The Restricted Isometry Property for Block Diagonal Matrices

Han Lun Yap,* Armin Eftekhari, Michael B. Wakin, and Christopher J. Rozell

***Abstract*—In compressive sensing (CS), the Restricted Isometry Property (RIP) is a powerful condition on measurement operators which ensures robust recovery of sparse vectors is possible from noisy, undersampled measurements via computationally tractable algorithms. Early papers in CS showed that Gaussian random matrices satisfy the RIP with high probability, but such matrices are usually undesirable in practical applications due to storage limitations, computational considerations, or the mismatch of such matrices with certain measurement architectures. To alleviate some or all of these difficulties, recent research efforts have focused on structured random matrices. In this paper, we study block diagonal measurement matrices where each block on the main diagonal is itself a Gaussian random matrix. The main result of this paper shows that such matrices can indeed satisfy the RIP but that the requisite number of measurements depends on the coherence of the basis in which the signals are sparse. In the best case—for signals that are sparse in the frequency domain—these matrices perform nearly as well as dense Gaussian random matrices despite having many fewer nonzero entries.**

***Index Terms*—Compressive Sensing, Block Diagonal Matrices, Restricted Isometry Property**

I. Introduction

The literature on compressive sensing (CS) has established that many matrices satisfy the Restricted Isometry Property (RIP), guaranteeing a stable (distance preserving) embedding of sparse signals when using an undersampled linear measurement system [1]. The RIP is valuable because it ensures a type of information preservation in the low-dimensional measurement space, and it is often used to establish guarantees on robust signal recovery and signal processing in the compressed domain [2]. Recently, it has also been shown that if a matrix satisfies the RIP, that matrix (with its column signs randomized) can also be used to stably embed a finite collection of points [3] or a low-dimensional submanifold [4].

*Corresponding Author. HLY and CJR are with the School of Electrical and Computer Engineering at the Georgia Institute of Technology. AE and MBW are with the Division of Engineering at the Colorado School of Mines. This work was partially supported by NSF grants CCF-0830456 and CCF-0830320 and by DSO National Laboratories of Singapore.

There has been significant recent interest in structured measurement systems because unstructured random measurements (e.g., dense Gaussian random matrices) may be undesirable due to memory limitations, computational costs, or specific constraints in the data acquisition architecture. Many structured systems have been studied, including random convolution systems (described by partial Toeplitz [5] and circulant matrices [6]) and deterministic matrix constructions [7].

In this paper, we consider the special case of random block diagonal matrices that are zero everywhere except along the main diagonal [8–10]. While such matrices have many fewer degrees of randomness (and therefore require fewer memory resources), they are also particularly useful for representing acquisition systems with architectural constraints that prevent global data aggregation. For example, this type of architecture arises in distributed sensing systems where communication constraints limit the dependence of each measurement to only a subset of the data and in streaming applications where signals have datarates that necessitate operating on local signal blocks rather than the entire signal simultaneously. In these scenarios, the data may be divided naturally into discrete blocks, with each block acquired via a local measurement operator.

To make things concrete, for some positive integers J, N, we model a signal $x \in \mathbb{R}^{JN}$ as being partitioned into J blocks $x_1, x_2, \ldots, x_J \in \mathbb{R}^N$. As an example, x can be a video sequence and $x_1, x_2, \ldots, x_J$ can be the individual frames in the video. For each $j \in \{1, 2, \ldots, J\}$, we suppose that a linear operator $\Phi_j : \mathbb{R}^N \to \mathbb{R}^M$ collects the measurements $y_j = \Phi_j x_j$. In our example, this means that each video frame x_j is measured with a different operator. Concatenating all of the measurements into a vector $y \in \mathbb{R}^{JM}$, we then have

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{bmatrix}}_{y:\ \widetilde{M}\times 1} = \underbrace{\begin{bmatrix} \Phi_1 & & & \\ & \Phi_2 & & \\ & & \ddots & \\ & & & \Phi_J \end{bmatrix}}_{\Phi:\ \widetilde{M}\times\widetilde{N}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_J \end{bmatrix}}_{x:\widetilde{N}\times 1}, \quad (1)$$

where $\widetilde{M} = JM$ and $\widetilde{N} = JN$. Thus we see that the overall measurement operator Φ will have a characteristic block diagonal structure.

In this paper, we suppose that each local measurement operator Φ_j is a random matrix populated with independent and identically distributed (i.i.d.) zero mean Gaussian random variables. Following [10], we say that the resulting Φ has a *Distinct Block Diagonal* (DBD) structure.

Our main result is in establishing that undersampled (i.e., $\widetilde{M} < \widetilde{N}$) DBD matrices can indeed satisfy the RIP but that the requisite number of measurements $\widetilde{M}$ depends on the basis in which the signals are sparse. Surprisingly, we show that when the signals of interest are sparse in the frequency domain, DBD matrices satisfy the RIP with approximately the same number of rows required in a dense Gaussian matrix (despite having many fewer nonzero entries).

II. Background and Related Work

In this section, we formally state the main concepts that are relevant for this work (including the RIP) and present notation that will be used throughout the paper. We also give a brief overview of work related to the study of block diagonal matrices.

A. RIP and Compressive Sensing

We say that a vector $\alpha \in \mathbb{C}^{\widetilde{N}}$ is S-sparse if $\|\alpha\|_0 \le S$, where $\|\cdot\|_0$ counts the number of nonzero entries in a vector. Matrices which satisfy the RIP preserve the norms of sparse vectors. This is codified in the following definition.

Definition II.1. *A linear operator* $\widetilde{\Phi} : \mathbb{C}^{\widetilde{N}} \to \mathbb{C}^{\widetilde{M}}$ *satisfies the* Restricted Isometry Property *of* order S *and* conditioning δ *(or RIP-(S, δ) in short) if for all S-sparse* $\alpha \in \mathbb{C}^{\widetilde{N}}$*, we have:*

$$(1-\delta)\|\alpha\|_2^2 \le \|\widetilde{\Phi}\alpha\|_2^2 \le (1+\delta)\|\alpha\|_2^2.$$

If an undersampled matrix $\widetilde{\Phi}$ satisfies RIP-($2S, \delta$), then we see that for any two S-sparse vectors u and v, $\|\widetilde{\Phi}u - \widetilde{\Phi}v\|_2^2 \approx \|u-v\|_2^2$. This distance preservation allows sparse vectors to be recovered from undersampled measurements using standard techniques such as ℓ_1-minimization [1]. For example, the following theorem quantifies guarantees on robust recovery of sparse (and nearly sparse) vectors from noisy measurements.

Theorem II.1. [11] *Assume a matrix* $\widetilde{\Phi}$ *satisfies RIP-($2S, \delta$) with $\delta < 0.4651$. Let $\alpha \in \mathbb{C}^{\widetilde{N}}$ be any vector and suppose we acquire the noisy measurements $y = \widetilde{\Phi}\alpha + e$ with $\|e\|_2 \le \eta$. Let $\widehat{\alpha}$ be the unique solution of:*

$$\min_z \|z\|_1 \text{ subject to } \|\widetilde{\Phi}z - y\|_2 \le \eta. \quad (2)$$

Then

$$\|\alpha - \widehat{\alpha}\|_2 \le C_1\eta + C_2\frac{\sigma_S(\alpha)_1}{\sqrt{S}}, \quad (3)$$

where $\sigma_S(\alpha)_1 := \inf\{\|\alpha - z\|_1 : z \text{ is } S\text{-sparse}\}$ *is the error (measured in the ℓ_1 norm) of the best S-term approximation of α, and C_1, C_2 are some constants.*

Thus for an input vector α and measurement operator $\widetilde{\Phi}$ satisfying the RIP, solving the ℓ_1-minimization program guarantees an output $\widehat{\alpha}$ whose distance from α is bounded both by the measurement noise level and by the distance from α to its best S-term approximation. When Φ is an $\widetilde{M} \times \widetilde{N}$ matrix populated with i.i.d. Gaussian random variables having mean zero and variance $\widetilde{M}^{-1}$, it is known that Φ satisfies RIP-(S, δ) with high probability when $\widetilde{M} = O\left(\delta^{-2} S \log(N/S)\right)$. Therefore nonadaptively measuring a vector α via a random matrix can be nearly as efficient as adaptively recording the S largest terms of the vector directly.

In this paper, we are interested in showing that random DBD matrices can also satisfy the RIP and can therefore be successfully used to measure sparse vectors. The proof techniques that we employ require an alternative (but equivalent) formulation of the RIP. Define the set of all (normalized) S-sparse vectors as:

$$T := \left\{\alpha \in \mathbb{C}^{\widetilde{N}} : \|\alpha\|_0 \le S, \ \|\alpha\|_2 \le 1\right\},$$

and define the following norm on Hermitian $\widetilde{N} \times \widetilde{N}$ matrices:

$$|||A||| := \sup_{\alpha \in T} |\alpha^H A \alpha|.$$

Note that the sparsity level S of the vectors under consideration is implicit in the definition of this norm. Then we see that a matrix $\widetilde{\Phi}$ satisfies RIP-(S, δ) if and only if

$$\left|\left|\left|\widetilde{\Phi}^H\widetilde{\Phi} - I\right|\right|\right| \le \delta.$$

In Section IV-A, we shall see that by writing the conditioning this way, we will be able to draw on powerful results for bounding the suprema of random processes.

B. Coherence and Universality

Many natural signals can be sparsified when transformed in an orthonormal basis. In particular, we say that a signal $x \in \mathbb{C}^{\widetilde{N}}$ is S-sparse in an orthonormal basis $U \in \mathbb{C}^{\widetilde{N} \times \widetilde{N}}$ if $\|U^H x\|_0 \le S$. Such signals can also be recovered from compressive measurements much like the sparse vectors described in Section II-A. In particular, given noisy measurements $y = \Phi x + e$ of a signal x that is exactly or approximately sparse in a basis U, it is straightforward to adapt (2) to recover an estimate of x: one simply sets $\widetilde{\Phi} := \Phi U$, solves for $\widehat{\alpha}$ as in (2), and sets $\widehat{x} = U\widehat{\alpha}$. Since U is assumed to be unitary, it follows that $\|x - \widehat{x}\|_2 = \|\alpha - \widehat{\alpha}\|_2$, where $\alpha = U^H x$, and (3) provides a bound on $\|\alpha - \widehat{\alpha}\|_2$. Therefore, if ΦU satisfies the RIP, one can ensure that $\|x - \widehat{x}\|_2$ is small for signals that are sparse (or nearly sparse) in the basis U.

Some distributions of random matrices Φ (such as dense Gaussian matrices) are known to be *universal* in that, for any fixed orthobasis U, ΦU is as likely to satisfy the RIP as Φ itself. Such matrices can therefore be used equally well for sensing signals that are sparse in any known basis. Based on a past analysis of the statistical properties of random DBD matrices [10], however, it is unlikely that these matrices will be universal. For a given orthobasis U, then, it will be important for us to understand the impact of U in determining whether ΦU satisfies the RIP.

The critical characterization of an orthobasis U will be its *coherence*, which is defined to be

$$\mu_U := \sqrt{\widetilde{N}} \max_{1 \le p,q \le \widetilde{N}} |u_{p,q}|,$$

where $\{u_{p,q}\}$ are the entries of the matrix U. To be precise, μ_U is a measure of the coherence between the basis U and the canonical basis, i.e., the maximum inner product between the rows of U and the columns of I. Further note that $1 \le \mu_U \le \sqrt{\widetilde{N}}$, where the minimum is attained by a basis that is perfectly *incoherent* with the canonical basis (e.g., the Fourier basis [1]), and the maximum is attained by the canonical basis itself.

C. Related Work

Block diagonal matrices have been the subject of a *concentration of measure* analysis in [8–10], where it was proved that the probability of ensuring $\|\Phi x\|_2^2 \approx \|x\|_2^2$ depends on the characteristics of the signal x itself, but that for the best case signals, DBD matrices can exhibit the same concentration performance as dense Gaussian matrices. Our paper extends these previous results by showing that DBD matrices populated with Gaussian entries can indeed satisfy the RIP for signals that are sparse in an arbitrary basis U, but that the requisite number of measurements depends on the coherence of U.

This work is also a generalization of previous results in which RIP bounds are essentially derived for DBD matrices that are populated with Rademacher[1] entries and that have only one measurement per block ($M = 1$) [12, 13]. In [12], the bound is specific to Fourier sparse signals and in [13], an additional step of randomly subsampling the measurements is considered.

III. RIP FOR DBD MATRICES

Our main RIP result for DBD matrices is encapsulated in the following theorem, which we prove in Section IV.

Theorem III.1. *Suppose $\widetilde{N}, S > 2$ and suppose $\Phi \in \mathbb{R}^{\widetilde{M} \times \widetilde{N}}$ (with $\widetilde{M} \le \widetilde{N}$) is a random DBD matrix as defined in* (1) *with i.i.d. Gaussian entries of zero mean and variance $\sigma^2 = \frac{1}{M}$. Also suppose $U \in \mathbb{C}^{\widetilde{N} \times \widetilde{N}}$ is a unitary matrix representing a basis. Let $\widetilde{\Phi} := \Phi U$ and choose an RIP conditioning $0 < \delta < 1$. If the total number of measurements satisfies*

$$\widetilde{M} \ge \frac{C_3}{\delta^2} S \, \widetilde{\mu}_U^2 \, \log^2(S) \, \log^4(\widetilde{N}),$$

where C_3 is a constant and where $\widetilde{\mu}_U := \min\left\{\sqrt{J}, \mu_U\right\}$, then with probability at least $1 - 8\widetilde{N}^{-1}$, $\widetilde{\Phi}$ satisfies RIP-(S, δ), i.e.,

$$\mathbb{P}\left\{ \left\| \widetilde{\Phi}^H \widetilde{\Phi} - I \right\| > \delta \right\} \le 8\widetilde{N}^{-1}.$$

Using the fact that $\log(S) \le \log(\widetilde{N})$, we first see that the number of measurements required to satisfy the RIP scales like $O\left(S\widetilde{\mu}_U^2 \log^6(\widetilde{N})\right)$. We also note that $1 \le \widetilde{\mu}_U^2 \le J$, where the lower bound is attained when U is the Fourier basis. Thus, when measuring signals that are sparse in the frequency domain, we see the surprising result that a DBD matrix performs nearly as well (within log factors) as a fully dense random matrix of an equivalent size. On the other hand, when the basis U is highly coherent with the canonical basis, the requisite number of measurements $\widetilde{M}$ is proportional to SJ (instead of S as desired in the best case). While possibly unfavorable, this number can still be parsimonious ($\widetilde{M} \ll \widetilde{N}$) if the sparsity level S of the signal x is much less than the length N of each signal block x_j.

In fact when $U = I$ (and possibly for all U such that $\mu_U^2 \ge J$), we cannot expect to satisfy the RIP with a number of measurements $\widetilde{M}$ much less than $O\left(SJ\log(\widetilde{N})\right)$. To see why, suppose we have a signal x whose S nonzero entries are located within a single block x_j. We then know that the number of rows M in Φ_j cannot be less than $O\left(\delta^{-2} S \log(N/S)\right)$ for the RIP to hold since Φ_j itself is a Gaussian random matrix [1].

[1]A Rademacher random variable takes a value of 1 or -1, each with probability $\frac{1}{2}$.

Since the RIP is a uniform bound over all S-sparse signals, we must take at least this many measurements in each block. Therefore, our results are tight (to within log factors) for the canonical basis.

IV. Proof of Theorem III.1

The proof of this theorem utilizes recent creative techniques for bounding isometry constants via the extrema of random processes [12–14]. In particular, our proof follows very closely the proof structure for Theorem 2 in [12]. In short, we first show that $\mathbb{E}\left\{\left|\left|\left|\widetilde{\Phi}^H\widetilde{\Phi} - I\right|\right|\right|\right\} \le \delta$ if we pick $\widetilde{M}$ large enough. Once this is established, we complete the proof by applying a tail bound on the random variable $\left|\left|\left|\widetilde{\Phi}^H\widetilde{\Phi} - I\right|\right|\right|$ to establish that "bad" matrices occur with very low probability.

A. Important Lemmas

In order to prove Theorem III.1, we draw mainly on two recent results from the literature. This section defines notation and states those results so they can be employed where appropriate.

First, for $1 \le \widetilde{m} \le \widetilde{M}$, let the $\widetilde{m}$-th row of $\widetilde{\Phi}$ be represented by $\widetilde{\phi}_{\widetilde{m}}^H$ where $\widetilde{\phi}_{\widetilde{m}} \in \mathbb{C}^{\widetilde{N}}$. It is not difficult to see that $\widetilde{\Phi}^H\widetilde{\Phi}$ can be written as a sum of independent rank-1 operators:

$$\widetilde{\Phi}^H\widetilde{\Phi} = \sum_{\widetilde{m}=1}^{\widetilde{M}} \widetilde{\phi}_{\widetilde{m}} \otimes \widetilde{\phi}_{\widetilde{m}}, \tag{4}$$

where for any vector $a \in \mathbb{C}^N$, $a \otimes a := aa^H$ is the *outer product* of the vector. The first lemma, originally stated in [14] and later refined in [12], shows that we can bound the expected norm of a random sum of rank-1 operators.

Lemma IV.1. [12, 14] *Let $\{\widetilde{\phi}_{\widetilde{m}}\}$ be a (fixed) sequence of $\widetilde{M}$ vectors in $\mathbb{C}^{\widetilde{N}}$ that have bounded entries, i.e., $\|\widetilde{\phi}_{\widetilde{m}}\|_\infty \le B_1$, and let $\{\xi_{\widetilde{m}}\}$ be a Rademacher sequence. Then*

$$\mathbb{E}\left\{\left|\left|\left|\sum_{\widetilde{m}} \xi_{\widetilde{m}}\widetilde{\phi}_{\widetilde{m}} \otimes \widetilde{\phi}_{\widetilde{m}}\right|\right|\right|\right\} \le \beta \left|\left|\left|\sum_{\widetilde{m}} \widetilde{\phi}_{\widetilde{m}} \otimes \widetilde{\phi}_{\widetilde{m}}\right|\right|\right|^{1/2},$$

where $\beta \le C_4 B_1 \sqrt{S}\log S \sqrt{\log \widetilde{N}}\sqrt{\log \widetilde{M}}$ for a universal constant C_4.

The second lemma establishes a tail bound for random variables in a Banach space.

Lemma IV.2. [12] *Let $Y_1, Y_2, \ldots, Y_R$ be* independent, symmetric[2] *random variables in a Banach space X such that $\|Y_r\|_X \le B_2$* almost surely. *Let $Y = \|\sum_r Y_r\|_X$. Then for any $u, t > 1$,*

$$\mathbb{P}\{Y > C_5\left(u\mathbb{E}\{Y\} + tB_2\right)\} \le e^{-u^2} + e^{-t}.$$

[2]A random variable Y is *symmetric* if Y and $-Y$ have the same probability distribution.

Using these two lemmas, the proof of our main result proceeds in two stages. In Section IV-B, we apply Lemma IV.1 to our quantity of interest $\left(\left|\left|\left|\widetilde{\Phi}^H\widetilde{\Phi} - I\right|\right|\right|\right)$ to establish a bound on its expectation. In Section IV-C, we then use Lemma IV.2 to confirm that this quantity is in fact small with high probability.

B. Bounding Expectation

The following lemma establishes a bound for $\left|\left|\left|\widetilde{\Phi}^H\widetilde{\Phi} - I\right|\right|\right|$ in expectation.

Lemma IV.3. *Pick an RIP conditioning $0 < \delta' < 1$. If*

$$\widetilde{M} \ge \frac{128C_4^2}{\delta'^2}\, S\, \widetilde{\mu}_U^2\, \log^2(S)\, \log^3(\widetilde{N}),$$

for C_4 appearing in Lemma IV.1, then

$$\mathbb{E}\left\{\left|\left|\left|\widetilde{\Phi}^H\widetilde{\Phi} - I\right|\right|\right|\right\} \le \delta'.$$

Proof: To begin, we note that $\mathbb{E}\left\{\widetilde{\Phi}^H\widetilde{\Phi}\right\} = \mathbb{E}\left\{U^H\Phi^H\Phi U\right\} = U^H\mathbb{E}\left\{\Phi^H\Phi\right\}U = U^HU = I$. Also, each entry $\widetilde{\phi}_{\widetilde{m},\widetilde{n}}$ of the matrix $\widetilde{\Phi}$ can be written in the form $\sum_k \phi_{\widetilde{m},k}u_{k,\widetilde{n}}$, which can be viewed as a sum of independent zero mean Gaussian random variables having different variances. Thus, by using a general form of the Khinchine inequality (Corollary 12 in [15]) and taking union bounds, it is possible to bound the maximum squared entry of $\widetilde{\Phi}$ in expectation:

$$\mathbb{E}\left\{\max_{\widetilde{m}} \|\widetilde{\phi}_{\widetilde{m}}\|_\infty^2\right\} \le \widetilde{\mu}_U^2 \frac{8\log \widetilde{N}}{\widetilde{M}}. \tag{5}$$

This simply means that the quantity B_1 in Lemma IV.1 can be bounded in expectation by $\mathbb{E}\left\{B_1^2\right\} \le \widetilde{\mu}_U^2 \frac{8\log \widetilde{N}}{\widetilde{M}}$.

Let $E := \mathbb{E}\left\{\left|\left|\left|\widetilde{\Phi}^H\widetilde{\Phi} - I\right|\right|\right|\right\}$ be the expectation of the quantity of interest and recall that $\widetilde{\Phi}^H\widetilde{\Phi}$ can be expressed as a sum of rank-1 operators as in (4). Then using a standard symmetrization procedure (take for example Lemma 69 in [15]), we can establish that

$$\begin{aligned} E &\le 2\mathbb{E}\left\{\left|\left|\left|\sum \xi_{\widetilde{m}}\widetilde{\phi}_{\widetilde{m}} \otimes \widetilde{\phi}_{\widetilde{m}}\right|\right|\right|\right\} \\ &= 2\mathbb{E}_{\{\widetilde{\phi}_{\widetilde{m}}\}}\left\{\mathbb{E}_{\{\xi_{\widetilde{m}}\}}\left\{\left|\left|\left|\sum \xi_{\widetilde{m}}\widetilde{\phi}_{\widetilde{m}} \otimes \widetilde{\phi}_{\widetilde{m}}\right|\right|\right|\right\} \mid \{\widetilde{\phi}_{\widetilde{m}}\}\right\}, \end{aligned}$$

where $\{\xi_{\widetilde{m}}\}$ is a Rademacher sequence independent of $\{\widetilde{\phi}_{\widetilde{m}}\}$ and the second statement simply uses the law of iterated expectation. Applying Lemma IV.1 to the inner expectation yields:

$$\begin{aligned} E &\leq 2\mathbb{E}_{\{\widetilde{\phi}_{\widetilde{m}}\}}\left\{\beta \left|\!\left|\!\left|\sum \widetilde{\phi}_{\widetilde{m}} \otimes \widetilde{\phi}_{\widetilde{m}}\right|\!\right|\!\right|^{1/2}\right\} \\ &\leq \sqrt{4C_4^2 S \log^2 S \log^2 \widetilde{N}} \cdot \sqrt{\mathbb{E}\{B_1^2\}} \cdot \\ &\quad \sqrt{\mathbb{E}\left\{\left|\!\left|\!\left|\sum_{\widetilde{m}} \widetilde{\phi}_{\widetilde{m}} \otimes \widetilde{\phi}_{\widetilde{m}}\right|\!\right|\!\right|\right\}}, \end{aligned} \tag{6}$$

where the second line follows from the Cauchy-Schwarz inequality. Applying (4), we observe that

$$\begin{aligned} \mathbb{E}\left\{\left|\!\left|\!\left|\sum \widetilde{\phi}_{\widetilde{m}} \otimes \widetilde{\phi}_{\widetilde{m}}\right|\!\right|\!\right|\right\} &= \mathbb{E}\left\{\left|\!\left|\!\left|\widetilde{\Phi}^H \widetilde{\Phi} - I + I\right|\!\right|\!\right|\right\} \\ &\leq E + 1, \end{aligned}$$

where the second step follows from the triangle inequality. Combining this statement with the bounds in (5) and (6), we have the following bound on the expectation:

$$E \leq \sqrt{\frac{32C_4^2 S \widetilde{\mu}_U^2 \log^2 S \log^3 \widetilde{N}}{\widetilde{M}}}\sqrt{E+1}.$$

By manipulating the above inequality, we find that (see for example [12]) because the conditions of Lemma IV.3 imply that $\frac{32C_4^2 S \widetilde{\mu}_U^2 \log^2 S \log^3 \widetilde{N}}{\widetilde{M}} \leq 1$, then:

$$E \leq 2\sqrt{\frac{32C_4^2 S \widetilde{\mu}_U^2 \log^2 S \log^3 \widetilde{N}}{\widetilde{M}}}.$$

Requiring the right hand side to be less than δ' concludes the proof. ∎

C. Tail Bound

We complete the proof of Theorem III.1 by arguing that $\left|\!\left|\!\left|\widetilde{\Phi}^H \widetilde{\Phi} - I\right|\!\right|\!\right|$ must be small with high probability.

First, using (4) and the fact that $\mathbb{E}\left\{\widetilde{\Phi}^H \widetilde{\Phi}\right\} = I$, we see that

$$\begin{aligned} Z &:= \left|\!\left|\!\left|\widetilde{\Phi}^H \widetilde{\Phi} - I\right|\!\right|\!\right| \\ &= \left|\!\left|\!\left|\sum_{\widetilde{m}=1}^{\widetilde{M}} \left(\widetilde{\phi}_{\widetilde{m}} \otimes \widetilde{\phi}_{\widetilde{m}} - \mathbb{E}_{\{\widetilde{\phi}_{\widetilde{m}}\}}\left\{\widetilde{\phi}_{\widetilde{m}} \otimes \widetilde{\phi}_{\widetilde{m}}\right\}\right)\right|\!\right|\!\right|, \end{aligned}$$

Although Z consists of a sum of independent random variables, we cannot apply Lemma IV.2 directly to this sum because it is unclear whether the summands are symmetric and we do not have a satisfactory bound on the norms of the summands. As such consider the random variable

$$Y := \left|\!\left|\!\left|\sum_{\widetilde{m}=1}^{\widetilde{M}} \left(\widetilde{\phi}_{\widetilde{m}} \otimes \widetilde{\phi}_{\widetilde{m}} - \widetilde{\phi}'_{\widetilde{m}} \otimes \widetilde{\phi}'_{\widetilde{m}}\right)\right|\!\right|\!\right|,$$

where $\{\widetilde{\phi}'_{\widetilde{m}}\}$ are independent copies of $\{\widetilde{\phi}_{\widetilde{m}}\}$. We will use Y as a proxy for Z since the summands of Y are symmetric random variables.[3] The random variables Y and Z are related by the following inequality (see for example (15) in [12]):

$$\mathbb{P}\{Z > 2\mathbb{E}\{Z\} + u\} \leq 2\mathbb{P}\{Y > u\}. \tag{7}$$

From the triangle inequality we have the additional relation:

$$\mathbb{E}\{Y\} \leq 2\mathbb{E}\{Z\}. \tag{8}$$

To bound the summands of Y, first define the events $\{F_{\widetilde{m}}\}$ and F:

$$\begin{aligned} F_{\widetilde{m}} &:= \left\{\max\{\|\widetilde{\phi}_{\widetilde{m}}\|_\infty^2, \|\widetilde{\phi}'_{\widetilde{m}}\|_\infty^2\} \leq \widetilde{\mu}_U^2 \frac{14 \log \widetilde{N}}{\widetilde{M}}\right\}, \\ F &:= \bigcap_{\widetilde{m}} F_{\widetilde{m}}. \end{aligned}$$

From basic norm inequalities, it can be shown that under the event F, all of the summands of Y are bounded in the $|\!|\!|\cdot|\!|\!|$ norm by $B_2 := \frac{28 S \widetilde{\mu}_U^2 \log \widetilde{N}}{\widetilde{M}}$. Thus, we define a truncated version of Y to which we may apply Lemma IV.2:

$$Y_t := \left|\!\left|\!\left|\sum_{\widetilde{m}=1}^{\widetilde{M}} \left(\widetilde{\phi}_{\widetilde{m}} \otimes \widetilde{\phi}_{\widetilde{m}} - \widetilde{\phi}'_{\widetilde{m}} \otimes \widetilde{\phi}'_{\widetilde{m}}\right) \mathbb{I}_{F_{\widetilde{m}}}\right|\!\right|\!\right|,$$

where $\mathbb{I}_{F_{\widetilde{m}}}$ is the indicator function of the event $F_{\widetilde{m}}$. The following relates the probability distributions of Y and Y_t:

$$\begin{aligned} \mathbb{P}\{Y > \nu\} &= \mathbb{P}\{Y > \nu | F\}\, \mathbb{P}\{F\} + \\ &\quad \mathbb{P}\{Y > \nu | F^c\}\, \mathbb{P}\{F^c\} \\ &\leq \mathbb{P}\{Y_t > \nu\} + \mathbb{P}\{F^c\}. \end{aligned} \tag{9}$$

Thus we need to calculate the probability of occurrence of event F^c. Because (5) provides a bound on $\mathbb{E}\left\{\max_{\widetilde{m}} \|\widetilde{\phi}_{\widetilde{m}}\|_\infty^2\right\}$, from an appropriate application of the Markov inequality, one can show that

$$\mathbb{P}\left\{\max_{\widetilde{m}} \|\widetilde{\phi}_{\widetilde{m}}\|_\infty^2 > \widetilde{\mu}_U^2 \frac{14 \log \widetilde{N}}{\widetilde{M}}\right\} \leq \widetilde{N}^{-1}.$$

It follows that

$$\mathbb{P}\{F^c\} \leq 2\widetilde{N}^{-1}. \tag{10}$$

[3] If z is a random variable and z' is an independent copy of z, then $y := z - z'$ is a symmetric random variable (called the *symmetrized version* of z).

From Lemma IV.3, we know that if $0 < \delta' < 1$ and

$$\widetilde{M} \geq \frac{128 C_4^2}{(\delta')^2} S \widetilde{\mu}_U^2 \log^2 S \log^3 \widetilde{N},$$

then $\mathbb{E}\{Z\} \leq \delta'$. Under this condition on $\widetilde{M}$, and supposing the event F occurs, then

$$B_2 \leq \frac{7}{32 C_4^2} \frac{(\delta')^2}{\log^2 S \log^2 \widetilde{N}}. \tag{11}$$

Thus, applying Lemma IV.2 to Y_t with $u, t > 1$ gives

$$\mathbb{P}\{Y_t > C_5 \left(u \mathbb{E}\{Y_t\} + t B_2\right)\} \leq e^{-u^2} + e^{-t}.$$

Recalling the bound on B_2 in (11) and noting that[4]

$$\mathbb{E}\{Y_t\} \leq \mathbb{E}\{Y\} \leq 2\mathbb{E}\{Z\},$$

we set $u = \sqrt{\log \widetilde{N}}$ and $t = \log \widetilde{N}$ and conclude that

$$\mathbb{P}\left\{ Y_t > C_5 \left(2\sqrt{\log \widetilde{N}} \delta' + \frac{7}{32 C_4^2} \frac{(\delta')^2}{\log^2 S \log \widetilde{N}} \right) \right\} \leq 2\widetilde{N}^{-1}.$$

Substituting $\delta' = \frac{C_6 \delta}{\sqrt{\log \widetilde{N}}}$ with $C_6 < 1$ to be chosen below, we get (after some manipulations):

$$\mathbb{P}\left\{ Y_t > C_5 \left(2 + \frac{7}{32 C_4^2} \right) C_6 \delta \right\} \leq 2\widetilde{N}^{-1}.$$

To remove the dependence on the event F, apply (9) together with (10) to give:

$$\mathbb{P}\left\{ Y > C_5 \left(2 + \frac{7}{32 C_4^2} \right) C_6 \delta \right\} \leq 4\widetilde{N}^{-1}.$$

Further applying (7) and the fact that $\mathbb{E}\{Z\} \leq \delta' \leq C_6 \delta$ gives us:

$$\mathbb{P}\left\{ Z > \left[C_5 \left(2 + \frac{7}{32 C_4^2} \right) + 2 \right] C_6 \delta \right\} \leq 8\widetilde{N}^{-1}.$$

Choosing $C_6 \leq \left[C_5 \left(2 + \frac{7}{32 C_4^2} \right) + 2 \right]^{-1}$ completes the proof of Theorem III.1.

V. Summary and Future Work

In this paper, we have showed that random DBD matrices populated with i.i.d. Gaussian entries can satisfy the RIP. Our main result indicates that the requisite number of measurements $\widetilde{M}$ scales linearly in the sparsity level S, quadratically in the coherence $\widetilde{\mu}_U$ of the sparse basis, and poly-logarithmically in the ambient dimension $\widetilde{N}$. We note that, when considering signals that are sparse in the frequency domain, a random DBD matrix requires essentially the same number of rows (up to a poly-logarithmic factor in the ambient dimension) as a Gaussian random matrix.

There are several interesting directions for future analysis of block diagonal matrices. For example, it would be useful to know what other random distributions could be used to populate these matrices. It would also be useful to study the effects of either generalizing this architecture (e.g., having a different number of measurements M_j in each block) or restricting this architecture (e.g., repeating a single block along the main diagonal).

[4]The proof of the first inequality requires us to multiply the summands of Y_t by Rademacher sequence, which does not change the distribution of Y (see [12] for details).

References

[1] E. J. Candès, "Compressive sampling," in *Proc. Int. Congr. Math.*, vol. 3, Madrid, Spain, August 2006, pp. 1433–1452.

[2] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, "Signal processing with compressive measurements," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 445–460, 2010.

[3] F. Krahmer and R. Ward, "New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property," *Arxiv preprint arXiv:1009.0744*, 2010.

[4] H. L. Yap, M. B. Wakin, and C. J. Rozell, "Stable manifold embeddings with operators satisfying the restricted isometry property," in *Proc. Conf. Information Sciences and Systems (CISS)*, March 2011.

[5] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inform. Theory*, 2008.

[6] H. Rauhut, J. K. Romberg, and J. A. Tropp, "Restricted isometries for partial random circulant matrices," *Arxiv preprint arXiv:1010.1847*, 2010.

[7] R. A. DeVore, "Deterministic constructions of compressed sensing matrices," *Journal of Complexity*, vol. 23, no. 4-6, pp. 918–925, 2007.

[8] M. B. Wakin, J. Y. Park, H. L. Yap, and C. J. Rozell, "Concentration of measure for block diagonal measurement matrices," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, March 2010.

[9] C. J. Rozell, H. L. Yap, J. Y. Park, and M. B. Wakin, "Concentration of measure for block diagonal matrices with repeated blocks," in *Proc. Conf. Information Sciences and Systems (CISS)*, February 2010.

[10] J. Y. Park, H. L. Yap, C. J. Rozell, and M. B. Wakin, "Concentration of measure for block diagonal matrices with applications to compressive sensing," 2010, preprint.

[11] H. Rauhut, "Compressive sensing and structured random matrices," *Theoretical Foundations and Numerical Methods for Sparse Recovery*, 2010.

[12] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 520–544, 2009.

[13] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "A restricted isometry property for structurally-subsampled unitary matrices," in *Proc. 47th Annual Allerton Conf. Communication, Control, and Computing (Allerton)*, 2009, pp. 1005–1012.

[14] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Communications on Pure and Applied Mathematics*, vol. 61, no. 8, pp. 1025–1045, 2008.

[15] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *Compressed sensing: theory and applications. Cambridge University Press. Submitted.*