Concentration of Measure for Block Diagonal Matrices with Applications to Compressive Signal Processing

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

October 2010; Revised June 2011

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. 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. We also discuss applications of these results in ensuring stable embeddings for various signal families and in establishing performance guarantees for solving various signal processing tasks (such as detection and classification) directly in the compressed domain.

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 [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

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 and to the anonymous reviewers for valuable feedback.

Index terms: Concentration of measure phenomenon, Block diagonal matrices, Compressive Sensing, Restricted Isometry Property.

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.

As we discuss in Section II of this paper, such concentration results have a number of favorable implications. Among these is 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 a broad variety of signal processing problems by permitting these problems to be solved in the low-dimensional observation space (such as finding the nearest neighbor to a point x in a database Q). Such concentration results have 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 Compressive Sensing (CS), the RIP is commonly used as a sufficient condition to guarantee that a sparse signal x can be recovered from the measurements Φx.

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 are often either impractical because of the resources required to store and work with a large unstructured matrix (e.g., one with i.i.d. entries), or unrealistic as models of acquisition devices with architectural constraints preventing such global data aggregation. For 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.

In such scenarios, the data may be divided naturally into discrete subsections (or blocks), with each block acquired via a local measurement operator. 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.

Starting from this block diagonal matrix structure, the main contributions of this paper are a derivation of concentration of measure bounds for DBD and RBD matrices and an extensive exploration of the implications and utility of these bounds for the signal processing community. Specifically, in Section III we derive concentration of measure bounds both for DBD matrices populated with i.i.d. subgaussian[1] random variables and for RBD matrices populated with i.i.d. Gaussian random variables. In contrast to the signal agnostic concentration of measure bounds for i.i.d. dense matrices, these bounds are signal dependent; in particular, the probability of concentration depends on the "diversity" of the component signals $x_1, x_2, \ldots, x_J$ being well-matched to the measurement matrix (we make this precise in Section III). As our analytic discussion and supporting simulations show, these measures of diversity have clear intuitive interpretations and indicate that, for signals with the most favorable characteristics, the concentration of measure probability for block diagonal matrices can scale exactly as for an i.i.d. dense random matrix.

Sections IV and V are devoted to a detailed investigation of the utility of these non-uniform concentration results for signal processing practitioners. Specifically, in Section IV, we extend our concentration results to formulate a modified version of the JL lemma appropriate for block diagonal matrices. We also explain how this lemma can be used to guarantee the performance of various compressive-domain signal inference and processing algorithms such as signal detection and estimation. Given the applicability of these results for providing performance guarantees in these tasks, a natural question is whether there are large classes of signals that have the diversity required to make block diagonal matrices perform well. In Section V we provide several examples of signal families that are particularly favorable for measurement via DBD or RBD matrices.

II. Background and Related Work

In this section, we begin with a definition of subgaussian random variables, describe more formally several existing concentration of measure results for random matrices, and review some standard applications of these results in the literature.

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 *subgaussian* distributions [14, 15]. Where possible, we state our results in terms of subgaussian random variables, which are defined below using standard notation from the literature.

Definition II.1. [15] *A random variable W is subgaussian if $\exists a > 0$ such that*

$$(\mathbf{E}|W|^p)^{1/p} \leq a\sqrt{p} \text{ for all } p \geq 1.$$

The quantity $\|W\|_{\psi_2} := \sup_{p\geq 1} p^{-1/2}(\mathbf{E}|W|^p)^{1/p}$ is known as the subgaussian norm of W.

We restrict our attention to zero-mean subgaussian random variables in this paper. Examples of such random variables include zero-mean Gaussian random variables, ± 1 Bernoulli random variables (each value with probability $\frac{1}{2}$), and uniform random variables on $[-1, 1]$.

[1]Subgaussian random variables [14, 15] 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.

For a given subgaussian random variable W, the variance $\mathrm{Var}(W)$ is a constant multiple of $\|W\|_{\psi_2}^2$ with $0 \leq \frac{\mathrm{Var}(W)}{\|W\|_{\psi_2}^2} \leq \sqrt{2}$; the exact value of $\frac{\mathrm{Var}(W)}{\|W\|_{\psi_2}^2}$ depends on the specific distribution for W under consideration (Gaussian, Bernoulli, etc.). We also note that in some cases, we will consider independent realizations of a subgaussian random variable W that are normalized to have different variances. However, it is useful to note that for any scalar $\alpha > 0$, $\frac{\mathrm{Var}(\alpha W)}{\|\alpha W\|_{\psi_2}^2} = \frac{\alpha^2 \mathrm{Var}(W)}{\alpha^2 \|W\|_{\psi_2}^2} = \frac{\mathrm{Var}(W)}{\|W\|_{\psi_2}^2}$.

B. Concentration Inequalities

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

$$P\left(\left|\|\Phi x\|_2^2 - \|x\|_2^2\right| > \epsilon \|x\|_2^2\right) \leq 2e^{-M c_0(\epsilon)}, \tag{2}$$

where $c_0(\epsilon)$ is some constant (depending on ϵ) that is typically on the order of ϵ^2. When discussing bounds such as (2) where the probability of failure scales as e^{-X}, we refer to X 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 uniform concentration result of the form (2) was originally derived for dense Gaussian matrices populated with entries having mean zero and variance $\frac{1}{M}$ [16]; 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 uniform concentration results of the form (2) in fact holds for *all* subgaussian distributions having variance $\frac{1}{M}$ [11, 17].[2] Moreover, it has been shown that a subgaussian distribution is actually necessary for deriving a uniform concentration result of the form (2) for a dense random matrix populated with i.i.d. entries [17].

Concentration inequalities have also been derived in the literature for certain structured (non-i.i.d.) dense random matrices. Examples include matrices populated with random orthogonal rows [18] or matrices constructed by combining a structured matrix with certain random operations [19]. A concentration bound also holds for the randomized RIP matrices [20, 21] that we discuss in the final paragraph of Section II-C below.

C. Applications of Concentration Inequalities

While nominally a concentration result of the form (2) appears to guarantee only that the norm of a particular signal x is preserved in the measurements Φx, in fact such a result can be used to guarantee that the information required to discriminate x from *other signals* may actually be preserved in Φx. Indeed, one of the prototypical applications of a concentration result of the form (2) is to prove that with high probability, Φ will preserve distances between various signals of interest.

Definition II.2. *Consider two sets $U, V \subset \mathbb{R}^N$. We say that a mapping Φ provides a* stable embedding of (U, V) with

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

conditioning δ *if*

$$(1-\delta)\|u-v\|_2^2 \leq \|\Phi(u-v)\|_2^2 \leq (1+\delta)\|u-v\|_2^2 \tag{3}$$

holds for all $u \in U$ *and* $v \in V$.

To relate this concept to norm preservation, note that for a fixed $x \in \mathbb{R}^N$, a randomized operator that satisfies (2) will, with high probability, provide a stable embedding of $(\{x\}, \{0\})$ with conditioning ϵ. As we discuss in Section IV, this fact is useful for studying the performance of a compressive-domain signal detector [22].

However, much richer embeddings may also be considered. For example, suppose that a signal family of interest $Q \subset \mathbb{R}^N$ consists of a finite number of points. If a randomized operator Φ satisfies (2) for each vector of the form $u - v$ for $u, v \in Q$, then it follows from a simple union bound that with probability at least $1 - 2|Q|^2 e^{-Mc_0(\epsilon)}$, Φ will provide a stable embedding of (Q, Q) with conditioning ϵ. From this fact one obtains the familiar JL lemma [6–8], which states that this stable embedding will occur with high probability if $M = O\left(\frac{\log(|Q|)}{c_0(\epsilon)}\right)$. Thus, the information required to discriminate between signals in Q is preserved in the compressive measurements. This fact is useful for studying the performance of a compressive-domain nearest-neighbor search [16], a compressive-domain signal classifier [22], and various other signal inference strategies that we discuss in Section IV.

In certain cases (particularly when dealing with randomized operators that satisfy (2) uniformly over all signals $x \in \mathbb{R}^N$), it is possible to significantly extend embedding results far beyond the JL lemma. For example, for a set Q consisting of all signals with sparsity K in some basis for $\mathbb{R}^N$, one can couple the above union bound approach with some elementary covering arguments [10, 11] to show that if $M = O\left(K \log(N/K)\right)$, then Φ will provide a stable embedding of (Q, Q) with high probability. This guarantee is known as the RIP in CS, and from the RIP, one can derive deterministic bounds on the performance of CS recovery algorithms such as ℓ_1 minimization [23]. Concentration of measure type results have also been used to prove the RIP for random matrices with subexponential columns [24], and a concentration result of the form (2) has also been used to probabilistically analyze the performance of ℓ_1 minimization [17]. Finally, we note that one can also generalize sparsity-based embedding arguments to the case where Q is a low-dimensional submanifold of $\mathbb{R}^N$ [25].

In a different direction of interest regarding stable embeddings of finite collections of points, we note that several authors have also recently shown that the direction of implication between the JL lemma and the RIP can be reversed. Specifically, it has been shown [20, 21] that randomizing the column signs of a matrix that satisfies the RIP results in a matrix that also satisfies the JL lemma. One of the implications of this is that a computationally efficient stable embedding can be achieved by randomizing the column signs of a partial Fourier matrix [26–28].

III. Non-uniform Concentration of Measure Inequalities

In this section we state our 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, we 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: $\mathbf{M} := \text{diag}(M_1, M_2, \ldots, M_J)$. Finally, for a given signal $x \in \mathbb{R}^{NJ}$ and measurement allocation $\mathbf{M}$, we define the quantities

$$\Gamma_2(x, \mathbf{M}) := \frac{\|\gamma\|_1^2}{\|\mathbf{M}^{-1/2}\gamma\|_2^2} = \frac{\left(\sum_{j=1}^J \|x_j\|_2^2\right)^2}{\sum_{j=1}^J \frac{\|x_j\|_2^4}{M_j}} \quad \text{and} \quad \Gamma_\infty(x, \mathbf{M}) := \frac{\|\gamma\|_1}{\|\mathbf{M}^{-1}\gamma\|_\infty} = \frac{\sum_{j=1}^J \|x_j\|_2^2}{\max_j \frac{\|x_j\|_2^2}{M_j}}. \tag{4}$$

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

Theorem III.1. *Suppose $x \in \mathbb{R}^{NJ}$, and for each $j \in \{1, 2, \ldots, J\}$ suppose that $M_j > 0$. Let ϕ denote a subgaussian random variable with mean 0, variance 1, and subgaussian norm $\|\phi\|_{\psi_2}$. Let $\{\Phi_j\}_{j=1}^J$ be random matrices drawn independently, where each Φ_j has size $M_j \times N$ and is populated with i.i.d. realizations of the renormalized random variable $\frac{\phi}{\sqrt{M_j}}$, 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*

$$P\left(\left|\|\Phi x\|_2^2 - \|x\|_2^2\right| > \epsilon \|x\|_2^2\right) \leq 2 \exp\left\{-C_1 \min\left(\frac{C_2^2 \epsilon^2}{\|\phi\|_{\psi_2}^4} \Gamma_2(x, \mathbf{M}), \frac{C_2 \epsilon}{\|\phi\|_{\psi_2}^2} \Gamma_\infty(x, \mathbf{M})\right)\right\}, \tag{5}$$

where C_1 and C_2 are absolute constants.

Proof: See Appendix A.

From the tail bound (5), it is easy to deduce that the concentration probability of interest decays exponentially as a function of $\epsilon^2 \Gamma_2(x, \mathbf{M})$ in the case where $0 \leq \epsilon \leq \frac{\|\phi\|_{\psi_2}^2 \Gamma_\infty(x, \mathbf{M})}{C_2 \Gamma_2(x, \mathbf{M})}$ and exponentially as a function of $\epsilon \Gamma_\infty(x, \mathbf{M})$ in the case where $\epsilon > \frac{\|\phi\|_{\psi_2}^2 \Gamma_\infty(x, \mathbf{M})}{C_2 \Gamma_2(x, \mathbf{M})}$. One striking thing about Theorem III.1 is that, in contrast to analogous concentration of measure results for dense matrices with i.i.d. subgaussian entries,[3] the concentration rate depends explicitly on 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 of (5), when the concentration exponent scales with $\Gamma_2(x, \mathbf{M})$. In this case, we see that larger values of $\Gamma_2(x, \mathbf{M})$ promote sharper concentration of $\|\Phi x\|_2^2$ about its mean $\|x\|_2^2$. Using elementary inequalities relating the ℓ_1 and ℓ_2 norms, one can bound the range of possible Γ_2 values by $\min_j M_j \leq \Gamma_2(x, \mathbf{M}) \leq \sum_{j=1}^J M_j$. The worst case, $\Gamma_2(x, \mathbf{M}) = \min_j M_j$, 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 intuitive sense, as this extreme case would correspond to only one block of the DBD matrix sensing all of the signal energy. On the other hand, the best case, $\Gamma_2(x, \mathbf{M}) = \sum_{j=1}^J M_j$, is achieved when the number of measurements collected for each block is proportional to the signal energy in that block. In other words, letting $\text{diag}(\mathbf{M})$ represent the diagonal of $\mathbf{M}$, when $\text{diag}(\mathbf{M}) \propto \gamma$ (i.e.,

[3]The uniformity of such concentration results comes not from the fact that these matrices are dense but rather that that are populated with i.i.d. random variables; certain structured dense random matrices (such as partial Fourier matrices) could have signal-dependent concentration inequalities.

when $\mathrm{diag}(\mathbf{M}) = C\gamma$ for some constant $C > 0$) 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. This is in spite of the fact that the DBD matrix has many fewer nonzero elements.

The probability of concentration behaves similarly in the second case of (5), where the concentration exponent scales with $\Gamma_\infty(x, \mathbf{M})$. One can bound the range of possible Γ_∞ values by $\min_j M_j \leq \Gamma_\infty(x, \mathbf{M}) \leq \sum_{j=1}^{J} M_j$. 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 again is achieved when $\mathrm{diag}(\mathbf{M}) \propto \gamma$.

The above discussion makes clear that the concentration performance of a DBD matrix can vary widely depending on the signal being measured. In particular, DBD matrices can perform as well as dense i.i.d. matrices if their measurement allocation is well matched to the energy distribution of the signal. Such a favorable event can occur either (i) by design, if a system designer has some operational knowledge of the energy distributions to expect, or (ii) by good fortune, if favorable signals happen to arrive that are well matched to a fixed system design. We note that even in the former situation when the general energy distribution across blocks is known, this does not imply that the designer has a priori knowledge of the signal being sensed. Furthermore, even when significant information about the signal (or a finite class of signals) is known, there may still be much to learn by actually measuring the signal. For example, Section IV outlines several interesting signal inference problems that benefit from a norm-preservation guarantee for a known signal (or finite signal family). Also, in the second of these situations, it may not be unreasonable to expect that a fixed measurement allocation will be well matched to an unknown signal most of the time. For example, in Section V we describe several realistic signal classes that are favorably matched to fixed systems that have equal measurement allocations ($M_1 = M_2 = \cdots = M_J$).

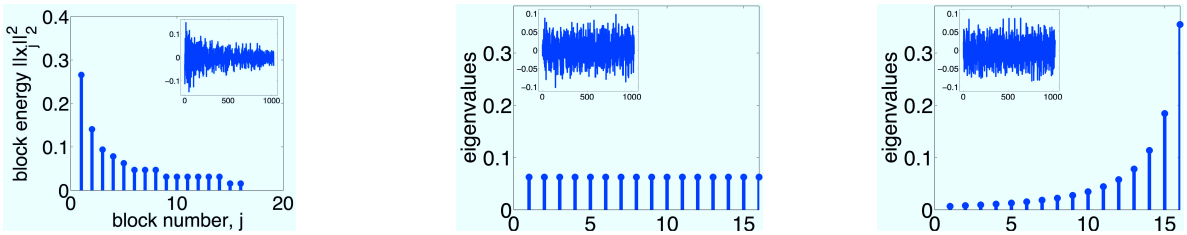
Two final comments are in order. First, while Theorem III.1 was derived by considering all signal blocks to be of equal length N, one can see by a close examination of the proof that the same theorem in fact holds for signals partitioned into blocks of unequal lengths. Second, it is instructive to characterize the range of ϵ for which the two cases of Theorem III.1 are relevant; we do so in the following lemma, which can be proved using standard manipulations of the ℓ_1, ℓ_2, and ℓ_∞ norms.

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

$$\frac{\|\phi\|_{\psi_2}^2 \cdot 2(\sqrt{J}-1)}{C_2(J-1)} \frac{\min_j \sqrt{M_j}}{\max_j \sqrt{M_j}} \leq \frac{\|\phi\|_{\psi_2}^2 \Gamma_\infty(x, \mathbf{M})}{C_2 \Gamma_2(x, \mathbf{M})} \leq \frac{\|\phi\|_{\psi_2}^2}{C_2}.$$

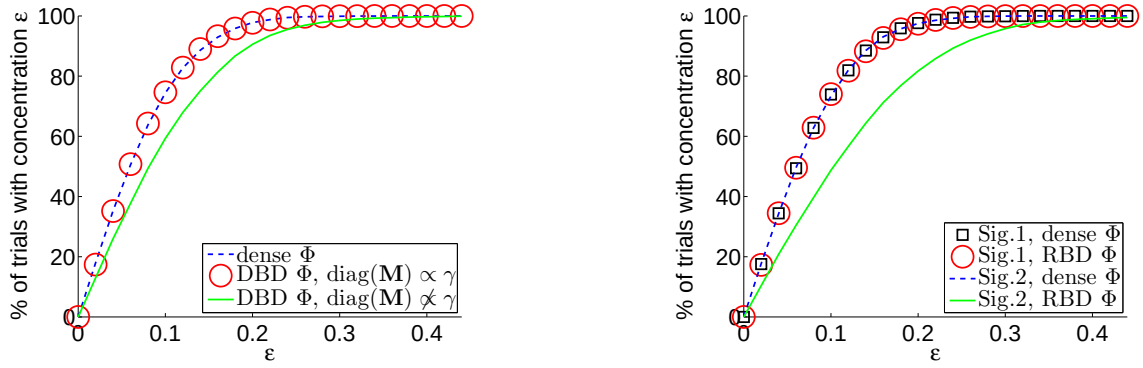
Examining the bound above, we note that for $J \geq 2$ it holds that $\frac{2(\sqrt{J}-1)}{J-1} \geq \frac{1}{\sqrt{J}}$. Thus, as an example, when $M_1 = M_2 = \cdots = M_J$, the first ("small ϵ") case of Theorem III.1 is guaranteed to at least permit $\epsilon \in \left[0, \frac{\|\phi\|_{\psi_2}^2}{C_2\sqrt{J}}\right]$. We note further that when the measurement matrix is well-matched to the signal characteristics, the first case of Theorem III.1 permits ϵ as large as $\frac{\|\phi\|_{\psi_2}^2}{C_2}$, which is independent of J.

2) Supporting Experiments: While the quantity $\Gamma_2(x, \mathbf{M})$ plays a critical role in our analytical tail bound (5), 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 energy distribution γ of the signal x is plotted in Figure 1(a) (and the signal x itself is plotted in the top right corner). For this simulation, to ensure $\mathrm{diag}(\mathbf{M}) \propto \gamma$ with integer values for the M_j, we begin by



(a) (b) (c)

Fig. 1. *(a) Test signal for concentration in a DBD matrix. The main panel plots the energy distribution γ when the signal is partitioned into $J = 16$ blocks of length $N = 64$; the subpanel plots the length-1024 signal x itself. (b),(c) Test signals for concentration in a RBD matrix. The main panels plot the eigenvalue distributions λ for Sig. 1 and Sig. 2, respectively, when partitioned into $J = 16$ blocks of length $N = 64$; the subpanels plot the length-1024 signals themselves.*



(a) (b)

Fig. 2. *Comparisons of concentration behavior for various matrix and signal types; each figure shows the percentage of trials for which $(1-\epsilon) \leq \|\Phi x\|_2/\|x\|_2 \leq (1+\epsilon)$ as a function of ϵ. (a) Performance of the dense and DBD matrices on the signal shown in Figure 1(a). (b) Performance of the dense and RBD matrices on the signals shown in Figure 1(b),(c).*

constructing $\mathbf{M}$ (populated with integers) and then normalize 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 matrices are DBD with $\mathrm{diag}(\mathbf{M}) \propto \gamma$ and the entries in each block have variance $1/M_j$. Thus, we have $\Gamma_2(x, \mathbf{M}) = \sum_{j=1}^{J} M_j$ 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. For each type of matrix, Figure 2(a) shows the percentage of trials for which $(1-\epsilon) \leq \|\Phi x\|_2/\|x\|_2 \leq (1+\epsilon)$ as a function of ϵ, and indeed, 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 $\Gamma_2(x, 4 \cdot I_{J \times J}) = 32.77 < \sum_{j=1}^{J} M_j$. Again, Figure 2(a) 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

$$X := [x_1 \; x_2 \; \cdots \; x_J]^T, \tag{6}$$

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. We let $M := M_1 = M_2 = \cdots = M_J$ denote the number of measurements to be taken in each block. Finally, for a given signal $x \in \mathbb{R}^{NJ}$ and per-block measurement rate M, we define the quantities

$$\Lambda_2(x, M) := \frac{M\|\lambda\|_1^2}{\|\lambda\|_2^2} \quad \text{and} \quad \Lambda_\infty(x, M) := \frac{M\|\lambda\|_1}{\|\lambda\|_\infty}. \tag{7}$$

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*

$$P\left(\left|\|\Phi x\|_2^2 - \|x\|_2^2\right| > \epsilon\|x\|_2^2\right) \leq 2\exp\left\{-C_1 \min\left(C_3^2\epsilon^2\Lambda_2(x, M),\; C_3\epsilon\Lambda_\infty(x, M)\right)\right\}, \tag{8}$$

where C_1 and C_3 are absolute constants.

Proof: See Appendix B.

From (8), one can deduce that the concentration probability of interest decays exponentially as a function of $\epsilon^2\Lambda_2(x, M)$ in the case where $0 \leq \epsilon \leq \frac{\Lambda_\infty(x,M)}{C_3\Lambda_2(x,M)}$ and exponentially as a function of $\epsilon\Lambda_\infty(x, M)$ in the case where $\epsilon > \frac{\Lambda_\infty(x,M)}{C_3\Lambda_2(x,M)}$. Thus, we see that the concentration rate again depends explicitly on the signal x being measured.

Again, since we are frequently concerned in practice with applications where ϵ is small, let us focus on the first case of (8), when the concentration exponent scales with $\Lambda_2(x, M)$. It follows from the standard relation between ℓ_1 and ℓ_2 norms that $M \leq \Lambda_2(x, M) \leq M\min(J, N)$. One extreme, $\Lambda_2(x, M)$, 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, $\Lambda_2(x, M) = M\min(J, N)$ 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 $M \leq \Lambda_2(x, M) \leq MK$. 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) [29] on the matrix X comprised of the J signal blocks. Said another way, an RBD matrix performs as well as a dense i.i.d. matrix of the same size when the signal has uniform energy distribution across

its blocks (as in the DBD case) *and* has sufficient variation in the directions exhibited by the blocks.

We note that there is a close connection between the diversity measures $\Gamma_2(x, \mathbf{M})$ and $\Lambda_2(x, M)$ that is not apparent at first glance. For a fair comparison, we assume in this discussion that $\mathbf{M} := \text{diag}(M, M, \ldots, M)$. Now, note that $\|\lambda\|_1^2 = \|\gamma\|_1^2$ and also that $\|\lambda\|_2^2 = \|A\|_F^2 = \|XX^T\|_F^2 = \sum_{i=1}^J \|x_i\|_2^4 + 2\sum_{i>j}(x_i^T x_j)^2 = \|\gamma\|_2^2 + 2\sum_{i>j}(x_i^T x_j)^2$. Using these two relationships, we can rewrite $\Lambda_2(x, M)$ as

$$\Lambda_2(x, M) = \frac{M\|\lambda\|_1^2}{\|\lambda\|_2^2} = \frac{M\|\gamma\|_1^2}{\|\gamma\|_2^2 + 2\sum_{i>j}(x_i^T x_j)^2} \leq \frac{M\|\gamma\|_1^2}{\|\gamma\|_2^2} = \Gamma_2(x, \mathbf{M}). \tag{9}$$

From this relationship we see that Λ_2 and Γ_2 differ only by the additional inner-product term in the denominator of Λ_2, and we also see that $\Lambda_2 = \Gamma_2$ if and only if the signal blocks are mutually orthogonal. This more stringent condition for RBD matrices—requiring more intrinsic signal diversity—is expected given the more restricted structure of the RBD matrices.

2) Supporting Experiments: While the quantity $\Lambda_2(x, M)$ 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 nonzero eigenvalues of $A = X^T X$ are shown in the plot of λ in Figure 1(b) (and the signal x itself, denoted "Sig. 1", is plotted in the top left corner).

As we have discussed above, for signals such as Sig. 1 we should have $\Lambda_2(x, M) = MJ$, 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 $\frac{1}{4}$. For each type of matrix, Figure 2(b) shows the percentage of trials for which $(1-\epsilon) \leq \|\Phi x\|_2/\|x\|_2 \leq (1+\epsilon)$ as a function of ϵ. As anticipated, we can see that 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 Figure 1(c). 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 $\Lambda_2(x, M) = 21.3284$ 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. As shown in Figure 2(b), 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.

IV. Applications

As discussed briefly in Section II-C, a concentration of measure inequality—despite nominally pertaining to the norm preservation of a single signal—can lead to a number of guarantees for problems involving multi-signal embeddings and signal discrimination. In this section, we extend our concentration bounds to formulate a modified version of the JL lemma appropriate

for block diagonal matrices. We also survey a collection of compressive-domain inference problems (such as detection and classification) in which such a result can be leveraged. For simplicity we will focus on DBD matrices in this section, but parallel results can be derived in each case for RBD matrices. Given the nonuniform nature of our concentration bounds, the performance of algorithms for solving these problems will depend on the signals under consideration, and so, in Section V we provide several examples of signal classes that are particularly favorable for measurement via DBD or RBD matrices.

A. Stable Embeddings and the Johnson-Lindenstrauss Lemma

For a given signal $x \in \mathbb{R}^{NJ}$ and measurement allocation $\mathbf{M}$, let us define $\widetilde{\Gamma}_2(x, \mathbf{M}) := \frac{\Gamma_2(x,\mathbf{M})}{\sum_{j=1}^J M_j}$ and $\widetilde{\Gamma}_\infty(x, \mathbf{M}) := \frac{\Gamma_\infty(x,\mathbf{M})}{\sum_{j=1}^J M_j}$. Note that both quantities are upper bounded by 1, with equality achieved for signals best matched to $\mathbf{M}$ as discussed in Section III-A. Using this notation, Theorem III.1 allows us to immediately formulate a version of the JL lemma appropriate for DBD matrices.

Theorem IV.1. *Let U, V be two finite subsets of $\mathbb{R}^{NJ}$, let Φ be a randomly generated DBD matrix as described in Theorem III.1 with measurement allocation $\mathbf{M}$, and let $\rho \in (0,1)$ be fixed. If*

$$\sum_{j=1}^J M_j \geq \frac{\log|U| + \log|V| + \log(2/\rho)}{C_1 \min\left(\frac{C_2^2\delta^2}{\|\phi\|_{\psi_2}^4} \min_{u\in U, v\in V} \widetilde{\Gamma}_2(u-v,\mathbf{M}), \frac{C_2\delta}{\|\phi\|_{\psi_2}^2} \min_{u\in U, v\in V} \widetilde{\Gamma}_\infty(u-v,\mathbf{M})\right)}, \tag{10}$$

then with probability exceeding $1-\rho$, Φ will provide a stable embedding of (U,V) with conditioning δ. Alternatively, under the same conditions, with probability exceeding $1-\rho$ the matrix Φ will provide a stable embedding of (U,V) with conditioning

$$\widetilde{\delta}(U,V,\mathbf{M},\rho) := \frac{\|\phi\|_{\psi_2}^2}{C_2} \max\left(\sqrt{\frac{\log|U| + \log|V| + \log(2/\rho)}{C_1 \min_{u\in U, v\in V} \Gamma_2(u-v,\mathbf{M})}}, \frac{\log|U| + \log|V| + \log(2/\rho)}{C_1 \min_{u\in U, v\in V} \Gamma_\infty(u-v,\mathbf{M})}\right). \tag{11}$$

Proof: Taking the union bound over all $u \in U$ and $v \in V$ and using (5), we then know that the desired (near) isometry holds over all difference vectors $u - v$ except with probability bounded by

$$2|U||V| \exp\left(-C_1 \min\left(\frac{C_2^2\delta^2}{\|\phi\|_{\psi_2}^4} \min_{u\in U, v\in V} \Gamma_2(u-v,\mathbf{M}), \frac{C_2\delta}{\|\phi\|_{\psi_2}^2} \min_{u\in U, v\in V} \Gamma_\infty(u-v,\mathbf{M})\right)\right). \tag{12}$$

Ensuring that (10) holds guarantees that the above failure probability is less than ρ. The bound in (11) follows from (10) and the observation that $\min(a\delta^2, b\delta) = c$ implies that $\delta = \max\left(\sqrt{\frac{c}{a}}, \frac{c}{b}\right)$. ∎

Similar theorems can be formulated for RBD matrices, as well as for stable embeddings of a signal subspace rather than just a finite family of signals. Equation (10) gives a lower bound on the total number of measurements to guarantee a stable embedding with conditioning δ. One can see that the denominator on the right-hand side will scale with $\delta^2 \cdot \min_{u\in U, v\in V} \widetilde{\Gamma}_2(u-v,\mathbf{M})$ when $0 \leq \delta \leq \frac{\|\phi\|_{\psi_2}^2 \min_{u\in U, v\in V} \widetilde{\Gamma}_\infty(u-v,\mathbf{M})}{C_2 \min_{u\in U, v\in V} \widetilde{\Gamma}_2(u-v,\mathbf{M})}$ and with $\delta \cdot \min_{u\in U, v\in V} \widetilde{\Gamma}_\infty(u-v,\mathbf{M})$ when $\delta > \frac{\|\phi\|_{\psi_2}^2 \min_{u\in U, v\in V} \widetilde{\Gamma}_\infty(u-v,\mathbf{M})}{C_2 \min_{u\in U, v\in V} \widetilde{\Gamma}_2(u-v,\mathbf{M})}$. Thus, focusing just on cases where δ is small, in order to guarantee a stable embedding with a moderate number of measurements, we require $\widetilde{\Gamma}_2(u-v,\mathbf{M})$ to be sufficiently close to 1 over all $u \in U$ and $v \in V$. Equivalently, if $\Gamma_2(u-v,\mathbf{M})$ is uniformly close to $\sum M_j$ over all $u \in U$ and $v \in V$, the conditioning $\widetilde{\delta}$ provided in (11) is comparable to what would be achieved with a dense i.i.d. random matrix of the same size. In Section V, we provide several examples of signal classes of U and V for which it may be reasonable to expect such uniformly favorable Γ_2 (or Λ_2) values.

The attentive reader may notice that the failure probability in (12) is in fact loose, since we have bounded the sum of the individual failure probabilities by $|U|\,|V|$ times the worse case failure probability. Due to the exponential form of these probabilities, however, it seems that the worse case probability—even if it is rare—will typically dominate this sum. Therefore, in most applications we do not expect that it is possible to significantly improve over the bounds provided in (12) and thus (10). Unfortunately, it appears that this fact would forbid the derivation of a sharp RIP bound for block diagonal matrices via the elementary covering arguments mentioned briefly in Section II-C.[4] However, such an RIP result, which would guarantee the stable embedding of an infinite family of signals that are sparse in a particular basis, is not necessary for problems that require embeddings of only finite signal families or appropriate for problems where the signals may not all be sparse in the same basis (e.g., using the RIP to derive embedding guarantees such as Theorem IV.1 could require many more measurements than using a concentration bound).

Indeed, ensuring a stable embedding of even a finite signal class is very useful for guaranteeing the performance of many types of compressive-domain signal inference and processing algorithms. In the following subsections, we explore two canonical tasks (detection and classification) in detail to show how signal characteristics affect one's ability to solve these problems using measurements acquired via a block diagonal matrix. Performing tasks such as these directly in the measurement space not only reduces the data acquisition burden but can also reduce the computational burden far below what is required to solve these problems in the high dimensional ambient signal space.

Before concluding this subsection, we briefly note that there are several other tasks (aside from detection and classification) that can be performed in the measurement space when a block diagonal matrix provides a stable embedding of a finite signal family. For one example, when a block diagonal matrix Φ provides a stable embedding of $(\mathcal{S}, \{x\})$ for some signal database $\mathcal{S}$ and query signal x, it is possible to solve the approximate nearest neighbor problem [16] (finding the closest point in $\mathcal{S}$ to x) in the compressed domain without much loss of precision. Another potential application in compressive signal processing involves a simple compressive-domain linear estimator [22]. When Φ provides a stable embedding of $(\mathcal{L}, \mathcal{X} \cup -\mathcal{X})$ for some sets $\mathcal{L}$ and $\mathcal{X}$, then for any $\ell \in \mathcal{L}$ and $x \in \mathcal{X}$, we can estimate the value of $\langle \ell, x \rangle$ from the measurements $\langle \Phi\ell, \Phi x \rangle$. Signal families $\mathcal{L}$ and $\mathcal{X}$ whose sum and difference vectors $\ell \pm x$ have favorable Γ_2 values will have favorable and predictable estimation performance. Finally, a similar result also discussed in [22] shows that *filtering* vectors in order to separate signal and interference subspaces is possible when the difference vectors between these subspaces are stably embedded by Φ.

B. Signal Detection in the Compressed Domain

While the canonical CS results center mostly around reconstructing signals from compressive measurements, there is a growing interest in forgoing this recovery process and answering certain signal processing questions directly in the compressed domain. One such problem that can be solved is binary detection, where one must decide whether or not a known template signal was present when the noisy compressive measurements were collected [22, 31–33]. In particular, let $s \in \mathbb{R}^{NJ}$ denote a

[4]As an aside, since the original submission of this manuscript, some of the authors (with an additional collaborator) have shown that using tools from the theory of empirical processes [28], it is possible to derive RIP bounds for DBD matrices [30] that are in fact dependent on the basis used for sparse representation of the signals. This does not make the present results obsolete, however: neither our concentration bounds nor our measurement bound (10) follow as a consequence of the RIP result.

known signal, and suppose that from the measurement vector y, we wish to decide between two hypotheses:

$$\mathcal{H}_0 : \ y = z \quad \text{or} \quad \mathcal{H}_1 : \ y = \Phi s + z,$$

where Φ is a known compressive measurement matrix, and z is a vector of i.i.d. zero-mean Gaussian noise with variance σ^2. If one were designing a measurement matrix specifically for the purpose of detecting this signal, then the optimal choice of Φ would be the matched filter, i.e., $\Phi = s^T$. However, implementing a measurement system that is designed specifically for a particular s restricts its capabilities to detecting that signal only, which could require a hardware modification every time s changes. A more generic approach would be to design Φ randomly (perhaps with a block diagonal structure out of necessity or due to efficiency considerations) and then use the acquired measurements y to test for one or more candidate signals s.

Given the measurements y, a Neyman-Pearson (NP) optimal detector [22] maximizes the probability of detection, $P_D := P\{\mathcal{H}_1 \text{ chosen } | \mathcal{H}_1 \text{ is true}\}$, subject to a given limitation on the probability of a false alarm, $P_F = P\{\mathcal{H}_1 \text{ chosen } | \mathcal{H}_0 \text{ is true}\}$. The optimal decision for this problem is made based on whether or not the sufficient statistic $t := y^T \Phi s$ surpasses a threshold κ, i.e., $t \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \kappa$, where κ is chosen to meet the constraint $P_F \leq \alpha$ for a specified α. As can be seen from the analysis in [22], the performance of such a detector depends on $\|\Phi s\|_2$. 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. A stable embedding of any signal the detector may encounter, however, guarantees consistent performance of the detector.

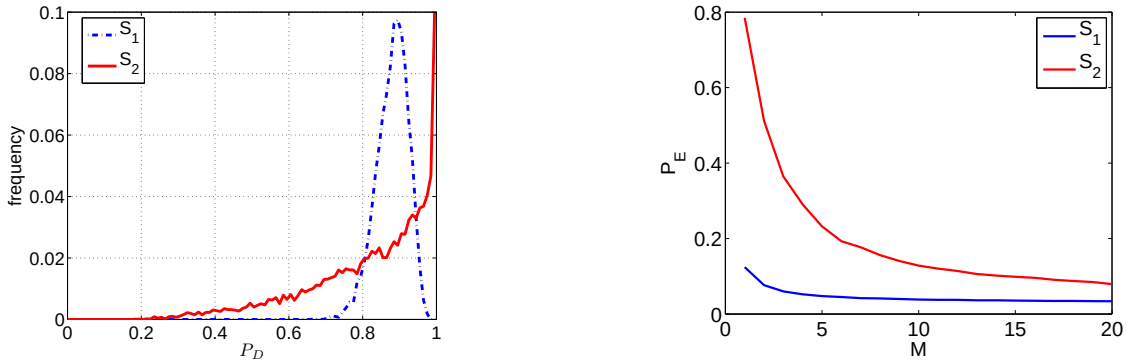
Theorem IV.2. *Suppose $\mathcal{S}$ is a finite set of signals and let Φ be a randomly generated DBD matrix as described in Theorem III.1 with a number of measurements denoted by the matrix $\mathbf{M}$. Fix $0 < \rho < 1$ and pick α such that $P_F \leq \alpha$. Then with probability exceeding $1 - \rho$, any signal $s \in \mathcal{S}$ can be detected with probability of detection bounded by*

$$Q\left(Q^{-1}(\alpha) - \sqrt{1 - \widetilde{\delta}(\mathcal{S}, \{0\}, \mathbf{M}, \rho)}\sqrt{\frac{\|s\|_2^2}{\sigma^2}}\right) \leq P_D(\alpha) \leq Q\left(Q^{-1}(\alpha) - \sqrt{1 + \widetilde{\delta}(\mathcal{S}, \{0\}, \mathbf{M}, \rho)}\sqrt{\frac{\|s\|_2^2}{\sigma^2}}\right),$$

where $Q(\alpha) = \frac{1}{\sqrt{2\pi}} \int_\alpha^\infty e^{-\frac{u^2}{2}} du$ and where $\widetilde{\delta}(\mathcal{S}, \{0\}, \mathbf{M}, \rho)$ is as defined in (11)*.*

The proof of this theorem follows by combining the fact that $P_D(\alpha) = Q\left(Q^{-1}(\alpha) - \frac{\|\Phi s\|_2}{\sigma}\right)$ (see [22]) with (11) and the monotonicity of the function $Q(\cdot)$. While achieving the best possible P_D for a given P_F is of course desirable for a detector, 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 to take action with a certain level of confidence. The theorem above tells us that the consistency of the detector performance is tied to how reliably Φ preserves the norm of signals in $\mathcal{S}$. Examining this relationship, it is clear that more favorable values of $\Gamma_2(s, \mathbf{M})$ for a signal class result in tighter bounds on $P_D(\alpha)$ and therefore in stronger consistency guarantees for the detector.

To illustrate this fact with an example, we create a single DBD measurement matrix $\Phi \in \mathbb{R}^{MJ \times NJ}$ having an equal number of measurements $M_j = M$ per block. We take $M = 4$, $J = 16$ and $N = 64$, and we draw the nonzero entries of Φ as i.i.d. Gaussian random variables with variance $1/M$. We test the detection performance of this matrix by drawing 10000 unit-norm test signals randomly from two classes: $\mathcal{S}_1$, in which signals have uniform energy across their blocks, and $\mathcal{S}_2$, in which signals have decaying energy across their blocks. We choose the noise variance σ^2 such that each test signal s has a constant signal-



(a) (b)

Fig. 3. *(a) Histogram of P_D for 10000 signals with uniform energy across blocks (signal class $\mathcal{S}_1$) and for 10000 signals with decaying energy across blocks (signal class $\mathcal{S}_2$) when measured with a DBD matrix. The compressive NP detector has the constraint $P_F \leq \alpha = 0.1$. (b) Plots of the probability of misclassification P_E over a range of values of $M = 1, \cdots, 20$. The first class of signals $\mathcal{S}_1$ are sparse in the frequency domain. The second class of signals $\mathcal{S}_2$ are nonzero only on a single block in the time domain. While P_E decreases with increasing M for both classes of signals, classification performs better for the signals in $\mathcal{S}_1$, which are more amenable to a stable embedding with a DBD matrix.*

to-noise ratio of $10\log_{10}\left(\frac{\|s\|_2^2}{\sigma^2}\right) = 8$dB. Because of the construction of $\mathcal{S}_1$, $\Gamma_2(s, \mathbf{M})$ attains its maximum value of MJ for all signals $s \in \mathcal{S}_1$, resulting a small conditioning $\widetilde{\delta}$ and a tight bound on P_D. In contrast, $\mathcal{S}_2$ will have a smaller value of $\Gamma_2(s, \mathbf{M})$, resulting in larger values of $\widetilde{\delta}$ and much looser bounds on P_D. We choose the maximum probability of failure to be $\alpha = 0.1$ and use the equation $P_D(\alpha) = Q\left(Q^{-1}(\alpha) - \frac{\|\Phi s\|_2}{\sigma}\right)$ to calculate the expected P_D for each signal.

Figure 3(a) shows the histogram of P_D for the signals in $\mathcal{S}_1$ and $\mathcal{S}_2$ when measured with a DBD matrix. We see that for the uniform energy signals in $\mathcal{S}_1$, the detector performance is indeed tightly clustered around $P_D = 0.9$; one can see that this is the point of concentration predicted by Theorem IV.2 since $Q\left(Q^{-1}(0.1) - \sqrt{10^{8/10}}\right) \approx 0.8907$. Thus for this class of signals, the detector performance is consistent and we can be assured of a favorable P_D when using the detector for all signals in $\mathcal{S}_1$. However, when using the DBD matrix on the signal class $\mathcal{S}_2$, the P_D values are widely spread out compared to those for $\mathcal{S}_1$, despite the fact that the average P_D is nearly the same. Although some individual signals may have above average performance because the measurement matrix happened to amplify their energies, other signals may have very poor performance because the measurement matrix significantly attenuated their energies. Thus this experiment shows how the signal statistics affect the performance reliability in compressive detection tasks when the measurements matrices have block diagonal structure.

C. Classification in the Compressed Domain

Rather than determining the presence or absence of a fixed candidate signal, some scenarios may require the classification of a signal among multiple hypotheses [22, 32]. In particular, let $s_1, s_2, \ldots, s_R \in \mathbb{R}^{NJ}$ denote known signals, and suppose that from the measurement vector y, we wish to decide between R hypotheses:

$$\mathcal{H}_i : y = \Phi s_i + z, \text{ for } i = 1, 2, \ldots, R,$$

where Φ is a known compressive measurement matrix, and z is a vector of i.i.d. zero-mean Gaussian noise with variance σ^2.

It is straightforward to show that when each hypothesis is equally likely, the classifier with minimum probability of error selects the hypothesis that minimizes the sufficient statistic $t_i := \|y - \Phi s_i\|_2^2$. As can be imagined, the performance of such a classifier depends on how well Φ preserves pairwise distances among the signals $\{s_i\}$. If a situation were to occur where

$\frac{\|\Phi s_p - \Phi s_q\|_2}{\|s_p - s_q\|_2}$ was small for some p, q, then s_p could easily be mistaken for s_q in the measurements y. Therefore, having a stable embedding can again be particularly useful for guaranteeing consistent and predictable performance.

Theorem IV.3. *Let $\mathcal{S}$ denote a fixed set of signals with $|\mathcal{S}| = R < \infty$ and fix $0 < \rho < 1$. Suppose Φ is a randomly generated DBD matrix as described in Theorem III.1 with a number of measurements denoted by the matrix $\mathbf{M}$. Assume we receive the measurements $y = \Phi s_{i^*} + z$ for some $i^* \in \{1, 2, \ldots, R\}$ and $z \sim \mathcal{N}(0, \sigma^2 I)$. Then, with probability at least*

$$1 - \left(\frac{R-1}{2}\right) \exp\left\{ -\frac{d^2\left(1 - \widetilde{\delta}(\mathcal{S}, \mathcal{S}, \mathbf{M}, \rho)\right)}{8\sigma^2} \right\} - 2\rho,$$

we have $i^* = \arg\min_{i \in \{1,\ldots,R\}} t_i$ and thus the signal s_{i^*} can be correctly classified. Here $d := \min_{i,j} \|s_i - s_j\|_2$ denotes the minimum separation among the signals in $\mathcal{S}$ and $\widetilde{\delta}(\mathcal{S}, \mathcal{S}, \mathbf{M}, \rho)$ is as defined in (11)*.*

The proof of this theorem again follows by combining bounds from [22] with (11). From this theorem it follows that, if Φ is a block diagonal matrix, signal families $\mathcal{S}$ whose difference vectors $s_p - s_q$ have favorable Γ_2 values will have consistent and predictable classification performance.

The following simulation demonstrates the potential for predictable classification of signals acquired using compressive block diagonal matrices. We again consider DBD matrices having an equal number of measurements $M_j = M$ per block, and we consider signals having $J = 16$ blocks of length $N = 64$. We first create a favorable class of unit-norm signals $\mathcal{S}_1$ with $R = J$ elements such that each signal has just 4 nonzero DFT coefficients at randomly chosen frequencies. To ensure that the signals are real, we restrict the coefficients on conjugate pairs of frequencies to have complex conjugate values. We also ensure that no frequencies are repeated amongst the signals in $\mathcal{S}_1$. As we show in Section V-B, frequency sparse signals with randomly selected frequency support will have large Γ_2 values with high probability; therefore the difference of any two signals from $\mathcal{S}_1$ will also have a large Γ_2 value with high probability. We also create a second class of unit-norm signals $\mathcal{S}_2$ with $R = J$ elements such that each signal s_r for $r = 1, 2, \ldots, R$ has nonzero (and randomly selected) values only on its r-th block and is zero everywhere else. Difference signals from this class will have small Γ_2 values since their energies across the blocks are not uniform.

For each M ranging from 1 to 20, we create 1000 instances of a random DBD matrix Φ of size $MJ \times NJ$. For each Φ and for each signal class $\mathcal{S}_1$ and $\mathcal{S}_2$, we identify the indices i_1, i_2 that minimize $\|\Phi s_{i_1} - \Phi s_{i_2}\|_2$. The signals corresponding to s_{i_1} and s_{i_2} will be among the most difficult to classify since they each have a close neighbor (either Φs_{i_2} or Φs_{i_1}, respectively) after projection by Φ. Then for each of these signals $\{s_{i_j}\}_{j=1,2}$, we create 1000 noisy measurement vectors $y = \Phi s_{i_j} + z$ with $z \sim \mathcal{N}(0, \sigma^2 I)$ and with σ chosen such that $10 \log_{10}\left(\frac{d^2}{\sigma^2}\right) = 15$dB. Finally, we let $p = \arg\min t_i$ be the output of the classifier and calculate the probability of misclassification, $P_E(M)$, for each M as the proportion of occurrences of $p \neq i_1$ or $p \neq i_2$, respectively, over the combined 1000 instances of noise z and 1000 instances of Φ.

Figure 3(b) plots $P_E(M)$ for both classes of signals. As expected, the curve for $\mathcal{S}_1$ is lower than that for $\mathcal{S}_2$ since the signals in $\mathcal{S}_1$ are expected to have a stable embedding with a tighter conditioning. Both curves also show a decreasing trend for increasing M (although it is much more obvious for signal class $\mathcal{S}_2$) as should be expected. Lastly, we see that $P_E(M)$ saturates at a certain level with increasing M. This is also predicted by Theorem IV.3, where the smallest upper bound that can

be provided for P_E is given by $\frac{R-1}{2}\exp\left(-\frac{d^2}{8\sigma^2}\right) > 0$. With the parameters used in this experiment, it can be calculated that the smallest theoretical upper bound for P_E is approximately 0.144. This shows that Theorem IV.3 may be slightly pessimistic.

V. Favorable Signal Classes

The various compressive signal processing guarantees presented in Section IV are built upon the premise that a DBD matrix provides a stable embedding of the signals of interest; as we have noted, these arguments can be extended to RBD matrices as well. Our analysis has also indicated that such stable embeddings are most easily realized with matrices that are well matched to the energy distribution (and sometimes orthogonality) of the signal blocks. In many applications, however—perhaps for the sake of generality, or because little advance knowledge of the signals is available—it may be most natural to use a fixed and equal allocation of measurements. Fortunately, there are a number of interesting signal families (and, in some cases, the corresponding difference signals) that provide favorable Γ_2 values for "uniform" DBD matrices where all M_j are equal (for all j, we suppose $M_j = M$ for some M) and in some cases also provide favorable Λ_2 values for RBD matrices. In this section we survey several such examples.

A. Delay Networks and Multiview Imaging

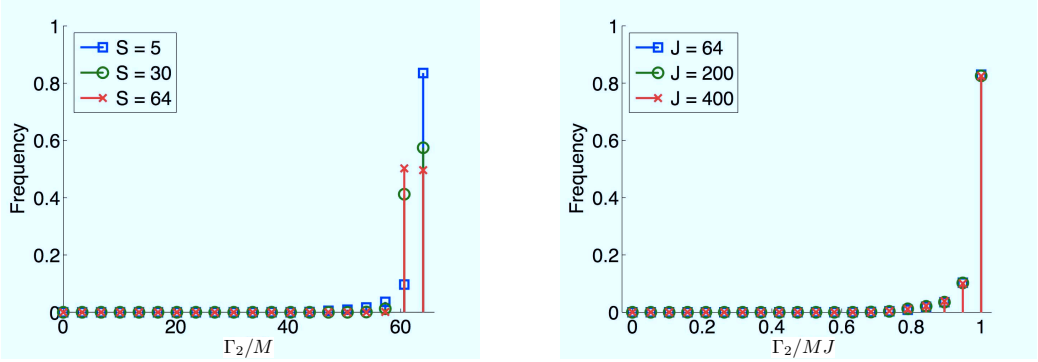
One favorable signal class for uniform DBD and RBD matrices can occasionally arise in certain sensor network or multi-view imaging scenarios where signals with steeply decaying autocorrelation functions are measured under small perturbations. Consider for example a distributed sensor network of J sensors where we would like to detect the presence of a known signal given the observations from each sensor. Due to limited resources, each sensor uses random measurement matrices $\Phi_1, \Phi_2, \ldots, \Phi_J$ to efficiently capture the underlying information with only a few random projections. Suppose that the received signals $x_1, x_2, \ldots, x_J \in \mathbb{R}^N$ represent observations of some common known underlying prototype signal $w \in \mathbb{R}^N$. 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, $x_j(n) = w(n-\delta_j)$. Then, denoting the measurements at sensor j as $y_j = \Phi_j x_j$ it is straightforward to see that the overall system of equations can be represented with a DBD matrix, or when $\Phi_1 = \Phi_2 = \cdots = \Phi_J$ with an RBD matrix.

Assuming w is suitably zero-padded so that border and truncation artifacts can be neglected, we will have $\|x_j\|_2 = \|w\|_2$ for all x_j; this gives $\Gamma_2([x_1^T\ x_2^T\ \cdots x_J^T]^T, \mathbf{M}) = MJ$, which is the ideal case for observation with a uniform DBD matrix. This suggests that the outputs from distributed network systems can be highly amenable to the sort of compressive-domain signal inference and processing tasks described Section IV.

Moreover, the correlations among the components x_j can be characterized in terms of the autocorrelation function R_w of w: we will have $\langle x_i, x_j\rangle = \sum_{n=1}^N x_i(n)x_j(n) = \sum_{n=1}^N w(n-\delta_i)w(n-\delta_j)$, which neglecting border and truncation artifacts will simply equal $R_w(|\delta_i - \delta_j|)$. Therefore, signals w that exhibit strong decay in their autocorrelation function will be natural candidates for observation with RBD matrices as well. For example, equation (9) gives

$$\Lambda_2([x_1^T\ x_2^T\ \cdots x_J^T]^T, M) = \frac{MJ^2\|w\|_2^4}{J\|w\|_2^4 + 2\sum_{i>j} R_w(|\delta_i - \delta_j|)^2}.$$

When $R_w(|\delta_i - \delta_j|)$ is small for most i and j, the quantity $\Lambda_2([x_1^T\ x_2^T\ \cdots x_J^T]^T, M)$ is near its optimal value of MJ.



(a) (b)

Fig. 4. *Histograms of the normalized quantity Γ_2 for frequency sparse signals. (a) The distribution of $\frac{\Gamma_2}{M}$ 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 $\frac{\Gamma_2}{M}$ accumulates near its upper bound of $J = 64$ for all three sparsity levels. (b) The distribution of $\frac{\Gamma_2}{MJ}$ for randomly generated frequency sparse signals with $S = 5$ and the number of signal blocks $J \in \{64, 200, 400\}$. Note that $\frac{\Gamma_2}{MJ}$ accumulates near its upper bound of 1.*

B. Frequency Sparse Signals

Signals having sparse frequency spectra arise in many different applications involving communications intelligence systems and RF sensor networks. Based on time-frequency uncertainty principles and the well-known incoherence of sinusoids and the canonical basis (i.e., "spikes and sines") [34, 35], it is natural to expect that most signals that are sparse in the frequency domain should have their energy distributed relatively uniformly across blocks in the time domain. In the following theorem, we make formal the notion that frequency sparse signals are indeed likely to be favorable for measurement via uniform DBD matrices, producing values of $\Gamma_2(x, \mathbf{M})$ within a log factor of its maximum possible.

Theorem V.1. *Let $N, \beta > 1$ be fixed, suppose $N' = NJ > 512$, and let $\mathbf{M} = \mathrm{diag}\{M, M, \ldots, M\}$ be a DBD measurement allocation with M fixed. 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:*[6]

$$\Gamma_2(x, \mathbf{M}) \geq MJ \cdot \min\left\{ \frac{0.0779}{(\beta+1)\log(N')}, \frac{1}{\left(\sqrt{6(\beta+1)\log N'} + \frac{(\log N')^2}{N}\right)^2} \right\}. \tag{13}$$

Proof: See Appendix C.

Note that as N' grows, the lower bound on $\Gamma_2(x, \mathbf{M})$ scales as $\frac{MJ}{\frac{1}{N^2}\log^4(N')}$, which (treating the fixed value N as a constant) is within $\log^4(N')$ of its maximum possible value of MJ. Thus the concentration exponent for *most* frequency sparse signals when measured by a uniform DBD matrix will scale with $\epsilon^2 MJ/\log^4(N')$ for small ϵ. 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 4(a) plots the ratio $\frac{\Gamma_2(x,\mathbf{M})}{M}$, showing that this quantity is indeed near the upper bound of $J = 64$, indicating

[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'.
[6]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.

a favorable energy distribution. This gives support to the fact that the theoretical value of $\Gamma_2(x, \mathbf{M})$ predicted in Theorem V.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 4(b) the distribution of $\frac{\Gamma_2(x,\mathbf{M})}{MJ}$. It is clear that this value concentrates near its upper bound of 1, showing the relative accuracy of the prediction that $\frac{\Gamma_2}{M}$ scales linearly with increasing J. While some of the quantities in Theorem V.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.

Because differences between frequency sparse signals are themselves sparse in the frequency domain, it follows immediately that not only do frequency sparse signals x have favorable $\Gamma_2(x, \mathbf{M})$ values for uniform DBD matrices, but also that most differences $x_1 - x_2$ between frequency sparse signals have favorable $\Gamma_2(x_1 - x_2, \mathbf{M})$ values. Thus, when measured by a uniform DBD matrix, many families of frequency sparse signals are likely to perform favorably and predictably in the compressive signal processing scenarios outlined in Section IV.

Importantly, Theorem V.1 can also allow us to guarantee the stable embedding of certain infinite collections of frequency sparse signals. In particular, for any sparse support Ω on which (13) holds uniformly, one can apply standard covering arguments (as discussed briefly in Section II-C) to guarantee that with a moderate total number of measurements $MJ = O\left(\frac{|\Omega|}{\epsilon^2}\log^4(N')\right)$, a uniform DBD matrix will simultaneously approximately preserve the norm of all frequency signals supported on Ω. This fact allows one to consider compressive-domain interference cancellation (as discussed in Section IV-A and in [22]) from a set of frequency sparse signals, where the set of possible interferers live in a known subspace of frequency sparse signals.

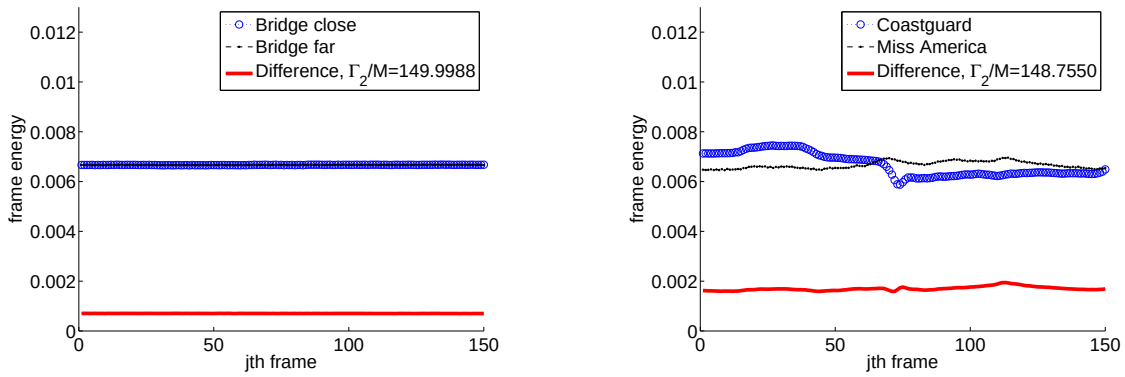
C. Difference Signals

In applications such as classification, we require a stable embedding of difference vectors between signals in a certain signal class. It is interesting to determine what signal families in addition to frequency sparse signals will give rise to difference signals that have favorable values of Γ_2 (for uniform DBD matrices) or Λ_2 (for RBD matrices).

We provide a partial answer to this question by briefly exemplifying a signal family that is favorable for measurement via uniform DBD matrices. 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 if $\mathbf{M} = \mathrm{diag}\{M, M, \ldots, M\}$, then each $x \in Q$ will have $\Gamma_2(x, \mathbf{M}) = MJ$ and thus be highly amenable to measurement by a uniform 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 uniform 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



(a) (b)

Fig. 5. *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* $\Gamma_2(x-y,\mathbf{M})/M = 149.9988$*. (b) The difference of the video pair, "Coastguard" and "Miss America", giving the worst value of* $\Gamma_2(x-y,\mathbf{M})/M = 148.7550$*.*

implication of this bound is that as $a \to 1$, $\Gamma_2(x-y,\mathbf{M}) \to MJ$.

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.[7] 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_2/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_2/M *for each video. The best possible value of* Γ_2/M *is* $J = 150$*.*

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 $\frac{\Gamma_2(x,\mathbf{M})}{M}$ 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 $\frac{\Gamma_2}{M}$ for all possible $\binom{8}{2}$ pairwise difference signals. As we are limited in space, we present in Figure 5 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 $\frac{\Gamma_2(x-y,\mathbf{M})}{M}$. We see that even the smallest $\frac{\Gamma_2}{M}$ value is quite close to the best possible value of $\frac{\Gamma_2}{M} = 150$. 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 uniform DBD matrix.

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

D. Random Signals

Our discussions above have demonstrated that favorable Γ_2 values (for uniform DBD matrices) and Λ_2 values (for RBD matrices) 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 Γ_2 and Λ_2. One way of formalizing this fact is with a probabilistic treatment such as that given in the following lemma.

Lemma V.1. *Let ϕ denote a subgaussian random variable with mean 0, variance σ^2, and subgaussian norm $\|\phi\|_{\psi_2}$, and suppose $x \in \mathbb{R}^{NJ}$ is populated with i.i.d. realizations of ϕ. Let $\mathbf{M} = \operatorname{diag}\{M, M, \ldots, M\}$ with M fixed. Pick $\epsilon \leq \frac{\|\phi\|_{\psi_2}^2}{C_2}$ and suppose that $J \leq \frac{C_1 C_2^2 N \epsilon^2}{2\|\phi\|_{\psi_2}^4 \log(12/\epsilon)}$, where C_1, C_2 are absolute constants as given in Theorem III.1. Then, with probability at least $1 - 2\exp\left(-\frac{1}{2}\frac{C_1 C_2^2 N \epsilon^2}{\|\phi\|_{\psi_2}^4}\right)$, we have*

$$\Gamma_2(x, \mathbf{M}) \geq \Lambda_2(x, M) \geq M + M\left(\frac{1-\epsilon}{1+\epsilon}\right)^2 (J-1).$$

Proof: See Appendix D.

We see from Lemma V.1 that when random vectors are partitioned into a sufficiently small number of blocks, these signals will have their norms preserved giving rise to $\Gamma_2(x, \mathbf{M})$ and $\Lambda_2(x, M)$ values close to their optimal value of MJ. To give some illustrative numbers, numerical simulations showed that over 10000 random draws of Gaussian i.i.d. signals with $J = 16$ and $N = 64$ the average value of $\Gamma_2(x, \mathbf{M})/M$ was 15.5 and the average value of $\Lambda_2(x, M)/M$ was 12.6, which are both large fractions of the maximum possible value of 16. We also note that by using the same argument we can show that the differences of random signals will exhibit large Γ_2 and Λ_2 values. 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 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.

VI. Conclusion

In this paper we have derived concentration of measure inequalities for compressive DBD and RBD matrices. Our experimental results confirm what our theoretical bounds suggest: that the actual probability of concentration depends on the degree of alignment between the allocation of the measurements and the energy distribution (and sometimes orthogonality) of the signal blocks. However, in situations where one can optimize the measurement allocation in anticipation of certain signal characteristics or where a fixed system may be measuring certain favorable classes of signals, we have shown that the highly structured DBD and RBD matrices can provide concentration performance that is on par with the dense i.i.d. matrices often used in CS. We have highlighted a number of compressive signal processing applications that benefit from having a stable embedding of a finite signal family, and we have presented a modified JL lemma for block diagonal matrices that reflects the number of measurements required to ensure such a stable embedding. Finally, we have surveyed a number of signal classes whose blockwise energy distribution and/or orthogonality makes them well-suited to measurement via uniform DBD matrices

or via RBD matrices. Despite not leading to state-of-the-art RIP bounds, we conclude that our nonuniform concentration results can provide a valuable tool for understanding and possibly mitigating the potential pitfalls of working with highly constrained block diagonal matrices.

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), or whether strong RIP results can be established for RBD matrices. 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. Additionally, it would be interesting to examine whether the concentration of measure result for RBD matrices could prove useful in the analysis in the multiple measurement vector (MMV) problem [36, 37] that arises, for example, in array signal processing.

As a final note, we briefly mention that our concentration bounds for block diagonal matrices can actually be useful for studying certain *other* types of structured matrices that arise in linear systems applications. In particular, these results can be applied to derive concentration bounds and RIP results for compressive Toeplitz matrices that arise in problems such as channel sensing and for compressive observability matrices that arise in the analysis of linear dynamical systems. Although space limitations prevent us from detailing these results here, we refer the interested reader to [38, 39] for more information.

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 $\left|\|y\|_2^2 - \|x\|_2^2\right| > \epsilon\|x\|_2^2$. Since $\mathbf{E}\|y\|_2^2 = \|x\|_2^2$, this is equivalent to the condition that $\left|\|y\|_2^2 - \mathbf{E}\|y\|_2^2\right| > \epsilon\mathbf{E}\|y\|_2^2$. 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 subgaussian norms equal to $\|\frac{\phi}{\sqrt{M_j}}\|_{\psi_2} = \frac{\|\phi\|_{\psi_2}}{\sqrt{M_j}}$. 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. As with Gaussian random variables, linear combinations of i.i.d. subgaussian random variables are also subgaussian. In particular, from [40, Lemma 9] it follows that each $y_j(i)$ is a subgaussian random variable with subgaussian norm $\|y_j(i)\|_{\psi_2} \leq c_1 \frac{\|\phi\|_{\psi_2}}{\sqrt{M_j}}\|x_j\|_2$, where c_1 is an absolute constant.

In order to obtain a concentration bound for $\|y\|_2^2$, we require the following important theorem regarding sums of squares of subgaussian random variables.

Theorem A.1. [40] *Let $X_1, \ldots, X_L$ be independent subgaussian random variables with subgaussian norms $\|X_i\|_{\psi_2}$ for all*

$i = 1, \ldots, L$. *Let* $T = \max_i \|X_i\|_{\psi_2}^2$. *Then for every* $t \geq 0$ *and every* $a = (a_1, \ldots, a_L) \in \mathbb{R}^L$, *we have*

$$P\left\{\left|\sum_{i=1}^{L} a_i(X_i^2 - \mathbf{E}X_i^2)\right| \geq t\right\} \leq 2\exp\left\{-C_1 \min\left(\frac{t^2}{16T^2\|a\|_2^2}, \frac{t}{4T\|a\|_\infty}\right)\right\},$$

where $C_1 > 0$ *is an absolute constant.*

Proof: From [40, Lemma 14], we know that each $X_1^2, \ldots, X_L^2$ is a subexponential random variable with subexponential norm $\|X_i^2\|_{\psi_1} \leq 2\|X_i\|_{\psi_2}^2$. For each $i = 1, 2, \ldots, L$, we define $Y_i = X_i^2 - \mathbf{E}X_i^2$ is a zero-mean subexponential random, and from [40, Remark 18], it follows that $\|Y_i\|_{\psi_1} \leq 2\|X_i^2\|_{\psi_1}$. The theorem follows by applying [40, Proposition 16] to the sum $\sum_{i=1}^{L} a_i Y_i$ with $K = 4T \geq \max_i \|Y_i\|_{\psi_1}$. ■

Now, let us define $\widetilde{y}_j(i) := \frac{y_j(i)}{\|y_j(i)\|_{\psi_2}}$ so that $\|\widetilde{y}_j(i)\|_{\psi_2} = 1$, and note that

$$P\left(\left|\|y\|_2^2 - \mathbf{E}\|y\|_2^2\right| > \epsilon\|x\|_2^2\right) = P\left(\left|\sum_j \sum_i \|y_j(i)\|_{\psi_2}^2 \left(\widetilde{y}_j^2(i) - \mathbf{E}\widetilde{y}_j^2(i)\right)\right| > \epsilon\|x\|_2^2\right).$$

We apply Theorem A.1 to the subgaussian random variables $\widetilde{y}_j(i)$ (over all i, j) with weights $a_j(i) = \|y_j(i)\|_{\psi_2}^2$. Letting a denote a vector of length $\sum_j M_j$ containing these weights, we have that $\|a\|_2^2 = \sum_j \sum_i a_j^2(i) = \sum_j \sum_i \|y_j(i)\|_{\psi_2}^4 \leq c_1^4\|\phi\|_{\psi_2}^4 \sum_j \sum_i \|x_j\|_2^4/M_j^2 = c_1^4\|\phi\|_{\psi_2}^4 \sum_j \|x_j\|_2^4/M_j = c_1^4\|\phi\|_{\psi_2}^4\|\mathbf{M}^{-1/2}\gamma\|_2^2$ and $\|a\|_\infty = \max_{i,j} a_j(i) = \max_{i,j} \|y_j(i)\|_{\psi_2}^2 \leq c_1^2\|\phi\|_{\psi_2}^2 \max_j \|x_j\|_2^2/M_j = c_1^2\|\phi\|_{\psi_2}^2\|\mathbf{M}^{-1}\gamma\|_\infty$. Further note that $\|x\|_2^2 = \|\gamma\|_1$ and $\|x\|_2^4 = \|\gamma\|_1^2$. We complete the proof by substituting these quantities into Theorem A.1 with $T = 1$ and $t = \epsilon\|x\|_2^2$ and by taking $C_2 = \frac{1}{4c_1^2}$. ■

APPENDIX B: PROOF OF THEOREM III.2

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

Lemma B.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* $A = X^T X$ *with* X *defined in (6).*

Proof of Lemma B.1: $\|y\|_2^2 = x^T\Phi^T\Phi x = \sum_{j=1}^{J} x_j^T\widetilde{\Phi}^T\widetilde{\Phi}x_j = \sum_{j=1}^{J} x_j^T\left(\sum_{i=1}^{M}\phi_i\phi_i^T\right)x_j = \sum_{i=1}^{M}\phi_i^T\left(\sum_{j=1}^{J}x_j x_j^T\right)\phi_i = \sum_{i=1}^{M}\phi_i^T A\phi_i$. ■

Lemma B.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 B.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 B.1 to y and Lemma B.2 with $z = \phi_i$ for each $i = 1, 2, \ldots, M$, we have $\|y\|_2^2 = \sum_{i=1}^{M}\phi_i^T A\phi_i = \sum_{i=1}^{M}\sum_{j=1}^{N}\frac{\lambda_j}{M}w_{i,j}^2$,

where each $\{w_{i,j}\}_{i,j}$ is an independent Gaussian random variable with zero mean and unit variance. After switching the order of the summations and observing that $\mathrm{Tr}(X^TX) = \mathrm{Tr}(XX^T)$ where $\mathrm{Tr}(\cdot)$ is the trace operator, we have $\mathbf{E}\|y\|_2^2 = \sum_{j=1}^N \frac{\lambda_j}{M} \sum_{i=1}^M \mathbf{E}w_{i,j}^2 = \sum_{j=1}^N \lambda_j = \mathrm{Tr}(XX^T) = \|x\|_2^2$.

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 $\left|\|y\|_2^2 - \mathbf{E}\|y\|_2^2\right| > \epsilon\mathbf{E}\|y\|_2^2$. We again apply Theorem A.1 to establish a concentration result. To do so, note that each $w_{i,j}$ is a subgaussian random variable with the same subgaussian norm $\|w\|_{\psi_2} := \|w_{i,j}\|_{\psi_2}$; because these variables are Gaussian with unit variance, it is also known [40] that there exists an absolute constant c_2 such that $\|w\|_{\psi_2} \leq c_2$. Let us define $\widetilde{w}_{i,j} := \frac{w_{i,j}}{\|w\|_{\psi_2}}$ so that $\|\widetilde{w}_{i,j}\|_{\psi_2} = 1$, and note that

$$P(|\|y\|_2^2 - \mathbf{E}\|y\|_2^2| > \epsilon\|x\|_2^2) = P\left(\left|\sum_j \sum_i \frac{\|w\|_{\psi_2}^2}{M}\lambda_j(\widetilde{w}_{i,j}^2 - \mathbf{E}\widetilde{w}_{i,j}^2)\right| > \epsilon\|x\|_2^2\right).$$

We apply Theorem A.1 to the subgaussian random variables $\widetilde{w}_{i,j}$ (over all i,j) with weights $a_j(i) = \frac{\|w\|_{\psi_2}^2}{M}\lambda_j$. Letting a denote a vector of length MJ containing these weights, we have that $\|a\|_2^2 = \sum_j \sum_i a_j^2(i) = \sum_j \sum_i \frac{\|w\|_{\psi_2}^4}{M^2}\lambda_j^2 = \frac{\|w\|_{\psi_2}^4}{M}\sum_j \lambda_j^2 = \frac{\|w\|_{\psi_2}^4}{M}\|\lambda\|_2^2 \leq \frac{c_2^4}{M}\|\lambda\|_2^2$ and $\|a\|_\infty = \max_{i,j} a_j(i) = \frac{\|w\|_{\psi_2}^2}{M}\max_j \lambda_j = \frac{\|w\|_{\psi_2}^2}{M}\|\lambda\|_\infty \leq \frac{c_2^2}{M}\|\lambda\|_\infty$. Note that $\|x\|_2^2 = \mathrm{Tr}(X^TX) = \|\lambda\|_1$ and $\|x\|_2^4 = \|\lambda\|_1^2$ since the eigenvalues $\{\lambda_j\}_{j=1}^N$ are non-negative. We complete the proof by substituting these quantities into Theorem A.1 with $T = 1$ and $t = \epsilon\|x\|_2^2$ and by taking $C_3 = \frac{1}{4c_2^2}$. ■

APPENDIX C: PROOF OF THEOREM V.1

Our result follows from an application of the following.

Theorem C.2. [35, Theorem 3.1] *Let $x \in \mathbb{C}^{N'}$ and $\beta > 1$. Suppose $N' > 512$ and choose N_T and N_Ω such that:*

$$N_T + N_\Omega \leq \frac{0.5583N'/q}{\sqrt{(\beta+1)\log(N')}} \text{ and } N_T + N_\Omega \leq \frac{\sqrt{2/3}N'\left(\frac{1}{q} - \frac{(\log N')^2}{N'}\right)}{\sqrt{(\beta+1)\log(N')}}. \tag{14}$$

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 V.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 C.2 with $N_\Omega = S$ and $N_T = N = N'/J$, being careful to select a value for q such that (14) is satisfied. In particular, we require

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

This is satisfied if we choose

$$q \leq \min\left\{\frac{0.5583N'}{(N+S)\sqrt{(\beta+1)\log N'}}, \frac{N'}{\frac{(N+S)}{\sqrt{2/3}}\sqrt{(\beta+1)\log N'} + (\log N')^2}\right\}. \tag{15}$$

Choosing any q satisfying (15), 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

$k = 1, 2, \ldots, J$, 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, $\frac{\Gamma_2}{M} = \frac{\|\gamma\|_1^2}{\|\gamma\|_2^2} \geq \frac{q^2}{J}$. Combining with (15) and using the fact that $S < N$, we thus have:

$$\begin{aligned}
\frac{\Gamma_2}{M} &\geq \min\left\{ \frac{0.5583^2 N^2 J}{(N+S)^2(\beta+1)\log N'}, \frac{N^2 J}{\left(\frac{(N+S)}{\sqrt{2/3}}\sqrt{(\beta+1)\log N' + (\log N')^2}\right)^2} \right\} \\
&\geq \min\left\{ \frac{(0.5583^2/2^2) J}{(\beta+1)\log(N')}, \frac{J}{\left(\frac{2}{\sqrt{2/3}}\sqrt{(\beta+1)\log N' + \frac{(\log N')^2}{N}}\right)^2} \right\}.
\end{aligned}$$

■

APPENDIX D: PROOF OF LEMMA V.1

Proof: Let X be the $J \times N$ matrix as defined in (6). 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_2 in terms of these extremal eigenvalues by writing

$$\Lambda_2 = \frac{M\|\lambda\|_1^2}{\|\lambda\|_2^2} = M\frac{\sum_i \lambda_i^2 + \sum_i \sum_{j\neq i} \lambda_i \lambda_j}{\sum_i \lambda_i^2} \geq M + M\frac{\lambda_{\min}}{\lambda_{\max}}\frac{\sum_i \sum_{j\neq i} \lambda_i}{\sum_i \lambda_i} = M + M\frac{\lambda_{\min}}{\lambda_{\max}}(J-1). \tag{16}$$

Assume that $\epsilon \leq \frac{\|\phi\|_{\psi_2}^2}{C_2}$, and let us define the following events:

$$A = \left\{ N\sigma^2(1-\epsilon)^2 \leq \frac{\|X^T z\|_2^2}{\|z\|_2^2} \leq N\sigma^2(1+\epsilon)^2,\ \forall z \in \mathbb{R}^J \right\}, \quad B = \left\{\lambda_{\max} \leq N\sigma^2(1+\epsilon)^2\right\} \bigcap \left\{\lambda_{\min} \geq N\sigma^2(1-\epsilon)^2\right\},$$
$$C = \left\{ \frac{\lambda_{\min}}{\lambda_{\max}} \geq \left(\frac{1-\epsilon}{1+\epsilon}\right)^2 \right\}, \qquad D = \left\{ \Lambda_2 \geq M + M\left(\frac{1-\epsilon}{1+\epsilon}\right)^2 (J-1) \right\}.$$

These events satisfy $A = B \subseteq C \subseteq D$, where the last relation follows from (16). 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 $\epsilon \leq \frac{\|\phi\|_{\psi_2}^2}{C_2}$, $P(\left|\|X^T z\|_2^2 - N\sigma^2\|z\|_2^2\right| > \epsilon N\sigma^2\|z\|_2^2) \leq 2\exp\left(-\frac{C_1 C_2^2 N\epsilon^2}{\|\phi\|_{\psi_2}^4}\right)$. Thus, for an upper bound for $P(A^c)$, we can follow the straightforward arguments in [10, Lemma 5.1] and conclude that $P(A^c) \leq 2\left(\frac{12}{\epsilon}\right)^J \exp\left(-\frac{C_1 C_2^2 N\epsilon^2}{\|\phi\|_{\psi_2}^4}\right)$. Thus by choosing $J \leq \frac{C_1 C_2^2 N\epsilon^2}{2\|\phi\|_{\psi_2}^4 \log(12/\epsilon)}$, we see that $P(D^c) \leq 2\exp\left(-\frac{1}{2}\frac{C_1 C_2^2 N\epsilon^2}{\|\phi\|_{\psi_2}^4}\right)$. Finally, the fact that $\Gamma_2 \geq \Lambda_2$ follows from (9). ■

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] V. V. Buldygin and Y. V. Kozachenko, "Sub-gaussian random variables," *Ukrainian Mathematical Journal*, vol. 32, pp. 483–489, Nov. 1980.

[15] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," in *Compressed Sensing: Theory and Applications*. Cambridge University Press, to appear.

[16] 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.

[17] 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.

[18] 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.

[19] N. Ailon and B. Chazelle, "The fast Johnson-Lindenstrauss transform and approximate nearest neighbors," *SIAM J. Comput.*, vol. 39, pp. 302–322, May 2009.

[20] N. Ailon and E. Liberty, "Almost optimal unrestricted fast Johnson-Lindenstrauss transform," *Arxiv preprint arXiv:1005.5513*, 2010.

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

[22] 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.

[23] 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.

[24] R. Adamczak, A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann, "Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling," *Arxiv preprint arXiv:0904.4723*, 2009.

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

[26] E.J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406 –5425, 2006.

[27] O. Guédon, S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Subspaces and orthogonal decompositions generated by bounded orthogonal systems," *Positivity*, vol. 11, no. 2, pp. 269–283, 2007.

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

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

[30] H. L. Yap, A. Eftekhari, M. B. Wakin, and C. J. Rozell, "The restricted isometry property for block diagonal matrices," in *Proc. Conf. Information Sciences and Systems (CISS)*, March 2011.

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

[32] 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.

[33] 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.

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

[35] 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.

[36] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 505–519, 2009.

[37] M. E. Davies and Y. C. Eldar, "Rank awareness in joint sparse recovery," *Arxiv preprint arXiv:1004.4529*, 2010.

[38] H. L. Yap and C. J. Rozell, "On the relation between block diagonal matrices and compressive Toeplitz matrices," Technical Report, 2011.

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

[40] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *Arxiv preprint arXiv:1011.3027*, 2010.