apply a LASSO on the covariance matrix of the data for recovering the sparse support. In future, we will be interested in studying the effect of this specific correlation prior in the SBL framework which has not been investigated so far.

### C. Comments on Our Contributions

Every existing algorithm for sparse reconstruction obeys the upper limit (3). In particular, no existing algorithm considers sparsity levels which can be in the region $\mid S_0 \mid > M$. One reason is that in current literature, no distinction is made between recovery of sparse support and that of the sparse unknown vector itself. Our work clearly distinguishes between the two and shows that using correlation priors on the unknown signal, it is possible to recover supports of size as large as $O(M^2)$ but not the unknown sparse vector itself. Since we are explicitly using additional information in the form of correlation and are interested in the recovery of support and not the sparse vector $\mathbf{x}_0$, the upper limit in (3) or in (4) does not apply to the results developed here. We summarize some important aspects of our contributions as follows:

*1) Correlation-Awareness and Recovery of Larger Sparse Support:* We will develop a novel approach to support recovery (i.e., recovery of $S_0$) incorporating a statistical framework in the SMV model. Rather than treating $\mathbf{x}_0$ as a deterministic unknown vector, we model it to be a sparse random vector with a fixed support. In particular, we will

1) Explicitly use statistical priors on the unknown sparse vector by imposing a structure on the covariance matrix of $\mathbf{x}_0$.
2) Propose novel conditions on the measurement matrix (which, in particular imply new strategies for structured sampling) and new reconstruction algorithms that can exploit the aforementioned priors.

Specifically, we make the following assumptions regarding the statistics of $\mathbf{x}_0$:

1) (A1): $E(\mathbf{x}_0) = \mathbf{0}$
2) (A2): $E([\mathbf{x_0}]_i[\mathbf{x_0}]_j^*) = 0,\ i \neq j, i, j \in S_0$

Assumption (A2) means that the non-zero elements of $\mathbf{x}_0$ are statistically uncorrelated. In what follows, we will develop a new method for recovering $S_0$ that exploits this fact. The main idea is to use the data covariance matrix $\mathbf{R_{yy}} = E(\mathbf{y}\mathbf{y}^H)$ to recover $S_0$. The first part of our results is based on the *exact* knowledge of the data covariance matrix $\mathbf{R_{yy}}$ whereby we will show it is possible to guarantee the recovery of sparsity levels as large as $\mid S_0 \mid = O(M^2)$. This becomes possible under certain conditions on the measurement matrix. Sparse sampling schemes such as nested and coprime sampling will be shown to satisfy these conditions and hence they become natural choices for correlation-aware support recovery. Without the knowledge of correlation priors and appropriate sampling strategies, existing performance bounds on sparse models [5], [12], [13], [35], [43], [44] can only guarantee recovery of $O(M)$ sparsity levels.

*2) Recovery of Sparse Support v/s Sparse Vector:* We want to highlight an important difference between the recovery of the sparse vector(s) $\mathbf{x}_0$ and the recovery of its support. In all existing approaches to sparse recovery, the recovery of support $S_0$ automatically leads to recovery of the sparse vector $\mathbf{x}_0$ since it is possible to solve $\mathbf{y} = \mathbf{A}_{S_0}[\mathbf{x}]_{S_0}$ uniquely for $[\mathbf{x}]_{S_0}$. This happens because $\mid S_0 \mid < \mathrm{Krank}(\mathbf{A})$. In our correlation-aware framework, we will demonstrate that it is *possible to recover support* $S_0$ where $\mid S_0 \mid \gg M$, possibly $\mid S_0 \mid = O(M^2)$. However, in this case, it is *not possible to recover the vector(s)* $\mathbf{x}_0$ *itself*. This can be easily seen from the fact that when $\mid S_0 \mid > M \geq \mathrm{rank}(\mathbf{A})$, there is no unique $[\mathbf{x}_0]_{S_0}$ satisfying $\mathbf{y} = \mathbf{A}_{S_0}[\mathbf{x}_0]_{S_0}$. As we will show later, it is possible to recover $E(|[\mathbf{x}_0]_i|^2), i \in S_0$ although it is not possible to recover $[\mathbf{x}_0]_{S_0}$ itself.

*3) Non-Ideal Conditions:* The results developed in the first half of the paper are based on the assumption that we know the data covariance matrix $\mathbf{R_{yy}}$ *exactly*. This is also referred to as the ideal covariance matrix. This is achievable as the number of measurement vectors become arbitrarily large. However, in practice, we need to rely upon estimated $\hat{\mathbf{R}}_{\mathbf{yy}}$ (computed from finite number of measurements). In this paper, we perform rigorous analysis of these non-ideal scenarios in Section VI and establish fundamental conditions as well as performance guarantees of practical algorithms working on a non-ideal model. The effect of exploiting only a finite number of measurements on the estimated covariance matrix leads to an additive noise-like term. We propose the use of LASSO [71] for support recovery from such a model derived from the covariance matrix and analyze its performance. We prove that without any specific assumption on the distribution of the sparse signal, the probability of failure decreases only as $\frac{1}{L}$. However, for Gaussian distributed signals, the probability of failure decays exponentially fast (i.e., as $\alpha^{-L}$ for $\alpha > 1$). Hence in this case, we can claim successful recovery with overwhelming probability with respect to $L$.

## III. The Khatri-Rao Product, and Support Recovery From Exact Covariance Matrix

The results developed in this section illustrate by how much it is possible to reduce the sampling rate if we have perfect knowledge of the second order statistics of the data. We make the following assumptions:

1) We know the ideal data covariance matrix $\mathbf{R_{yy}} = E(\mathbf{y}\mathbf{y^H})$.
2) Assumptions (A1) and (A2) hold.

*Lemma 1:* Consider a SMV model (1) where $S_\mathbf{0}$ denotes the support of $\mathbf{x}_0$. Under assumptions (A1) and (A2), recovery of $S_\mathbf{0}$ is equivalent to recovering Supp($\mathbf{p}$) from

$$\mathbf{z} = \Big(\mathbf{A}^* \odot \mathbf{A}\Big)\mathbf{p} \qquad (Co-SMV) \tag{5}$$

where $\mathbf{z} = \mathrm{vec}(\mathbf{R_{yy}}), \mathbf{p} = \mathrm{diag}(\mathbf{R_{xx}})$, and $\mathbf{R_{xx}} = \mathrm{E}(\mathbf{x}_0\mathbf{x}_0^H)$. ♦

*Proof:* The covariance matrix of $\mathbf{y}$ is given by $\mathbf{R_{yy}} = \mathbf{A}\mathbf{R_{xx}}\mathbf{A}^H$. Upon vectorization, we obtain

$$\mathbf{z} = (\mathbf{A}^* \otimes \mathbf{A})\mathrm{vec}(\mathbf{R_{xx}}) \tag{6}$$

Under assumptions (A1)-(A2), $[\mathbf{R_{xx}}]_{i,j} = \sigma_i^2\delta(i-j)$ where $\sigma_i^2 = \mathrm{E}\Big(|[\mathbf{x}_0]_i|^2\Big), i \in S_0$. Therefore, the non-zero elements of

$\text{vec}(\mathbf{R_{xx}})$ are contained in the vector $\text{diag}(\mathbf{R_{xx}}) \in \mathbb{R}^{N\times 1}$. It is readily seen that $\text{Supp}\Big(\text{diag}(\mathbf{R_{xx}})\Big) = S_0$. Using the definition of Khatri-Rao product [23], (6) reduces to (5) and this concludes the proof. ■

The lemma suggests we can equivalently recover the sparse support if we can recover $\mathbf{p}$ satisfying (5). The problem formulation for recovery of $S_\mathbf{0}$ then becomes

$$\begin{aligned} &\min_{\mathbf{p}} \|\mathbf{p}\|_0 \quad (P0)_{Co-SMV} \\ &\text{subject to} \quad \Big(\mathbf{A}^* \odot \mathbf{A}\Big)\mathbf{p} = \mathbf{z} \end{aligned} \tag{7}$$

Henceforth, we will denote $\mathbf{A}_{KR} = \mathbf{A}^* \odot \mathbf{A}$. Notice that the size of $\mathbf{A}_{KR}$ is $M^2 \times N$, i.e., it has the same number of columns as $\mathbf{A}$ but the number of rows has increased $M$ times. So, intuitively, there is a possibility that solving (7) might lead us to recover supports of larger size, if $\mathbf{A}_{KR}$ satisfies suitable conditions. In order to understand what is gained by solving $(P0)_{Co-SMV}$ instead of $(P0)_{SMV}$, we now turn to investigate the matrix $\mathbf{A}_{KR}$ and what advantages it offers over $\mathbf{A}$ in terms of support recovery.

To study how $\mathbf{A}_{KR}$ improves the recovery limits, we define the following quantity called the "Degree of Krank Expansion" (denoted $\alpha_\mathbf{A}$) of $\mathbf{A}$:

*Definition 2:* The Degree of Krank Expansion of $\mathbf{A} \in \mathbb{C}^{M\times N}$ is defined as $\alpha_\mathbf{A} = \frac{\text{Krank}(\mathbf{A}_{KR})}{\text{Krank}(\mathbf{A})}$ ♦

*Lemma 2:* Let $s_{SMV}$ and $s_{Co-SMV}$ denote the maximum size of $S_0$ that can be uniquely recovered by solving $(P0)_{SMV}$ and $(P0)_{Co-SMV}$ respectively. Then $s_{Co-SMV} = \alpha_\mathbf{A} s_{SMV}$ ♦

*Proof:* Applying Theorem 1 to $(P0)_{Co-SMV}$, we obtain that all sparse supports upto size $|\ S_0\ | \leq \frac{\text{Krank}(\mathbf{A}_{KR})}{2}$ can be uniquely recovered. The lemma then follows from definition of $\alpha_\mathbf{A}$. ■

Hence, for $\alpha_\mathbf{A}$-krank expandable matrix $\mathbf{A}$, by solving $(P0)_{Co-SMV}$ we can recover sparsity levels which are $\alpha_\mathbf{A}$ times as large as the sparsity level that can be recovered by solving $(P0)_{SMV}$. A matrix $\mathbf{A}$ is said to have full Krank if $\text{Krank}(\mathbf{A}) = \text{rank}(\mathbf{A})$. The following theorem precisely characterizes how large this $\alpha_\mathbf{A}$ can be when $\mathbf{A}$ has full Krank.:

*Theorem 3:* The Degree of Krank expansion of $\mathbf{A}$ with full Krank, satisfies $1 < \alpha_\mathbf{A} \leq M$ ♦

*Proof:* We use the following result from [33]

$$\text{Krank}(\mathbf{A} \odot \mathbf{B}) \geq \min(N, \text{Krank}(\mathbf{A}) + \text{Krank}(\mathbf{B}) - 1). \tag{8}$$

This implies $\text{Krank}(\mathbf{A}_{KR}) > \text{Krank}(\mathbf{A})$ and hence $\alpha_\mathbf{A} > 1$. Let $\text{rank}(\mathbf{A}) = r$. From properties of Kronecker Products [20], $\text{rank}(\mathbf{A}^* \otimes \mathbf{A}) = r^2$. Now, $\text{Krank}(\mathbf{A}_{KR}) \leq \text{rank}(\mathbf{A}_{KR}) \leq \text{rank}(\mathbf{A}^* \otimes \mathbf{A}) = r^2$. Hence, we obtain $\alpha_\mathbf{A} \leq r \leq M$. This concludes the proof. ■

Theorem 3 implies the following:

1) Since, $\alpha_\mathbf{A} > 1$, we can always recover sparse support *of larger size* by using the "Co-SMV" framework.
2) If $\mathbf{A}$ has full Krank, the maximum size of recoverable support can be $M$ times larger than that obtained from solving $(P0)_{SMV}$, i.e., it can be $O(M^2)$.

This leads us to the following question which we shall address next:

*1) Question:*

What are deterministic and random designs of $\mathbf{A}$ which can guarantee Kruskal-Rank of $\mathbf{A}_{KR} = O(M^2)$

## *A. Random Measurement Matrix*

The following Theorem proves that random $\mathbf{A}$ with independently drawn entries has $\alpha_\mathbf{A} = O(M)$ almost surely:

*Theorem 4:* Let $\mathbf{A} \in \mathbb{R}^{M\times N}$ be a random matrix whose elements are drawn *independently* from continuous distributions over $\mathbb{R}$. Assuming $N > \frac{M^2+M}{2}$, the following holds with probability 1: $\text{Krank}(\mathbf{A}_{KR}) = \frac{M^2+M}{2}$ ♦

*Proof:* We will show that any collection of $\frac{M^2+M}{2}$ columns in $\mathbf{A}_{KR}$ has rank $\frac{M^2+M}{2}$ with probability 1 over $\mathbb{R}^{\frac{M(M^2+M)}{2}}$. This will establish that $\mathbf{A}_{KR}$ has Kruskal Rank $\frac{M^2+M}{2}$ with probability 1 over $\mathbb{R}^{MN}$ by using the union bound over $\binom{N}{\frac{M^2+M}{2}}$ sets. Consider a subset $S \subseteq \{1, 2, \ldots, N\}$, where $|\ S\ | = \frac{M^2+M}{2}$ and let $[\mathbf{A}_{KR}]_S$ denote the submatrix consisting of columns indexed by the elements of $S$. Denote the elements of $\mathbf{A}$ as $a_{i,j}, \quad 1 \leq i \leq M, 1 \leq j \leq N$. The elements in the $j$th column of $\mathbf{A}_{KR}$ are given by $a_{i,j}a_{l,j}, \quad 1 \leq i, l \leq M$. Given integers $i_0, l_0 \in \{1, 2, \ldots, M\}, i_0 \neq l_0$,and $j \in S$ there are exactly two rows in $[\mathbf{A}_{KR}]_S$ that are identical and consist of the elements $a_{i_0,j}a_{l_0,j}, \quad j \in S$. Overall, there are $\frac{M^2-M}{2}$ such identical rows, upon removing which we obtain an $\frac{M^2+M}{2} \times \frac{M^2+M}{2}$ submatrix of $[\mathbf{A}_{KR}]_S$ which we will denote as $[\mathbf{A}_{KR}]_S^{\text{sub}}$. We will show that this matrix is full rank with probability 1. In order to establish this, denote $\mathbf{a}_S = \{a_{i,j}, 1 \leq i \leq M, j \in S\}$. Since the variables $a_{i,j}$ are drawn independently from continuous distributions on $\mathbb{R}$, the function $f(\mathbf{a}_S) \triangleq \det([\mathbf{A}_{KR}]_S^{\text{sub}})$ is a multivariate polynomial in $M\ |\ S\ |$ variables given by the elements of $\mathbf{a}_S$ and hence it is analytic in $\mathbb{R}^{M|S|}$. By the properties of analytic functions, if $f(\mathbf{a}_S)$ is not the zero polynomial, its zero set will be of measure 0 in $\mathbb{R}^{M|S|}$. Given S, if we can prove that $f(\mathbf{a}_S)$ is not a trivial polynomial, then it will imply that the random matrix $[\mathbf{A}_{KR}]_S^{\text{sub}}$ is full rank with probability 1 with respect to any continuous distribution over $\mathbb{R}^{M|S|}$ and it will conclude our proof. In order to prove $f(\mathbf{a}_S)$ is non-trivial, it suffices to find a specific point $[\mathbf{a}_S]_0$ such that $f([\mathbf{a}_S]_0) \neq 0$. Such a $[\mathbf{a}_S]_0$ can be found by the following construction: Let us divide the $\frac{M^2+M}{2}$ columns indexed by $S$ into $M$ unequal groups $\{S_m\}_{i=1}^M$ with the $|\ S_m\ | = M - m + 1$. Relabel the columns in $S_m$ as $\{j_k^m, 1 \leq k \leq M - m + 1\}$. The elements of $\mathbf{A}$ in columns indexed by $S_m, 1 \leq m \leq M$ are then constructed as:

$$a_{i,j_k^m} = \begin{cases} 1 & \text{if } i = m, k = 1 \\ 1 & \text{if } i = m, m + k - 1, k > 1 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

The resulting matrix $[\mathbf{A}_S]^{\text{sub}}$ can be transformed, through column and row exchanges, to an upper triangular matrix, with each diagonal element being equal to 1. Hence it is full rank and this concludes the proof. ■

The result shows that Khatri-Rao product of random matrices continue to exhibit very high Kruskal Rank, as large as $O(M^2)$.

### B. Deterministic Designs

In many applications, such as imaging or source localization, the physics of the problem imposes structure on the measurement matrix $\mathbf{A}$. In many applications, the matrix $\mathbf{A}$ can represent the adjacency matrix of a network graph and can have further deterministic constraints. Usually in such situations, it is more difficult to come up with a deterministic design whose Khatri-Rao product continues to have $O(M^2)$ Kruskal Rank, and common sampling strategies often fail. In this context we would like to revisit the nested [30] and coprime sampling [42] strategies and show how these deterministic designs naturally lead to $O(M^2)$ Kruskal Rank for their Khatri-Rao products.

When the application of interest is DOA estimation in antenna array processing, assumptions (A1) and (A2) naturally fit the context of direction finding because (A2) simply implies that the signals coming from different directions are statistically uncorrelated. Consider an antenna array with $M$ elements, with $z_i$ denoting the distance of the $i$th sensor from the origin. The manifold of the array $\mathbf{A}$ has a special structure where the $(i,j)$th element has the form $\alpha_j^{z_i}, |\alpha_j| = 1$. Then, the elements of $\mathbf{A}_{KR}$ are given by

$$\alpha_j^{z_i - z_l}, \quad i,l \in \{1,2,\ldots M\}, \quad j = 1,2,\ldots N \tag{10}$$

The rows of $\mathbf{A}_{KR}$ are therefore characterized by the "difference set", also known as the difference co-array [19]:

$$S_{ca} = \{z_i - z_j, \quad 1 \le i,j \le M\} \tag{11}$$

The following lemma makes an explicit connection between the difference set and the Kruskal Rank of $\mathbf{A}_{KR}$:

*Lemma 3:* The Kruskal Rank of $\mathbf{A}_{KR}$ where $\mathbf{A}$ represents an array manifold, satisfies $\text{Krank}(\mathbf{A}_{KR}) \le \mid S_{ca} \mid$ where $S_{ca}$ is given by (11) ♦

*Proof:* It is clear from (10) that the number of *distinct* rows in $\mathbf{A}_{KR}$ is equal to the number of distinct elements in $S_{ca}$. Hence $\text{Krank}(\mathbf{A}_{KR}) \le \text{rank}(\mathbf{A}_{KR}) \le |S_{ca}|$. ■ The number of distinct elements in $S_{ca}$ therefore fundamentally limits the Krank of $\mathbf{A}_{KR}$. We now revisit the traditional Uniform Linear Array (ULA), the nested array and the coprime array [42] in the light of this fact. It is also worthwhile to mention that the family of sparse rulers or Minimum Redundancy Arrays [55] form an optimal (but analytically intractable) class of samplers which produce the maximum number of elements in $S_{ca}$ in a continuous range.

1) **ULA**: For a uniform linear array, $z_i = i, \quad 1 \le i \le M$. It can be easily seen that in this case $S_{ca} = \{m, -M + 1 \le m \le M-1\}$ and hence $\mid S_{ca} \mid = 2M - 1$. In fact, it can be verified that in this case $\text{Krank}(\mathbf{A}_{KR}) = 2M - 1$. Since for a ULA, $\text{Krank}(\mathbf{A}) = M$, this shows from (8) that ULA actually achieves the minimum bound for the Kruskal Rank of its Khatri-Rao product. This observation was also recently made in [74] in the context of wideband spectrum sensing.

2) **Nested Array**: For a nested array [30], the set of sensor locations is given by

$$\{z_i\}_{i=1}^M = \{m, 1 \le m \le M/2\} \cup \{(M/2+1)n, \\ 1 \le n \le M/2\}$$

The difference set is $S_{ca} = \{m, -M^2/4 - M/2 + 1 \le m \le M^2/4 + M/2 - 1\}$. Since $S_{ca}$ contains all consecutive integers within the above range, $\mathbf{A}_{KR}$ is a Vandermonde matrix with $\frac{M^2}{2} + M - 1$ distinct rows. This shows that $\text{Krank}(\mathbf{A}_{KR}) = \frac{M^2}{2} + M - 1$. Hence the nested array corresponds to a structured spatial sampling scheme that produces a deterministic measurement matrix with $\text{Krank}(\mathbf{A}_{KR}) = O(M^2)$.

3) **Coprime Array**: A coprime array is an array of $M = M_1 + 2M_2 - 1$ sensors, where $M_1$ and $M_2$ are coprime integers. The location of the sensors are given by

$$\{z_i\}_{i=1}^M \\ = \{M_1 m, 0 \le m \le 2M_2 - 1\} \cup \{M_2 n, 1 \le n \le M_1 - 1\}.$$

The difference set in this case contains all integers in the range $\{n, -M_1M_2 \le n \le M_1M_2\}$. Thus, the rows of $\mathbf{A}_{KR}$ contain a Vandermonde sub matrix with $2M_1M_2+1$ distinct rows. Hence $\text{Krank}(\mathbf{A}_{KR}) \ge 2M_1M_2 + 1$. For $M_2 = M_1 - 1$, we have $M = 3M_2$ and $\text{Krank}(\mathbf{A}_{KR}) \ge O(M^2)$.

### C. Coherence Based Properties of the Khatri-Rao Product

Coherence of the measurement matrix plays an important role to ensure the recovery of sparse support when it is cast as an appropriate convex problem. We now derive a few properties related to the coherence of the Khatri-Rao product.

*Definition 3:* The coherence of a matrix $\mathbf{A}$ is defined as $\mu_{\mathbf{A}} = \max_{i \ne j} \frac{|\mathbf{a}_i^H \mathbf{a}_j|}{\|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2}$

*Lemma 4:* The coherence of $\mathbf{A}_{KR}$ and $\mathbf{A}$ are related as:

$$\mu_{\mathbf{A}_{KR}} = \mu_{\mathbf{A}}^2 \tag{12}$$

♦

*Proof:* It can be verified from the definition of Kronecker products, that

$$(\mathbf{a}_i \otimes \mathbf{b}_i)^H (\mathbf{a}_j \otimes \mathbf{b}_j) = \left(\mathbf{a}_i^H \mathbf{a}_j\right)\left(\mathbf{b}_i^H \mathbf{b}_j\right)$$

Hence $\mid [\mathbf{a}_{KR}]_i^H [\mathbf{a}_{KR}]_j \mid = \mid (\mathbf{a}_i^* \otimes \mathbf{a}_i)^H (\mathbf{a}_j^* \otimes \mathbf{a}_j) \mid = \left(\mid \mathbf{a}_i^H \mathbf{a}_j \mid\right)^2$. This concludes the proof. ■

A related property of the measurement matrix, known as the cumulative coherence, plays an important role to guarantee the success of a family of greedy algorithms (Orthogonal Matching Pursuit or OMP) [39] in recovering the sparse support. It is defined as below:

*Definition 4:* [39] Given an integer $s$, the cumulative coherence function of a matrix $\mathbf{A}$ is defined as $\mu_1^{\mathbf{A}}(s) = \max_{S:|S|=s} \max_{j \notin S} \|\mathbf{A}_S^H \mathbf{a}_j\|_1$

The following lemma relates $\mu_1^{\mathbf{A}_{KR}}(s)$ and $\mu_1^{\mathbf{A}}(s)$

*Lemma 5:* Given an integer $s$, The cumulative coherence functions of $\mathbf{A}$ and $\mathbf{A}_{KR}$ are related as $\mu_1^{\mathbf{A}_{KR}}(s) < \left(\mu_1^{\mathbf{A}}(s)\right)^2$ ♦

*Proof:* Let $\mathbf{B}$ denote $\mathbf{A}_{KR}$ and $\mathbf{b}_i$ denote its $i$th column for ease of reference. Given $S \subseteq [1,2,\ldots,N]$ with $\mid S \mid = s$, and integer $j \notin S$,

$$\|\mathbf{B}_S^H \mathbf{b}_j\|_1 = \sum_{i \in S} \mid \mathbf{b}_i^H \mathbf{b_j} \mid = \sum_{i \in S} \left( \mid \mathbf{a}_i^H \mathbf{a}_j \mid \right)^2 \\ < \left( \sum_{i \in S} \mid \mathbf{a}_i^H \mathbf{a}_j \mid \right)^2 \le \left(\mu_1^{\mathbf{A}}(s)\right)^2$$

Hence $\mu_1^{\mathbf{A}_{KR}}(s) = \max_{S:|S|=s} \max_{j\notin S} \|\mathbf{B}_S^H \mathbf{b}_j\|_1 < \left(\mu_1^{\mathbf{A}}(s)\right)^2$. ■

### D. Convex Relaxation for Support Recovery From Ideal Covariance Matrix

So far, we have demonstrated that it is possible to design both random and deterministic measurement matrices so that supports of size $O(M^2)$ can be recovered from the data covariance matrix $\mathbf{R_{yy}}$, say, via exhaustive search. However, one practical approach to recover the support is to formulate a convex problem:

$$\min_{\mathbf{p}} \|\mathbf{p}\|_1 \quad (P1)_{Co-SMV}$$
$$\text{subject to} \quad \left(\mathbf{A}^* \odot \mathbf{A}\right)\mathbf{p} = \mathbf{z} \tag{13}$$

There are various ways to characterize conditions under which the solution to $(P0)_{Co-SMV}$ is identical to that of $(P1)_{Co-SMV}$. In what follows, we will first describe a coherence based guarantee since the coherence for $\mathbf{A}$ and $\mathbf{A}_{KR}$ can be easily related as given by (12). In Section III-E2, we will further exploit additional structure (involving positive solutions) in the covariance based support recovery problem, and develop a set of necessary and sufficient conditions for successful support recovery.

*1) New Coherence Based Guarantee:* The coherence based guarantee for finding the support $S_0$ by solving $(P1)_{Co-SMV}$ is given by

*Theorem 5:* Consider the SMV model (1) where $S_0$ denotes the support of $\mathbf{x}_0$. Under assumptions (A1) and (A2), solving $(P1)_{Co-SMV}$ can uniquely recover supports of size $|\ S_0\ | < \frac{1}{2}\left(1+\frac{1}{\mu_{\mathbf{A}}^2}\right)$ ♦

*Proof:* It follows from the coherence based condition in [11] which states that if $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ with the support of $\mathbf{x}_0$ given by $S_0$, then solving $\min_{\mathbf{x},\mathbf{A}\mathbf{x}=\mathbf{y}} \|\mathbf{x}\|_1$ yields $\mathbf{x}_0$ if $|\ S_0\ | < \frac{1}{2}\left(1+\frac{1}{\mu_A}\right)$. ■

Since, by definition $\mu_A < 1$, (12) implies that $\mu_{\mathbf{A}_{KR}} < \mu_{\mathbf{A}}$. Theorem 5 therefore shows that an argument based on coherence alone indicates that it is possible to recover larger supports by solving $(P1)_{Co-SMV}$. On one hand, coherence of a matrix is easy to compute which makes it a popular way to analyze recovery performance in many situations [1], [2], [12], [40]. On the other hand, it is only a sufficient condition for sparse recovery and suffers from the Welch bound [45], which indicates, for $\mathbf{A}_{KR}$, $\mu_{\mathbf{A}_{KR}} \geq \frac{1}{M}\sqrt{\frac{N-M^2}{N-1}}$. For given $N$, this implies that coherence based conditions on $\mathbf{A}_{KR}$ can guarantee recovery of supports of size only as large as $O(M)$. Hence coherence fails to indicate if it is at all possible to recover $S_0$ with $|\ S_0\ | = O(M^2)$ by solving $(P1)_{Co-SMV}$. In what follows next, we will exploit additional structures in the Co-SMV model to provide much sharper guarantees.

*2) The Case of Positive Solution:* In the previous discussion, we ignored the fact that $\mathbf{p}$ in (5) contains the diagonal entries of $\mathbf{R_{xx}}$ and hence is a non-negative vector. In this section, we seek to recover the support $S_0$ by using this additional constraint on the unknown vector $\mathbf{p}$ to be non-negative by solving

$$\min_{\mathbf{p}} \|\mathbf{p}\|_1 \qquad (P1+)_{Co-SMV}$$
$$\text{subject to} \quad \mathbf{A}_{KR}\mathbf{p} = \mathbf{z}, \quad \mathbf{p} \succeq \mathbf{0} \tag{14}$$

It is to be noted that even when we have the sample autocorrelation matrix estimated from a finite number of measurements, the solution we are seeking is still non-negative since it contains the estimated power of the sources. The recovery of sparse non-negative vectors has been of considerable interest [10], [14], [15] in compressive sensing. The necessary and sufficient conditions for perfect recovery of $s$-sparse non-negative vectors relate to the theory of neighborly polytopes [15]. We shall first state a result (proved in [15]) when the measurement matrix is a Vandermonde matrix (with elements on the unit circle).

*Theorem 6:* Let $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ where the elements of $\mathbf{A}$ are given by $[\mathbf{A}]_{m,n} = e^{jmt_n}, 1 \leq m \leq M, 0 < t_1 < \cdots < t_N < 2\pi$ and $\mathbf{x}_0 \succeq \mathbf{0}$. If $D \triangleq \|\mathbf{x}_0\|_0 \leq M$ and $M < 2N$, then $\mathbf{x}_0$ can be uniquely recovered by solving $(P1+)_{Co-SMV}$. ♦

It is to be noted that the necessary and sufficient condition for $(P1+)_{Co-SMV}$ to recover the sparse positive vector *does not* involve a coherence or Restricted Isometry condition on the measurement matrix $\mathbf{A}$. Rather, it requires the measurement matrix to be associated with a so-called outwardly $|\ S_0\ |$-neighborly polytope to ensure perfect recovery. Also, notice that as long as $N > 2M$, the maximum level of sparsity that can be recovered has no dependence on $N$.

## IV. A Unified Framework for Imperfect Correlation-Awareness: Preliminaries

The results developed so far are based on the *exact knowledge* of the covariance matrix of the observations and it uses the assumption that the non zero elements of the unknown sparse vector are statistically uncorrelated. In practice, we can only estimate the covariance matrix from multiple measurements which share a common sparse support. The assumption of non zero elements being uncorrelated might not always hold in a strict sense. We now develop a practical approach to correlation-aware sparse support recovery by explicitly accounting for such deviations from the ideal model, and analyzing its performance.

### A. A Common Underlying Model

We represent the "non-ideal" data correlation matrix as $\hat{\mathbf{R}}_{yy}$, which can either represent the estimated correlation matrix from finite number of samples, or it can also correspond to a non-diagonal $\mathbf{R_{xx}}$. To address the challenge of recovering support under such non-ideal conditions in a unified manner, we propose to solve the following problem:

$$\min_{\mathbf{x}\succcurlyeq\mathbf{0}} \left\|\mathrm{vec}(\hat{\mathbf{R}}_{\mathbf{yy}}) - \mathbf{A}_{KR}\mathbf{x}\right\|_2^2 + h\,\|\mathbf{x}\|_1 \quad (Co-LASSO) \tag{15}$$

### B. The Case of Finite Number of Measurements

In this case, the formulation in (15) captures the effect of a finite number (say, $L$) of multiple measurement vectors. This formulation is largely based on the Multiple Measurement Vector (MMV) model, where $L$ vectors $\mathbf{Y}_L = [\mathbf{y}_1\mathbf{y}_2\cdots\mathbf{y}_L]$ are observed and the corresponding $L$ unknown sparse vectors $\mathbf{x}_l$ share the *same* sparse support $S_0$ ($\mathrm{Supp}(\mathbf{x}_l) = S_0, \forall l$):

$$\mathbf{Y}_L = \mathbf{A}\underbrace{[\mathbf{x}_1 \quad \mathbf{x}_2\cdots\mathbf{x}_L]}_{\mathbf{X}_L} \qquad \text{(MMV)} \tag{16}$$

Since we do not assume to know the actual $\mathbf{R}_{\mathbf{yy}}$, we will estimate it from these $L$ vectors as $\hat{\mathbf{R}}_{yy} = \frac{1}{L}\mathbf{y}_l\mathbf{y}_l^H$. Our goal would be to recover the support $S_0$ using $\hat{\mathbf{R}}_{yy}$.

### C. Analysis of Support Recovery in Presence of Cross Correlation

So far we have assumed that the non zero elements of the unknown sparse vector are statistically uncorrelated. We will now investigate the case where such an assumption in violated. We examine if it is still possible to have unique representation for sparse supports of size larger than $M$ and what role is played by the non zero cross correlation terms. In this case the non-ideal correlation matrix $\hat{\mathbf{R}}_{yy}$ is represented as $\hat{\mathbf{R}}_{yy} = \mathbf{A}\hat{\mathbf{R}}_{xx}\mathbf{A}^H$ where $\hat{\mathbf{R}}_{xx}$ is allowed to be non-diagonal. We prove that the presence of these cross correlation terms leads to an additional condition, reminiscent of a Signal-to-Noise Ratio (SNR)-like condition which needs to be satisfied for unambiguous recovery of the support. Under this condition, it is indeed possible to recover supports of larger size ($O(M^2)$) from the covariance matrix.

We now proceed to treat these two cases separately, under the common framework of "imperfect correlation-awareness". As we pointed out earlier, in both cases, the support can be recovered by solving the common problem (Co-LASSO).

## V. MMV Model and the Estimated Correlation Matrix

Let $\mathbf{y}[l] \in \mathcal{R}^{M\times 1}, 1 \leq l \leq L$ denote a set of $L$ multiple measurement vectors. Here $\mathbf{A} \in \mathcal{R}^{M\times N} = [\mathbf{a}_1 \quad \mathbf{a}_2 \cdots \mathbf{a}_N]$ is the measurement matrix, and $\mathbf{x}_s[l] \in \mathcal{R}^{N\times 1}, 1 \leq l \leq L$ are $L$ unknown vectors with the same sparsity $\|\mathbf{x}_s[l]\|_0 = D$ and common sparsity support denoted by the set $S_0 = \{i_0, i_1, \ldots, i_{D-1}\}$. We have

$$\mathbf{y}[l] = \mathbf{A}\mathbf{x}_s[l], \quad 1 \leq l \leq L. \tag{17}$$

Analogous to the assumptions (A1) and (A2), we now have the following assumptions:

1) (A1)' : $\mathrm{E}(\mathbf{x}_s[l]) = \mathbf{0}, \forall l$.
2) (A2)' : $\mathbf{x}_s[l]$ are i.i.d random vectors whose nonzero entries are zero mean continuous random variables that are statistically uncorrelated. In particular, $E([\mathbf{x}_s[l]]_i[\mathbf{x}_s[m]]_k) = \sigma_i^2\delta(i-k)\delta(l-m)$.

As we show shortly, the use of finite number of samples leads to an additive noise like term which becomes arbitrarily small as $L \to \infty$ under the assumptions (A1)'-(A2)'. We will analyze the effect of such a noise-like term on the uniqueness of sparse representation of the estimated covariance matrix. We will also study the performance of LASSO [71] used for support recovery from such a model.

The estimated sample correlation matrix is given by

$$\hat{\mathbf{R}}_{\mathbf{yy}}^{[L]} \triangleq \frac{1}{L}\sum_{l=1}^{L}\mathbf{y}[l]\mathbf{y}[l]^T = \mathbf{A}\hat{\mathbf{R}}_{xx}^{[L]}\mathbf{A}^T$$

where $\hat{\mathbf{R}}_{xx}^{[L]} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{x}_s[l]\mathbf{x}_s[l]^T$. On vectorization, we can write

$$\mathbf{z}^{[L]} \triangleq \mathrm{vec}(\hat{\mathbf{R}}_{yy}^{[L]}) = \underbrace{(\mathbf{A}\odot\mathbf{A})}_{\mathbf{A}_{KR}}\mathbf{r}_{xx}^{[L]} + \mathbf{e}^{[L]} \tag{18}$$

where $\mathbf{r}_{xx}^{[L]} \in \mathcal{R}^{N\times 1}$ is a $D$ sparse vector *with the same support* $S_0$ and non-zero elements given by

$$\sigma_i^{2[L]} \triangleq \frac{1}{L}\sum_{l=1}^{L}[\mathbf{x}_s[l]]_i^2, \quad i \in S_0 \tag{19}$$

Hence $\mathbf{r}_{xx}^{[L]}$ is a non-negative vector for all values of $L$. The imperfect knowledge of the ideal covariance matrix thus leads to an additive noise-like term $\mathbf{e}^{[L]}$. The elements of the vector $\mathbf{e}^{[L]} \in \mathcal{R}^{M^2\times 1}$ are given by

$$\frac{1}{L}\sum_{i\neq j}\sum_{l=1}^{L}\mathbf{A}_{m,i}\mathbf{A}_{n,j}[\mathbf{x}_s[l]]_i[\mathbf{x}_s[l]]_j, \;\; 1 \leq m,n \leq M. \tag{20}$$

They denote the estimates of cross-correlation between the elements of $\mathbf{y}[l]$. The fact that the sample cross-correlations are (unfortunately) nonzero for finite $L$ is represented by these terms. We can think of $\mathbf{z}^{[L]}$ as our effective Single Measurement Vector (derived from the MMV), $\mathbf{A}\odot\mathbf{A}$ as the effective measurement matrix and $\mathbf{e}^{[L]}$ as the effective additive noise term. We aim to recover the support $S_0$ of the unknown vector $\mathbf{r}_{xx}^{[L]}$ from this model.

## VI. Performance Analysis of LASSO for MMV Model

Since the imperfect knowledge of the ideal covariance matrix leads to an additive noise-like term $\mathbf{e}^{[L]}$, this provides the ideal framework for applying LASSO ($l_1$ norm regularized quadratic programming) [71] for sparse support recovery. Additionally, there is a constraint requiring the unknown sparse vector to be positive. So the problem of recovering the sparse support becomes finding the support of a vector $\hat{\mathbf{r}} \in \mathcal{R}^{N\times 1}$ which is the solution to the following constrained LASSO:

$$\begin{aligned}&\min_{\mathbf{r}}\left(\frac{1}{2}\|(\mathbf{A}\odot\mathbf{A})\mathbf{r} - \mathbf{z}^{[L]}\|_2^2 + h\|\mathbf{r}\|_1\right) \quad (LASSO)\\ &\text{subject to } \mathbf{r}\succeq\mathbf{0}\end{aligned} \tag{21}$$

In the analysis that follows, we will assume that the columns of $\mathbf{A}$ are unit-norm. Notice that this does not affect the conditions for recovering the support, since we can always think of multiplying $\mathbf{A}$ by an appropriate diagonal matrix on the right and modify the powers in $\mathbf{r}_{xx}^{[L]}$.

### A. Analysis of LASSO Independent of Distribution

In [43], [60], the performance of unconstrained LASSO for support recovery has been analyzed for both bounded and sub-Gaussian noise (for deterministic as well as stochastic measurement matrix $\mathbf{A}$). However, unlike [43], [60], we do not assume $\mathbf{e}^{[L]}$ to be bounded or sub-Gaussian. We only have assumptions on the second order statistics of $\mathbf{e}^{[L]}$ (which results from our original assumption about uncorrelatedness of the vectors) and not on their specific distributions. With this lone assumption, we compute a lower bound on the probability of successful recovery of the support. As a first step, the following theorem provides a set of sufficient conditions for the solution of the LASSO (21) to yield the true support of the sparse vectors. The theorem is inspired from [60]. In the remainder of the paper, we denote the coherence of $\mathbf{A}$ by $\mu_A$.

*Theorem 7:* Consider the model (18) where $S_0$ denotes the sparse support of $\mathbf{r}_{xx}^{[L]}$. If $\mid S_0 \mid < \frac{1}{2}\left(1+\frac{1}{\mu_A^2}\right)$, and the following two conditions hold:

$$\|\mathbf{e}^{[L]}\|_2 < h\frac{1+\mu_A^2-2\mu_A^2 \mid S_0 \mid}{1+\mu_A^2-\mu_A^2 \mid S_0 \mid}, \qquad \text{(Event } E1) \quad (22)$$

$$\sigma_i^{2[L]} > \frac{\|\mathbf{e}^{[L]}\|_2+h}{1+\mu_A^2-\mu_A^2 \mid S_0 \mid}, \quad \forall i \in S_0 \qquad \text{(Event } E2) \quad (23)$$

then the optimal solution $\mathbf{r}^*$ to the constrained LASSO (21) is unique and it satisfies $\text{Supp}(\mathbf{r}^*) = S_0$.

*Proof:* See Appendix A. ■

Some important remarks follow from the above theorem and its proof.

1) ***Conditions for Optimality:*** The conditions (22) and (23) in Theorem 7 are based on the coherence $\mu_{\mathbf{A}}$. A more general set of sufficient conditions for the same can be derived following the proof in Appendix A. These are:

$$h(1-\mathbf{a}_i^{(KR)^T}\mathbf{A}_{S_0}^{(KR)^{\dagger^T}}\mathbf{1}_{S_0}) > \mathbf{a}_i^{(KR)^T}\mathbf{e}^{[L]\perp}, \;\; i \in S_0^c, \;\; (24)$$

and

$$\left\|\mathbf{A}_{S_0}^{(KR)^\dagger}\mathbf{e}^{[L]} - h\left(\mathbf{A}_{S_0}^{(KR)^T}\mathbf{A}_{S_0}^{(KR)}\right)^{-1}\mathbf{1}_{S_0}\right\|_\infty < \sigma_{\min}^{2\;[L]} \quad (25)$$

As shown in Appendix A, if $\mid S_0 \mid < \frac{1}{2}\left(1+\frac{1}{\mu_A^2}\right)$, then (22) implies (24) and (23) implies (25). In future, we will be interested in deriving a different set of guarantees to satisfy (24) and (25) using properties of $\mathbf{A}$ which are less restrictive than its coherence.

2) ***Optimal Solution:*** If (24) and (25) hold, the optimal $\mathbf{r}^*$ solving the constrained LASSO can be shown to have the following form (as derived in Appendix A)

$$\mathbf{r}^*_{S_0^c} = 0$$

and

$$\mathbf{r}^*_{S_0} = \mathbf{r}_{xx\,S_0}^{[L]} + \mathbf{A}_{S_0}^{(KR)^\dagger}\mathbf{e}^{[L]} - h\left(\mathbf{A}_{S_0}^{(KR)^T}\mathbf{A}_{S_0}^{(KR)}\right)^{-1}\mathbf{1}_{S_0}$$

Also, (25) further implies $\mathbf{r}^*_{S_0} \succ \mathbf{0}$ and hence $\text{Supp}(\mathbf{r}^*) = S_0$.

Notice that conditions (22) and (23) involve the random variables $\sigma_i^{2[L]}$ and $\mathbf{e}^{[L]}$. Since events $E_1$ and $E_2$ are sufficient conditions for support recovery, if we can develop a lower bound on the probability of joint occurrence of the events $E_1$ and $E_2$, then the probability of successful support recovery by solving LASSO can also be lower bounded by the same quantity. Please note that success is defined as the event when the recovered support exactly equals the true support $S_0$. Denote $\mu_i^{(4)} \triangleq E([\mathbf{x}_s[l]]_i^4)$ and $\sigma_{\min}^2 = \min\{\sigma_i^2, i \in S_0\}$

*Lemma 6:* Consider the model (18) and let (A1)' and (A2)' hold. If $0 < h < \sigma_{\min}^2\frac{(1+\mu_A^2-\mu_A^2|S_0|)^2}{2(1+\mu_A^2)-3\mu_A^2|S_0|}$, the probability of the event $E_1 \bigcap E_2$ is bounded below by $\mathcal{P}(E_1 \bigcap E_2) > 1 - \frac{C(h)}{L} + O\left(\frac{1}{L^2}\right)$ where the constant $C(h)$ is independent of $L$ and depends on $h$ as well as on coherence $\mu_A$, sparsity $D$ and the matrix $\mathbf{A}$ as

$$C(h) = \sum_{i=1}^{D}\frac{\mu_i^{(4)}-\sigma_i^4}{\left(\sigma_i^2 - h\frac{2(1+\mu_A^2)-3\mu_A^2D}{1+\mu_A^2-\mu_A^2D}\right)^2} + \frac{M^2(1+\mu_A^2-\mu_A^2D)^2}{h^2(1+\mu_A^2-2\mu_A^2D)^2}\sum_{i\neq j}^{D}\sigma_i^2\sigma_j^2\Big(\|\mathbf{a}_i\|_2^2\|\mathbf{a}_j\|_2^2 + (\mathbf{1}_{M\times1}^T[\mathbf{a}_i \circ \mathbf{a}_j])^2\Big)$$

*Proof:* See Appendix B. ■

Combining Theorem 7 and Lemma 6, we obtain the following theorem which sums up the desired result on the probability of successful recovery:

*Theorem 8:* Consider the MMV model (17) with the assumptions given by (A1)'-(A2)'. If

$$\mid S_0 \mid < \frac{1}{2}\left(1+\frac{1}{\mu_A^2}\right), \quad \text{and}$$
$$0 < h < \sigma_{\min}^2\frac{(1+\mu_A^2-\mu_A^2 \mid S_0 \mid)^2}{2(1+\mu_A^2)-3\mu_A^2 \mid S_0 \mid},$$

then the common support $S_0$ can be recovered by solving LASSO given by (21), with probability greater than $1-\frac{1}{L}C(h)+O\left(\frac{1}{L^2}\right)$ ♦

The following points are worth noticing:

1) When $L \to \infty$, one can guarantee the recovery of any sparsity level upto $\frac{1}{2}\left(1+\frac{1}{\mu_A^2}\right)$ with probability 1. Since we used our assumption only on the second order statistics of the random vectors (and not their distribution), the above probability of successful recovery was bounded using Chebyshev Inequality. Consequently, the upper bound on probability of failure decays as $\frac{1}{L}$ which can be rather loose for smaller values of $L$. However, with more assumptions on the distribution of nuisance terms (such as sub-Gaussian tail etc), the bound can be tightened by using Chernoff-like bounds as we demonstrate next. It is also to be noted that unlike most analyses (including that in [43]) on sparse recovery which considers probability of successful recovery over an ensemble of random measurement matrices, the current analysis has been performed for a fixed measurement matrix $\mathbf{A}$ and the probability is considered with respect to the distribution of the random source vectors $\mathbf{x}_s[l]$. Our goal is to characterize how it varies with $L$ (snapshots), rather than with the size of $\mathbf{A}$ as done in compressive sensing literature.

2) *Role of cross-correlation*: The non-zero cross-correlation terms arise from the finite number of snapshots and their norm decreases with high probability as $L$ increases. Hence, we can bound their norm at any given value of $L$ with a probability that increases with $L$, and this in turn dictates the probability of successful recovery using LASSO. There is another interpretation of the role played by the cross-correlation in terms of the number of identifiable sources. For a given $L$ (and hence a given value of the non-zero cross-correlation and estimated signal power), the right hand side of the inequalities (22) and (23) can always be reduced by *decreasing* $|S_0|$ until one satisfies the

inequalities. Hence, for a given level of cross-correlation and signal power, one can always guarantee the recovery of a possibly smaller number of sources.

### B. Analysis of LASSO for Gaussian Model

In the previous subsection, the analysis of LASSO (21) was performed without assuming any specific probability distribution for the source vectors. In this subsection, we assume a specific distribution on the unknown source vectors (namely, multivariate Gaussian Distribution) and show that it is possible to guarantee support recovery with overwhelming probability by increasing $L$. In particular, we will show that the probability of support recovery increases as $\mathcal{P}_s \geq 1 - C\beta^{-L}, \quad C > 0, \beta > 1$.

*1) Concentration Inequalities:* We first state a concentration inequality ([61], Lemma 6) that will be directly used in our derivation:

*Lemma 7:* Let each of $x_i$ and $y_i$, $i = 1, \ldots, k$ be real sequences of uncorrelated zero mean Gaussian random variables with variance $\sigma_x^2$ and $\sigma_y^2$ respectively. Then

$$\mathcal{P}\left(\left|\sum_{i=1}^{k} x_i y_i\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\sigma_x\sigma_y(2\sigma_x\sigma_y k + t)}\right)$$

♦

Using this lemma, we can get the following result. The notation $\sigma_{\max}^{(k)}$ denotes the $k$th largest element in the set of non-negative numbers $\{\sigma_i\}_{i=1}^{|S_0|}$.

*Lemma 8:* Consider the model (18). Assuming that (A1)',(A2)' hold and that $\mathbf{x}_s[l]$ follows Gaussian distribution, the following holds $\forall 1 \leq i \leq M^2$ (where $t \triangleq \frac{C}{\|\mathbf{A}\|_{\infty,\infty}^2 |S_0|(|S_0|-1)}$):

$$\mathcal{P}\left(|[\mathbf{e}^{[L]}]_i| \geq C\right) \leq 2\exp\left(-\frac{Lt^2}{2\sigma_{\max}^{(1)}\sigma_{\max}^{(2)}(2\sigma_{\max}^{(1)}\sigma_{\max}^{(2)} + t)}\right),$$

♦

*Proof:* See Appendix C ∎

The following concentration inequality directly provides a bound on the probability of occurrence of the event $E_2$.

*Lemma 9:* Let $x_i, i = 1, \ldots, L$ denote i.i.d zero mean Gaussian random variables with variance $\sigma_x^2$. Assume $0 < C < \sigma_x^2$. Then, there exists $\beta > 1$ such that $\mathcal{P}\left(\frac{1}{L}\sum_{i=1}^{L} x_i^2 > C\right) \geq 1 - \beta^{-L}$ ♦

*Proof:* See Appendix D ∎

*2) Probability of Support Recovery by Solving the LASSO:* Armed with the inequalities provided by Lemmas 8 and 9, we now state our main result on the probability of support recovery by solving the LASSO (21):

*Theorem 9:* Consider the MMV model given by (17) which satisfies the assumptions (A1)'-(A2)'. If the columns of $\mathbf{A}$ are unit norm and $|\ S_0 \ | < \frac{1}{2}\left(1 + \frac{1}{\mu_A^2}\right)$, and $0 < h < \sigma_{\min}^2 \frac{(1+\mu_A^2 - \mu_A^2|S_0|)^2}{2(1+\mu_A^2) - 3\mu_A^2|S_0|}$, then the common support $S_0$ can be recovered by solving the LASSO (21), with probability greater than $1 - \alpha\gamma^{-L}$ for some $\gamma > 1$. ♦

*Proof:* Using Theorem 7, we can say that the probability of successful recovery of sparse support by solving the LASSO (21), denoted as $P_s$, satisfies: $P_s \geq \mathcal{P}(E_1 \bigcap E_2)$ where events $(E1)$ and $(E2)$ are defined in (22) and (23). Using a technique similar to the proof of Lemma 6 in Appendix B (see (43)), it can be shown that

$$\mathcal{P}\left(E_1 \bigcap E_2\right) \geq \prod_{i=1}^{D} \mathcal{P}(\sigma_i^{2\,[L]} > c_2) - \sum_{i=1}^{M^2} \mathcal{P}\left(|\ [\mathbf{e}^{[L]}]_i \,| \geq \frac{c_1}{M}\right)$$

Here $c_1$ and $c_2$ are given by (42) in Appendix B Now using the expression for $\sigma_i^{2\,[L]}$ from (19), we can say, from Lemma 9 that

$$\prod_{i=1}^{|S_0|} \mathcal{P}(\sigma_i^{2\,[L]} > c_2) \geq \prod_{i=1}^{|S_0|} (1 - \beta_i^{-L}) \tag{26}$$

and using Lemma 8, we get

$$\sum_{i=1}^{M^2} \mathcal{P}\left(|\ [\mathbf{e}^{[L]}]_i \,| \geq \frac{c_1}{M}\right) \leq 2M^2 e^{-\delta L} \tag{27}$$

where $\delta \triangleq \frac{t^2}{2\sigma_{\max}^{(1)}\sigma_{\max}^{(2)}(2\sigma_{\max}^{(1)}\sigma_{\max}^{(2)} + t)}$ and $t \triangleq \frac{c_1}{M\|\mathbf{A}\|_{\infty,\infty}^2 |S_0|(|S_0|-1)}$. Using (26) and (27) in (43), we obtain

$$\mathcal{P}_s \geq \prod_{i=1}^{|S_0|} (1 - \beta_i^{-L}) - 2M^2 e^{-\delta L} \tag{28}$$

which proves the desired result since each $\beta_i > 1$. ∎

It is to be noted that in Theorems 8 and 9, the size of the recoverable support $|S_0|$ is bounded by a quantity that depends on the coherence $\mu_{\mathbf{A}}$.It is known that the Welch bound [45] dictates that $\mu_{\mathbf{A}}^2 = O\left(\frac{1}{M}\right)$. Hence, the guarantees on the performance of LASSO developed in this section are valid for supports only as large as $O(M)$. However, in our experiments in Section VII, we consider sparsity levels satisfying $|S_0| \gg M$ and show that the proposed LASSO can indeed successfully recover the support as $L$ becomes large. In future, we will be interested in developing stronger guarantees on the success of LASSO to recover supports of size $O(M^2)$, possibly using sufficient conditions such as the Restricted Isometry Property (RIP) [54].

## VII. Numerical Results

We conducted several experiments to test the proposed correlation-aware support recovery. We choose $\mathbf{A} \in \mathcal{R}^{M\times N}$ with elements $\mathbf{a}_{ij}$ being zero mean i.i.d Gaussian with unit variance. We fix $M = 20, N = 256$ and consider a sparsity level of $|\ S_0 \ | = 63 \gg M$. We consider a fixed realization of a random $\mathbf{A}$ and generate i.i.d. Gaussian $\mathbf{x}_s[l], \quad 1 \leq l \leq L$.

We compare the proposed LASSO on the covariance matrix with two traditional MMV algorithms:

1) MMV BP: Basis pursuit for MMV model. We used the SPGL1 solver for MATLAB from the website: http://www.cs.ubc.ca/ mpf/spgl1/.
2) item MFOCUSS [58]: Modification of FOCUSS for MMV model. We set the p-norm to 0.8 for our simulations.

Furthermore, we consider two interpretations of percentage of successful recovery as follows:

### A. Case I

In the first case we define the metric as follows.

- Percentage of successful recovery (Case I): This represents what percentage of the recovered indices contain the entire true support. If the true support is not entirely included in the support of the recovered vector, we consider the recovery to have failed.

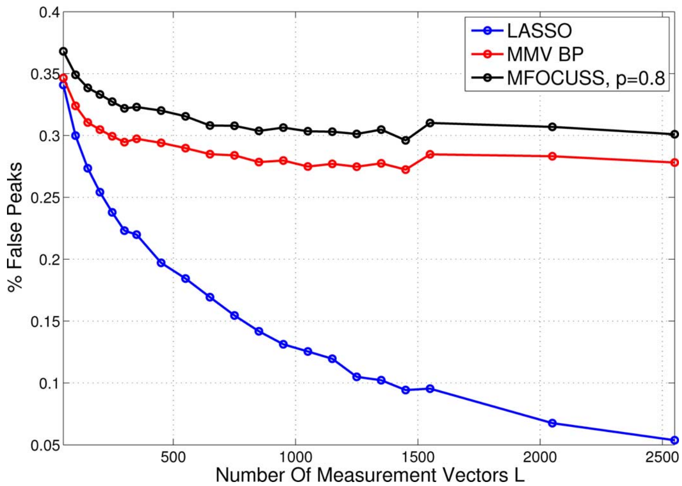



Fig. 1. Probability of False Detection as a function of $L$, in the regime $|S_0| > M$. Here $M = 20$, $N = 156$, $|S_0| = 63$.

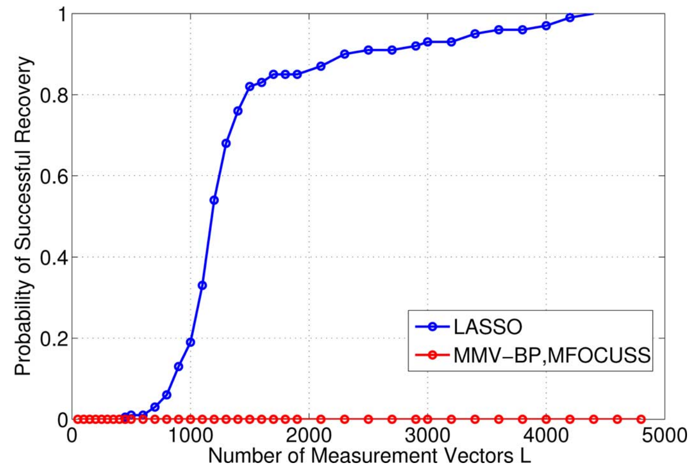



Fig. 3. Probability of Successful Recovery (Case II) as a function of $L$ in the regime $|S_0| > M$. Here $M = 20$, $N = 156$, $|S_0| = 63$.

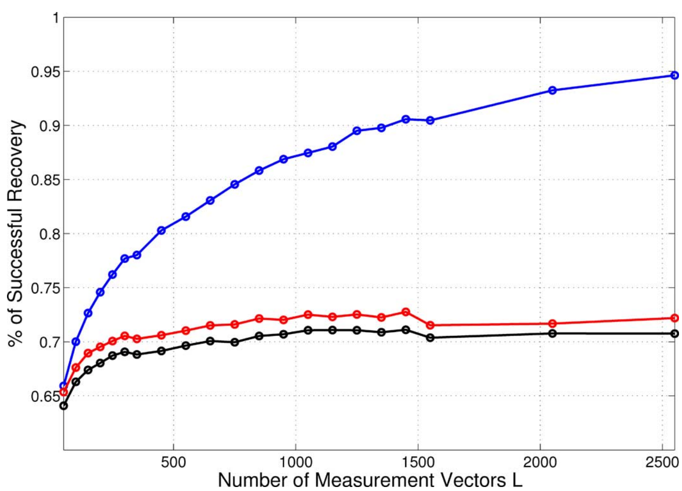



Fig. 2. Percentage of true support recovered (Case I) as a function of $L$, in the regime $|S_0| > M$. Here $M = 20$, $N = 156$, $|S_0| = 63$.

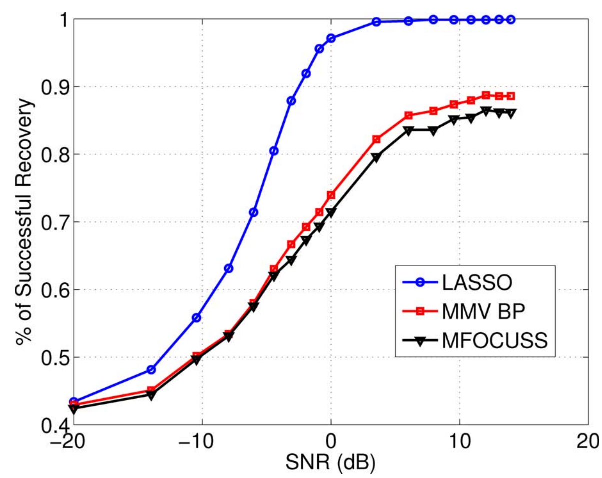



Fig. 4. Percentage of true support recovered (Case I), as a function of SNR, in the regime $|S_0| > M$. Here $M = 20$, $N = 156$, $|S_0| = 50$, $L = 2000$.

- Percentage of false peaks: This represents what percentage of the recovered indices correspond to incorrect indices.

In Figs. 1 and 2, we plot the above two performance metrics as a function of $L$.

The above two quantities add up to 100% as is evident from the plots. It can be seen that, for this sparsity level, there is a drastic difference between the performance of the proposed method and the traditional approaches. With increasing $L$, the proposed LASSO can recover most of the 63 indices correctly whereas the traditional approaches show little improvement. It can also be seen that the proposed framework outperforms MMV-BP by a significant margin. This proves its effectiveness in dealing with sparsity levels larger than $M$.

### B. Case II

In this case, we have a stricter interpretation for probability of successful recovery (as in Theorems 8 and 9). We define it as the probability with which the recovered support is *exactly equal* to the true support. We plot this metric as a function of $L$ in Fig. 3 for $M = 20$, $N = 156$, $|S_0| = 63 \gg M$. The observation is that given this level of sparsity and this particular interpretation of probability of success, MMV-BP and MFOCUSS are incapable of recovering the entire true support at all values of $L$. However, the proposed LASSO can successfully recover the exact support, especially as $L$ becomes large. This is evidently because our approach is based on the covariance matrix and can hence successfully exploit the correlation priors.

### C. Performance Evaluation With Varying SNR

So far in the simulations, we did not consider the effect of additive white noise. We now proceed to evaluating the preceding three metrics, viz., probability of false detection, percentage of successful recovery (Case I), and probability of success (Case II) by adding white Gaussian noise to the signal model (17). Consider the noisy model (??) where $\mathbf{v}[l]$ represents additive white Gaussian noise with power $\sigma_n^2$. We consider the same experimental setup as before and fix $M = 20$, $N = 156$, $|S_0| = 50$, $L = 2000$. We vary the signal to noise ratio (SNR) from $-20$ dB to 14 dB. For each value of SNR, the probability estimates are computed by averaging over 500 Monte Carlo Runs. Fig. 4 shows the percentage of true support recovered (Case I), Fig. 5 shows the probability of success and Fig. 6 shows the probability of false detection, all as functions of SNR. While the performance of LASSO, MMVBP

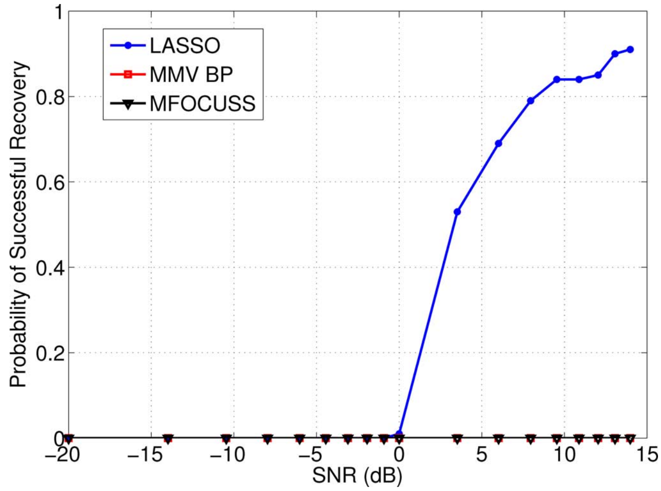



Fig. 5. Probability of success (Case II) as a function of SNR, in the regime $|S_0| > M$. Here $M = 20$, $N = 156$, $|S_0| = 50$, $L = 2000$.

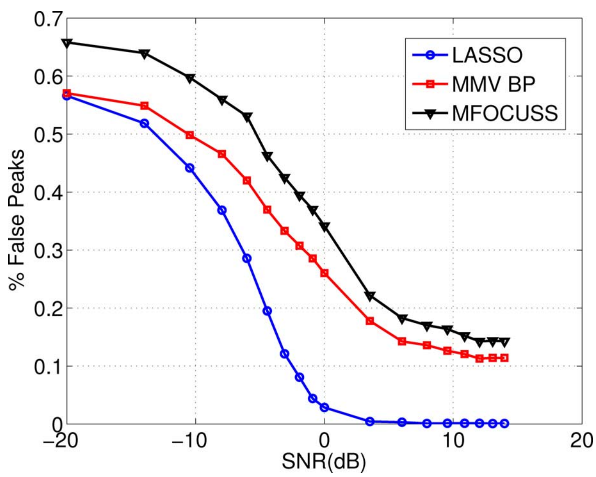



Fig. 6. Probability of false detection as a function of SNR, in the regime $|S_0| > M$. Here $M = 20$, $N = 156$, $|S_0| = 50$, $L = 2000$.

and MFOCUSS improve with SNR, LASSO outperforms the other methods throughout the entire range of SNR considered. LASSO is the only method that recovers the true support with very high probability, without producing false peaks, as evident from Fig. 5. The other methods invariably result in false peaks and hence their probability of success (Case II) is always 0.

## VIII. Conclusion

In this paper, we proposed a correlation-aware framework for sparse support recovery. Using priors on the unknown sparse signal which assume that the non-zero entries are statistically uncorrelated, we showed that it is possible to have unique representation for the support even when the sparsity level exceeds the dimension of the measured vector(s). We developed new conditions based on the Khatri-Rao product of the measurement matrix to allow unique representation. We also developed coherence based guarantees for a modified $l_1$ problem for support recovery from our proposed framework. We further exploited the non negativity of the solution and provided necessary and sufficient conditions for support recovery from the covariance matrix. These conditions were shown to naturally justify the choice of sparse sampling schemes such as nested and coprime arrays. The results developed in this paper point to the effectiveness of our proposed formulation for recovering sparse supports of much larger size than what is considered in existing literature.

We further evaluated the performance of our proposed algorithm when we only have estimates of correlation obtained from a finite number ($L$) of measurements in the MMV model. Finite number of measurements was also shown to have a noise-like effect on the proposed model for support recovery and hence the use of LASSO was proposed. We analyzed the performance of LASSO for support recovery by probabilistically characterizing the effect of estimated correlation values. We first performed a general analysis without considering any specific distribution for the unknown sparse signal and then extended this analysis to the case where it belongs to a multivariate Gaussian distribution with diagonal covariance matrix. We showed that it is possible to recover the support with a probability which grows overwhelmingly fast with the number of measurements. The theory presented in this paper leads to several interesting problems for future. For instance, it gives rise to many interesting questions about how to design measurement matrices, which can exploit the correlation priors (via its Khatri-Rao product). The presence of limited cross-correlation and effect of additive noise also need to be analyzed in detail. Other kind of priors on the distribution (such as higher order moments) can also be explored to possibly further increase the size of the recoverable sparse support. In conclusion, this paper provides a first step towards explicit use of correlation priors in sparse recovery which can lead to many future directions for prior-aware sparse recovery techniques.

## Appendix

*Proof of Theorem 7:*

We introduce Lagrange multipliers $\lambda_i$, $1 \leq i \leq N$ corresponding to the constraints $-\mathbf{r} \preceq \mathbf{0}$, with $\lambda_i \geq 0$ for each $i$. Using the KKT conditions [52], we obtain that $(\mathbf{r}^*, \boldsymbol{\lambda}^*)$ constitute a primal-dual optimal solution pair if and only if

$$\left(\mathbf{A}^{(KR)^T}\mathbf{A}^{(KR)}\right)\mathbf{r}^* - \mathbf{A}^{(KR)^T}\mathbf{z}^{[L]} + h\mathbf{1}_{N\times 1} - \boldsymbol{\lambda}^* = 0 \tag{29}$$

$$[\boldsymbol{\lambda}^*]_i[\mathbf{r}^*]_i = 0, \ 1 \leq i \leq N \tag{30}$$

$$\boldsymbol{\lambda}^* \succeq \mathbf{0}, \mathbf{r}^* \succeq \mathbf{0} \tag{31}$$

Here $\mathbf{A}^{(KR)} = \mathbf{A} \odot \mathbf{A}$. Recall that $S_0$ denotes the set containing the indices indicating the true support of the sparse vector $\mathbf{r}_{xx}^{[L]}$ that generates the observed $\mathbf{z}^{[L]}$. Given a vector $\mathbf{v}$, and a set $S$ of integers, let us denote $\mathbf{v}_S$ as the $|S| \times 1$ subvector of $\mathbf{v}$ containing elements of $\mathbf{v}$ whose indices are given by the set $S$. Similary for a matrix $\mathbf{A}$, $\mathbf{A}_S$ is the submatrix composed of $|S|$ columns indexed by $S$. Let us define $\mathbf{r}^* = \hat{\mathbf{r}}$ and $\boldsymbol{\lambda}^* = \hat{\boldsymbol{\lambda}}$ as below:

$$\hat{\mathbf{r}}_{S_0^c} \triangleq \mathbf{0}, \tag{32}$$

$$\hat{\boldsymbol{\lambda}}_{S_0} \triangleq \mathbf{0} \tag{33}$$

$$\hat{\mathbf{r}}_{S_0} \triangleq \mathbf{r}_{xx\,S_0}^{[L]} + \mathbf{A}_{S_0}^{(KR)^\dagger}\mathbf{e}^{[L]} - h\left(\mathbf{A}_{S_0}^{(KR)^T}\mathbf{A}_{S_0}^{(KR)}\right)^{-1}\mathbf{1}_{S_0} \tag{34}$$

$$\hat{\boldsymbol{\lambda}}_{S_0^c} \triangleq h\mathbf{1}_{S_0^c} - \mathbf{A}_{S_0^c}^{(KR)^T}\left(\mathbf{e}^{[L]^\perp} + h\mathbf{A}_{S_0}^{(KR)^{\dagger^T}}\mathbf{1}_{S_0}\right) \tag{35}$$

Here $S_0^c = [N] \setminus S_0$ and $\mathbf{e}^{[L]\perp} = \mathbf{e}^{[L]} - \mathbf{A}_{S_0}^{(KR)} \mathbf{A}_{S_0}^{(KR)^\dagger} \mathbf{e}^{[L]}$. Now, if $h$ is such that:

$$h(1 - \mathbf{a}_i^{(KR)^T} \mathbf{A}_{S_0}^{(KR)^{\dagger^T}} \mathbf{1}_{S_0}) > \mathbf{a}_i^{(KR)^T} \mathbf{e}^{[L]\perp}, \;\; i \in S_0^c, \quad (36)$$

then $\hat{\boldsymbol{\lambda}} \succeq \mathbf{0}$. Similarly, if

$$\|\mathbf{A}_{S_0}^{(KR)^\dagger} \mathbf{e}^{[L]} - h\left(\mathbf{A}_{S_0}^{(KR)^T} \mathbf{A}_{S_0}^{(KR)}\right)^{-1} \mathbf{1}_{S_0}\|_\infty < \sigma_{\min}^{2\;[L]} \quad (37)$$

then $\hat{\mathbf{r}}_{S_0} \succ \mathbf{0}$ which implies

$$\mathrm{Supp}(\hat{\mathbf{r}}) = S_0 \quad (38)$$

Hence if (36) and (37) hold, then $(\hat{\mathbf{r}}, \hat{\boldsymbol{\lambda}})$ as defined by (32)–(35) satisfy (29)–(31) and thereby constitute an optimal primal-dual pair.

Now, if (36) holds, then it implies from (35)

$$[\hat{\boldsymbol{\lambda}}]_i > 0, \quad \forall i \in S_0^c \quad (39)$$

This condition has an important implication in recovery of true support. Any optimal $\mathbf{r}^*$ must now satisfy (30) with the particular optimal dual variable $\hat{\boldsymbol{\lambda}}$ satisfying (39) and this immediately implies: $\mathrm{Supp}(\mathbf{r}^*) \subseteq S_0$

We assume now that (36) and (37) hold true. Then, for any primal optimal $\mathbf{r}^*$ and the particular dual optimum $\hat{\boldsymbol{\lambda}}$, we obtain from (29) that $\mathbf{r}^*_{S_0} = \hat{\mathbf{r}}_{S_0}$ where $\hat{\mathbf{r}}_{S_0}$ is as in (34).

Hence, any optimal $\mathbf{r}^*$ is necessarily equal to $\hat{\mathbf{r}}$, thereby making the primal solution unique. Since (37) holds, we obtain from (38), $\mathrm{Supp}(\mathbf{r}^*) = S_0$.

In conclusion, we can say that the optimal solution of the LASSO (21) is unique and its support will be equal to the true support $S_0$ provided (36) and (37) hold. Following an approach similar to [60], it can be shown that if $D = \mid S_0 \mid < \frac{1}{2}\left(1 + \frac{1}{\mu_A^2}\right)$, then (36) is satisfied provided

$$\|\mathbf{e}^{[L]}\|_2 < h \frac{1 + \mu_A^2 - 2\mu_A^2 D}{1 + \mu_A^2 - \mu_A^2 D} \quad (40)$$

and (37) is satisfied if

$$\sigma_{\min}^{2\;[L]} > \frac{\|\mathbf{e}^{[L]}\|_2 + h}{1 + \mu_A^2 - \mu_A^2 D} \quad (41)$$

Summarizing, if (22) and (23) hold, then $\mathrm{Supp}(\mathbf{r}^*) = S_0$ for any optimal $\mathbf{r}^*$.

*Proof of Lemma 6:*

Let $D = \mid S_0 \mid$. The probability of successful recovery of sparse support by solving the LASSO (21), denoted as $P_s$, satisfies:

$$\begin{aligned} P_s &\geq \mathcal{P}\left(E_1 \bigcap E_2\right) \\ &= \mathcal{P}\left( \|\mathbf{e}^{[L]}\|_2 < h \frac{1 + \mu_A^2 - 2\mu_A^2 \mid S_0 \mid}{1 + \mu_A^2 - \mu_A^2 \mid S_0 \mid}, \right. \\ &\qquad \left. \sigma_{\min}^{2\;[L]} > \frac{\|\mathbf{e}^{[L]}\|_2 + h}{1 + \mu_A^2 - \mu_A^2 D} \right) \end{aligned}$$

If $(E_1)$ holds, then $\sigma_{\min}^{2\;[L]} > \frac{h\frac{1+\mu_A^2 - 2\mu_A^2 D}{1+\mu_A^2-\mu_A^2 D} + h}{1+\mu_A^2 - \mu_A^2 D} \Rightarrow \sigma_{\min}^{2\;[L]} > \frac{\|\mathbf{e}^{[L]}\|_2 + h}{1+\mu_A^2-\mu_A^2 D}$. Using this, we get $P_s \geq \mathcal{P}\Big(\|\mathbf{e}^{[L]}\|_2 < c_1, \sigma_{\min}^{2\;[L]} > c_2\Big)$ where

$$c_1 \triangleq h \frac{1 + \mu_A^2 - 2\mu_A^2 D}{1 + \mu_A^2 - \mu_A^2 D}, \quad c_2 \triangleq h \frac{2(1 + \mu_A^2) - 3\mu_A^2 D}{(1 + \mu_A^2 - \mu_A^2 D)^2}. \quad (42)$$

Now

$$\begin{aligned} P_s &\geq \mathcal{P}(\|\mathbf{e}^{[L]}\|_2 < c_1, \sigma_{\min}^{2\;[L]} > c_2) \\ &\geq \mathcal{P}\left( \bigcap_{i=1}^{M^2} \left\{ \mid [\mathbf{e}^{[L]}]_i \mid < \frac{c_1}{M} \right\}, \sigma_{\min}^{2\;[L]} > c_2 \right) \\ &= \mathcal{P}(\sigma_{\min}^{2\;[L]} > c_2) - \mathcal{P}\left( \sigma_{\min}^{2\;[L]} > c_2, \bigcup_{i=1}^{M^2} \left\{ \mid [\mathbf{e}^{[L]}]_i \mid \geq \frac{c_1}{M} \right\} \right) \\ &\geq \mathcal{P}(\sigma_{\min}^{2\;[L]} > c_2) - \mathcal{P}\left( \bigcup_{i=1}^{M^2} \left\{ \mid [\mathbf{e}^{[L]}]_i \mid \geq \frac{c_1}{M} \right\} \right) \\ &\geq \mathcal{P}(\sigma_{\min}^{2\;[L]} > c_2) - \sum_{i=1}^{M^2} \mathcal{P}\left( \mid [\mathbf{e}^{[L]}]_i \mid \geq \frac{c_1}{M} \right) \\ &= \prod_{i=1}^{D} \mathcal{P}(\sigma_i^{2[L]} > c_2) - \sum_{i=1}^{M^2} \mathcal{P}\left( \mid [\mathbf{e}^{[L]}]_i \mid \geq \frac{c_1}{M} \right) \end{aligned} \quad (43)$$

Now, recall from (20) that each element $\mathbf{e}_k^{[L]}$ of the vector $\mathbf{e}^{[L]}$ is of the form $\frac{1}{L}\sum_{l=1}^{L} \sum_{i \neq j} \mathbf{A}_{m,i} \mathbf{A}_{n,j} [\mathbf{x}_s[l]]_i [\mathbf{x}_s[l]]_j$ for some $m, n \in \{1, 2, \ldots, M\}$. Therefore, $E(\mathbf{e}_k^{[L]}) = \frac{1}{L}\sum_{l=1}^{L} \sum_{i \neq j} \mathbf{A}_{m,i} \mathbf{A}_{n,j} E\Big([\mathbf{x}_s[l]]_i [\mathbf{x}_s[l]]_j\Big) = 0$ due to assumption (A1). Furthermore, using (A1) it can be shown that

$$E(\mathbf{e}_k^{[L]^2}) = \frac{\sigma_i^2 \sigma_j^2}{L} \sum_{i \neq j} \left( \mathbf{A}_{m,i}^2 \mathbf{A}_{n,j}^2 + \mathbf{A}_{m,i} \mathbf{A}_{m,j} \mathbf{A}_{n,i} \mathbf{A}_{n,j} \right) \quad (44)$$

If $h > 0$ which implies $c_1 > 0$, applying Chebyshev's Inequality, we obtain

$$\mathcal{P}\left( \mid \mathbf{e}_k^{[L]} \mid \geq \frac{c_1}{M} \right) \leq \frac{E(\mathbf{e}_k^{[L]^2}) M^2}{c_1^2} \quad (45)$$

Substituting (44) in (45), we obtain

$$\begin{aligned} &\sum_{k=1}^{M^2} \mathcal{P}\left( \mid \mathbf{e}_k^{[L]} \mid \geq \frac{c_1}{M} \right) \\ &\quad \leq \frac{M^2}{c_1^2} \sum_{k=1}^{M^2} E(\mathbf{e}_k^{[L]^2}) \\ &\quad = \frac{M^2}{c_1^2 L} \sum_{i \neq j} \sigma_i^2 \sigma_j^2 \Big( \|\mathbf{a}_i\|_2^2 \|\mathbf{a}_j\|_2^2 + (\mathbf{1}^T [\mathbf{a}_i \circ \mathbf{a}_j])^2 \Big) \end{aligned} \quad (46)$$

On the other hand, from (19), we obtain $E(\sigma_i^{2[L]}) = \sigma_i^2$. Under the assumption of the theorem $h < \sigma_{\min}^2 \frac{(1+\mu_A^2-\mu_A^2 D)^2}{2(1+\mu_A^2)-3\mu_A^2 D}$, we have $\sigma_i^2 > c_2 = h\frac{2(1+\mu_A^2)-3\mu_A^2 D}{(1+\mu_A^2-\mu_A^2 D)^2}, \quad 1 \leq i \leq D$. Hence we can say that

$$\mathcal{P}(\sigma_i^{[L]^2} > c_2) \geq \mathcal{P}(\mid \sigma_i^{[L]^2} - \sigma_i^2 \mid < \sigma_i^2 - c_2) \quad (47)$$

Applying Chebyshev Inequality, we get

$$\mathcal{P}(\mid {\sigma_i^{[L]}}^2 - \sigma_i^2 \mid \geq \sigma_i^2 - c_2) \leq \frac{\mu_i^{(4)} - \sigma_i^4}{L(\sigma_i^2 - c_2)^2} \tag{48}$$

since $E\Big(\boldsymbol{\sigma}_i^{2[L]} - E(\boldsymbol{\sigma}_i^{2[L]})\Big)^2 = \frac{1}{L}(\mu_i^{(4)} - \sigma_i^4)$. From (47) and (48), we get

$$\mathcal{P}({\sigma_i^{[L]}}^2 > c_2) \geq 1 - \frac{\mu_i^{(4)} - \sigma_i^4}{L(\sigma_i^2 - c_2)^2} \tag{49}$$

Using (46) and (49) in (43), we obtain the desired bound on the probability of success.

*Proof of Lemma 8:*

Let $D = \mid S_0 \mid$. Notice from (20) that $|[\mathbf{e}^{[L]}]_i| \leq \frac{\|\mathbf{A}\|_{\infty,\infty}^2}{L} \sum_{k\neq j}\sum_{l=1}^{L} \Big|[\mathbf{x}_s[l]]_k[\mathbf{x}_s[l]]_j\Big|$ Therefore $|[\mathbf{e}^{[L]}]_i| \geq C$ implies

$$\sum_{k\neq j}\sum_{l=1}^{L} \Big|[\mathbf{x}_s[l]]_k[\mathbf{x}_s[l]]_j\Big| \geq \frac{CL}{\| \mathbf{A} \|_{\infty,\infty}^2} \tag{50}$$

Also, notice that

$$\sum_{k\neq j}\sum_{l=1}^{L} \Big|[\mathbf{x}_s[l]]_k[\mathbf{x}_s[l]]_j\Big| \geq \frac{CL}{\| \mathbf{A} \|_{\infty,\infty}^2} \implies$$
$$\sum_{l=1}^{L} \Big|[\mathbf{x}_s[l]]_{k_0}[\mathbf{x}_s[l]]_{j_0}\Big| > \frac{CL}{\| \mathbf{A} \|_{\infty,\infty}^2 D(D-1)} \tag{51}$$

for some $k_0, j_0 \in \{1,\ldots,D\}, k_0 \neq j_0$. Using (50) and (51), we can say that

$$\mathcal{P}\Big(|[\mathbf{e}^{[L]}]_i| \geq C\Big) \leq \mathcal{P}\left(\sum_{l=1}^{L} \Big|[\mathbf{x}_s[l]]_{k_0}[\mathbf{x}_s[l]]_{j_0}\Big| > \frac{CL}{\| \mathbf{A} \|_{\infty,\infty}^2 D(D-1)}\right) \tag{52}$$

Under assumption (A1), $\{[\mathbf{x}_s[l]]_j\}_{l=1}^L$ and $\{[\mathbf{x}_s[l]]_k\}_{l=1}^L$ denote sequences of i.i.d. zero mean Gaussian random variables which are independent of each other. Hence, given $k \neq j$, we can use Lemma 7 to obtain

$$\mathcal{P}\left(\sum_{l=1}^{L} \Big|[\mathbf{x}_s[l]]_{k_0}[\mathbf{x}_s[l]]_{j_0}\Big| > \frac{CL}{\| \mathbf{A} \|_{\infty,\infty}^2 D(D-1)}\right)$$
$$\leq 2\exp\left(-\frac{Lt^2}{2\sigma_{k_0}\sigma_{l_0}(2\sigma_{k_0}\sigma_{l_0} + t)}\right)$$
$$\leq 2\exp\left(-\frac{Lt^2}{2\sigma_{\max}^{(1)}\sigma_{\max}^{(2)}(2\sigma_{\max}^{(1)}\sigma_{\max}^{(2)} + t)}\right)$$

The desired result directly follows from above and (52).

*Proof of Lemma 9:*

The proof is based upon Chernoff Bound. Denote $p_x^{[L]} \triangleq \frac{1}{L}\sum_{i=1}^{L} x_i^2$. Then $\mathcal{P}(p_x^{[L]} > C) = \mathcal{P}\Big(\sum_{i=1}^{L} x_i^2 > CL\Big) = \mathcal{P}\Big(\sum_{i=1}^{L} z_i^2 > \frac{CL}{\sigma_x^2}\Big)$ where $z_i$ denote i.i.d zero mean standard normal variables. Therefore $\sum_{i=1}^{L} z_i^2$ is a Chi-Squared random variable with $L$ degrees of freedom. Observe

$$\mathcal{P}(p_x^{[L]} > C) = 1 - \mathcal{P}\left(\sum_{i=1}^{L} z_i^2 \leq \frac{CL}{\sigma_x^2}\right) \geq 1 - \exp\left(\frac{sCL}{\sigma_x^2}\right)(1+2s)^{-L/2} \tag{53}$$

for $s > 0$. Equation (53) follows from the Chernoff Bound and also from the fact that the Moment Generating function of a Chi-Squared random variable with $L$ degrees of freedom is given by $(1-2s)^{-L/2}, s < 1/2$. Now define the function $\beta(s) \triangleq \exp\left(-\frac{2sC}{\sigma_x^2}\right)(1+2s)$ we can write from (53) that

$$\mathcal{P}(p_x^{[L]} > C) \geq 1 - \Big(\beta(s)\Big)^{-L/2}. \tag{54}$$

We want to ensure that $\exists s > 0$, such that $\beta(s) > 1$. Now $\beta(s) > 1 \Leftrightarrow C < \gamma(s)$ where $\gamma(s) \triangleq \frac{\sigma_x^2}{2s}\log(1+2s)$. It can be verified that $\gamma(s)$ is a decreasing function in $s$ for $s > 0$ and $\gamma(0) = \sigma_x^2$. Since, it is given that $C < \sigma_x^2$, then indeed $\exists s_0 > 0$ such that $C < \gamma(s_0)$. This in turn implies, $\beta(s_0) > 1$. Hence, we conclude from (54) that $\mathcal{P}\left(\frac{1}{L}\sum_{i=1}^{L} x_i^2 > C\right) \geq 1 - \beta^{-L}$ for $\beta = \sqrt{\beta(s_0)} > 1$. ■

## References


[1] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5030–5043, Oct. 2010.

[2] T. T. Cai, L. Wang, and G. Xu, "Stable recovery of sparse signals and an oracle inequality," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3516–3522, Jul. 2010.

[3] E. J. Candes, "Compressive sampling," in *Proc. Int. Congr. Mathematicians*, 2006, pp. 1433–1452.

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

[5] E. J. Candes 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.

[6] S. Chen, D. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.

[7] A. Cohen, D. Wolfgang, and R. DeVore, "Compressed sensing and best *k*-term approximation," *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 211–231, 2009.

[8] M. Davenport, M. Wakin, and R. Baraniuk, "Detection and Estimation with Compressive Measurements," Tech. Rep. Rice ECE Dept., 2006.

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

[10] D. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete Comput. Geomet.*, vol. 102, no. 27, pp. 617–652, 2006.

[11] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via $l1$ minimization," in *Proc. Nat. Acad. Sci.*, 2003, vol. 100, pp. 2197–2202, 5.

[12] D. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[13] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Jul. 2001.

[14] D. Donoho and J. Tanner, "Neighborlyness of randomly-projected simplices in high dimensions," in *Proc. Nat. Acad. Sci.*, 2005, vol. 102, pp. 9452–9457, 27.

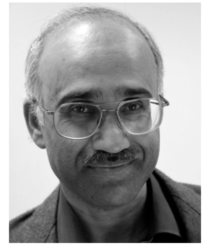
[15] D. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," in *Proc. Nat. Acad. Sci.*, 2005, vol. 102, pp. 9446–9451, 27.

[16] M. F. Duarte and R. G. Baraniuk, "Kronecker compressive sensing," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 494–504, Feb. 2012.

[17] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via $lq$ -minimization for $0 < q \leq 1$," *Appl. Comput. Harmonic Anal.*, vol. 26, no. 3, pp. 395–407, 2009.
[18] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted norm minimization algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.
[19] R. T. Hoctor and S. A. Kassam, "The unifying role of the coarray in aperture synthesis for coherent and incoherent imaging," *Proc. IEEE*, vol. 78, no. 4, pp. 735–752, Apr. 1990.
[20] R. A. Horn and C. R. Johnson*, Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[21] M. M. Hyder and K. Mahata, "A robust algorithm for joint-sparse recovery," *IEEE Signal Process. Lett.*, vol. 16, no. 12, pp. 1091–1094, Dec. 2009.
[22] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.
[23] C. G. Khatri and C. R. Rao, "Solutions to some functional equations and their applications to characterization of probability distributions," *Indian J. Stat. Series A*, vol. 30, no. 2, pp. 167–180, 1968.
[24] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Appl.*, vol. 18, 1977.
[25] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3613–3641, Jun. 2012.
[26] M. Lustig, D. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, Apr. 2008.
[27] D. Malioutov, M. Cetin, and A. 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.
[28] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas Propag.*, vol. 16, no. 2, pp. 172–175, Mar. 1968.
[29] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmonic Anal.*, vol. 26, pp. 301–321, 2008.
[30] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4167–4181, Aug. 2010.
[31] P. Pal and P. P. Vaidyanathan, "Correlation aware techniques for sparse support recovery," in *IEEE Statist. Signal Process. Workshop (SSP)*, Aug. 5–8, 2012, pp. 53–56.
[32] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus, "Compressive wide-band spectrum sensing," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Apr. 2009, pp. 2337–2340.
[33] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *J. Chemometrics*, vol. 14, no. 3, pp. 229–239, May 2000.
[34] P. Stoica, P. Babu, and J. Li, "Spice: A sparse covariance-based estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629–638, Feb. 2011.
[35] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1383–1399, Mar. 2010.
[36] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Apr. 2007, vol. 4, pp. 1357–1360.
[37] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, pp. 211–244, 2001.
[38] S. Shamsunder and G. B. Giannakis, "Modeling of non-Gaussian array data using cumulants: DOA estimation of more sources with less sensors," *EURASIP Signal Process.*, vol. 30, pp. 279–297, Feb. 1993.
[39] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[40] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.
[41] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[42] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 573–586, Feb. 2011.
[43] M. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using $l1$-constrained quadratic programming (LASSO)," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2183–2202, May 2009.
[44] M. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inf. Theory*, no. 12, pp. 5728–5741, Dec. 2009.
[45] L. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 397–399, May 1974.
[46] D. Wipf and S. Nagarajan, "A unified Bayesian framework for MEG/EEG source imaging," *NeuroImage*, pp. 947–966, 2008.
[47] D. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.
[48] D. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, Jul. 2007.
[49] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, Sep. 2011.
[50] Z. Zhang and B. D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, Apr. 15, 2013.
[51] J. Zheng and M. Kaveh, "Sparse spatial spectral estimation: A covariance fitting algorithm, performance and regularization," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2767–2777, Jun. 1, 2013.
[52] S. V. BoydL*, Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[53] R. Levanda and A. Leshem, "Radio astronomical image formation using sparse reconstruction techniques," in *Proc. IEEE 25th Convention Electr. Electron. Engineers Israel*, Dec. 2008, pp. 716–720.
[54] E. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, 2008.
[55] S. Shakeri, D. D. Ariananda, and G. Leus, "Direction of arrival estimation using sparse ruler array design," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.*, Cesme, Turkey, Jun. 2012.
[56] D. D. Ariananda and G. Leus, "Direction of arrival estimation of correlated signals using a dynamic linear array," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2012.
[57] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.
[58] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.
[59] M. E. Davies and Y. Eldar, "Rank awareness in joint sparse recovery," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1135–1146, Feb. 2012.
[60] J. J. Fuchs, "Recovery of exact sparse representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3601–3608, Oct. 2005.
[61] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5862–5875, Nov. 2010.
[62] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, Jun. 2009.
[63] R. A. Horn and C. R. Johnson*, Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.
[64] M. Mishali and Y. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signal," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.
[65] G. Obozinski, B. Taskar, and M. Jordan, "Joint covariate selection and joint subspace selection for multiple classification problems," *Statist. Comput.*, vol. 20, no. 2, pp. 231–252, 2010.
[66] G. Obozinski, M. Wainwright, and M. Jordan, "Support union recovery in high-dimensional multivariate regression," *Annu. Statist.*, vol. 39, no. 1, pp. 1–47, 2011.
[67] P. Pal and P. P. Vaidyanathan, "On application of LASSO for sparse support recovery with imperfect correlation awareness," presented at the 46th Asilomar Conf. Signals, Syst. an Comput., 2012, Nov. 4–7, 2012.
[68] P. Pal and P. P. Vaidyanathan, "Correlation-aware sparse support recovery: Gaussian sources," presented at the Internat. Conf. Acoustics, Speech, Signal Process. (ICASSP), 2013, Vancouver, BC, Canada, May 26–31, 2013.
[69] P. Pal and P. P. Vaidyanathan, "Conditions for identifiability in sparse spatial spectrum sensing," in *21st Eur. Signal Process. Conf., EUSIPCO*, Marrakech, Morocco, Sep. 9–13, 2013, (Invited Paper).
[70] D. D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4775–4789, Sep. 2012.
[71] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Statist. Soc.*, ser. B, vol. 58, no. 1, pp. 267–288, 1996.

[72] J. Tropp, “Algorithms for simultaneous sparse approximation. Part II: Convex relaxation,” *Signal Process.*, vol. 86, no. 3, pp. 589–602, 2006.
[73] J. Tropp, A. Gilbert, and M. Strauss, “Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit,” *Signal Process.*, vol. 86, no. 3, pp. 572–588, 2006.
[74] T. Yingming, C. Yen, and X. Wang, “Wideband spectrum sensing based on sub-Nyquist sampling,” *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3028–3040, Jun. 15, 2013.
[75] P. Zhao and B. Yu, “On model selection consistency of Lasso,” *J. Mach. Learning Res.*, vol. 7, pp. 2541–2567, 2006.

**Piya Pal** (S'08) received her Bachelors in Technology degree in Electronics and Electrical Communication Engineering from Indian Institute of Technology, Kharagpur, India in 2007, and her Ph.D in Electrical Engineering from Caltech in 2013, where her Ph.D thesis won the Charles and Ellen Wilts Prize for Outstanding Doctoral Thesis in Electrical Engineering. In January 2014, she joined the Department of Electrical and Computer Engineering at the University of Maryland, College Park, as an Assistant Professor of Electrical and Computer Engineering. She is also affiliated with the Institute for Systems Research. During the course of her Ph.D, she developed novel techniques for non-uniform sparse sampling which received Best Student Paper Awards at 14th IEEE DSP Workshop, and 45th Asilomar Conference on Signals, Systems and Computers, 2011. Her research interests span compressive and structured sampling, statistical signal processing with applications in radar and array processing, tensor methods, and statistical learning.



**P. P. Vaidyanathan** (S'80–M'83–SM'88–F'91) was born in Calcutta, India on Oct. 16, 1954. He received the B.Sc. (Hons.) degree in physics and the B.Tech. and M.Tech. degrees in radiophysics and electronics, all from the University of Calcutta, India, in 1974, 1977 and 1979, respectively, and the Ph.D degree in electrical and computer engineering from the University of California at Santa Barbara in 1982. He was a post doctoral fellow at the University of California, Santa Barbara from Sept. 1982 to March 1983. In March 1983 he joined the electrical engineering department of the Calfornia Institute of Technology as an Assistant Professor, and since 1993 has been Professor of electrical engineering there. His main research interests are in digital signal processing, multirate systems, wavelet transforms, signal processing for digital communications, genomic signal processing, radar signal processing, and sparse array signal processing.

Dr. Vaidyanathan served as Vice-Chairman of the Technical Program committee for the 1983 IEEE International symposium on Circuits and Systems, and as the Technical Program Chairman for the 1992 IEEE International symposium on Circuits and Systems. He was an Associate editor for the IEEE Transactions on Circuits and Systems for the period 1985–1987, and is currently an associate editor for the journal IEEE Signal Processing letters, and a consulting editor for the journal Applied and computational harmonic analysis. He has been a guest editor in 1998 for special issues of the IEEE Trans. on Signal Processing and the IEEE Trans. on Circuits and Systems II, on the topics of filter banks, wavelets and subband coders.

Dr. Vaidyanathan has authored more than 425 papers in journals and conferences, and is the author/coauthor of the four books Multirate systems and filter banks, Prentice Hall, 1993, Linear Prediction Theory, Morgan and Claypool, 2008, and (with Phoong and Lin) Signal Processing and Optimization for Transceiver Systems, Cambridge University Press, 2010, and Filter Bank Transceivers for OFDM and DMT Systems, Cambridge University Press, 2010. He has written several chapters for various signal processing handbooks. He was a recipient of the Award for excellence in teaching at the California Institute of Technology for the years 1983–1984, 1992–93 and 1993–94. He also received the NSF's Presidential Young Investigator award in 1986. In 1989 he received the IEEE ASSP Senior Award for his paper on multirate perfect-reconstruction filter banks. In 1990 he was recepient of the S. K. Mitra Memorial Award from the Institute of Electronics and Telecommuncations Engineers, India, for his joint paper in the IETE journal. In 2009 he was chosen to receive the IETE students' journal award for his tutorial paper in the IETE Journal of Education. He was also the coauthor of a paper on linear-phase perfect reconstruction filter banks in the IEEE SP Transactions, for which the first author (Truong Nguyen) received the Young outstanding author award in 1993. Dr. Vaidyanathan was elected Fellow of the IEEE in 1991. He received the 1995 F. E. Terman Award of the American Society for Engineering Education, sponsored by Hewlett Packard Co., for his contributions to engineering education. He has given several plenary talks including at the IEEE ISCAS-04, Sampta-01, Eusipco-98, SPCOM-95, and Asilomar-88 conferences on signal processing. He has been chosen a distinguished lecturer for the IEEE Signal Processing Society for the year 1996–97. In 1999 he was chosen to receive the IEEE CAS Society's Golden Jubilee Medal. He is a recepient of the IEEE Signal Processing Society's Technical Achievement Award for the year 2002, and the IEEE Signal Processing Society's Education Award for the year 2012.