STRUCTURED SAMPLING OF STRUCTURED SIGNALS

Bo Li and Athina P. Petropulu
ECE Department
Rutgers, The State University of New Jersey, NJ, 08854
Email: {paul.bo.li, athinap}@rutgers.edu

***Abstract*—The paper considers structured sampling of structured signals, more specifically, using block diagonal (BD) measurement matrices to sense signals with uniform partitions that share the same sparsity profile. This model arises in distributed compressive sensing systems. In general, the fact that the number of nonzero entries in the measurement matrix is smaller than in a dense matrix leads to the need for more measurements. However, taking advantage of a certain structure in the sparse signal allows one to relax the conditions on the measurement matrix for the restricted isometry property (RIP) to hold, thus allowing for higher compression rate. We systematically provide guarantees for a unique solution, and also an efficient recovery method. The analysis relies on the RIP of the random BD matrix for signals in a particular union of subspaces. Also, we show how our theoretical results can be used to analyze the multiple measurement vector (MMV) problem.**

***Index Terms*—Compressive Sensing, Multiple Measurement Vectors, Block Diagonal Matrices, Restricted Isometry Property, Block Sparsity.**

I. INTRODUCTION

While random measurement matrices have dominated the studies of compressive sensing [1], [2], block diagonal measurement matrices have only recently started attracting attention. Such matrices arise due to physical constraints of the sensing system ([3] and references therein) [4]. Meanwhile, recent works have investigated structure in the measured signals in addition to sparsity, which can be exploited to reduce the number of required measurements. Examples of such structures include multiple measurement vectors [5]–[7] and union of subspaces [8]–[10].

To the best of our knowledge, the use of structured measurement matrices for structured signals has received limited attention [12], and is the point of this paper. In particular, we consider the following general model for the measured signal:

$$\underbrace{\begin{bmatrix}\mathbf{y}_1\\ \vdots \\ \mathbf{y}_J\end{bmatrix}}_{\mathbf{y}\in\mathbb{R}^{\tilde{M}}} = \underbrace{\begin{bmatrix}\mathbf{\Phi}_1 & & \\ & \ddots & \\ & & \mathbf{\Phi}_J\end{bmatrix}}_{\mathbf{\Psi}\in\mathbb{R}^{\tilde{M}\times\tilde{N}}} \underbrace{\begin{bmatrix}\mathbf{x}_1\\ \vdots \\ \mathbf{x}_J\end{bmatrix}}_{\mathbf{x}\in\mathbb{R}^{\tilde{N}}} \tag{1}$$

where $\mathbf{y}_j \in \mathbb{R}^M$, $\mathbf{\Phi}_j \in \mathbb{R}^{M\times N}$, $\mathbf{x}_j \in \mathbb{R}^N$, $\tilde{M} = JM$ and $\tilde{N} = JN$. Several applications can be described by this model. For example, in the context of distributed compressive sensing systems [4], [12], $\mathbf{y}_j$ corresponds to the compressed signal obtained at node j based on the measurement matrix $\mathbf{\Phi}_j$. Other applications include time-variant sparse channel estimation, hyperspectral imaging, cognitive radio and sensor networks ([3] and references therein). Later in this paper, we will show that the multiple measurement vector (MMV) problem also fits the model of (1). In general, the fact that the number of nonzero entries in $\mathbf{\Psi}$ is smaller than in a dense matrix leads to the need for more measurements. However, taking advantage of a certain structure in the signal allows one to relax the conditions which $\mathbf{\Psi}$ has to meet in order to satisfy the restricted isometry property, thus enabling a higher compression rate.

The contribution of this paper is as follows. Under the assumption that the $\mathbf{\Phi}_j$'s in (1) contain i.i.d. Gaussian entries, and at the same time, the signals $\{\mathbf{x}_j\}_{j=1}^J$, share the same sparse support, we establish the RIP of the random block diagonal (BD) matrix $\mathbf{\Psi}$ for structured signals $\mathbf{x}$ based on the concentration of measure (CoM) of random BD matrices (see Theorem 1). By leveraging results in [10], we provide guarantees for a unique solution (see Theorem 2), and also for an efficient recovery method (see Theorem 3). Also, we reformulate the MMV problem to fit the model of (1) and specialize our analytical results for the MMV problem (see Theorem 4).

This work was supported by ONR under grant N00014-12-1-0036.

II. MAIN RESULTS

Let us rewrite (1) as

$$\mathbf{y} = \mathbf{\Psi}\mathbf{x} \tag{2}$$

where $\mathbf{x}$ lies in the union of subspaces $\mathcal{A}_K \triangleq \{\mathbf{x} \in \mathbb{R}^{\tilde{N}} | \text{supp}(\mathbf{x}_1) = \cdots = \text{supp}(\mathbf{x}_J), |\text{supp}(\mathbf{x}_j)| \leq K\}$. The operator supp$(\cdot)$ and $|\cdot|$ denote the index set of nonzero entries of a vector and the cardinality of a set, respectively. Thus, $\mathcal{A}_K$ contains vectors of size $\tilde{N}$, in which all blocks $\mathbf{x}_j$ share the same sparse support. Here, we assume that $\{\mathbf{\Phi}_j \in \mathbb{R}^{M\times N}\}_{j=1}^J$ are matrices with i.i.d. entries, distributed as $\mathcal{N}(0, 1/M)$.

Let us define a special case of the $\mathcal{A}$-restricted isometry introduced in [8], in which the union of subspaces is the $\mathcal{A}_K$ defined above.

Definition 1: Matrix $\mathbf{\Psi} \in \mathbb{R}^{\tilde{M}\times\tilde{N}}$ satisfies the RIP over $\mathcal{A}_K$ with constant δ_K, or equivalently the $\mathcal{A}$-RIP(K, δ_K), if for every $\mathbf{x} \in \mathcal{A}_K$ it holds that

$$(1-\delta_k)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Psi}\mathbf{x}\|_2^2 \leq (1+\delta_k)\|\mathbf{x}\|_2^2, \tag{3}$$

Successful signal recovery requires that the block diagonal matrix $\mathbf{\Psi}$ satisfies the $\mathcal{A}$-RIP with proper constants, which is considered in the following subsection.

A. Random BD Matrices Satisfying $\mathcal{A}$-RIP

The subspace RIP analysis for BD matrices is based on the concentration inequality of random BD matrices which has been studied in [11]. However, the result in [11] depends explicitly on $\mathbf{x}$ and thus cannot be directly used to analyze the RIP of BD matrices. In the following lemma, we eliminate the dependence on $\mathbf{x}$ by using the upper bound of the result in [11, Theorem III.1 and III.2].

Lemma 1 (Upper bound of result in [11]): Let $\mathbf{\Psi}$ be a BD matrix as defined in (2) with entries of distribution $\mathcal{N}(0, \frac{1}{M})$. Then for any fixed $\mathbf{x} \in \mathbb{R}^{\tilde{N}}$

$$\mathbf{P}(|\|\mathbf{\Psi x}\|_2^2 - \|\mathbf{x}\|_2^2| \geq \epsilon \|\mathbf{x}\|_2^2) \leq 2e^{-cM\epsilon^2} \tag{4}$$

where c is an absolute constant.

Based on the above CoM result, we have the following $\mathcal{A}$-RIP for random BD matrices.

Theorem 1: Suppose that J, M and N are given and $\tilde{M} = JM$, $\tilde{N} = JN$. Then, the block diagonal matrix $\mathbf{\Psi} \in \mathbb{R}^{\tilde{M} \times \tilde{N}}$, as stated in Lemma 1, satisfies the $\mathcal{A}$-RIP(K, δ_K) for some $\delta_K \in (0, 1)$, $t > 0$ and

$$M \geq \frac{c_1}{\delta_K^2} \left(K \log \frac{N}{K} + JK \log(e(1 + \frac{12}{\delta_K})) + \log 2 + t \right) \tag{5}$$

with probability at least $(1 - e^{-t})$, where $c_1 \leq 9/c$.

Proof: The concentration inequality for the random BD matrix $\mathbf{\Psi}$ was given in Lemma 1. Given any set of indices T with $|T| \leq JK$, let us denote by X_T the subspace of all vectors in $\mathbb{R}^{\tilde{N}}$ that are zero outside T. Based on Lemma 5.1 in [2], it holds that for all $\mathbf{x}$ in any subspace X_T and $\delta_K \in (0, 1)$, matrix $\mathbf{\Psi}$ will fail to satisfy

$$(1 - \delta_K)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Psi x}\|_2^2 \leq (1 + \delta_K)\|\mathbf{x}\|_2^2 \tag{6}$$

with probability no larger than $2(1 + 12/\delta_K)^{JK} e^{-cM\delta_K^2/9}$. Here, the $\{\mathbf{x}_j\}_{j=1}^J$ in $\mathbf{x}$ have some additional structure, i.e., share the same sparsity profile. There are $\binom{N}{K} \leq (eN/K)^K$ such subspaces, instead of $\binom{\tilde{N}}{JK}$ in the scenario of sparse vectors with no additional structure. Hence, the union bound can be used to generate the total failure probability, i.e.,

$$\mathbb{P}((6) \text{ fails}) \leq 2(eN/K)^K (1 + 12/\delta_K)^{JK} e^{-cM\delta_K^2/9}.$$

On setting the above right-hand-side term to less than e^{-t}, the claim of (5) follows. ■

Next, based on the $\mathcal{A}$-RIP result, we introduce recovery methods with performance guarantees.

B. Recovery Methods and Guarantees

Let the sets $\mathcal{I}_n, n = 1, \ldots, N$, with cardinality J, contain the indices of the nth entries from all blocks $\{\mathbf{x}_j\}_{j=1}^J$. We can recover $\mathbf{x}$ by solving the optimization problem

$$\min_{\mathbf{x}} \sum_{n=1}^{N} I\left(\|\mathbf{x}[\mathcal{I}_n]\|_2\right) \quad \text{s.t.} \quad \mathbf{y} = \mathbf{\Psi x} \tag{7}$$

where $\mathbf{x}[\mathcal{I}_n]$ denotes a vector containing entries with indices $\mathcal{I}_n$ of $\mathbf{x}$, $I(a)$ is the indicator function, i.e., if $a = 0$, $I(a) = 0$, else $I(a) = 1$. Let the permutation $\mathcal{P}_v : \mathbb{R}^{\tilde{N}} \to \mathbb{R}^{\tilde{N}}$ rearrange the entries of $\mathbf{x}$ as $\mathcal{P}_v(\mathbf{x}) = [\mathbf{x}[\mathcal{I}_1]; \ldots; \mathbf{x}[\mathcal{I}_N]]$. One can see that $\mathcal{P}_v(\mathbf{x})$ is a block sparse vector. The cost function in (7), denoted by $\|\mathbf{x}\|_{0,\mathcal{I}}$, counts the number of nonzero blocks in $\mathcal{P}_v(\mathbf{x})$. If $\|\mathbf{x}\|_{0,\mathcal{I}} \leq K$, we should have $\mathbf{x} \in \mathcal{A}_K$.

Theorem 2: $\mathbf{x} \in \mathcal{A}_K$ can be uniquely recovered by solving (7), if $\mathbf{\Psi}$ satisfies the $\mathcal{A}$-RIP$(2K, \delta_{2K})$ with $\delta_{2K} < 1$.

The proof follows along the lines of the proof when the standard RIP considered and is omitted. The solution to (7) can be found by a combinatorial search, usually referred to as the ℓ_0-norm approach. For efficient recovery, we can solve the following convex optimization problem

$$\min_{\mathbf{x}} \sum_{n=1}^{N} \|\mathbf{x}[\mathcal{I}_n]\|_2 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{\Psi x} \tag{8}$$

The cost function in (8) first calculates the ℓ_2-norms inside the blocks of $\mathcal{P}_v(\mathbf{x})$, and then sums the results up to get the ℓ_1-norm across the blocks $\mathbf{x}[\mathcal{I}_n]$. Thus, the entries in a given index set would be forced to be either zero or nonzero. As for $\mathbf{x}$, all the J subvectors would have the same sparsity profile. The above problem can be formulated into a standard second-order cone programming problem, which could be solved efficiently using standard software packages.

Note that the optimization problem in (8) is equivalent to that in eqn. (26) of [10]. Therefore, if we can establish a certain equivalence for the conditions on the measurement matrices, then the guarantees of exact recovery in [10, Theorem 1] can be leveraged for our case.

In [10], the block-RIP condition was defined as follows.

Definition 2: Matrix $\mathbf{D} \in \mathbb{R}^{\tilde{M} \times \tilde{N}}$ satisfies the block RIP with constant δ_K, if for any $\mathbf{c} \in \mathbb{R}^{\tilde{N}}$ that has at most K nonzero blocks out of its N uniform partition blocks we have

$$(1 - \delta_k)\|\mathbf{c}\|_2^2 \leq \|\mathbf{Dc}\|_2^2 \leq (1 + \delta_k)\|\mathbf{c}\|_2^2, \tag{9}$$

abbreviated by block-RIP(K, δ_K).

Let $\mathcal{P}_M$ denote the column permutation of $\mathbf{\Psi}$, so that $\mathbf{\Psi x} = \mathcal{P}_M(\mathbf{\Psi})\mathcal{P}_v(\mathbf{x})$. Clearly, if $\{\mathbf{x}_j\}_{j=1}^J$ in $\mathbf{x}$ are K-sparse and have the same support, then $\mathcal{P}_v(\mathbf{x})$ is block K-sparse.

Proposition 1: $\mathbf{\Psi}$ satisfies the $\mathcal{A}$-RIP(K, δ_K) *if and only if* $\mathcal{P}_M(\mathbf{\Psi})$ satisfies the block-RIP(K, δ_K).

Proof: $\mathbf{\Psi}$ satisfying the $\mathcal{A}$-RIP(K, δ_K) means that (3) holds for every $\mathbf{x} \in \mathcal{A}_K$. We know that $\mathcal{P}_v(\mathbf{x})$ is block K-sparse and $\mathbf{\Psi x} = \mathcal{P}_M(\mathbf{\Psi})\mathcal{P}_v(\mathbf{x})$. By the fact that $\|\mathbf{x}\|_2^2 = \|\mathcal{P}_v(\mathbf{x})\|_2^2$, (9) holds for $\mathcal{P}_M(\mathbf{\Psi})$ and every $\mathcal{P}_v(\mathbf{x})$ that is block K-sparse, which means that $\mathcal{P}_M(\mathbf{\Psi})$ satisfies the block-RIP(K, δ_K). Since only permutation involved in the transformation between these two RIPs, the strict equivalence also holds when we consider the probabilistic nature of the RIP condition. ■

Based on the above proposition, we can easily arrive at the following theorem regarding the exact noise-free recovery of $\mathbf{x}$ by leveraging Theorem 1 of [10].

Theorem 3: For the system model as in (2), let $\mathbf{y} = \mathbf{\Psi x}_0$ be the measurement of the sparse vector $\mathbf{x}_0 \in \mathcal{A}_K$. If the block diagonal matrix $\mathbf{\Psi}$ satisfies the $\mathcal{A}$-RIP$(2K, \delta_{2K})$ with $\delta_{2K} < \sqrt{2} - 1$, then there is a unique sparse vector $\mathbf{x}$ in $\mathcal{A}_K$ consistent with $\mathbf{y}$ and the solution of (8) is equal to $\mathbf{x}_0$.

Proof: From Proposition 1, we obtain that $\mathcal{P}_M(\mathbf{\Psi})$ satisfies the block-RIP$(2K, \delta_{2K})$ also with $\delta_{2K} < \sqrt{2} - 1$. The optimization problem in [10, (26)] is actually

$$\min_{\mathcal{P}_v(\mathbf{x})} \sum_{n=1}^{N} \|\mathbf{x}[\mathcal{I}_n]\|_2 \quad \text{s.t.} \quad \mathbf{y} = \mathcal{P}_M(\mathbf{\Psi})\mathcal{P}_v(\mathbf{x})$$

which will find the unique solution $\mathcal{P}_v(\mathbf{x}_0)$ based on [10, Theorem 1]. Thus, the uniqueness of the solution of (8) is guaranteed since its solution is only the inverse permutation of the solution $\mathcal{P}_v(\mathbf{x}_0)$. ■

It is practical to consider the situation in which the measurements are noisy and the sparse vector is not exactly in $\mathcal{A}_K$. Suppose we have corrupted measurements as $\mathbf{y} = \mathbf{\Psi}\mathbf{x} + \mathbf{n}$ where bounded noise is with $\|\mathbf{n}\|_2 \leq \epsilon$. The robust recovery of $\mathbf{x}$ can be achieved by

$$\min_{\mathbf{x}} \sum_{n=1}^{N} \|\mathbf{x}[\mathcal{I}_n]\|_2 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{\Psi}\mathbf{x}\|_2 \leq \epsilon \tag{10}$$

Similarly, by leveraging the conditions and recovery performance of [10, Theorem 2] we get the corresponding theorem regarding the performance of the above model.

Relation to Literature: One of our main results is the $\mathcal{A}$-RIP analysis of BD matrix $\mathbf{\Psi}$, given in Theorem 1, for which there is no prior literature. The work in [10] only provided the block-RIP analysis for dense measurement matrices. Note that applying [10, Theorem 1 and 2] to $\mathcal{P}_M(\mathbf{\Psi})$ requires the block-RIP of $\mathcal{P}_M(\mathbf{\Psi})$, which has a special structure. Adopting the approach of [10, Section VII] to analyze the block-RIP of $\mathcal{P}_M(\mathbf{\Psi})$ would require its CoM inequality, for which there are no results in the literature. However, this difficulty can be bypassed by our result on the $\mathcal{A}$-RIP analysis for $\mathbf{\Psi}$ along with Proposition 1.

Based on the $\mathcal{A}$-RIP analysis and the equivalence result in Preposition 1, Theorem 3 establishes the recovery performance by leveraging results from [10].

The standard RIP for random BD matrices was considered in [4] based on an improved bound on the suprema of chaotic random processes. Their approach does not take into account the special structure of $\mathbf{x}$, and thus cannot be directly used to generate the results in our paper.

The work of [12] studied the same model of (2), referred to as distributed compressive sensing (DCS) from a distributive coding perspective. A graphical model was introduced to provide the bound on the number of *noiseless* measurements required for signal recovery. However, the achieved bound is akin to that for the ℓ_0-norm approach. In this paper, we have provided both uniqueness guarantees for the ℓ_0-norm approach, and the equivalence results for the efficient ℓ_1-norm recovery approach.

III. Revisit the MMV Models

In this section, we specialize our model to the MMV problem and provide the first equivalence results based on strict RIP analysis.

A. Background

MMV problems have attracted interest due to their wide applicability [5]–[7]. The most detailed theoretical analysis establishes equivalence results for the mixed ℓ_p/ℓ_1 approach based on mutual coherence [5], which shows no improvement over the single measurement vector case.

The MMV model observes a $M \times J$ dimensional matrix of measurements $\mathbf{Y} = \mathbf{M}\mathbf{X}$, where $\mathbf{M} \in \mathbb{R}^{M \times N}$ is a matrix with i.i.d. zero-mean Gaussian entries and variance $1/M$, and $\mathbf{X}$ is an unknown $N \times J$ matrix that has at most K nonzero rows. Each column of $\mathbf{X}$ is one measurement vector.

In [10], the multiple measurement vectors problem was transformed into the single vector model as

$$\text{vec}(\mathbf{Y}^T) = (\mathbf{M} \otimes \mathbf{I}_J)\text{vec}(\mathbf{X}^T), \tag{11}$$

in which block-sparsity arises in $\text{vec}(\mathbf{X}^T)$. The new measurement matrix $(\mathbf{M} \otimes \mathbf{I}_J)$ is not a dense matrix and has a special structure. The authors could not provide its accurate block-RIP based on their method in [10, Section VII], and claimed that the block-RIP of $(\mathbf{M} \otimes \mathbf{I}_J)$ is equivalent to the standard RIP of $\mathbf{M}$. However, this might not be accurate. In the first phase of the proof, $\mathbf{X}$ was assumed to have identical columns, which might not be the case in practical settings. In the second phase of the proof, the probabilistic nature of the RIP condition (3) for *random matrices* was omitted, while the union operation of probabilities was essentially involved in the summation of RIP inequalities. Actually, as shown in Proposition 1, the block-RIP of $(\mathbf{M} \otimes \mathbf{I}_J)$ is equivalent to the $\mathcal{A}$-RIP of $\text{diag}(\mathbf{M}, \ldots, \mathbf{M})$; this is because only permutation operations are involved in the transformation from one to the other. Also in [10], it was proposed to use the unstructured measurement matrix to improve the performance of MMV. However, this approach is at the cost of higher computational complexity and not applicable for the measurement vectors at hand. In short, strict RIP-based analysis for the MMV model has not been previously addressed.

B. Equivalence Results based on RIP Analysis

Here, we propose an alternative reformulation of the MMV model, which can enable a thorough RIP-based analysis via the results of Section II. Instead of vectorizing the transpose of $\mathbf{Y}$, let us concatenate the columns of matrix $\mathbf{Y}$ into one long vector, i.e.,

$$\text{vec}(\mathbf{Y}) = \text{diag}(\mathbf{M}, \ldots, \mathbf{M})\text{vec}(\mathbf{X}) \tag{12}$$

where $\text{diag}(\mathbf{M}, \ldots, \mathbf{M})$ is BD with repeated blocks, and $\text{vec}(\mathbf{X})$ belongs to the $\mathcal{A}_K$. The above formulation fits the model of (2), therefore, a the uniqueness of solution and the recovery method with performance guarantees can be readily established based on the results of Section II.

Theorem 4: Given the MMV model $\mathbf{Y} = \mathbf{M}\mathbf{X}$, if M satisfies the bound in (5) with $\delta_{2K} < \sqrt{2} - 1$, then applying the ℓ_1 approach of (8) to (12), will uniquely recover the signal matrix $\mathbf{X}$ with probability at least $(1 - e^{-t})$.

The proof can be achieved simply by combining Theorem 1 and Theorem 3 on the model of (12). A few remarks are

TABLE I
COMPARISON: BOUNDS ON $\tilde{M}$ UNDER DIFFERENT STRUCTURES IN $\mathbf{\Psi}$ AND $\mathbf{x}$

Structure in $\mathbf{x}$	Structure in $\mathbf{\Psi}$	
	Full populated	Block diagonal
Arbitrary JK-sparse [1]	Standard CS [1], [2] $\tilde{M} = \mathcal{O}(JK\log(N/K))$	Standard RIP for BD in Corollary 1 $\tilde{M} = \mathcal{O}(J^2K\log(N/K))$
$\mathbf{x} \in \mathcal{A}_K$	Model-based CS [9], [10] $\tilde{M} = \mathcal{O}(K\log(N/K) + JK)$	Our model $\tilde{M} = \mathcal{O}(JK\log(N/K) + J^2K)$

in order. The number of measurements required for robust recovery scales as $M = \mathcal{O}(K\log(N/K) + JK)$. Compared to $M = \mathcal{O}(K\log(N/K))$ that would be required by the single measurement vector case, no improved performance is theoretically guaranteed as discovered in previous works [5], [6]. Conversely, additional measurements proportional to JK are needed in order to recover multiple signal vectors. It is clear that the recovery ability will not be improved when all columns in $\mathbf{X}$ are the same. We cannot know what particular signals would be the most difficult to recover. In the worst case, it is quite possible that the benefit brought by one additional measurement vector cannot remedy its incidental estimation burden. Above all, the above analysis for the MMV problem is the first strict RIP-based worst case analysis and provides tighter bound than existing coherence-based analysis, which usually require $M = \mathcal{O}(K^2\log^2 N)$ [13]. For methods with theoretically guaranteed performance improvement, we refer to the average case analysis of [6] and the rank-aware method of [7].

IV. DISCUSSIONS AND COMPARISONS

In this paper we have provide the first RIP analysis for block diagonal matrices over the union of subspaces $\mathcal{A}_K$, i.e., $\mathcal{A}$-RIP of BD matrices.

For comparison, we list in Table I the number of measurements bounds in four different models. The system model in Corollary 1 of the appendix provides the bound for BD matrices satisfying the standard RIP for arbitrary sparse signals with JK nonzero entries. Compared with the bound for model-based CS [9], [10], which is not for distributed sensing applications, our bound is J times higher. This is due to the fact that the measurement matrix in our model is BD with much fewer nonzero random entries. Recall that for each node, the robust recovery requires $M = \mathcal{O}(K\log(N/K))$. One may say that the total number for separate recovery scales as $\tilde{M} = \mathcal{O}(JK\log(N/K))$, which is smaller than our result. However, the simple addition for the number of measurements above is not accurate, because the overall recovery probability degrades.

APPENDIX
STANDARD RIP FOR RANDOM BD MATRICES

Based on the CoM inequality for the random BD matrix $\mathbf{\Psi}$ in Lemma 1, it is easy to prove the standard RIP for $\mathbf{\Psi}$.

Corollary 1: The BD matrix $\mathbf{\Psi} \in \mathbb{R}^{\tilde{M}\times\tilde{N}}$, as stated in Theorem 1, satisfies the RIP(K, δ_K) for some $\delta_K \in (0, 1)$, $t > 0$ with probability at least $(1 - e^{-t})$ if

$$M \geq c_1\delta_K^{-2}\left(K\log\frac{\tilde{N}}{K} + K\log(e(1+\frac{12}{\delta_K})) + \log 2 + t\right).$$

Let us compare with the standard RIP for random BD matrices in [4]. Particularly for signals which are arbitrarily K sparse in the *canonical basis*, the authors in [4] claimed that BD measurement matrix $\mathbf{\Psi} \in \mathbb{R}^{\tilde{M}\times\tilde{N}}$ with independent blocks satisfies the RIP(K, δ_K) with high probability for $\tilde{M} = \mathcal{O}(JK\log^2 K\log^2\tilde{N})$ (here we substitute $\tilde{\mu}^2(I_{\tilde{N}})$ by J in their original result). Our bound above, i.e., $\tilde{M} = \mathcal{O}(JK\log(\tilde{N}/K) + J\log J)$, is tighter than that in [4] because it does not have the $\log^2 K$ term and scales linear-logarithmically with $\tilde{N}$. Similarly, we observe that our result on the RIP for the BD matrix with repeated blocks attains tighter bounds than the corresponding result in [4].

REFERENCES

[1] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21-30, Mar. 2008.

[2] R. Baraniuk, M. Davenport, et al., "A simple proof of the restricted isometry property for random matrices," *Construct. Approx.*, vol. 28, no. 3, pp. 253-263, 2008.

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

[4] A. Eftekhari, H. L. Yap, et al., "The restricted isometry property for random block diagonal matrices," *arXiv preprint*, arXiv:1210.3395v1, Oct. 2012.

[5] C. Jie and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Signal Process.*, vol. 54, no. 12, pp. 4634-4643, Dec. 2006.

[6] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 505-519, Jan. 2010.

[7] M. E. Davies and Y. C. Eldar, "Rank awareness in joint sparse recovery," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1135-1146, Feb. 2012.

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

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

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

[11] J. Y. Park, H. L. Yap, C. J. Rozell, and M. B. Wakin, "Concentration of measure for block diagonal matrices with applications to compressive signal processing," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5859-5875, Dec. 2011.

[12] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed compressive sensing," *arXiv preprint*, arXiv:0901.3403v1, Jan. 2009.

[13] J. Tropp and A. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[1] For comparison fairness, we consider vectors with JK nonzero entries instead of K.