Concentration of Measure for Block Diagonal Matrices with Applications to Compressive Sensing

Jae Young Park, Han Lun Yap, Christopher J. Rozell, and Michael B. Wakin

October 2010

Abstract

Theoretical analysis of randomized, compressive operators often depends on a concentration of measure inequality for the operator in question. Typically, such inequalities quantify the likelihood that a random matrix will preserve the norm of a signal after multiplication. When this likelihood is very high for any signal, the random matrices have a variety of known uses in dimensionality reduction and Compressive Sensing. Concentration of measure results are well-established for unstructured compressive matrices, populated with independent and identically distributed (i.i.d.) random entries. Many real-world acquisition systems, however, are subject to architectural constraints that make such matrices impractical. In this paper we derive concentration of measure bounds for two types of block diagonal compressive matrices, one in which the blocks along the main diagonal are random and independent, and one in which the blocks are random but equal. For both types of matrices, we show that the likelihood of norm preservation depends on certain properties of the signal being measured, but that for the best case signals, both types of block diagonal matrices can offer concentration performance on par with their unstructured, i.i.d. counterparts. We support our theoretical results with illustrative simulations as well as (analytical and empirical) investigations of several signal classes that are highly amenable to measurement using block diagonal matrices. Finally, we discuss applications of these results in establishing performance guarantees for solving signal processing tasks in the compressed domain (e.g., signal detection), and in establishing the Restricted Isometry Property for the Toeplitz matrices that arise in compressive channel sensing.

Index Terms

Concentration of measure phenomenon, Block diagonal matrices, Compressive Sensing, Restricted Isometry Property, Toeplitz matrices

I. Introduction

Recent technological advances have enabled the sensing and storage of massive volumes of data from a dizzying array of sources. While access to such data has revolutionized fields such as signal processing, the limits of some computing and storage resources are being tested, and front-end signal acquisition devices are not always able to support the desire to measure in increasingly finer detail. To confront these challenges, many signal processing researchers have begun investigating compressive linear operators $\Phi : \mathbb{R}^N \to \mathbb{R}^M$ for high resolution signals $x \in \mathbb{R}^N$ ($M < N$), either as a method for simple dimensionality reduction or as a model for novel data acquisition devices. In settings such as these, the data x is often thought to belong to some concise model class; for example, in the field of Compressive Sensing (CS) [3, 4], one assumes that x has

JYP and HLY contributed equally to this work. JYP is with the Department of Electrical Engineering and Computer Science at the University of Michigan. HLY and CJR are with the School of Electrical and Computer Engineering at the Georgia Institute of Technology. MBW is with the Division of Engineering at the Colorado School of Mines. Preliminary versions of portions of this work appeared in [1] and [2]. This work was partially supported by NSF grants CCF-0830456 and CCF-0830320, DARPA grant HR0011-08-1-0078, and AFOSR grant FA9550-09-1-0465. The authors are grateful to Tyrone Vincent, Borhan Sanandaji, Armin Eftekhari, and Justin Romberg for valuable discussions about this work.

a *sparse* representation (having few nonzero coefficients) in the time domain or in some transform basis. If the number of "measurements" M is sufficient relative to the complexity of the model, the signal x can be recovered from Φx by solving an inverse problem (often cast as a convex optimization program [3, 4]). Because of their universality and amenability to analysis, randomized compressive linear operators (i.e., random matrices with $M < N$) have drawn particular interest.

The theoretical analysis of random matrices often relies on the general notions that these matrices are well-behaved most of the time, and that we can bound the probability with which they perform poorly. Frequently, these notions are formalized using some form of the *concentration of measure phenomenon* [5], a powerful characterization of the tendency of certain functions of high-dimensional random processes to concentrate sharply around their mean. As one important example of this phenomenon, it is known that for any fixed signal $x \in \mathbb{R}^N$, if Φ is an $M \times N$ matrix populated with independent and identically distributed (i.i.d.) random entries drawn from a suitable distribution, then with high probability Φ will approximately preserve the norm of x. More precisely, for many random distributions for Φ, the probability that $\left| \|\Phi x\|_2^2 - \|x\|_2^2 \right|$ will exceed a small fraction of $\|x\|_2^2$ decays exponentially in the number of measurements M (see Section II-B for additional details).

This likely preservation of signal norms makes random matrices remarkably useful. For example, the above mentioned concentration result has been used to prove the Johnson-Lindenstrauss (JL) lemma [6–8], which states that when applied to a finite set of points $Q \subset \mathbb{R}^N$, a randomized compressive operator Φ can provide a stable, distance preserving embedding of Q in the measurement space $\mathbb{R}^M$. This enables the efficient solution of problems such as finding the nearest neighbor to a point x in a database Q by permitting these problems to be solved in the low-dimensional observation space. The same concentration result has also been used to prove that certain families of random matrices can satisfy the Restricted Isometry Property (RIP) [9–11], which concerns the stable, distance preserving embedding of families of sparse signals. In the field of CS, the RIP is commonly used as a sufficient condition to guarantee that a sparse signal x can be recovered from the measurements Φx. In Section II, after providing a brief introduction to some preliminary ideas and notation, we survey related concentration analysis results and applications in the literature.

Despite the utility of norm preservation in dimensionality reduction, concentration analysis to date has focused almost exclusively on dense matrices that require each measurement to be a weighted linear combination of all entries of x. Dense random matrices with i.i.d. entries are often either impractical because of the resources required to store and work with a large unstructured matrix, or unrealistic as models of acquisition devices with architectural constraints preventing such global data aggregation. For one example, in a distributed sensing system, communication constraints may limit the dependence of each measurement to only a subset of the data. For a second example, applications involving streaming signals [12, 13] often have datarates that necessitate operating on local signal blocks rather than the entire signal simultaneously. For a third example, recent work has shown the utility of measuring x by convolving it with a random pulse and downsampling. These convolution measurement systems lead to computationally efficient designs and have been shown to work almost as well as dense randomized operators [14–22]. This paradigm is particularly relevant for compressive channel sensing [16–18], where a sparse wireless channel is estimated from its convolution with a random probe signal.

The common thread in each example above is that the data is divided naturally into discrete subsections (or blocks), and each

block is acquired via a local measurement operator.[1] To see the implications of this, let us model a signal $x \in \mathbb{R}^{NJ}$ as being partitioned into J blocks $x_1, x_2, \ldots, x_J \in \mathbb{R}^N$, and for each $j \in \{1, 2, \ldots, J\}$, suppose that a local measurement operator $\Phi_j : \mathbb{R}^N \to \mathbb{R}^{M_j}$ collects the measurements $y_j = \Phi_j x_j$. Concatenating all of the measurements into a vector $y \in \mathbb{R}^{\sum_j M_j}$, we then have

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{bmatrix}}_{y:\ (\sum_j M_j)\times 1} = \underbrace{\begin{bmatrix} \Phi_1 & & & \\ & \Phi_2 & & \\ & & \ddots & \\ & & & \Phi_J \end{bmatrix}}_{\Phi:\ (\sum_j M_j)\times NJ} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_J \end{bmatrix}}_{x: NJ\times 1}. \tag{1}$$

In cases such as these, we see that the overall measurement operator Φ will have a characteristic block diagonal structure. In some scenarios, the local measurement operator Φ_j may be unique for each block, and we say that the resulting Φ has a *Distinct Block Diagonal* (DBD) structure. In other scenarios it may be appropriate or necessary to repeat a single operator across all blocks (such that $\Phi_1 = \Phi_2 = \cdots = \Phi_J$); we call the resulting Φ a *Repeated Block Diagonal* (RBD) matrix.

In Section III of this paper, we derive concentration of measure bounds both for DBD matrices populated with i.i.d. subgaussian[2] random variables and for RBD matrices populated with i.i.d. Gaussian random variables. Our main results essentially state that the probability of concentration depends on the "diversity" of the component signals $x_1, x_2, \ldots, x_J$, where this notion of signal diversity depends whether the matrix is DBD or RBD (we make this precise in Section III). Such nonuniform concentration behavior is markedly unlike that of dense matrices, for which concentration probabilities are signal agnostic. At one extreme, for the most favorable classes of component signals, the concentration of measure probability for block diagonal matrices scales exactly as for a fully dense random matrix. In other words, the concentration probability decays exponentially with the total number of measurements, which in this case equals $\sum_j M_j$. At the other extreme, when the signal characteristics are less favorable, the measurement operator effectiveness is diminished possibly to the point that it is no more effective than measuring a single block. As we discuss analytically and support experimentally, our measures of diversity have clear intuitive interpretations and intimately capture the relevant phenomena dictating whether a signal x is well matched to the block diagonal structure of Φ. In Section IV we further discuss potential signal classes that are well-behaved for DBD and RBD matrices, and finally in Section V we discuss several possible applications of our results to tasks such as detection in the compressed space and proving RIP guarantees for the Toeplitz measurement matrices that arise in channel sensing.

II. Background and Related Work

In this section, we begin with a brief overview of subgaussian random variables and a survey of some of their key properties that will be of relevance in our analysis. We then describe more formally several existing concentration of measure results for random matrices, and we conclude by reviewing some standard applications of these results in the literature.

[1] See Section V-B1 for details on how the channel sensing problem fits into this viewpoint.

[2] Subgaussian random variables [23] are precisely defined in Section II-A, and can be thought of as random variables from a distribution with tails that can be bounded by a Gaussian.

A. Subgaussian Random Variables

In fields such as CS, the Gaussian distribution is often invoked for probabilistic analysis thanks to its many convenient properties. Gaussian random variables, however, are just one special case in a much broader class of so-called *subgaussian* distributions [23]. As we discuss below, subgaussian random variables actually retain some of the same properties that make Gaussians particularly well suited to concentration of measure analysis in CS. Thus, for the sake of generality, we will state our main results in terms of subgaussian random variables where possible.

Before proceeding, however, we present a brief definition of subgaussian random variables and a short overview of their key properties.

Definition II.1. [23] *A random variable w is* subgaussian *if $\exists a \geq 0$ such that*

$$\mathbf{E}e^{tw} \leq \exp\left(\frac{1}{2}a^2t^2\right) \text{ for all } t \in \mathbb{R}.$$

The quantity

$$\tau(w) := \inf\{a \geq 0 : \mathbf{E}e^{tw} \leq \exp\left(\frac{1}{2}a^2t^2\right) \text{ for all } t \in \mathbb{R}\}$$

is known as the Gaussian standard *of w.*

From this definition and Jensen's inequality, it follows that subgaussian random variables must always be centered, i.e., $\mathbf{E}w = 0$, and with some simple functional analysis one can also see that the variance $\mathbf{E}w^2 \leq \tau^2(w)$. The class of subgaussian random variables that achieve this bound with equality (i.e., for which $\mathbf{E}w^2 = \tau^2(w)$) are known as *strictly subgaussian* [23]. Examples of strictly subgaussian random variables include zero-mean Gaussian random variables, ± 1 Bernoulli random variables ($p = \frac{1}{2}$) and uniform random variables on $[-1, 1]$.

As with Gaussian random variables, linear combinations of i.i.d. subgaussian random variables are also subgaussian. This fact will be useful to us when studying the matrix-vector products that arise when a compressive linear operator is applied to a signal. We provide a more formal statement in the following lemma.

Lemma II.1. [23, Theorem 1 and Lemma 3] *Let $\beta \in \mathbb{R}^Z$ be a fixed vector, and suppose $w(1), w(2), \ldots, w(Z)$ are a collection of i.i.d. subgaussian random variables with Gaussian standards all equal to $\tau(w)$. Then the quantity $v := \sum_{i=1}^{Z} \beta(i)w(i)$ is a subgaussian random variable with Gaussian standard $\tau(v) \leq \tau(w)\|\beta\|_2$.*

Finally, we conclude with some important concentration results involving subgaussian random variables. The first result gives a standard bound on the tail distribution of a subgaussian random variable.

Lemma II.2. [23, Lemma 4] *Suppose that w is a subgaussian random variable with Gaussian standard $\tau(w)$. Then*

$$P(|w|^2 > t) \leq 2\exp\left(-\frac{t}{2\tau^2(w)}\right) \tag{2}$$

for all $t \geq 0$.

This property will also allow us to use subgaussian random variables in the following important theorem capturing the

concentration of measure properties of sums of random variables.

Theorem II.1. [24, Theorem 1.4] *Let $X_1, \ldots, X_L$ be independent Banach space valued random variables with $P\{\|X_i\| > t\} \leq a\exp\{-\alpha_i t\}$ for all t and i. Let $d \geq \max_i \alpha_i^{-1}$ and $b \geq a\sum_{i=1}^L \alpha_i^{-2}$. Then setting $S = \sum_{i=1}^L X_i$ we have*

$$P\{|\|S\| - \mathbf{E}\|S\|| > t\} \leq \begin{cases} 2\exp\{-t^2/32b\}, & 0 \leq t \leq \frac{4b}{d} \\ 2\exp\{-t/8d\}, & t \geq \frac{4b}{d} \end{cases}.$$

B. Concentration Inequalities for Dense Matrices with I.I.D. Random Entries

Concentration analysis to date has focused almost exclusively on dense random matrices populated with i.i.d. entries drawn from some distribution. Commonly, when Φ has size $M \times N$ and the entries are drawn from a suitably normalized distribution, then for any fixed signal $x \in \mathbb{R}^N$ the goal is to prove for any $\epsilon \in (0,1)$ that

$$P(\left|\|\Phi x\|_2^2 - \|x\|_2^2\right| > \epsilon\|x\|_2^2) \leq 2e^{-Mc_0(\epsilon)}, \tag{3}$$

where $c_0(\epsilon)$ is some constant (depending on ϵ) that is typically on the order of ϵ^2. When discussing bounds such as (3) where the probability of failure scales as e^{-X}, we refer to X as the as the *concentration exponent*.

The past several years have witnessed the derivation of concentration results for a variety of (ultimately related) random distributions for Φ. A concentration result of the form (3) was originally derived for dense Gaussian matrices populated with entries having mean zero and variance $\frac{1}{M}$ [25]; one straightforward derivation of this uses standard tail bounds for chi-squared random variables [7]. Using slightly more complicated arguments, similar concentration results were then derived for Bernoulli matrices populated with random $\frac{\pm 1}{\sqrt{M}}$ entries (each with probability $\frac{1}{2}$) and for a "database-friendly" variant populated with random $\{\frac{3}{\sqrt{M}}, 0, -\frac{3}{\sqrt{M}}\}$ entries (with probabilities $\{\frac{1}{6}, \frac{2}{3}, \frac{1}{6}\}$) [7]. Each of these distributions, however, is itself subgaussian, and more recently it has been shown that concentration results of the form (3) in fact hold for *all* subgaussian distributions having variance $\frac{1}{M}$ [11, 26].[3] Moreover, it has been shown that a subgaussian tail bound of the form (2) is actually necessary for deriving a concentration result of the form (3) for a dense random matrix populated with i.i.d. entries [26]. However, other forms of dense random matrices, such as those populated with random orthogonal rows, can also be considered [27].

C. Applications of Concentration Inequalities

One of the prototypical applications of a concentration result of the form (3) is to prove that with high probability, Φ will provide a stable, distance preserving embedding of some particular high-dimensional signal family $Q \subset \mathbb{R}^N$ in the low-dimensional measurement space $\mathbb{R}^M$. For example, supposing that Q consists of a finite number of points and that Φ is an $M \times N$ matrix populated with i.i.d. subgaussian random variables having variance $\frac{1}{M}$, it is possible to apply (3) to each vector of the form $u - v$ for $u, v \in Q$. Using the favorable probability of distance preservation for each pair, one can use a union bound to conclude that approximate isometry must hold simultaneously for all of these difference vectors with probability at

[3]This fact also follows from our Theorem III.1 by considering the special case where $J = 1$.

least $1 - 2|Q|^2 e^{-Mc_0(\epsilon)}$. From this fact one obtains the familiar JL lemma [6–8], which states that

$$(1-\epsilon)\|u-v\|_2^2 \leq \|\Phi(u-v)\|_2^2 \leq (1+\epsilon)\|u-v\|_2^2 \tag{4}$$

holds for all $u, v \in Q$ with high probability supposing that $M = O\left(\frac{\log(|Q|)}{c_0(\epsilon)}\right)$. The stable embedding of one or more discrete signals can be useful for solving various inference problems in the compressive domain. Potential problems of interest include nearest neighbor search [25], binary detection [28], multi-signal classification [28], and so on. We revisit the problem of compressive domain signal detection in Section V-A.

It is possible to significantly extend embedding results far beyond the JL lemma. For example, by coupling the above union bound approach with some elementary covering arguments, one can prove the RIP in CS [10, 11], which states that if $M = O\left(K \log(N/K)\right)$ (with a mild dependence on ϵ), the inequality (4) can hold with high probability for an (infinite) set Q containing all signals with sparsity K in some basis for $\mathbb{R}^N$. Supposing a matrix satisfies the RIP, one can derive deterministic bounds on the performance of CS recovery algorithms such as ℓ_1 minimization [29]. Alternately, a concentration result of the form (3) has also been used to probabilistically analyze the performance of ℓ_1 minimization [26]. Finally, we note that one can also generalize these covering/embedding arguments to the case where Q is a low-dimensional submanifold of $\mathbb{R}^N$ [30].

III. Main Concentration of Measure Inequalities

In this section we state our main concentration of measure results for Distinct Block Diagonal (DBD) and Repeated Block Diagonal (RBD) matrices. For each type of matrix we provide a detailed examination of the derived concentration rates and use simulations to demonstrate that our results do indeed capture the salient signal characteristics that affect the concentration probability. We also discuss connections between the concentration probabilities for the two matrix types.

A. Distinct Block Diagonal (DBD) Matrices

1) Analytical Results: Before stating our first result, let us define the requisite notation. For a given signal $x \in \mathbb{R}^{NJ}$ partitioned into J blocks of length N as in (1), we define a vector describing the energy distribution across the blocks of x:

$$\gamma = \gamma(x) := \begin{bmatrix} \|x_1\|_2^2 & \|x_2\|_2^2 & \cdots & \|x_J\|_2^2 \end{bmatrix}^T \in \mathbb{R}^J.$$

Also, letting $M_1, M_2, \ldots, M_J$ denote the number of measurements to be taken of each block, we define a $J \times J$ diagonal matrix containing these numbers along the diagonal:

$$\mathbf{M} := \begin{bmatrix} M_1 & & & \\ & M_2 & & \\ & & \ddots & \\ & & & M_J \end{bmatrix}.$$

Using this notation, our first main result concerning the concentration of DBD matrices is captured in the following theorem.

Theorem III.1. *Suppose $x \in \mathbb{R}^{NJ}$. For each $j \in \{1, 2, \ldots, J\}$, suppose that $M_j > 0$, and let Φ_j be a random $M_j \times N$ matrix populated with i.i.d. subgaussian entries having variance $\sigma_j^2 = \frac{1}{M_j}$ and Gaussian standard $\tau(\phi_j) \in [\sigma_j, \sqrt{c} \cdot \sigma_j]$ for some constant $c \geq 1$. Suppose that the matrices $\{\Phi_j\}_{j=1}^J$ are drawn independently of one another (though it is not necessary that they have the same subgaussian distribution), and let Φ be a $\left(\sum_{j=1}^J M_j\right) \times NJ$ DBD matrix composed of $\{\Phi_j\}_{j=1}^J$ as in (1). Then,*

$$P(\left|\|\Phi x\|_2^2 - \|x\|_2^2\right| > \epsilon \|x\|_2^2) \leq \begin{cases} 2\exp\{-\frac{\epsilon^2 \|\gamma\|_1^2}{256c^2 \|\mathbf{M}^{-1/2}\gamma\|_2^2}\}, & 0 \leq \epsilon \leq \frac{16c\|\mathbf{M}^{-1/2}\gamma\|_2^2}{\|\mathbf{M}^{-1}\gamma\|_\infty \|\gamma\|_1} \\ 2\exp\{-\frac{\epsilon\|\gamma\|_1}{16c\|\mathbf{M}^{-1}\gamma\|_\infty}\}, & \epsilon \geq \frac{16c\|\mathbf{M}^{-1/2}\gamma\|_2^2}{\|\mathbf{M}^{-1}\gamma\|_\infty \|\gamma\|_1} \end{cases}. \tag{5}$$

Proof: See Appendix A.

As we can see from the tail bound (5), the concentration probability of interest decays exponentially as a function of $\frac{\epsilon^2\|\gamma\|_1^2}{\|\mathbf{M}^{-1/2}\gamma\|_2^2}$ in the case of small ϵ and $\frac{\epsilon\|\gamma\|_1}{\|\mathbf{M}^{-1}\gamma\|_\infty}$ in the case of larger ϵ. One striking thing about Theorem III.1 is that, in contrast to analogous results for dense matrices, the concentration rate depends explicitly on characteristics of the signal x being measured. To elaborate on this point, since we are frequently concerned in practice with applications where ϵ is small, let us focus on the first case listed in (5). We see that in this case, the concentration exponent scales with

$$\Gamma = \Gamma(x, \mathbf{M}) := \frac{\|\gamma\|_1^2}{\|\mathbf{M}^{-1/2}\gamma\|_2^2} = \frac{\left(\sum_{j=1}^J \|x_j\|_2^2\right)^2}{\sum_{j=1}^J \frac{\|x_j\|_2^4}{M_j}}, \tag{6}$$

where larger values of Γ promote sharper concentration of $\|\Phi x\|_2^2$ about its mean $\|x\|_2^2$. We can bound the range of possible Γ values as follows.

Lemma III.1. *Let $\Gamma = \Gamma(x, \mathbf{M})$ be as defined in (6). Then $\min_j M_j \leq \Gamma \leq \sum_{j=1}^J M_j$.*

Proof: See Appendix B.

It is not difficult to see that the worst case, $\Gamma = \min_j M_j$, is achieved when all of the signal energy is concentrated into exactly one signal block where the fewest measurements are collected, i.e., when $\|x_j\|_2^2 = 0$ except for a single index $j' \in \{\arg\min_j M_j\}$ (where $\{\arg\min_j M_j\}$ is the set of indices where $\{M_j\}$ is minimum). In this case the DBD matrix exhibits significantly worse performance than a dense i.i.d. matrix of the same size $(\sum_{j=1}^J M_j) \times NJ$, for which the concentration exponent would scale with $\sum_{j=1}^J M_j$. This makes some intuitive sense, as this extreme case would correspond to only one block of the DBD matrix sensing all of the signal energy. The best case, $\Gamma = \sum_{j=1}^J M_j$, is achieved when the number of measurements collected for each block is proportional to the signal energy in that block. In other words, letting $\mathrm{diag}(\mathbf{M})$ represent the diagonal of $\mathbf{M}$, the concentration exponent scales with $\sum_{j=1}^J M_j$ just as it would for a dense i.i.d. matrix of the same size when $\mathrm{diag}(\mathbf{M}) \propto \gamma$ (i.e., when $\mathrm{diag}(\mathbf{M}) = C\gamma$ for some constant $C > 0$). This is in spite of the fact that the DBD matrix has many fewer nonzero elements.

The probability of concentration in the second case of (5) behaves similarly. For the ratio $\frac{\|\gamma\|_1}{\|\mathbf{M}^{-1}\gamma\|_\infty}$ appearing in the concentration exponent, one can show that $\min_j M_j \leq \frac{\|\gamma\|_1}{\|\mathbf{M}^{-1}\gamma\|_\infty} \leq \sqrt{J\sum_{j=1}^J M_j^2}$. The lower bound is again achieved when $\|x_j\|_2^2 = 0$ except for a single index $j' \in \{\arg\min_j M_j\}$, and the upper bound is achieved when the signal energy is uniformly distributed across the blocks and the measurement rates are constant, i.e., when $\|x_1\|_2^2 = \|x_2\|_2^2 = \cdots = \|x_J\|_2^2$ and

(a) (b)

Fig. 1. *Test signal for concentration in a DBD matrix. (a) Test signal x with length 1024. (b) Energy distribution $\gamma(x)$ when signal is partitioned into $J = 16$ blocks of length $N = 64$.*

$M_1 = M_2 = \cdots = M_J$.

The above discussion makes clear that the concentration performance a DBD matrix can vary widely depending on how well matched it is to the energy distribution of the measured signal. In particular, DBD matrices can perform as well as dense i.i.d. matrices if they are constructed to match the number of measurements in each block to the distribution of signal energy.

Three final comments concerning Theorem III.1 are in order. First, the bounds in (5) are most favorable for matrices populated with strictly subgaussian random variables because this allows one to set $c = 1$. As mentioned in Section II-A, zero-mean Gaussian random variables, ± 1 Bernoulli random variables with $p = \frac{1}{2}$, and uniform random variables on $[-1, 1]$ are all examples of strictly subgaussian random variables. Second, we may further characterize the range of ϵ for which the two cases of Theorem III.1 are relevant.

Lemma III.2. *If $J \geq 2$, the point of demarcation between the two cases of Theorem III.1 obeys*

$$\frac{16c \cdot 2(\sqrt{J}-1)}{J-1} \frac{\min_j \sqrt{M_j}}{\max_j \sqrt{M_j}} \leq \frac{16c\|\mathbf{M}^{-1/2}\gamma\|_2^2}{\|\mathbf{M}^{-1}\gamma\|_\infty \|\gamma\|_1} \leq 16c.$$

Proof: See Appendix C.

Examining the bound above, we note that for $J \geq 2$ it holds that $\frac{2(\sqrt{J}-1)}{J-1} \geq \frac{1}{\sqrt{J}}$. Thus, as an example, when $M_1 = M_2 = \cdots = M_J$, the first ("small ϵ") case of Theorem III.1 is guaranteed to at least permit $\epsilon \in [0, \frac{16c}{\sqrt{J}}]$. Finally, Theorem III.1 was derived by considering all signal blocks to be of equal length N. By close examination of the proof, one can see that the same theorem in fact holds for signals partitioned into blocks of unequal lengths.

2) Supporting Experiments: While the quantity Γ plays a critical role in our analytical upper bound (5) on the concentration tail probabilities, it is reasonable to ask whether this quantity actually plays a central role in the empirical concentration performance of DBD matrices. We explore this question with a series of simulations. To begin, we randomly construct a signal of length 1024 partitioned into $J = 16$ blocks of length $N = 64$. The signal x and its energy distribution γ are plotted in Figures 1(a) and 1(b), respectively. For this simulation, in order to be able to ensure $\mathrm{diag}(\mathbf{M}) \propto \gamma$ with integer values for the M_j, we began by constructing $\mathbf{M}$ (populated with integers) and then normalized each block of a randomly generated signal to set γ accordingly.

Fixing this signal x, we generate a series of 10000 random 64×1024 matrices Φ using zero-mean Gaussian random variables for the entries. In one case, the matrices are fully dense and the entries of each matrix have variance $1/64$. In another case, the

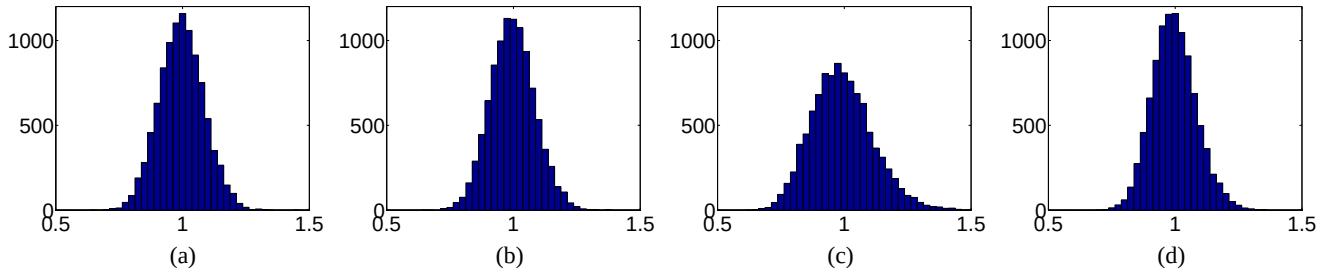


Fig. 2. *Histogram of* $\|\Phi x\|_2/\|x\|_2$ *for fixed* x *across* 10000 *randomly generated matrices* Φ. *(a) Dense* 64×1024 Φ. *(b) DBD* 64×1024 Φ *where each* Φ_j *has a number of rows such that diag*$(\mathbf{M}) \propto \gamma$. *(c) DBD* 64×1024 Φ *where each* Φ_j *has* 4 *rows so that diag*$(\mathbf{M}) \not\propto \gamma$. *This corresponds to* $\Gamma = 32.77 \leq 64$. *(d) Modified DBD* 128×1024 Φ *where each* Φ_j *has* 8 *rows. This corresponds to* $\Gamma \approx 64$.

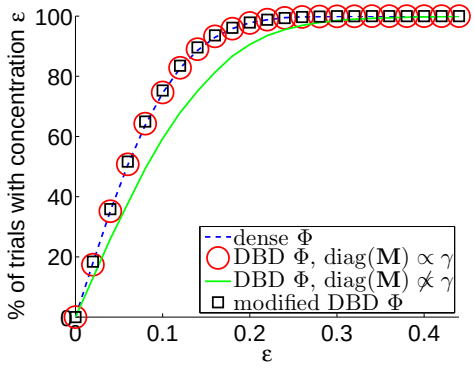


Fig. 3. *The percentage of trials for which* $(1-\epsilon) \leq \|\Phi x\|_2/\|x\|_2 \leq (1+\epsilon)$ *as a function of* ϵ. *Note that all curves overlap except for that corresponding to the DBD matrix* Φ *with diag*$(\mathbf{M}) \not\propto \gamma$.

matrices are DBD with $\text{diag}(\mathbf{M}) \propto \gamma$ and the entries in each block have variance $1/M_j$. Thus, we have $\Gamma(x, \mathbf{M}) = \sum_{j=1}^{J} M_j$ and our Theorem III.1 gives the same concentration bound for this DBD matrix as for the dense i.i.d. matrix of the same size. Indeed, Figures 2(a) and 2(b) show histograms of $\|\Phi x\|_2/\|x\|_2$ for these two types of matrices and, despite the drastically different structure of the matrices, both histograms are tightly concentrated around 1 as anticipated. For each type of matrix, Figure 3 also shows the percentage of trials for which $(1-\epsilon) \leq \|\Phi x\|_2/\|x\|_2 \leq (1+\epsilon)$ as a function of ϵ. The curves for the dense and DBD matrices are indistinguishable.

Finally, we consider a third scenario in which we construct 10000 random 64×1024 DBD matrices as above but with an equal number of measurements in each block. In other words, we set all $M_j = 4$, and obtain measurement matrices that are no longer matched to the signal energy distribution. We quantify this mismatch by noting that $\Gamma(x, 4 \cdot I_{J \times J}) = 32.77 < \sum_{j=1}^{J} M_j$. Figure 2(c) shows the histogram of $\|\Phi x\|_2/\|x\|_2$ and Figure 3 shows the concentration success probability over these 10000 random matrices. It is evident that these mismatched DBD matrices provide decidedly less sharp concentration of $\|\Phi x\|_2$.

B. Repeated Block Diagonal (RBD) Matrices

1) Analytical Results: We now turn our attention to the concentration performance of the more restricted RBD matrices. Before stating our result, let us again define the requisite notation. Given a signal $x \in \mathbb{R}^{NJ}$ partitioned into J blocks of length N, we define the $J \times N$ matrix of concatenated signal blocks

$$X := [x_1 \; x_2 \; \cdots \; x_J]^T, \tag{7}$$

and we denote the non-negative eigenvalues of the $N \times N$ symmetric matrix $A = X^T X$ as $\{\lambda_i\}_{i=1}^N$. We let

$$\lambda = \lambda(x) := [\lambda_1, \ldots, \lambda_N]^T \in \mathbb{R}^N$$

be the vector composed of these eigenvalues. Also, we let $M := M_1 = M_2 = \cdots = M_J$ denote the same number of measurements to be taken in each block, and we refer to the measurement matrices as $\widetilde{\Phi} := \Phi_1 = \Phi_2 = \cdots = \Phi_J$.

Equipped with this notation, our main result concerning the concentration of RBD matrices is as follows.

Theorem III.2. *Suppose $x \in \mathbb{R}^{NJ}$. Let $\widetilde{\Phi}$ be a random $M \times N$ matrix populated with i.i.d. zero-mean Gaussian entries having variance $\sigma^2 = \frac{1}{M}$, and let Φ be an $MJ \times NJ$ block diagonal matrix as defined in* (1)*, with $\Phi_j = \widetilde{\Phi}$ for all j. Then*

$$P(\left|\|\Phi x\|_2^2 - \|x\|_2^2\right| > \epsilon \|x\|_2^2) \leq \begin{cases} 2\exp\{-\frac{M\epsilon^2\|\lambda\|_1^2}{256\|\lambda\|_2^2}\}, & 0 \leq \epsilon \leq \frac{16\|\lambda\|_2^2}{\|\lambda\|_\infty \|\lambda\|_1} \\ 2\exp\{-\frac{M\epsilon\|\lambda\|_1}{16\|\lambda\|_\infty}\}, & \epsilon \geq \frac{16\|\lambda\|_2^2}{\|\lambda\|_\infty \|\lambda\|_1} \end{cases}. \tag{8}$$

Proof: See Appendix D.

As we can see from (8), the concentration probability of interest again decays with a rate that depends on the characteristics of the signal x being measured. Focusing on the first case listed in (8), we see that the concentration exponent scales with

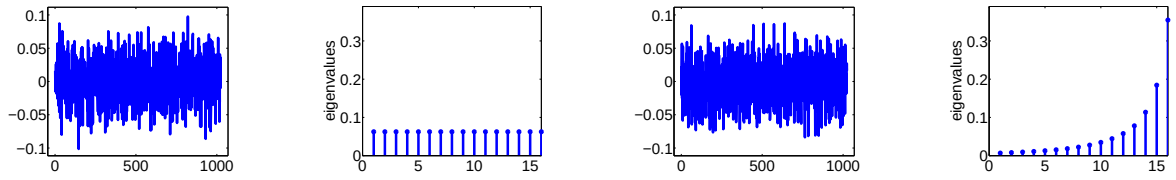
$$\Lambda = \Lambda(x, M) := \frac{M\|\lambda\|_1^2}{\|\lambda\|_2^2}. \tag{9}$$

From the standard relation between ℓ_1 and ℓ_2 norms, it follows that $M \leq \Lambda \leq M \min(J, N)$. The worst case, $\Lambda = M$, is achieved when $A = \sum_j x_j x_j^T$ has only one nonzero eigenvalue, implying that the blocks x_j are the same modulo a scaling factor. In this case, the RBD matrix exhibits significantly worse performance than a dense i.i.d. matrix of the same size $MJ \times NJ$, for which the concentration exponent would scale with MJ rather than M. However, this diminished performance is to be expected since the same $\widetilde{\Phi}$ is used to measure each identical signal block.

The other extreme case, $\Lambda = M \min(J, N)$ is favorable as long as $J \leq N$, in which case the concentration exponent scales with MJ just as it would for a dense i.i.d. matrix of the same size. For this case to occur, A must have J nonzero eigenvalues and they must all be equal. By noting that the nonzero eigenvalues of $A = X^T X$ are the same as those of the Grammian matrix $G = XX^T$, we conclude that this most favorable case can occur only when the signal blocks are mutually orthogonal and have the same energy. Alternatively, if the signal blocks span a K-dimensional subspace of $\mathbb{R}^N$ we will have $M \leq \Lambda \leq MK$. All of this can also be seen by observing that calculating the eigenvalues of $A = X^T X$ is equivalent to running Principal Component Analysis (PCA) [31] on the matrix X comprised of the J signal blocks.

We note that there is a close connection between Γ and Λ that is not apparent at first glance. For a fair comparison, we assume in this discussion that $M_1 = M_2 = \cdots = M_J = M$. Now, note that $\|\lambda\|_1^2 = \|\gamma\|_1^2$ and also that

$$\|\lambda\|_2^2 = \|A\|_F^2 = \|XX^T\|_F^2 = \sum_{i=1}^J \|x_i\|_2^4 + 2\sum_{i>j}(x_i^T x_j)^2 = \|\gamma\|_2^2 + 2\sum_{i>j}(x_i^T x_j)^2.$$



(a) (b) (c) (d)

Fig. 4. *Test signals for concentration in a RBD matrix. (a) Sig. 1 with $J = 16$ orthogonal blocks. (b) λ for Sig. 1. (c) Sig. 2 with non-orthogonal blocks. (d) λ for Sig. 2.*

Using these two relationships, we can rewrite Λ as

$$\Lambda = \frac{M\|\lambda\|_1^2}{\|\lambda\|_2^2} = \frac{M\|\gamma\|_1^2}{\|\gamma\|_2^2 + 2\sum_{i>j}(x_i^T x_j)^2} \leq \frac{M\|\gamma\|_1^2}{\|\gamma\|_2^2} = \Gamma. \tag{10}$$

From this relationship we see that Λ and Γ differ only by the additional inner-product term in the denominator of Λ, and we also see that $\Lambda = \Gamma$ if and only if the signal blocks are mutually orthogonal.

In relation to Theorem III.1, the conditions for achieving the optimal concentration exponent are more stringent in Theorem III.2. This is not surprising given the restricted nature of the RBD matrix. Indeed, it is remarkable that for some signals, an RBD matrix can yield the same concentration rate as a dense i.i.d. matrix, though as we have discussed above this happens only for signals containing enough intrinsic diversity to compensate for the lack of diversity in the measurement system.

2) Supporting Experiments: While the quantity Λ plays a critical role in our analytical upper bound (8) on the concentration tail probabilities, it is reasonable to ask whether this quantity actually plays a central role in the empirical concentration performance of RBD matrices. We explore this question with a series of simulations. To begin, we randomly construct a signal of length 1024 partitioned into $J = 16$ blocks of length $N = 64$, and we perform Gram-Schmidt orthogonalization to ensure that the J blocks are mutually orthogonal and have equal energy. The signal x (denoted "Sig. 1") is plotted in Figure 4(a), and the nonzero eigenvalues of $A = X^T X$ are shown in the plot of λ in Figure 4(b).

As we have discussed above, for signals such as Sig. 1 we should have $\Lambda = MJ$, and Theorem III.2 suggests that an RBD matrix can achieve the same concentration rate as a dense i.i.d. matrix of the same size. Fixing this signal, we generate a series of 10000 random 64×1024 matrices Φ populated with zero-mean Gaussian random variables. In one case, the matrices are dense and each entry has variance $1/64$. In another case, the matrices are RBD, with a single 4×64 block repeated along the main diagonal, comprised of i.i.d. Gaussian entries with variance $\frac{1}{4}$. Figures 5(a) and 5(b) show histograms of $\|\Phi x\|_2/\|x\|_2$ for these two types of matrices and, despite the drastically different structure of the matrices, both histograms are tightly concentrated around 1 as anticipated. For each type of matrix, Figure 6 also shows the percentage of trials for which $(1-\epsilon) \leq \|\Phi x\|_2/\|x\|_2 \leq (1+\epsilon)$ as a function of ϵ. The curves for the dense and RBD matrices are indistinguishable.

In contrast, we also construct a second signal x (denoted "Sig. 2") that has equal energy between the blocks but has non-orthogonal components (resulting in non-uniform λ); see Figures 4(c) and 4(d). This signal was constructed to ensure that the sorted entries of λ exhibit an exponential decay. Due to the non-orthogonality of the signal blocks, we see for this signal that $\Lambda = 21.3284$ which is approximately 3 times less than the best possible value of $MJ = 64$. Consequently, Theorem III.2 provides a much weaker concentration exponent when this signal is measured using an RBD matrix than when it is measured using a dense i.i.d. matrix. Fixing this signal, we plot in Figures 5(c) and 5(d) the histograms of $\frac{\|\Phi x\|_2}{\|x\|_2}$ for the fully dense and

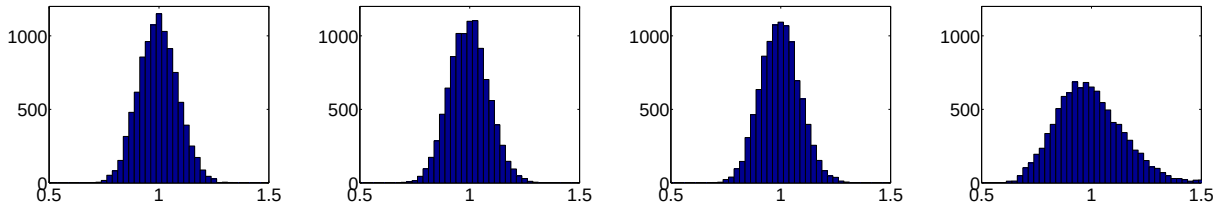


Fig. 5. *Histogram of* $\|\Phi x\|_2/\|x\|_2$ *for fixed* x *across 10000 randomly generated matrices* Φ*. (a) Sig. 1 with orthogonal blocks, dense* 64×1024 Φ*. (b) Sig. 1 with orthogonal blocks, RBD* 64×1024 Φ*. (c) Sig. 2 with non-orthogonal blocks, dense* 64×1024 Φ*. (d) Sig. 2 with non-orthogonal blocks, RBD* 64×1024 Φ*. This corresponds to* $\Lambda = 21.3284 < 64$*.*

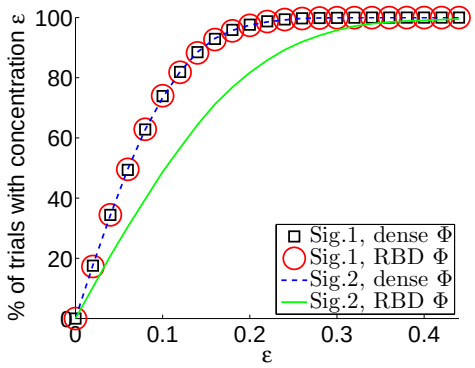


Fig. 6. *The percent of trials for which* $(1-\epsilon) \leq \|\Phi x\|_2/\|x\|_2 \leq (1+\epsilon)$ *as a function of* ϵ*. Note that all curves overlap except for that corresponding to the signal with non-orthogonal blocks being measured with RBD matrices.*

RBD matrices constructed above, respectively. As shown in these histograms and in Figure 6, we see that the concentration performance of the full dense matrix is agnostic to this new signal structure, while the concentration is clearly not as sharp for the RBD matrix.

C. Increasing Measurement Rates to Protect Against Signal Uncertainty

As we have seen both analytically and experimentally, given a fixed number of total measurements, the concentration performance of a DBD matrix can be optimized if there is a suitable match between the measurement allocations and the structure of the observed signal x. In some cases, however, it may not be possible for a system designer to match the measurement ratios to the signal structure; for example, without knowledge of the exact signal to be acquired, one may wish to design a fixed measurement system with an equal number of measurements per block. In such cases, it may still be possible to guarantee suitable concentration performance for a variety of possible received signals by increasing the *total* number of measurements collected.

For example, recall the experiment from Section III-A involving the signal shown in Figure 1. Using an unmatched DBD matrix with $M = 4$ measurements per block, we obtained $\Gamma(x, \mathbf{M}) = \Gamma(x, M \cdot I_{J\times J}) = 32.77$, which gave rise to a smaller concentration exponent and worse empirical concentration performance than a dense i.i.d. matrix with the same total number of $MJ = 64$ rows. However, suppose that we had the resources available to acquire more measurements from each block (while keeping the number of measurements equal across the blocks). In particular, if we were to collect $M^{'} = \frac{MJ}{\Gamma(x, M\cdot I_{J\times J})} \cdot M$ measurements from each block, we would obtain a DBD matrix whose concentration exponent scales with $\Gamma(x, \mathbf{M}^{'}) = \Gamma(x, M' \cdot I_{J\times J}) = \frac{MJ}{\Gamma(x, M\cdot I_{J\times J})} \cdot \Gamma(x, M \cdot I_{J\times J}) = MJ$, the same as a dense matrix with MJ rows. For this specific signal

x, we illustrate this by taking $M' = 8 \approx \frac{64}{32.77} \cdot M$; as shown in Figure 2(d) and Figure 3, the concentration performance of a 128×1024 DBD matrix with $M' = 8$ measurements per block (we call this a "modified DBD" matrix due to its larger size) is indistinguishable from that of a dense 64×1024 Gaussian matrix. Thus, we see that it is possible to guarantee a certain performance from DBD matrices by increasing M to protect against signals with small ratios $\frac{\|\gamma\|_1^2}{\|\gamma\|_2^2}$.

Note that this apparent empirical equivalence between the concentration performance of the modified DBD matrix and the dense matrix does not mean that the concentration statistics are formally the same. It is straightforward to verify that in general, while the probability distributions of $\|\Phi x\|_2^2$ will not be the same for a dense and a modified DBD matrix, the variance of $\|\Phi x\|_2^2$ will be equal in both cases. This explains why the empirical concentration performance is very similar. In the special case that the signal blocks have equal energy and the matrices are populated with Gaussian random variables, the distributions will in fact be the same under both matrices. As a final remark, we note that similar arguments can be made for RBD matrices by increasing M to protect against signals with small ratios $\frac{\|\lambda\|_1^2}{\|\lambda\|_2^2}$.

IV. Favorable Signal Classes

In Section III we demonstrated that block diagonal matrices have concentration exponents that vary significantly depending on the exact characteristics of the signal being measured. It is only natural to ask whether there are any realistic signal *classes* where we expect most of the constituent signals to have the properties that allow favorable concentration exponents from DBD or RBD matrices. In this section, we briefly explore some example scenarios where we demonstrate empirically or analytically that signals in a restricted family often have these favorable characteristics.

A. Frequency Sparse Signals

One of the primary signal characteristics affecting the concentration exponents in the results of Section III is the distribution of signal energy across blocks. Supposing that $M_1 = M_2 = \cdots = M_J =: M$, this effect is most easily seen in the quantity $1 \leq \frac{\Gamma}{M} \leq J$, where the larger the value of $\frac{\Gamma}{M}$, the better the concentration exponent when using a DBD matrix. Because of existing results on time-frequency uncertainty principles and the well-known incoherence of sinusoids and the canonical basis (i.e., "spikes and sines") [32, 33], we have intuition that most signals that are sparse in the frequency domain should have their energy distributed relatively uniformly across blocks in the time domain. In this section, we make formal the notion that frequency sparse signals are indeed likely to have a favorable energy distribution, producing values of Γ that scale to within a log factor of its maximum value.

To be concrete, let $x \in \mathbb{C}^{N'}$ be a signal of interest[4] that is split into $J = N'/N$ blocks of fixed length N each. For simplicity, we assume that N divides N' and we will consider the case where N' grows (implying the number of blocks J is increasing, since the block size N is fixed). The signal x is further assumed to be frequency sparse, with S nonzero coefficients in the discrete Fourier transform (DFT). In other words, if $\widehat{x}$ is the DFT of x and $\Omega \subset [1,\ N']$ denotes the support of $\widehat{x}$, then $|\Omega| \leq S$. We assume that the frequency locations are chosen randomly (i.e., Ω is chosen uniformly at random from $[1,\ N']$), but the

[4]We consider complex-valued signals for simplicity and clarity in the exposition. A result with the same spirit that holds with high probability can be derived for strictly real-valued signals, but this comes at the cost of a more cumbersome derivation.

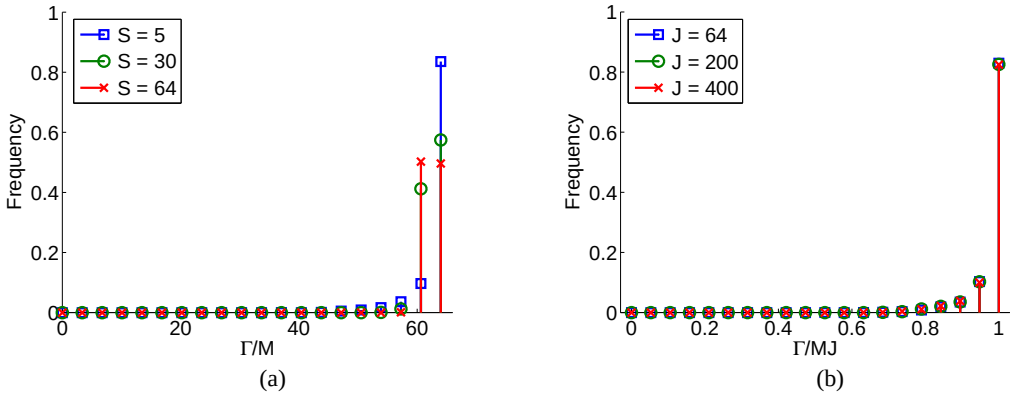


Fig. 7. *Histograms of the normalized quantity Γ for frequency sparse signals. (a) The distribution of $\frac{\Gamma}{M}$ for randomly generated frequency sparse signals of length $N' = N \times J = 64 \times 64$ for sparsity levels $S \in \{5, 30, 64\}$. Note that $\frac{\Gamma}{M}$ accumulates near its upper bound of $J = 64$ for all three sparsity levels. (b) The distribution of $\frac{\Gamma}{MJ}$ for randomly generated frequency sparse signals with $S = 5$ and the number of signal blocks $J \in \{64, 200, 400\}$. Note that $\frac{\Gamma}{MJ}$ accumulates near its upper bound of 1.*

values taken by $\widehat{x}$ on Ω are completely arbitrary with no probability distribution assumed. The following theorem then gives a lower bound on the ratio $\frac{\Gamma}{M}$ for frequency sparse signals.

Theorem IV.1. *Let $N, \beta > 1$ be fixed and suppose $N' = NJ > 512$. Let $\Omega \subset [1,\ N']$ be of size $S \leq N$ generated uniformly at random. Then with probability at least $1 - O(J(\log(N'))^{1/2}(N')^{-\beta})$,[5] every signal $x \in \mathbb{C}^{N'}$ whose DFT coefficients are supported on Ω will have Γ/M lower bounded by:*

$$\frac{\Gamma}{M} \geq \min\left\{ \frac{0.0779J}{(\beta+1)\log(N')}, \frac{J}{\left(\sqrt{6(\beta+1)\log N'} + \frac{(\log N')^2}{N}\right)^2} \right\}.$$

Proof: See Appendix E.

Note that as N' grows, the lower bound on $\frac{\Gamma}{M}$ scales as $\frac{J}{\frac{1}{N^2}\log^4(N')}$, which (treating the fixed value N as a constant) is within $\log^4(N')$ of its maximum possible value of J. Thus the concentration exponent for most frequency sparse signals when measured by a DBD matrix will scale with $\epsilon^2 MJ/\log^4(N')$ for small ϵ, which is close to the concentration exponent resulting from the application of a dense, i.i.d. random matrix of the same size. Also note that the failure probability in the above theorem can be bounded by $O(\frac{1}{N'^{\beta-2}})$ since both J and $\sqrt{\log(N')}$ are less than N'.

To explore the analysis above we use two illustrative simulations. For the first experiment, we generate 5000 signals with length $N' = NJ = 64 \times 64 = 4096$, using three different sparsity levels $S \in \{5, 30, 64\}$. The DFT coefficient locations are selected uniformly at random, and the corresponding nonzero coefficient values are drawn from the i.i.d. standard Gaussian distribution. Figure 7(a) plots the ratio $\frac{\Gamma}{M}$, showing that this quantity is indeed near the upper bound of $J = 64$, indicating a favorable energy distribution. This gives support to the fact that the theoretical value of $\frac{\Gamma}{M}$ predicted in Theorem IV.1 does not depend strongly on the exact value of S. For the second experiment, we fix the sparsity at $S = 5$ and vary the signal block length $J \in \{64, 200, 400\}$ (and thus the total signal length $N' = NJ$ changes as well). For each J we generate 5000 random signals in the same manner as before and plot in Figure 7(b) the distribution of $\frac{\Gamma}{MJ}$ (note the normalization by J).

[5]The $O(\cdot)$ notation is with respect to N'. With the component length N fixed, this means that only the number of blocks J is growing with increasing N'.

Again, it is clear that this value concentrates near the upper bound of 1, showing the relative accuracy of the prediction that $\frac{\Gamma}{M}$ scales linearly with J. While some of the quantities in Theorem IV.1 appear pessimistic (e.g., the scaling with $\log^4(N')$), these simulations confirm the intuition derived from the theorem that frequency sparse signals should indeed have favorable energy distributions, and therefore favorable concentration properties when measured with DBD matrices.

B. Delay Networks and Multi-view Imaging

Additional favorable signal classes for DBD and RBD matrices can occasionally arise in certain sensor network or multi-view imaging scenarios where signals with steeply decaying autocorrelation functions are measured under small perturbations. For example, consider a network of J sensors characterized by measurement operators $\Phi_1, \Phi_2, \ldots, \Phi_J$, and suppose that the received signals $x_1, x_2, \ldots, x_J \in \mathbb{R}^N$ represent observations of some common underlying prototype signal $z \in \mathbb{R}^N$. However, due to the configurations of the sensors, these observations occur with different delays or translations. More formally, we might consider the one-dimensional delay parameters $\delta_1, \delta_2, \ldots, \delta_J \in \mathbb{Z}$ and have that for each j, $x_j(n) = z(n - \delta_j)$. A similar two-dimensional translation model could be used for imaging scenarios.

The characteristics of z may make it well suited to observation using block diagonal matrices. Assuming z is suitably zero-padded so that border and truncation artifacts can be neglected, we will have $\|x_j\|_2 = \|z\|_2$ for all x_j; this gives $\Gamma = \sum_j M_j$, which is the ideal case for observation with a DBD matrix. Moreover, the correlations among the components x_j can be characterized in terms of the autocorrelation function R_z of z: we will have $\langle x_i, x_j \rangle = \sum_{n=1}^{N} x_i(n)x_j(n) = \sum_{n=1}^{N} z(n - \delta_i)z(n - \delta_j)$, which neglecting border and truncation artifacts will simply equal $R_z(|\delta_i - \delta_j|)$. Therefore, signals z that exhibit strong decay in their autocorrelation function will be natural candidates for observation with RBD matrices as well. For example, if we assume that all $M_j = M$, equation (10) gives

$$\Lambda = \frac{MJ^2\|z\|_2^4}{J\|z\|_2^4 + 2\sum_{i>j} R_z(|\delta_i - \delta_j|)^2}.$$

Note in the expression above that when $R_z(|\delta_i - \delta_j|)$ is small for most i and j, the quantity Λ is near its optimal value of MJ. Finally, we note that the repeated observation of a signal from multiple translations gives rise to a structure not unlike that which arises when considering Toeplitz matrices in problems such as channel sensing. In Section V-B1, we explore this related problem in much more depth.

C. Difference Signals

Some applications of our main results could require considering difference vectors of the form $x - y$ where $x, y \in \mathbb{R}^{NJ}$. For example, as suggested by the discussion in Section II-C, in order to guarantee that a block diagonal measurement matrix Φ provides a stable embedding of a signal family $Q \subset \mathbb{R}^{NJ}$, it may be necessary to repeatedly invoke bounds such as (5) for many or all of the difference vectors between elements of Q. It is interesting to determine what signal families will give rise to difference vectors that have favorable values of Γ or Λ.

We provide a partial answer to this question by briefly exemplifying such a favorable signal family. For the sake of simplicity, let us restrict our attention to DBD matrices of the form in (1) where $M_1 = M_2 = \cdots = M_J =: M$. Consider a set $Q \subset \mathbb{R}^{JN}$

of signals that, when partitioned into J blocks of length N, satisfy both of the following properties: (i) each $x \in Q$ has uniform energy across the J blocks, i.e., $\|x_1\|_2^2 = \|x_2\|_2^2 = \cdots = \|x_J\|_2^2 = \frac{1}{J}\|x\|_2^2$, and (ii) each $x \in Q$ has highly correlated blocks, i.e., for some $a \in (0, 1]$, $\langle x_i, x_j \rangle \geq a\frac{1}{J}\|x\|_2^2$ for all $i, j \in \{1, 2, \ldots, J\}$. The first of these conditions ensures that each $x \in Q$ will have $\Gamma = MJ$ and thus be highly amenable to measurement by a DBD matrix. The second condition, when taken in conjunction with the first, ensures that all difference vectors of the form $x - y$ where $x, y \in Q$ will also be highly amenable to measurement by a DBD matrix. In particular, for any $i, j \in \{1, 2, \ldots, J\}$, one can show that

$$\left| \|x_i - y_i\|_2^2 - \|x_j - y_j\|_2^2 \right| \leq \frac{4\sqrt{2}\|x\|_2\|y\|_2\sqrt{1-a}}{J},$$

meaning that the energy differences in each block of the difference signals can themselves have small differences. One implication of this bound is that as $a \to 1$, $\Gamma(x - y) \to MJ$.

Signal families of the form specified above—with uniform energy blocks and high inter-block correlations—may generally arise as the result of observing some phenomenon that varies slowly as a function of time or of sensor position. As an empirical demonstration, let us consider a small database of eight real-world video signals frequently used as benchmarks in the video compression community, where we will treat each video frame as a signal block.[6] We truncate each video to have $J = 150$ frames, each of size $N = 176 \times 144 = 25344$ pixels, and we normalize each video (not each frame) to have unit energy. Because the test videos are real-world signals, they do not have perfectly uniform energy distribution across the frames, but we do observe that most frame energies are concentrated around $\frac{1}{J} = 0.00667$.

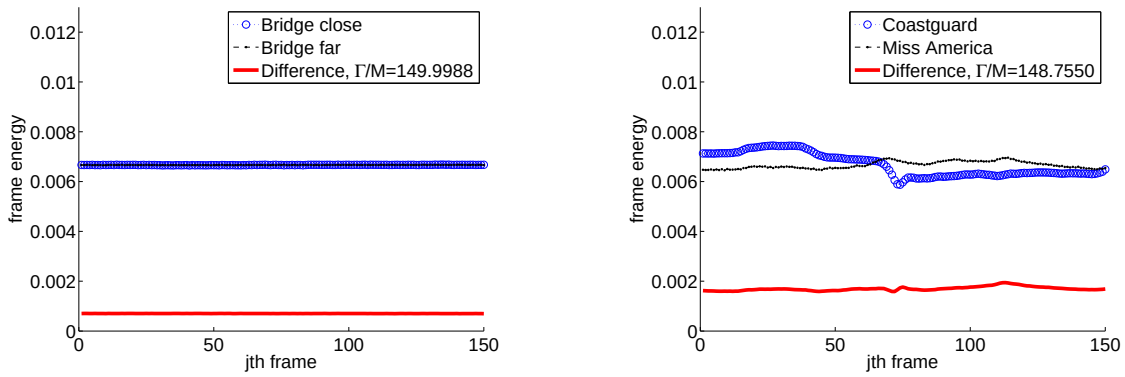
Video name	Akiyo	Bridge close	Bridge far	Carphone	Claire	Coastguard	Foreman	Miss America
$\max\langle x_i, x_j \rangle$	0.00682	0.00668	0.00668	0.00684	0.00690	0.00742	0.00690	0.00695
$\min\langle x_i, x_j \rangle$	0.00655	0.00664	0.00665	0.00598	0.00650	0.00562	0.00624	0.00606
Γ/M	149.9844	149.9998	149.9999	149.9287	149.9782	149.2561	149.9329	149.9301

TABLE I

The maximum and minimum inner products between all pairs of distinct frames in each video, and the quantity Γ/M for each video. The best possible value of Γ/M is $\Gamma/M = J = 150$.

For each video, we present in Table I the minimum and maximum inner products $\langle x_i, x_j \rangle$ over all $i \neq j$, and we also list the quantity Γ/M for each video. As we can see, the minimum inner product for each video is indeed quite close to 0.00667, suggesting from the arguments above that the pairwise differences between various videos are likely to have highly uniform energy distributions. To verify this, we compute the quantity Γ/M for all possible $\binom{8}{2}$ pairwise difference signals. As we are limited in space, we present in Figure 8 plots of the energies $\|x_j\|_2^2$, $\|y_j\|_2^2$, and $\|x_j - y_j\|_2^2$ as a function of the frame index j for the pairs of videos x, y that give the best (highest) and the worst (smallest) values of $\Gamma(x - y)/M$. We see that even the smallest Γ/M value is quite close to the best possible value of $\Gamma/M = 150$. All of this suggests that the information required to discriminate or classify various signals within a family such as a video database may be well preserved in a small number of random measurements collected by a DBD matrix.

[6]Videos were obtained from http://trace.eas.asu.edu/yuv/.



(a) (b)

Fig. 8. *Plots of the energy distributions of individual videos and of their differences for the best video pair and the worst video pair among all possible* $\binom{8}{2}$ *possible video pairs. (a) The difference of the video pair, "Bridge close" and "Bridge far", giving the best value of* $\Gamma(x-y)/M = 149.9988$. *(b) The difference of the video pair, "Coastguard" and "Miss America", giving the worst value of* $\Gamma(x-y)/M = 148.7550$.

D. Random Signals

Our discussions above have demonstrated that favorable Λ and Γ values can arise for signals in a variety of practical scenarios. This is no accident; indeed, as a blanket statement, it is true that a large majority of all signals $x \in \mathbb{R}^{JN}$, when partitioned into a sufficiently small number of blocks J and measured uniformly, will have favorable values of both Λ and Γ. One way of formalizing this fact is with a probabilistic treatment such as that given in the following lemma.

Lemma IV.1. *Suppose* $x \in \mathbb{R}^{NJ}$ *is populated with i.i.d. subgaussian entries having mean zero and variance* σ^2, *and let* $\tau = \sqrt{c} \cdot \sigma$ *denote the Gaussian standard of this distribution for a constant* $c \geq 1$. *Let* Γ *and* Λ *be defined as in (6) and (9) respectively, where we assume that* $M_1 = M_2 = \cdots = M_J =: M$. *Then, supposing that* $J \leq N\left(\frac{\epsilon^2}{4 \cdot 256 c^2} - c_1\right) \log^{-1}\left(\frac{12}{\epsilon}\right)$ *for some constant* $c_1 > 0$,

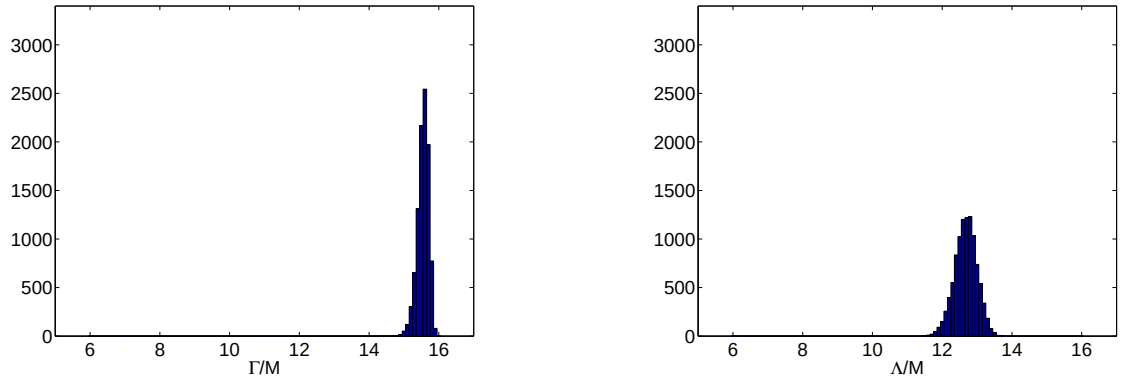
$$\Gamma \geq \Lambda \geq M + M\left(\frac{1-\epsilon}{1+\epsilon}\right)^2 (J-1),$$

with probability at least $1 - 2e^{-c_1 N}$.

Proof: See Appendix F.

We see from Lemma IV.1 that when random vectors are partitioned into a sufficiently small number of blocks, these signals will have Λ and Γ values close to their optimal value of MJ. One possible use of this lemma could be in studying the robustness of block diagonal measurement systems to noise in the signal. The lemma above tells us that when the restrictions are met on the number of blocks, random noise will tend to yield blocks that are nearly orthogonal and have highly uniform energies, thereby guaranteeing that they will not have their energy amplified by the matrix.

To illustrate this phenomenon with an example, we set $J = 16$ and $N = 64$ and construct signals $x \in \mathbb{R}^{JN}$ populated with i.i.d. Gaussian entries. Over 10000 random draws of the signal, Figures 9(a) and 9(b) plot histograms of the resulting quantities Γ/M and Λ/M, respectively. The average value of Γ/M is 15.5, while the average value of Λ/M is 12.6; both represent a large fraction of the maximum possible value of 16. In addition, we see that the two histograms concentrate sharply around their means, suggesting that a large majority of all signals will indeed be favorable for measurement with DBD and RBD matrices.



(a) (b)

Fig. 9. *The histograms of (a) Γ/M and (b) Λ/M for 10000 randomly drawn signals with $J = 16$ and $N = 64$. In this setting, the maximum possible value of both Γ/M and Λ/M is 16 and we see that both histograms concentrate relatively close to this maximum value.*

V. Applications

A. Detection in the Compressed Domain

In some settings, the ability of a measurement system to preserve the norm of an input signal is enough to make performance guarantees when certain signal processing tasks are performed directly on the measurements. For example, while the canonical CS results mostly center around reconstructing signals from compressive measurements, there is a growing interest in forgoing this recovery process and performing the desired signal processing directly in the compressed domain. One specific example of this involves a detection task where one is interested in detecting the presence of a signal from corrupted compressive linear measurements. It has been shown previously [22, 28, 34, 35] that the performance of such a compressive-domain detector depends on the norm of the signal being preserved in the compressed domain. Here we make explicit the role that concentration of measure results play in this task by guaranteeing consistency in the detector performance, and we demonstrate that we can determine favorable signal classes for this task when the compressive matrices are block diagonal.

Specifically, consider a binary hypothesis test where we attempt to detect the presence of a known signal by comparing the two hypotheses:

$$\begin{aligned} \mathcal{H}_0 &: \quad y = z \\ \mathcal{H}_1 &: \quad y = \Phi x + z \end{aligned}$$

where Φ is a compressive measurement matrix and z is a vector of i.i.d. zero-mean Gaussian noise with variance σ^2. We construct a Neyman-Pearson (NP) style detector which maximizes the probability of detection, $P_D = P\{\mathcal{H}_1 \text{ chosen } | \mathcal{H}_1 \text{ is true}\}$, given that we can tolerate a certain probability of false alarm, $P_F = P\{\mathcal{H}_1 \text{ chosen } | \mathcal{H}_0 \text{ is true}\}$. Starting from the usual likelihood ratio test, the decision for this problem is made based on whether or not the sufficient statistic $t := y^T \Phi x$ surpasses a threshold κ, i.e.,

$$t \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \kappa,$$

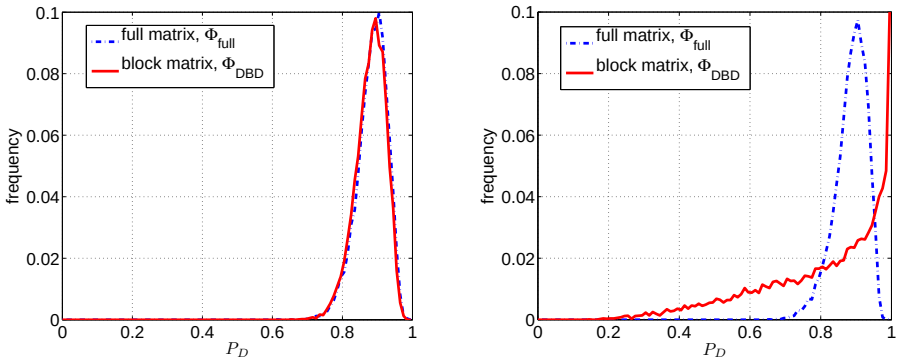


Fig. 10. Histogram of P_D for 10000 random signals with a compressive Neyman-Pearson detector with constraint $P_F \leq \alpha = 0.1$. (a) Signals with uniform energy across blocks (signal class $\mathcal{S}_1$) with both block measurement matrix Φ_{DBD} and full measurement matrix Φ_{full}. (b) Signals with decaying energy across blocks (signal class $\mathcal{S}_2$) with the same matrices Φ_{DBD} and Φ_{full}.

where κ is chosen to meet the constraint $P_F \leq \alpha$ for a specified α. It can be shown that for such a detector,

$$P_D(\alpha) = Q\left(Q^{-1}(\alpha) - \frac{\|\Phi x\|_2}{\sigma}\right), \tag{11}$$

with $Q(\alpha) = \int_\alpha^\infty e^{-\frac{u^2}{2}} du$. Notice from (11) that the performance of the detector depends on the norm of the signal x in the compressed domain, $\|\Phi x\|_2$. In effect, if Φ "loses" signal energy for some signals the detector performance will suffer, and if Φ "amplifies" signal energy for some signals the detector performance will improve.

While achieving the best possible P_D for a given P_F is of course desirable, another very important consideration for a system designer is the reliability and consistency of that system. Large fluctuations in performance make it difficult to ascribe meaning to a particular detection result and use it to take action based on a known probability of being incorrect. Therefore, we see that the concentration of measure of a matrix Φ is an important factor in the consistency of this detection system by guaranteeing that $\|\Phi x\|_2^2 \approx \|x\|_2^2$ with high probability. Of course, for block diagonal matrices these concentration results depend on the statistics of the signal class. Therefore, one important role of our results in this paper is to make explicit which signal classes can lead to reliable detectors for the detection problem above when Φ is constrained to be block diagonal.

As an example, we consider both a DBD measurement matrix $\Phi_{\text{DBD}} \in \mathbb{R}^{MJ \times NJ}$ having an equal number of measurements $M_j = M$ per block and a dense measurement matrix $\Phi_{\text{full}} \in \mathbb{R}^{MJ \times NJ}$. In the experiment below, $M = 4$, $J = 16$ and $N = 64$. For both matrices the nonzero entries are drawn as zero-mean i.i.d. Gaussian random variables; for Φ_{full} we use variance $\frac{1}{MJ}$ and for Φ_{DBD} we use variance $\frac{1}{M}$. After generating one instance of each matrix randomly, we then fix it. We test the detection performance of these two measurement matrices by drawing 10000 unit-norm test signals randomly from two classes: $\mathcal{S}_1$ having uniform energy across blocks, and $\mathcal{S}_2$ having decaying energy across blocks. The noise variance σ^2 is chosen such that each test signal x has a constant signal-to-noise ratio of $SNR = 10\log_{10}\left(\frac{\|x\|_2^2}{\sigma^2}\right) = 8dB$. Using $\alpha = 0.1$, we use (11) to calculate the expected P_D for each random signal and both measurement matrices.

Figure 10 shows the histogram of P_D for the signal/measurement combinations. We see that for uniform energy signals $\mathcal{S}_1$, using both Φ_{DBD} and Φ_{full} results in detector performance tightly clustered around $P_D = 0.9$. Thus for this class of signals, block diagonal and dense measurement matrices have the same consistency in their performance. However, when using a block

measurement matrix Φ_{DBD}, the P_D for signal class $\mathcal{S}_2$ is very spread out compared to using a full matrix Φ_{full} despite having nearly the same average performance. Although some individual signals might enjoy above average performance because the measurement matrix happens to amplify the energy of those particular signals, other signals will have very poor performance because the measurement matrix severely attenuates their energy, and moreover, the composition of these "above average" and "below average" signal sets will vary depending on the random draw of the measurement matrix. Thus, this experiment illustrates how the signal statistics can affect performance reliability in compressive detection tasks when the measurement matrices have block diagonal structure.

B. The Restricted Isometry Property in Linear Systems Applications

Interestingly, the concentration results of this paper can also be used as an analytic tool to prove the RIP for certain structured matrices that arise in linear systems applications.

1) RIP for Toeplitz Matrices: Our primary example of such an application involves the sub-sampled compressive and non-compressive Toeplitz measurement matrices that arise in problems such as channel sensing [14–22]. The convolution operation defining linear time-invariant systems is equivalent to multiplication by a Toeplitz matrix, and Toeplitz matrices in turn are closely related to the RBD matrices that are the subject of this paper. Consider the channel sensing problem, where a channel has an unknown impulse response $a \in \mathbb{R}^N$ of length N and we want to estimate this channel by probing the system with a known length-P signal $\phi \in \mathbb{R}^P$ and examining the system output $\widetilde{y}$:

$$\widetilde{y} = \phi * a = \widehat{\Phi}_{\mathrm{large}} a = \begin{bmatrix} \phi_1 & 0 & 0 \\ \vdots & \ddots & 0 \\ \phi_N & \cdots & \phi_1 \\ \vdots & & \vdots \\ \phi_P & \cdots & \phi_{(P-N+1)} \\ 0 & \ddots & \vdots \\ 0 & 0 & \phi_P \end{bmatrix} a, \tag{12}$$

where $\widehat{\Phi}_{\mathrm{large}}$ is an $(N+P-1) \times N$ Toeplitz matrix whose columns consist of shifts of ϕ.

In some applications, the channel can be assumed to be sparse (e.g., when there are only a few multipath reflections), and resource constraints lead to the desire to estimate a using as few measurements (i.e., values of $\widetilde{y}$) as possible. In these cases, one possible strategy is to use a random probe signal ϕ and only measure a subset of size $J \leq N+P-1$ of the coefficients in $\widetilde{y}$. In this case, the resulting measurement vector $y \in \mathbb{R}^J$ can be written as $y = \widehat{\Phi}_{\mathrm{small}} a$, where $\widehat{\Phi}_{\mathrm{small}}$ is a $J \times N$ matrix whose rows are a subset of those from $\widehat{\Phi}_{\mathrm{large}}$. In particular, we are interested in the case when $J \leq N$, making the matrix $\widehat{\Phi}_{\mathrm{small}}$ compressive.

While the Toeplitz structure in the convolution matrix shown in (12) does not immediately appear to fall into the block diagonal structure detailed in our concentration results, careful examination reveals that it can be written in such a format. Consider that every output value of $\widetilde{y}$ is the result of multiplying the same probe vector ϕ by a version of the (time-reversed)

impulse response a that has been shifted by an amount depending on on the measurement index. The intuition here is that this computation can be written as an RBD matrix with each block equal to the probe ϕ^T (i.e., $M = 1$), multiplied by a signal x where each block x_k is a time-reversed, shifted, and windowed version of the impulse response a. To make things concrete, suppose $J \leq N$ and let $i_1, i_2, \cdots, i_J$ denote the indices of the measured coefficients of $\widetilde{y}$. For simplicity, we assume that the probe length $P \geq N + J - 1$ and that $i_1, i_2, \cdots, i_J \in [N, P]$, but note that we will not assume that these indices are contiguous.[7] Using the notation that $\vec{\mathbf{0}}_L$ is a row vector of L zeros, we precisely define one block x_k corresponding to the i_k measurement index by appropriately shifting and zero padding the time-reversed impulse response,

$$x_k = \left[\vec{\mathbf{0}}_{(i_k - N)} \;\; a_N \;\; \ldots \;\; a_1 \;\; \vec{\mathbf{0}}_{(P - i_k)}\right]^T. \tag{13}$$

With this definition, we now see that y can be written as the multiplication of a $J \times PJ$ block diagonal matrix Φ times a length-PJ signal x with J blocks:

$$y = \widehat{\Phi}_{\text{small}} a = \Phi x = \begin{bmatrix} \phi^T & & \\ & \ddots & \\ & & \phi^T \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_J \end{bmatrix}. \tag{14}$$

In the upcoming lemma, we will use the earlier results concerning the concentration of measure for RBD matrices to develop concentration results relating $\|y\|_2^2$ to $\|a\|_2^2$. To simplify the statement of this lemma, it will be beneficial to use operator matrices to define the shifting and windowing operations creating the signal blocks. Specifically, let e_i denote the i^{th} canonical basis vector of $\mathbb{R}^{N+P-1}$ and define the $(N + P - 1) \times J$ matrix $R = [e_{i_1} \;\; e_{i_2} \;\; \cdots \;\; e_{i_J}]$ that removes measurements from the convolution operation to isolate just the selected measurements such that $y = R^T \widetilde{y}$. Furthermore, define the windowing matrix $W = \begin{bmatrix} e_N & \ldots & e_{(N+P-1)} \end{bmatrix}^T$ to be the $P \times (N + P - 1)$ matrix that keeps only the last P coefficients of a length-$(N + P - 1)$ vector. Finally, we note that we can now write the matrix of concatenated signal blocks $X \in \mathbb{R}^{J \times P}$ defined in (7) as $X^T = WAR$, where A is the $(N + P - 1) \times (N + P - 1)$ circulant matrix given by

$$A = \begin{bmatrix} a_N & 0 & \ldots & 0 & a_1 & \ldots & a_{N-1} \\ \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 & & \ddots & a_1 \\ a_1 & & & a_N & \ddots & & 0 \\ 0 & \ddots & & \vdots & a_N & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & 0 \\ 0 & \ldots & 0 & a_1 & a_2 & \ldots & a_N \end{bmatrix}.$$

The following lemma establishes the concentration of measure results for the subsampled output of the convolution operation, and follows directly from applying Theorem III.2 to the current problem formulation.

[7] The assumptions on the probe length and index locations ensure that each measurement in y depends on all entries of a. We make these assumptions merely to simplify the subsequent computation of the value around which $\|\widehat{\Phi}_{\text{small}} a\|_2^2$ concentrates; removing them would change only the point of concentration.

Lemma V.1. *Let $a \in \mathbb{R}^N$ be an arbitrary vector, $\phi \in \mathbb{R}^P$ be a random vector with i.i.d. Gaussian entries having variance $\sigma^2 = \frac{1}{J}$ with $J \leq N$, and $\widetilde{y}$ be the convolution of a and ϕ as defined in* (12)*. Suppose that $P \geq J + N - 1$ and let $i_1, i_2, \cdots, i_J \in [N, P]$ denote the indices of the available measurements from this convolution, such that $y_k = \widetilde{y}_{i_k}$. Also, define X as a concatenation of signal blocks as in* (7)*, with the individual signal blocks x_k (depending on the measurement indices i_k) defined as in* (13)*. Then $\mathbf{E}\|y\|_2^2 = \|a\|_2^2$, and*

$$P\left\{\left|\|y\|_2^2 - \|a\|_2^2\right| > \epsilon\|a\|_2^2\right\} \leq \begin{cases} 2\exp\{-\frac{\epsilon^2 J}{256(\|\lambda\|_\infty/\|a\|_2^2)}\}, & 0 \leq \epsilon \leq \frac{16\|\lambda\|_2^2}{\|\lambda\|_\infty\|\lambda\|_1} \\ 2\exp\{-\frac{\epsilon J}{16(\|\lambda\|_\infty/\|a\|_2^2)}\}, & \epsilon \geq \frac{16\|\lambda\|_2^2}{\|\lambda\|_\infty\|\lambda\|_1} \end{cases}, \tag{15}$$

where $\{\lambda_i\}$ are the eigenvalues of XX^T and $\lambda = [\lambda_1\ \lambda_2\ \cdots\ \lambda_J]^T$.

Proof: See Appendix G.

We first note that (15) relates the probability of concentration for a vector $a \in \mathbb{R}^N$ to the quantity $\|\lambda\|_\infty$ (which is defined in terms of a). If the vector a is sparse, it is possible to derive a useful upper bound for $\|\lambda\|_\infty$. In particular, suppose that a has no more than S nonzero components. Then letting $\|D\|$ denote the standard operator norm of a matrix D (i.e., the largest singular value of D), we have

$$\|\lambda\|_\infty = \|XX^T\| = \|X\|^2 \leq \|R\|^2\|A\|^2\|W\|^2 = \|A\|^2 = \|A^T\|^2,$$

since the largest singular values of both W and R are 1. Because A^T is a circulant matrix, its eigenvalues are equal to the un-normalized discrete Fourier transform (DFT) of its first row $\widetilde{a} = [a_N \ldots a_1\ \ 0 \ldots 0]$. Denoting the un-normalized DFT matrix by $F \in \mathbb{R}^{(N+P-1)\times(N+P-1)}$, we see that $\|A^T\|^2 = \|F\widetilde{a}\|_\infty^2$. It is shown in [22] that for S-sparse vectors a, we have $\|F\widetilde{a}\|_\infty^2 \leq S\|a\|_2^2$. Thus, restricting our consideration to small values of ϵ (so that we are in the first case of (15)), a concentration rate that holds for any S-sparse vector a is:

$$P\left\{\left|\|y\|_2^2 - \|a\|_2^2\right| > \epsilon\|a\|_2^2\right\} \leq 2\exp\left(-\frac{\epsilon^2 J}{256S}\right).$$

Lemma V.1 basically states that arbitrary vectors a can have favorable concentration properties when multiplied by compressive Toeplitz matrices. The analysis and bound above establishes that when a is a sparse vector, the concentration exponent can be stated simply in terms of the sparsity S and number of blocks J, making the concentration bounds suitable as an analysis tool for establishing results in the CS literature. In particular, one useful application of these results is to prove the RIP for compressive Toeplitz matrices. Using standard covering arguments and following the same steps as in [10], we arrive at the following theorem establishing RIP for the compressive Toeplitz matrices relevant for the channel sensing problem.

Theorem V.1 (RIP)**.** *Suppose $\widehat{\Phi}_{\text{small}} \in \mathbb{R}^{J\times N}$ is a compressive Toeplitz matrix as defined in* (14) *(with either contiguous or non-contiguous measurement indices). Then there exist constants c_1, c_2 such that if $J \geq c_1 S^2 \log(N/S)$, $\widehat{\Phi}_{\text{small}}$ will satisfy the RIP of order S with probability at least $1 - 2\exp(-c_2 J)$.*

The theorem above establishes the RIP for compressive Toeplitz matrices with a number of measurements J proportional to $S^2 \log N$. We make a special note here that there are no restrictions in the above arguments on the measurement locations

being contiguous, so the above theorem holds equally well for non-contiguous measurement locations. This result is thus equivalent to the state-of-the-art RIP results for contiguous compressive Toeplitz matrices [17, 18, 20, 22] and non-contiguous compressive Toeplitz matrices [21] in the literature. Among these, we note that [22] also establishes the RIP by first deriving a concentration bound, but via different arguments. However, in addition to the very different technical tools we have employed, the novelty in this approach is that it provides a unified framework for proving the RIP for compressive Toeplitz matrices with both contiguous and non-contiguous measurements.

Using the same general approach, we can make a similar statement about the RIP of non-compressive Toeplitz matrices. In such cases, we are not concerned with the number of measurements J but rather the length of the probe signal P necessary to ensure the RIP with high probability. With the methods presented here, one can guarantee the RIP with high probability by taking P proportional to $S^2 \log N$. This result is comparable to the results in [18, 22] but less favorable than the state-of-the-art result of $P \sim S \log^5 N$ implied by [19].

2) RIP for Observability Matrices: As evidence of one other potential application, a recent paper in the control systems literature [36] reveals that our results for DBD and RBD matrices also have immediate extensions to the observability matrices that arise in the analysis of linear dynamical systems. Suppose that x_j denotes the state of a system at time j, a matrix A describes the evolution of the system such that $x_j = Ax_{j-1}$, and a matrix C describes the observation of the system such that $y_j = Cx_j$. A standard problem in observability theory is to ask whether some initial state x_1 can be identified based on a collection of successive observations $[y_1^T \quad y_2^T \quad \cdots \quad y_J^T]^T$; the answer to this question is affirmative if the corresponding *observability matrix*

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{J-1} \end{bmatrix}$$

is full rank. A recent paper [36], however, details how $\mathcal{O}$ in fact has a straightforward relationship to a block diagonal matrix; how in certain settings (such as when A is unitary and C is random) $\mathcal{O}$ can satisfy the RIP; and how this can enable the recovery of sparse initial states x_1 from far fewer measurements than conventional rank-based observability theory suggests. We refer the reader to [36] for a complete discussion.

VI. Conclusion

Our main technical contributions in this paper consist of novel concentration of measure bounds for random block diagonal matrices. Our results cover two cases: DBD matrices, in which each block along the main diagonal is distinct and populated with i.i.d. subgaussian random variables, and RBD matrices, in which one block of i.i.d. Gaussian random variables is repeated along the diagonal. For each type of matrix, the likelihood of norm preservation depends on certain characteristics of the signal being measured, with the relevant characteristics essentially relating to how much intrinsic diversity the signal has to compensate for the constrained structure of the measurement matrix. For DBD matrices, suitable diversity arises when the signal energy is distributed across its component blocks proportionally to the number of measurements allotted to each

block. For RBD matrices, suitable diversity requires not only that the signal energy be evenly distributed across its component blocks, but also that these blocks be mutually orthogonal. Remarkably, our main theorems show that for signals exhibiting these desirable properties, block diagonal matrices can have concentration exponents that scale at the same rate as a fully dense matrix, despite having a fraction of the complexity (i.e., block diagonal matrices require generating many fewer random numbers). Our simulation results confirm that the main diversity measures arising in our theorems (Γ and Λ) do appear to capture the most prominent effects governing the ability of a block diagonal matrix to preserve the norm of a signal.

In addition to the main theoretical results above, this paper addresses the important question of what signal classes can have favorable diversity characteristics and thus be highly amenable to measurement with block diagonal matrices. As it turns out, many signal classes can have such favorable characteristics, including frequency sparse signals, signals with multiple related measurements (e.g., as arise in delay networks or multi-view imaging), difference vectors between highly correlated signals (e.g., between successive video frames), and random signals. Moreover, the compressive detection and compressive channel sensing problems we discuss in depth illustrate the potential for finding applications of our main results in a variety of interesting applications.

The sum total of these results and investigations lead us to conclude that block diagonal matrices can often have sufficiently favorable concentration properties. Thus, subject to the caveat that one must be mindful of the diversity characteristics of the signals to be measured, block diagonal matrices can be exploited in a number of scenarios as a sensing architecture or as an analytical tool. These results are particularly relevant because of the broad scope that block diagonal matrices can have in describing constrained systems, including communication constraints in distributed systems and convolutional systems such as wireless channels.

There are many natural questions that arise from these results and are suitable topics for future research. For example, it would be natural to consider whether the concentration results for Gaussian RBD matrices could be extended to more general subgaussian RBD matrices (to match the distribution used in our DBD analysis). Also, as more applications are identified in the future, it will be important to examine the diversity characteristics of a broader variety of signal classes to determine their favorability for measurement via block diagonal matrices. Finally, one could derive embedding results akin to (4) when certain signal families Q are measured with block diagonal matrices. The primary challenge in extending the covering arguments from Section II-C would be in accounting for the fact the difference vectors $u - v$ could potentially have highly variable concentration exponents.

APPENDIX A
PROOF OF THEOREM III.1

Proof: Let $y = \Phi x$. For each matrix Φ_j, we let $[\Phi_j]_{i,n}$ denote the n^{th} entry of the i^{th} row of Φ_j. Further, we let $y_j(i)$ denote the i^{th} component of measurement vector y_j, and we let $x_j(n)$ denote the n^{th} entry of signal block x_j.

We begin by characterizing the point of concentration. One can write $y_j(i) = \sum_{n=1}^{N} [\Phi_j]_{i,n} x_j(n)$, and so it follows that $\mathbf{E}y_j^2(i) = \mathbf{E}\left(\sum_{n=1}^{N} [\Phi_j]_{i,n} x_j(n)\right)^2$. Since the $[\Phi_j]_{i,n}$ are zero-mean and independent, all cross product terms are equal to zero, and therefore we can write $\mathbf{E}y_j^2(i) = \mathbf{E}\sum_{n=1}^{N} [\Phi_j]_{i,n}^2 x_j^2(n) = \sigma_j^2 \|x_j\|_2^2 = \frac{1}{M_j}\|x_j\|_2^2$. Combining all of the measurements,

we then have $\mathbf{E}\|y\|_2^2 = \sum_{j=1}^{J}\sum_{i=1}^{M_j}\mathbf{E}y_j^2(i) = \sum_{j=1}^{J}\sum_{i=1}^{M_j}\frac{\|x_j\|_2^2}{M_j} = \sum_{j=1}^{J}\|x_j\|_2^2 = \|x\|_2^2$.

Now, we are interested in the probability that $\big|\|y\|_2^2 - \|x\|_2^2\big| > \epsilon\|x\|_2^2$. Since $\mathbf{E}\|y\|_2^2 = \|x\|_2^2$, this is equivalent to the condition that $\big|\|y\|_2^2 - \mathbf{E}\|y\|_2^2\big| > \epsilon\mathbf{E}\|y\|_2^2$. For a given $j \in \{1,2,\ldots,J\}$ and $i \in \{1,2,\ldots,M_j\}$, all $\{[\Phi_j]_{i,n}\}_{n=1}^{N}$ are i.i.d. subgaussian random variables with Gaussian standards equal to $\tau(\phi_j)$. From above, we know that $y_j(i)$ can be expressed as a linear combination of these random variables, with weights given by the entries of x_j. Thus, from Lemma II.1 we conclude that each $y_j(i)$ is a subgaussian random variable with Gaussian standard $\tau(y_j(i)) \leq \tau(\phi_j)\|x_j\|_2$. Lemma II.2 then implies that

$$P(y_j^2(i) > t) \leq 2\exp\left(-\frac{t}{2\tau^2(y_j(i))}\right) \leq 2\exp\left(-\frac{t}{2\tau^2(\phi_j)\|x_j\|_2^2}\right) \leq 2\exp\left(-\frac{M_j t}{2c\|x_j\|_2^2}\right) \tag{16}$$

for any $t \geq 0$. We recall that $\|y\|_2^2 = \sum_{j=1}^{J}\sum_{i=1}^{M_j} y_j(i)^2$ and note that all entries of this summation will be independent. Therefore, to obtain a concentration bound for $\|y\|_2^2$, we can invoke Theorem II.1 for the random variables $y_j^2(i)$ across all j,i. From (16), we see that the assumptions of the theorem are satisfied with $a = 2$ and $\alpha_{j,i}^{-1} = \frac{2c}{M_j}\|x_j\|_2^2$. Note that $\alpha_{j,i}^{-1}$ is constant for a fixed j. Hence, for $d \geq \max_{j,i}\alpha_{j,i}^{-1} = 2c\max_j \frac{1}{M_j}\|x_j\|_2^2$ and $b \geq a\sum_{j=1}^{J}\sum_{i=1}^{M_j}\alpha_{j,i}^{-2} = 8c^2\sum_{j=1}^{J}\frac{1}{M_j}\|x_j\|_2^4$,

$$P(\big|\|y\|_2^2 - \|x\|_2^2\big| > \epsilon\|x\|_2^2) \leq \begin{cases} 2\exp\{-\frac{\epsilon^2\|x\|_2^4}{32b}\}, & 0 \leq \epsilon \leq \frac{4b}{d\|x\|_2^2} \\ 2\exp\{-\frac{\epsilon\|x\|_2^2}{8d}\}, & \epsilon \geq \frac{4b}{d\|x\|_2^2} \end{cases}. \tag{17}$$

Note that $\|x\|_2^2 = \|\gamma\|_1$ and $\|x\|_2^4 = \|\gamma\|_1^2$. Substituting $d = 2c\max_j \frac{1}{M_j}\|x_j\|_2^2 = 2c\|\mathbf{M}^{-1}\gamma\|_\infty$ and $b = 8c^2\sum_{j=1}^{J}\frac{1}{M_j}\|x_j\|_2^4 = 8c^2\|\mathbf{M}^{-1/2}\gamma\|_2^2$ into (17) completes the proof. ■

APPENDIX B
PROOF OF LEMMA III.1

Proof: To prove the lower bound, observe that since $\|\gamma\|_1^2 \geq \|\gamma\|_2^2$, we can write

$$\Gamma = \frac{\|\gamma\|_1^2}{\|\mathbf{M}^{-1/2}\gamma\|_2^2} \geq \frac{\|\gamma\|_2^2}{\|\mathbf{M}^{-1/2}\gamma\|_2^2} \geq \frac{1}{\|\mathbf{M}^{-1/2}\|_2^2} = \frac{1}{\max_j \frac{1}{M_j}} = \min_j M_j.$$

To prove the upper bound let us define $\gamma' := \mathbf{M}^{-1/2}\gamma$ and $m = \mathrm{diag}(\mathbf{M}^{1/2})$, and observe that

$$\Gamma = \frac{\|\gamma\|_1^2}{\|\mathbf{M}^{-1/2}\gamma\|_2^2} = \frac{\|\mathbf{M}^{1/2}\gamma^{'}\|_1^2}{\|\gamma^{'}\|_2^2} = \frac{\langle m,\gamma^{'}\rangle^2}{\|\gamma^{'}\|_2^2} \leq \frac{\|m\|_2^2\cdot\|\gamma^{'}\|_2^2}{\|\gamma^{'}\|_2^2} = \sum_{j=1}^{J} M_j.$$

■

APPENDIX C
PROOF OF LEMMA III.2

Proof: We prove the lower bound first. For simplicity and without loss of generality, we assume that $\|\mathbf{M}^{-1/2}\gamma\|_\infty = \max_j \frac{\gamma_j}{\sqrt{M_j}} = \frac{\gamma_1}{\sqrt{M_1}}$. Ignoring the constant factor of $16c$, we can then rewrite the numerator as $\|\mathbf{M}^{-1/2}\gamma\|_2^2 = \frac{\gamma_1^2}{M_1} + \sum_{j=2}^{J}\frac{\gamma_j^2}{M_j} = \|\mathbf{M}^{-1/2}\gamma\|_\infty^2 + \|\widehat{\mathbf{M}}^{-1/2}\widehat{\gamma}\|_2^2$, where $\widehat{\gamma} = [\gamma_2,\ldots,\gamma_J]^T \in \mathbb{R}^{J-1}$ and $\widehat{\mathbf{M}}$ is a $J-1\times J-1$ diagonal matrix containing the values $M_2, M_3, \ldots, M_J$ along the main diagonal. For a lower bound on the numerator, we then have $\|\mathbf{M}^{-1/2}\gamma\|_2^2 \geq$

$\|\mathbf{M}^{-1/2}\gamma\|_\infty^2 + \frac{1}{J-1}\|\widehat{\mathbf{M}}^{-1/2}\widehat{\gamma}\|_1^2$, and so

$$\frac{\|\mathbf{M}^{-1/2}\gamma\|_2^2}{\|\mathbf{M}^{-1}\gamma\|_\infty\|\gamma\|_1} \geq \frac{\|\mathbf{M}^{-1/2}\gamma\|_\infty^2 + \frac{1}{J-1}\|\widehat{\mathbf{M}}^{-1/2}\widehat{\gamma}\|_1^2}{\|\mathbf{M}^{-1}\gamma\|_\infty\|\gamma\|_1} = \frac{\|\mathbf{M}^{-1/2}\gamma\|_\infty^2}{\|\mathbf{M}^{-1}\gamma\|_\infty\|\gamma\|_1} + \frac{1}{J-1}\frac{\|\widehat{\mathbf{M}}^{-1/2}\widehat{\gamma}\|_1^2}{\|\mathbf{M}^{-1}\gamma\|_\infty\|\gamma\|_1}. \tag{18}$$

Focusing on the second term in (18),

$$\begin{aligned}\frac{1}{J-1}\frac{\|\widehat{\mathbf{M}}^{-1/2}\widehat{\gamma}\|_1^2}{\|\mathbf{M}^{-1}\gamma\|_\infty\|\gamma\|_1} &= \frac{1}{J-1}\frac{\left(\|\mathbf{M}^{-1/2}\gamma\|_1 - \|\mathbf{M}^{-1/2}\gamma\|_\infty\right)^2}{\|\mathbf{M}^{-1}\gamma\|_\infty\|\gamma\|_1}, \\ &= \frac{1}{J-1}\frac{\left(\|\mathbf{M}^{-1/2}\gamma\|_1^2 - 2\|\mathbf{M}^{-1/2}\gamma\|_1\|\mathbf{M}^{-1/2}\gamma\|_\infty + \|\mathbf{M}^{-1/2}\gamma\|_\infty^2\right)}{\|\mathbf{M}^{-1}\gamma\|_\infty\|\gamma\|_1}.\end{aligned}$$

Combining this with the first term in (18), we obtain

$$\frac{\|\mathbf{M}^{-1/2}\gamma\|_2^2}{\|\mathbf{M}^{-1}\gamma\|_\infty\|\gamma\|_1} \geq \frac{1}{J-1}\frac{\left(\|\mathbf{M}^{-1/2}\gamma\|_1^2 + J\|\mathbf{M}^{-1/2}\gamma\|_\infty^2 - 2\|\mathbf{M}^{-1/2}\gamma\|_1\|\mathbf{M}^{-1/2}\gamma\|_\infty\right)}{\|\mathbf{M}^{-1}\gamma\|_\infty\|\gamma\|_1}. \tag{19}$$

To combine the first two terms in the numerator of (19), we use the fact that for any $x, y \geq 0$, $x+y \geq 2\sqrt{xy}$, and we conclude that

$$\begin{aligned}\frac{\|\mathbf{M}^{-1/2}\gamma\|_2^2}{\|\mathbf{M}^{-1}\gamma\|_\infty\|\gamma\|_1} &\geq \frac{1}{J-1}\frac{\left(2\sqrt{J}\|\mathbf{M}^{-1/2}\gamma\|_1\|\mathbf{M}^{-1/2}\gamma\|_\infty - 2\|\mathbf{M}^{-1/2}\gamma\|_1\|\mathbf{M}^{-1/2}\gamma\|_\infty\right)}{\|\mathbf{M}^{-1}\gamma\|_\infty\|\gamma\|_1} \\ &= \frac{2\left(\sqrt{J}-1\right)}{J-1}\frac{\|\mathbf{M}^{-1/2}\gamma\|_1\|\mathbf{M}^{-1/2}\gamma\|_\infty}{\|\mathbf{M}^{-1}\gamma\|_\infty\|\gamma\|_1}.\end{aligned} \tag{20}$$

Now, we define $\gamma^{'} := \mathbf{M}^{-1/2}\gamma$ and observe that

$$\frac{\|\mathbf{M}^{-1/2}\gamma\|_1\|\mathbf{M}^{-1/2}\gamma\|_\infty}{\|\mathbf{M}^{-1}\gamma\|_\infty\|\gamma\|_1} = \frac{\|\gamma^{'}\|_\infty}{\|\mathbf{M}^{-1/2}\gamma^{'}\|_\infty}\cdot\frac{\|\gamma^{'}\|_1}{\|\mathbf{M}^{1/2}\gamma^{'}\|_1}, \geq \frac{1}{\|\mathbf{M}^{-1/2}\|_\infty}\cdot\frac{1}{\|\mathbf{M}^{1/2}\|_1}, = \min_j\sqrt{M_j}\cdot\frac{1}{\max_j\sqrt{M_j}}. \tag{21}$$

Combining (20) and (21) completes our proof of the lower bound. The upper bound follows simply because

$$\|\mathbf{M}^{-1/2}\gamma\|_2^2 = \sum_{j=1}^{J}\frac{\gamma_j^2}{M_j} \leq \left(\max_j\frac{\gamma_j}{M_j}\right)\cdot\sum_{j=1}^{J}\gamma_j = \|\mathbf{M}^{-1}\gamma\|_\infty\|\gamma\|_1,$$

keeping in mind that $\gamma_j \geq 0$ and $M_j > 0$. ■

APPENDIX D
PROOF OF THEOREM III.2

In order to prove Theorem III.2, we will require the following two lemmas.

Lemma D.1. *Suppose $x \in \mathbb{R}^{NJ}$ and $\widetilde{\Phi}$ is an $M \times N$ matrix where $\widetilde{\Phi}^T = [\phi_1\ \phi_2\ \cdots\ \phi_M]$ with each $\phi_i \in \mathbb{R}^N$. Let Φ be an $MJ \times NJ$ RBD matrix as defined in (1) with all $\Phi_j = \widetilde{\Phi}$. If $y = \Phi x$, then $\|y\|_2^2 = \sum_{i=1}^{M}\phi_i^T A\phi_i$, where $A = X^TX$ with X defined in (7).*

Proof of Lemma D.1: We first rewrite the block measurement equations as $y' = X'\phi'$, where y' consists of a rearrangement of the elements of y, i.e., $y' := [y_1(1)\ \cdots\ y_J(1)\ y_1(2)\ \cdots\ y_J(M)]^T \in \mathbb{R}^{MJ}$, $\phi' = [\phi_1^T\ \phi_2^T\ \cdots\ \phi_M^T]^T \in \mathbb{R}^{MN}$, and $X' \in \mathbb{R}^{MJ\times MN}$ is a block diagonal measurement matrix containing replicas of X along the main diagonal and zeros elsewhere.

It follows that $\|y\|_2^2 = \|y'\|_2^2 = \phi'^T X'^T X' \phi' = \sum_{i=1}^M \phi_i^T A \phi_i$. ■

Lemma D.2. *Suppose $z \in \mathbb{R}^N$ is a random vector with i.i.d. Gaussian entries each having zero-mean and variance σ^2. For any symmetric $N \times N$ matrix A with eigenvalues $\{\lambda_i\}_{i=1}^N$, there exists a collection of independent, zero-mean Gaussian random variables $\{w_i\}_{i=1}^N$ with variance σ^2 such that $z^T A z = \sum_{i=1}^N \lambda_i w_i^2$.*

Proof of Lemma D.2: Because A is symmetric, it has an eigen-decomposition $A = V^T D V$, where D is a diagonal matrix of its eigenvalues $\{\lambda_i\}_{i=1}^N$ and V is an orthogonal matrix of eigenvectors. Then we have $z^T A z = (Vz)^T D (Vz) = \sum_{i=1}^N \lambda_i w_i^2$, where $w = Vz$ and $w = [w_1,\ w_2,\ \cdots,\ w_N]^T$. Since V is an orthogonal matrix, $\{w_i\}_{i=1}^N$ are i.i.d. Gaussian random variables with zero-mean and variance σ^2. ■

Proof of Theorem III.2: Let $y = \Phi x$. We first calculate $\mathbf{E}\|y\|_2^2$ to determine the point of concentration. Applying Lemma D.1 to y and Lemma D.2 with $z = \phi_i$ for each $i = 1, 2, \ldots, M$, we have $\|y\|_2^2 = \sum_{i=1}^M \phi_i^T A \phi_i = \sum_{i=1}^M \sum_{j=1}^N \lambda_j w_{i,j}^2$, where each $\{w_{i,j}\}_{i,j}$ is an independent Gaussian random variable with zero-mean and variance $\sigma^2 = \frac{1}{M}$. After switching the order of the summations and observing that $\mathrm{Tr}(X^T X) = \mathrm{Tr}(X X^T)$ where $\mathrm{Tr}(\cdot)$ is the trace operator, we have $\mathbf{E}\|y\|_2^2 = \sum_{j=1}^N \lambda_j \sum_{i=1}^M \mathbf{E} w_{i,j}^2 = \sum_{j=1}^N \lambda_j = \mathrm{Tr}(X X^T) = \|x\|_2^2$.

Having established the point of concentration for $\|y\|_2^2$, let us now compute the probability that $\left| \|y\|_2^2 - \|x\|_2^2 \right| > \epsilon \|x\|_2^2$. Since $\mathbf{E}\|y\|_2^2 = \|x\|_2^2$, this is equivalent to the condition that $\left| \|y\|_2^2 - \mathbf{E}\|y\|_2^2 \right| > \epsilon \mathbf{E}\|y\|_2^2$. Let $\widetilde{w}_{i,j} = \sqrt{\lambda_j} w_{i,j}$. Then $\widetilde{w}_{i,j}$ is a Gaussian random variable with variance $\lambda_j \sigma^2$, and we have $\|y\|_2^2 = \sum_{i=1}^M \sum_{j=1}^N \widetilde{w}_{i,j}^2$. By Lemma II.2 we have

$$P(\widetilde{w}_{i,j}^2 > t) \leq 2 \exp\left(-\frac{t}{2\lambda_j \sigma^2} \right), \quad \forall t \geq 0. \tag{22}$$

To obtain a concentration bound for $\|y\|_2^2$, we invoke Theorem II.1 for the random variables $\widetilde{w}_{i,j}^2$ across all i, j. From (22), we see that the assumptions of the theorem are satisfied with $a = 2$ and $\alpha_{i,j}^{-1} = 2\lambda_j \sigma^2 = \frac{2}{M} \lambda_j$. Note that $\alpha_{i,j}^{-1}$ is constant for a fixed j. Hence, for $d \geq \max_{i,j} \alpha_{i,j}^{-1} = \frac{2}{M} \max_j \lambda_j$ and $b \geq a \sum_{i,j} \alpha_{i,j}^{-2} = \frac{8}{M} \sum_j \lambda_j^2$, the concentration of measure inequality as shown in (17) holds. Note that $\|x\|_2^2 = \mathrm{Tr}(X^T X) = \|\lambda\|_1$ and $\|x\|_2^4 = \|\lambda\|_1^2$ since the eigenvalues $\{\lambda_j\}_{j=1}^N$ are non-negative. Substituting $d = \frac{2}{M} \max_j \lambda_j = \frac{2}{M} \|\lambda\|_\infty$ and $b = \frac{8}{M} \sum_j \lambda_j^2 = \frac{8}{M} \|\lambda\|_2^2$ into (17) completes the proof. ■

APPENDIX E
PROOF OF THEOREM IV.1

Our result follows from an application of the following.

Theorem E.1. [33, Theorem 3.1] *Let $x \in \mathbb{C}^{N'}$ and $\beta > 1$. Suppose $N' > 512$ and choose N_T and N_Ω such that:*

$$N_T + N_\Omega \leq \frac{0.5583 N'/q}{\sqrt{(\beta+1)\log(N')}} \text{ and } N_T + N_\Omega \leq \frac{\sqrt{2/3} N' \left(\frac{1}{q} - \frac{(\log N')^2}{N'} \right)}{\sqrt{(\beta+1)\log(N')}}. \tag{23}$$

Fix a subset T of the time domain with $|T| = N_T$. Let Ω be a subset of size N_Ω of the frequency domain generated uniformly at random. Then with probability at least $1 - O((\log(N'))^{1/2} N'^{-\beta})$, every signal x supported on Ω in the frequency domain has most of its energy in the time domain outside of T. In particular, $\|x_T\|_2^2 \leq \frac{\|x\|_2^2}{q}$,where x_T denotes the restriction of x to the support T.

Proof of Theorem IV.1: First, observe that $\|\gamma\|_1^2 = \|x\|_2^4$ and $\|\gamma\|_2^2 = \sum_{k=1}^J \|x_k\|_2^4$. Next, apply Theorem E.1 with $N_\Omega = S$ and $N_T = N = N'/J$, being careful to select a value for q such that (23) is satisfied. In particular, we require

$$\frac{1}{q} \geq \frac{(N+S)\sqrt{(\beta+1)\log N'}}{0.5583N'} \quad \text{and} \quad \frac{1}{q} \geq \frac{\frac{(N+S)}{\sqrt{2/3}}\sqrt{(\beta+1)\log N'} + (\log N')^2}{N'}.$$

This is satisfied if we choose

$$q \leq \min\left\{ \frac{0.5583N'}{(N+S)\sqrt{(\beta+1)\log N'}}, \frac{N'}{\frac{(N+S)}{\sqrt{2/3}}\sqrt{(\beta+1)\log N'} + (\log N')^2} \right\}. \tag{24}$$

Choosing any q satisfying (24), we have that with *failure probability* at most $O((\log(N'))^{1/2}(N')^{-\beta})$, $\|x_k\|_2^2 \leq \frac{\|x\|_2^2}{q}$ for each $k = 1, \cdots, J$, implying that each block individually is favorable. Taking a union bound for all k to cover each block, we have that with total failure probability at most $O(J(\log(N'))^{1/2}(N')^{-\beta})$, $\|\gamma\|_2^2 = \sum_{k=1}^J \|x_k\|_2^4 \leq \frac{J\|x\|_2^4}{q^2}$. Thus with this same failure probability, $\frac{\Gamma}{M} = \frac{\|\gamma\|_1^2}{\|\gamma\|_2^2} \geq \frac{q^2}{J}$. Combining with (24) and using the fact that $S < N$, we thus have:

$$\begin{aligned}
\frac{\Gamma}{M} &\geq \min\left\{ \frac{0.5583^2 N^2 J}{(N+S)^2(\beta+1)\log N'}, \frac{N^2 J}{\left(\frac{(N+S)}{\sqrt{2/3}}\sqrt{(\beta+1)\log N'} + (\log N')^2\right)^2} \right\} \\
&\geq \min\left\{ \frac{(0.5583^2/2^2)J}{(\beta+1)\log(N')}, \frac{J}{\left(\frac{2}{\sqrt{2/3}}\sqrt{(\beta+1)\log N'} + \frac{(\log N')^2}{N}\right)^2} \right\}.
\end{aligned}$$

■

APPENDIX F
PROOF OF LEMMA IV.1

Proof: Let X be the $J \times N$ matrix as defined in (7). Without loss of generality, we suppose the nonzero eigenvalues $\{\lambda_i\}_{i=1}^{\min(J,N)}$ of $X^T X$ are sorted in order of decreasing magnitude, and we let $\lambda_{\max} := \lambda_1$ and $\lambda_{\min} := \lambda_{\min(J,N)}$. We can lower bound Λ in terms of these extremal eigenvalues by writing

$$\Lambda = \frac{M\|\lambda\|_1^2}{\|\lambda\|_2^2} = M\frac{\sum_i \lambda_i^2 + \sum_i \sum_{j\neq i} \lambda_i\lambda_j}{\sum_i \lambda_i^2} \geq M + M\frac{\lambda_{\min}}{\lambda_{\max}}\frac{\sum_i \sum_{j\neq i} \lambda_i}{\sum_i \lambda_i} = M + M\frac{\lambda_{\min}}{\lambda_{\max}}(J-1). \tag{25}$$

For some $0 \leq \epsilon \leq 1$, let us define the following events:

$$A = \left\{ N\sigma^2(1-\epsilon)^2 \leq \frac{\|X^T z\|_2^2}{\|z\|_2^2} \leq N\sigma^2(1+\epsilon)^2, \ \forall z \in \mathbb{R}^J \right\}, \quad B = \left\{\lambda_{\max} \leq N\sigma^2(1+\epsilon)^2\right\} \bigcap \left\{\lambda_{\min} \geq N\sigma^2(1-\epsilon)^2\right\},$$
$$C = \left\{ \frac{\lambda_{\min}}{\lambda_{\max}} \geq \left(\frac{1-\epsilon}{1+\epsilon}\right)^2 \right\}, \qquad D = \left\{ \Lambda \geq M + M\left(\frac{1-\epsilon}{1+\epsilon}\right)^2 (J-1) \right\}.$$

These events satisfy $A = B \subseteq C \subseteq D$, where the last relation follows from (25). It follows that $P(D^c) \leq P(A^c)$, where A^c represents the complement of event A. Because X^T is populated with i.i.d. subgaussian random variables, it follows as a corollary of Theorem III.1 (by setting $M \leftarrow N$ and $J \leftarrow 1$ in the context of that theorem) that for any $z \in \mathbb{R}^J$ and

$\epsilon \in (0,1)$, $P(\left|\|X^T z\|_2^2 - N\sigma^2\|z\|_2^2\right| > \epsilon N\sigma^2\|z\|_2^2) \leq 2e^{-\frac{N\epsilon^2}{256c^2}}$. Thus, for an upper bound for $P(A^c)$, we can follow the straightforward arguments in [10, Lemma 5.1] and conclude that $P(A^c) \leq 2\left(\frac{12}{\epsilon}\right)^J e^{-\frac{N\epsilon^2}{4\cdot 256c^2}}$. For any constant $c_1 > 0$, we note that $P(D^c) \leq 2\left(\frac{12}{\epsilon}\right)^J e^{-\frac{N\epsilon^2}{4\cdot 256c^2}} = 2e^{-\frac{N\epsilon^2}{4\cdot 256c^2} + J\log\left(\frac{12}{\epsilon}\right)} \leq 2e^{-c_1 N}$ holds whenever $J \leq N\left(\frac{\epsilon^2}{4\cdot 256c^2} - c_1\right)\log^{-1}\left(\frac{12}{\epsilon}\right)$. Finally, the fact that $\Gamma \geq \Lambda$ follows from (10). ■

APPENDIX G
PROOF OF LEMMA V.1

Proof: To apply Theorem III.2 directly, we first suppose the entries of the random probe ϕ have variance $\sigma^2 = 1$ (since $M = 1$ here). In this case, we have $\mathbf{E}\|y\|_2^2 = \|x\|_2^2 = J\|a\|_2^2$, and Theorem III.2 implies that

$$P(\left|\|y\|_2^2 - J\|a\|_2^2\right| > \epsilon J\|a\|_2^2) \leq \begin{cases} 2\exp\{-\frac{\epsilon^2\|\lambda\|_1^2}{256\|\lambda\|_2^2}\}, & 0 \leq \epsilon \leq \frac{16\|\lambda\|_2^2}{\|\lambda\|_\infty\|\lambda\|_1} \\ 2\exp\{-\frac{\epsilon\|\lambda\|_1}{16\|\lambda\|_\infty}\}, & \epsilon \geq \frac{16\|\lambda\|_2^2}{\|\lambda\|_\infty\|\lambda\|_1} \end{cases}. \tag{26}$$

Since $J \leq N$, the nonzero eigenvalues of XX^T equal those of X^TX, and so we could equivalently define $\{\lambda_j\}$ as the eigenvalues of XX^T. Note that XX^T is a symmetric matrix, and so all of its eigenvalues $\{\lambda_i\}$ are non-negative. Also, all of the diagonal entries of XX^T are equal to $\|a\|_2^2$. Consequently, it follows that $\|\lambda\|_1 = \sum_{i=1}^J |\lambda_i| = \mathrm{tr}(XX^T) = J\|a\|_2^2$. Using the norm inequality $\|\lambda\|_2^2 \leq \|\lambda\|_1\|\lambda\|_\infty$, we have $\frac{\|\lambda\|_1^2}{\|\lambda\|_2^2} \geq \frac{J}{\|\lambda\|_\infty/\|a\|_2^2}$ and it is easy to see that $\frac{\|\lambda\|_1}{\|\lambda\|_\infty} = \frac{J}{\|\lambda\|_\infty/\|a\|_2^2}$.

By plugging these inequalities into (26), we obtain the probabilities specified in (15). Finally, we note that

$$\left|\|y\|_2^2 - J\|a\|_2^2\right| > \epsilon J\|a\|_2^2 \iff \left|\|(1/\sqrt{J})\Phi_{\mathrm{small}}a\|_2^2 - \|a\|_2^2\right| > \epsilon\|a\|_2^2,$$

and so if we suppose the entries of the random probe ϕ actually have variance $\sigma^2 = \frac{1}{J}$, we complete our derivation of (15). ■

REFERENCES

[1] 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.

[2] 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.

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

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

[5] M. Ledoux, "The concentration of measure phenomenon," *Mathematical Surveys and Monographs, AMS*, 2001.

[6] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," in *Proc. Conf. Modern Analysis and Probability*, 1984.

[7] D. Achlioptas, "Database-friendly random projections," in *Proc. 20th ACM SIGMOD-SIGACT-SIGART Symp. Principles of Database Systems (PODS)*, New York, NY, USA, 2001, pp. 274–281, ACM.

[8] S. Dasgupta and A. Gupta, "An elementary proof of the Johnson-Lindenstrauss lemma," *Random Struct. Algor.*, vol. 22, no. 1, pp. 60–65, 2002.

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

[10] R. G. Baraniuk, M. A. Davenport, R. A. DeVore, and M. B. Wakin, "A simple proof of the restricted isometry property for random matrices," *Const. Approx.*, vol. 28, no. 3, pp. 253–263, Dec. 2008.

[11] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Uniform uncertainty principle for Bernoulli and subgaussian ensembles," *Const. Approx.*, vol. 28, no. 3, pp. 277–289, 2008.

[12] M. S. Asif, D. Reddy, P. T. Boufounos, and A. Veeraraghavanc, "Streaming compressive sensing for high-speed periodic videos," in *Proc. Int. Conf. Image Processing (ICIP)*, 2009.

[13] P. Boufounos and M. S. Asif, "Compressive sampling for streaming signals with sparse frequency content," in *Proc. Conf. Information Sciences and Systems (CISS)*, 2010.

[14] J. A. Tropp, M. B. Wakin, M. F. Duarte, D. Baron, and R. G. Baraniuk, "Random filters for compressive sampling and reconstruction," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2006.

[15] J. Romberg, "Compressive sensing by random convolution," *SIAM J. Imaging Sciences*, vol. 2, no. 4, pp. 1098–1128, 2009.

[16] W. U. Bajwa, J. D. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak, "Toeplitz-structured compressed sensing matrices," in *Proc. IEEE/SP 14th Workshop Statistical Signal Processing (SSP)*, 2007, pp. 294–298.

[17] W. U. Bajwa, J. D. Haupt, G. M. Raz, and R. D. Nowak, "Compressed channel sensing," in *Proc. Conf. Information Sciences and Systems (CISS)*, 2008.

[18] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," Preprint, 2008.

[19] J. Romberg, "A uniform uncertainty principle for Gaussian circulant matrices," in *Proc. Int. Conf. Digital Signal Processing (DSP)*, July 2009.

[20] V. Saligrama, "Deterministic designs with deterministic guarantees: Toeplitz compressed sensing matrices, sequence designs and system identification," 2008, Preprint.

[21] H. Rauhut, "Circulant and Toeplitz matrices in compressed sensing," in *Proc. Signal Processing with Adaptive Sparse Structured Representations (SPARS)*, 2010, vol. 9.

[22] B. M. Sanandaji, T. L. Vincent, and M. B. Wakin, "Concentration of measure inequalities for compressive Toeplitz matrices with applications to detection and system identification," in *Proc. IEEE Conf. Decision and Control (CDC)*, 2010.

[23] V. V. Buldygin and Y. V. Kozachenko, "Sub-gaussian random variables," *Ukrainian Mathematical Journal*, vol. 32, Nov. 1980.

[24] R. Vershynin, "On large random almost Euclidean bases," *Acta Math. Univ. Comenianae*, vol. 69, no. 2, pp. 137–144, 2000.

[25] P. Indyk and R. Motwani, "Approximate nearest neighbors: Towards removing the curse of dimensionality," in *Proc. ACM Symp. Theory of Computing*, 1998, pp. 604–613.

[26] R. DeVore, G. Petrova, and P. Wojtaszczyk, "Instance-optimality in probability with an l1-minimization decoder," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 275–288, 2009.

[27] P. Frankl and H. Maehara, "The Johnson-Lindenstrauss lemma and the sphericity of some graphs," *J. Combinatorial Theory, Series B*, vol. 44, no. 3, pp. 355–362, 1988.

[28] 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.

[29] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[30] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Found. Comp. Math.*, vol. 9, no. 1, pp. 51–77, 2009.

[31] A. J. Izenman, *Modern Multivariate Statistical Techniques*, Springer, 2008.

[32] J. A. Tropp, "On the linear independence of spikes and sines," *J. Fourier Analysis and Applications*, vol. 14, no. 5, pp. 838–858, 2008.

[33] E. J. Candès and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Found. Comp. Math.*, vol. 6, no. 2, pp. 227–254, 2006.

[34] J. D. Haupt and R. D. Nowak, "Compressive sampling for signal detection," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2007.

[35] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Proc. IS&T/SPIE Symp. Electronic Imaging: Computational Imaging*, 2007.

[36] M. B. Wakin, B. M. Sanandaji, and V. L. Tyrone, "On the observability of linear systems from random, compressive measurements," *IEEE Conf. Decision and Control (CDC)*, 2010.