CONCENTRATION OF MEASURE FOR BLOCK DIAGONAL MATRICES WITH REPEATED BLOCKS

Christopher J. Rozell and Han Lun Yap
School of Electrical and Computer Engineering, Georgia Institute of Technology

Jae Young Park
Department of Electrical Engineering and Computer Science, University of Michigan

Michael B. Wakin
Division of Engineering, Colorado School of Mines

***Abstract*—The theoretical analysis of randomized compressive operators often relies on the existence of a concentration of measure inequality for the operator of interest. Though commonly studied for unstructured, dense matrices, matrices with more structure are often of interest because they model constraints on the sensing system or allow more efficient system implementations. In this paper we derive a concentration of measure bound for block diagonal matrices where the nonzero entries along the main diagonal are a single repeated block of i.i.d. Gaussian random variables. Our main result states that the concentration exponent, in the best case, scales as that for a fully dense matrix. We also identify the role that the signal diversity plays in distinguishing the best and worst cases. Finally, we illustrate these phenomena with a series of experiments.**

***Index Terms*—Compressive Sensing, concentration of measure, Johnson-Lindenstrauss lemma, block diagonal matrices.**

I. Introduction

Signal processing applications today demand ever greater amounts of high-resolution data. While computing power has mostly been able to support this trend, the same cannot always be said for front-end signal acquisition devices. Recently, there has been a great deal of interest in compressive data acquisition schemes where the high resolution data $x \in \mathbb{R}^N$ is acquired using a compressive linear operator $\Phi : \mathbb{R}^N \to \mathbb{R}^M$. In such schemes, the data x is thought to be *sparse* (i.e., it has few nonzero coefficients in the time domain or in some transform basis) and if the number of measurements M is sufficient, we can recover the signal x by solving a convex optimization program [1], [2].

Randomized designs for compressive linear operators are of particular interest. One reason is that the resulting operators work universally with different signal classes. For example, when a linear operator represented by an $M \times N$ matrix has its entries made up of independent and identically distributed (i.i.d.) Gaussian random variables, it can be used to acquire signals x that are sparse in most bases.

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 theoretical analysis of such randomized operators often relies on the phenomenon known as concentration of measure [3]. Put simply, for any fixed signal $x \in \mathbb{R}^N$ a suitable randomized Φ will approximately preserve the norm of x with high probability. More formally, a typical result is the following:[1]

Lemma I.1. [4] *Let Φ be an $M \times N$ matrix with i.i.d. Gaussian entries having variance $\frac{1}{M}$. Then there exists a constant $C > 0$ such that, for any $\epsilon \in (0, 1)$,*

$$P(\left|\|\Phi x\|_2^2 - \|x\|_2^2\right| > \epsilon \|x\|_2^2) \leq 2\exp(-CM\epsilon^2). \quad (1)$$

For example, concentration of measure results such as (1) lead to the Johnson-Lindenstrauss (JL) lemma [5], which states that for a finite cloud of points $Q \subset \mathbb{R}^N$,

$$(1-\epsilon)\|u-v\|_2 \leq \|\Phi(u-v)\|_2 \leq (1+\epsilon)\|u-v\|_2 \quad (2)$$

holds for all $u, v \in Q$ with high probability supposing that $M = O(\log(|Q|)\epsilon^{-2})$. Concentration of measure results can also lead to the Restricted Isometry Property (RIP) [6] in Compressive Sensing (CS), which states that (2) holds for all pairs u, v of K-sparse signals in $\mathbb{R}^N$ with high probability supposing that $M = O(K\log(N/K)\epsilon^{-2})$ [4], [7].

However, the randomized operators that we have discussed in Lemma I.1 are dense, implying that the acquisition process must form each measurement as a weighted linear combination of all of the data. If N is large, the storage and use of such operators in any practical system will be computationally expensive. Additionally, many acquisition systems may have architectural constraints that restrict how the measurements can be formed from the data. For one example, in a distributed system, communication constraints may limit each measurement to being a weighted linear combination of only a subset of the data. Furthermore, to reduce the customization necessary for such a distributed system, it may be necessary for the measurements to be formed by applying the same weightings to different subsets of the data. For a second

[1] Lemma I.1 follows from Theorem II.2 by setting $J = 1$.

example, there has recently been interest in measuring x by convolving it with a random pulse and downsampling. These convolution measurement systems lead to computationally efficient designs and they have been shown to work almost as well as dense randomized operators [8], [9]. For a third example, we may have streaming data which can only be accessed in chunks and for the sake of efficiency we may need to apply the same measurement operator to each block.

In each of these cases, the common theme is that rather than acquiring the whole data vector via one dense measurement system, the data can be divided naturally into discrete subsections, and we may acquire this data via repeated application of a single (possibly unchanging) measurement system to each of these subsections. In these types of settings (for example, with a convolution system where the downsampling interval is the length of the random pulse), we may envision the signal $x \in \mathbb{R}^{NJ}$ as representing the concatenation $x = [x_1^T \; x_2^T \; \cdots \; x_J^T]^T$ of J component signals $\{x_j\}_{j=1}^J \subset \mathbb{R}^N$. The measurement vector $y \in \mathbb{R}^{MJ}$ thus is the concatenation $y = [y_1^T \; y_2^T \; \cdots \; y_J^T]^T$ of the vectors $y_j = \Phi_j x_j$, where Φ_j is a matrix of size $M \times N$ that measures only signal j. It follows that $y = \Phi x$, where the resulting $MJ \times NJ$ measurement matrix Φ is zero everywhere except for blocks along the main diagonal:

$$\Phi = \begin{pmatrix} \Phi_1 & & \\ & \ddots & \\ & & \Phi_J \end{pmatrix}. \tag{3}$$

In a previous paper [10] we derived a concentration of measure bound for such matrices where the nonzero entries of the block diagonal matrix are i.i.d. subgaussian random variables.[2] In this paper, we will analyze the case when the block diagonals are repeated: $\Phi_1 = \Phi_2 = \cdots = \Phi_J = \widetilde{\Phi}$, and $\widetilde{\Phi}$ is now an i.i.d. Gaussian random matrix. Our main result, detailed in Theorem II.2, essentially states that the probability of concentration behaves as

$$P(\left|\|y\|_2^2 - \|x\|_2^2\right| > \epsilon\|x\|_2^2) \leq 2\exp(-CM\Lambda\epsilon^2),$$

where $\Lambda = \Lambda(x)$ is a term that depends on the "diversity" of the signal components of x (we will explain the meaning of "diversity" in Section II). At one extreme, when the component signals $\{x_j\}$ of x are mutually orthogonal and have equal norm, we have $\Lambda(x) = J$ and the concentration of measure exponent scales exactly as in (1) for a fully dense random matrix; that is, it scales with the total number of measurements, which in this case equals MJ. At the other extreme, however, when the signal components are all the same modulo a constant, we have $\Lambda(x) = 1$ and the measurement operator effectiveness is diminished.

II. Main result

A concentration of measure result such as Lemma I.1 can be viewed as a statement about the tail probability bound of a random variable. As such, we require the following established results regarding the tail distribution of random variables.

Lemma II.1. [11] *Suppose that w is a Gaussian random variable with variance σ^2. Then*

$$P(|w|^2 > t) \leq 2\exp\left(-\frac{t}{2\sigma^2}\right) \text{ for all } t \geq 0.$$

This property of individual Gaussian random variables will allow us to use the following important theorem characterizing the tail distributions of sums of random variables.

Theorem II.1. [12] *Let $R_1, \ldots, R_L$ be independent real valued random variables with $P(|R_i| > t) \leq a\exp(-\alpha_i t)$ for all t and i. Let $d \geq \max_i \alpha_i^{-1}$ and $b \geq a\sum_{i=1}^L \alpha_i^{-2}$. Then setting $S = \sum_{i=1}^L R_i$ we have*

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

Before introducing the main theorem, let us present some notation that will be useful. Let Φ be as in (3), where each $\Phi_j \in \mathbb{R}^{M\times N}$ is the same and equal to $\widetilde{\Phi}$ (i.e., $\Phi_j = \widetilde{\Phi}, j = 1, \ldots, J$). For $i = 1, 2, \ldots, M$, let ϕ_i^T denote row i of $\widetilde{\Phi}$. As before, x is a concatenation of J component signals $\{x_j\}_{j=1}^J \subset \mathbb{R}^N$. Let X denote the $J \times N$ matrix of the component signals such that

$$X = \begin{pmatrix} x_1^T \\ \vdots \\ x_J^T \end{pmatrix}, \tag{4}$$

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

$$\lambda = \lambda(x) := [\lambda_1 \; \lambda_2 \; \ldots \; \lambda_N]^T \in \mathbb{R}^N$$

be the vector made up of these eigenvalues.

Equipped with this notation, we now present our main result.

Theorem II.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 (3), with $\Phi_j = \widetilde{\Phi}, j = 1, \ldots, J$. Then*

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

Note that (5) has two cases, each with roughly the same form as (1) but with the key difference being an extra factor in the rate of concentration that depends on the eigenstructure of the signal being measured.

For purpose of the proof of this theorem, we will require the following two lemmas whose proofs can be found in the appendix.

[2] [11] A random variable w is *subgaussian* if there exists $a \geq 0$ such that $\mathbf{E}\; e^{tw} \leq \exp\left(\frac{1}{2}a^2t^2\right)$ for all $t \in \mathbb{R}$.

Lemma II.2. *Suppose $x \in \mathbb{R}^{NJ}$ and $\widetilde{\Phi}$ is an $M \times N$ matrix where:*

$$\widetilde{\Phi} = \begin{pmatrix} \phi_1^T \\ \vdots \\ \phi_M^T \end{pmatrix}.$$

Let Φ be an $MJ \times NJ$ block diagonal matrix as defined in (3), but with all the Φ_j being the same and equal to $\widetilde{\Phi}$. 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 (4).

Lemma II.3. *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 Theorem II.2:

Let $y = \Phi x$. The proof will proceed in two parts. First we will calculate $\mathbf{E}\|y\|_2^2$ to determine the point of concentration for the norm of the measurements. Then, we will calculate tail bounds on the deviation of the measurement norm from this expected value.

Applying Lemma II.2 to y and Lemma II.3 with $z = \phi_i$ for each $i = 1, \ldots, M$, we have:

$$\|y\|_2^2 = \sum_{i=1}^{M} \phi_i^T A \phi_i = \sum_{i=1}^{M} \sum_{j=1}^{N} \lambda_j w_{i,j}^2$$

where each $\{w_{i,j}\}_{i,j}$ is an independent Gaussian random variable with zero mean and variance σ^2. After switching the order of the summations and observing that $\mathrm{Tr}(X^T X) = \mathrm{Tr}(X X^T)$ where $\mathrm{Tr}(\cdot)$ is the trace operator, we have:

$$\mathbf{E}\|y\|_2^2 = \sum_{j=1}^{N} \lambda_j \sum_{i=1}^{M} \mathbf{E} w_{i,j}^2 = \sum_{j=1}^{N} \lambda_j = \mathrm{Tr}(X X^T) = \|x\|_2^2.$$

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

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

We apply Theorem II.1 for the random variables $\widetilde{w}_{i,j}^2$, for all $i = 1, \ldots, M$ and $j = 1, \ldots, N$, with $a = 2$ and $\alpha_{i,j}^{-1} = 2\lambda_j \sigma^2 = \frac{2}{M} \lambda_j$, to compute the concentration result for $\|y\|_2^2$. Note that $\alpha_{i,j}^{-1}$ is constant for a fixed j. Hence, for $d \geq \max_{i,j} \alpha_{i,j}^{-1} = \frac{2}{M} \max_j \lambda_j$ and $b \geq a \sum_{i,j} \alpha_{i,j}^{-2} = \frac{8}{M} \sum_j \lambda_j^2$,

$$P\left(\left| \|y\|_2^2 - \|x\|_2^2 \right| > \epsilon \|x\|_2^2 \right) \leq \begin{cases} 2\exp\{-\frac{\epsilon^2 \|x\|_2^4}{32b}\}, & 0 \leq \epsilon \leq \frac{4b}{d\|x\|_2^2} \\ 2\exp\{-\frac{\epsilon \|x\|_2^2}{8d}\}, & \epsilon \geq \frac{4b}{d\|x\|_2^2}. \end{cases} \quad (6)$$

Note that $\|x\|_2^2 = \mathrm{Tr}(X^T X) = \|\lambda\|_1$ since the eigenvalues $\{\lambda_j\}_{j=1}^N$ are non-negative and $\|x\|_2^4 = \|\lambda\|_1^2$. Substituting $d = \frac{2}{M} \max_j \lambda_j = \frac{2}{M} \|\lambda\|_\infty$ and $b = \frac{8}{M} \sum_j \lambda_j^2 = \frac{8}{M} \|\lambda\|_2^2$ into (6) completes the proof.

■

To gain intuition about these types of constrained randomized measurement systems we examine the form of (5) more closely. As we will be frequently concerned with applications where ϵ is small, let us consider the first of the two cases given in (5). Define $\Lambda(x) = \frac{\|\lambda\|_1^2}{\|\lambda\|_2^2}$. From the standard relation between l_1 and l_2 norms, for any $x \in \mathbb{R}^{NJ}$,

$$1 \leq \Lambda(x) \leq \min(J, N).$$

The first extreme case ($\Lambda = 1$) is unfavorable as it implies that compared to a full dense Φ of size $MJ \times NJ$ (for which the concentration exponent would scale with MJ), we would diminish the effectiveness of the measurements by a factor of J. This happens when A has only one non-zero eigenvalue. Noting that $A = \sum_j x_j x_j^T$, this case happens when the component signals x_j are the same modulo a scaling factor. In this case, it is understandable that the concentration exponent only scales with the number of measurements per signal block, M, as we are basically measuring the same signal J times.

The other extreme ($\Lambda = \min(J, N)$) case is favorable as long as $J \leq N$, in which case the concentration exponent scales at the same rate as for a full dense Φ of size $MJ \times NJ$. For this case to occur, A must have J non-zero eigenvalues and they must all be equal. By noting that the non-zero eigenvalues of $A = X^T X$ are the same as those of the Grammian matrix $G = X X^T$, the only situation where this is true is when the component signals are orthogonal and of the same l_2 norm.

Before going further, it is interesting to point out that finding the eigenvalues of $A = X^T X$ is equivalent to running Principal Component Analysis (PCA) on the J component signals. If the component signals are of equal energy and are all orthogonal, then PCA would give us J equal eigenvalues whose eigenvectors correspond to the normalized component signals themselves. Now if the component signals are still orthogonal but instead have different energy, then PCA will give us eigenvectors that still correspond to the normalized component signals but the eigenvalues will not be equal anymore. This will result in $1 \leq \Lambda(x) < J$. The same conclusion holds if the component signals are no longer orthogonal but still span a J-dimensional subspace. Now if the J component signals lie on a K-dimensional subspace with $K < J$, then only K eigenvalues will be non-zero and $1 \leq \Lambda(x) \leq K$.

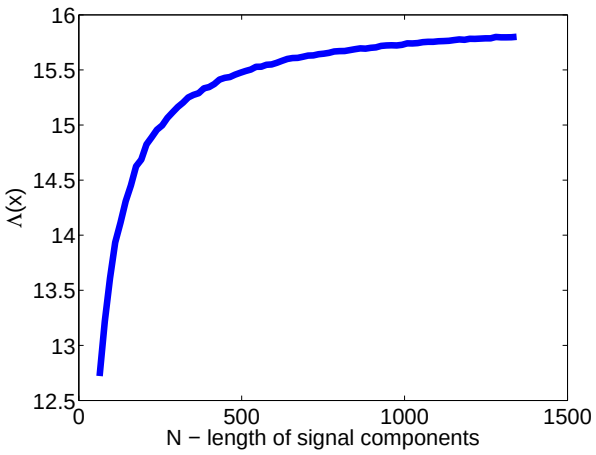

Fig. 1. *The average value of $\Lambda(x)$ (over 100 trials) of a vector $x \in \mathbb{R}^{NJ}$ composed of i.i.d. Gaussian random variables of mean 0 and equal variance as the length N of signal components increases for a fixed number of components $J = 16$.*

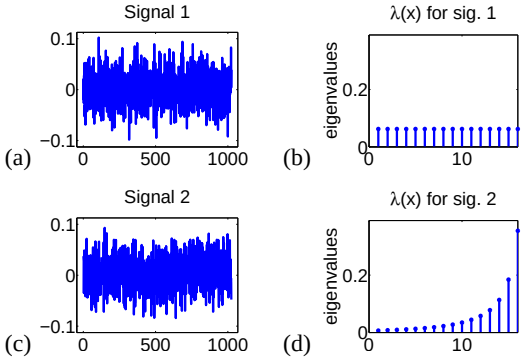

Fig. 2. *(a) Signal 1 with $J = 16$ orthogonal signal components. (b) $\lambda(x)$ for signal 1. (c) Signal 2 with non-orthogonal signal components. (d) $\lambda(x)$ for signal 2.*

A third interesting case worth highlighting is of a class of signals of length NJ whose entries are zero mean i.i.d. Gaussian random variables of equal variance. It is known that if $J \ll N$, any J vectors of length N composed of i.i.d. Gaussian random variables of zero mean and equal variance will form an almost orthogonal set in $\mathbb{R}^N$ [6]. This implies that if the number of signal components J is much less than the length of each signal component N, a signal x from this class will have $\Lambda(x)$ close to J. This phenomenon is depicted in Figure 1, where we experimentally obtained the average $\Lambda(x)$ of random signals x for different signal component length N but all with a fixed number of signal components J. This suggests that a random signal will concentrate almost as well with a block diagonal matrix with repeated blocks (provided that $J \ll N$) as with a full dense measurement matrix.

One final comment about Theorem II.2 is in order regarding the demarcation between the two cases in (5). One can show that $\frac{2(\sqrt{J}-1)}{J-1} \leq \frac{\|\lambda\|_2^2}{\|\lambda\|_\infty \|\lambda\|_1} \leq 1$. For $J \geq 2$, the left hand term obeys $\frac{2(\sqrt{J}-1)}{J-1} \geq \frac{1}{\sqrt{J}}$ implying that the first case in equation (5) is guaranteed to include at least $0 \leq \epsilon \leq \frac{16}{\sqrt{J}}$.

III. Experiments

We will demonstrate the concentration phenomenon by firstly considering two specific signals. The plots we show are typical of our experiments with other similar signals.

To begin, we randomly construct a signal of length 1024 made up of $J = 16$ components, with each signal component of length $N = 64$. Then we perform Gram-Schmidt orthogonalization on the J component vectors to obtain a signal made up of J orthogonal components. The signal x is plotted in Figure 2(a) and the non-zero eigenvalues of $A = X^T X$ are shown in the plot of $\lambda(x)$ in Figure 2(b). Fixing this signal, we next generate a series of random dense Φ matrices having size 64×1024, where all the entries of each matrix are i.i.d. zero mean, Gaussian random variables with variance $\frac{1}{64}$. 10000 of these randomly generated Φ are used to measure this fixed signal x and we plot in Figure 3(a) a histogram of $\frac{\|\Phi x\|_2}{\|x\|_2}$. This quantity should ideally be concentrated around 1.

Using the same fixed signal x, we now consider block diagonal matrices Φ of the form (3) with each of the J blocks being equal to $\widetilde{\Phi}$. The blocks $\widetilde{\Phi}$ are of size $M \times N = 4 \times 64$ and are populated with i.i.d. zero mean Gaussian random variables with variance $\frac{1}{4}$. The resulting matrix Φ is thus of size $MJ \times NJ = 64 \times 1024$, but contains many fewer nonzero entries (and even fewer independently generated random numbers) than the dense measurement matrix. Since our signal has orthogonal signal components, from Theorem II.2 we know that the concentration exponent scales at the same rate as for a full dense Φ of size 64×1024. We plot in Figure 3(b) a histogram of $\frac{\|\Phi x\|_2}{\|x\|_2}$ formed with 10000 randomly generated block diagonal Φ. We see that despite the dramatically reduced complexity of the measurement matrix Φ, the concentration of measure behavior is still identical to that of a fully dense matrix. For each type of matrix, we plot in Figure 4 as a function of ϵ the percentage of trials for which $(1-\epsilon) \leq \frac{\|\Phi x\|_2}{\|x\|_2} \leq (1+\epsilon)$. As expected, the curves for the above two types of matrices are indistinguishable.

We now focus instead on a signal x whose signal components have the same energy but whose $\lambda(x)$ is non-uniform. This signal would have a less desirable concentration of measure when using block diagonal matrices. An example of such a signal is shown in Figure 2(c) and its $\lambda(x)$ is plotted in Figure 2(d). It can be calculated from $\lambda(x)$ that $\Lambda(x) = 5.3341$, which is approximately 3 times less than the desired $J = 16$. Fixing this signal, we plot in Figure 3(c) a histogram of $\frac{\|\Phi x\|_2}{\|x\|_2}$ when the matrices Φ are fully dense. As shown in this histogram and in Figure 4, it is not surprising that the concentration of measure behavior for this signal with fully dense matrices is identical to a signal with orthogonal components.

Again fixing this new signal, we plot in Figure 3(d) a histogram of $\frac{\|\Phi x\|_2}{\|x\|_2}$ where the measurement matrices Φ are block diagonal of size $MJ \times NJ = 64 \times 1024$ with $J = 16$ identical $M \times N = 4 \times 64$ blocks made up of i.i.d. zero

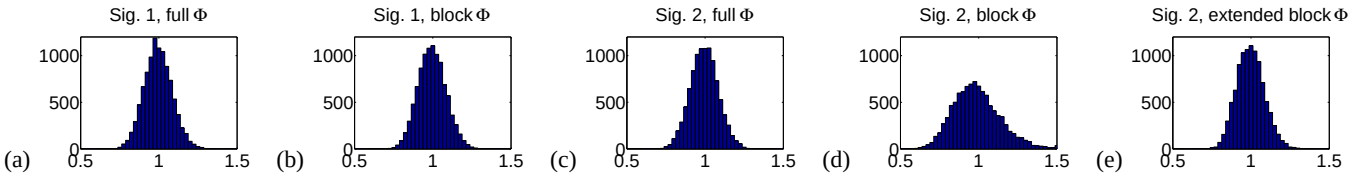

Fig. 3. *Histogram of $\|\Phi x\|_2/\|x\|_2$ for fixed x across 10000 randomly generated matrices Φ. (a) Orthogonal signal components, fully dense 64×1024 Φ. (b) Orthogonal signal components, block diagonal 64×1024 Φ with repeated blocks. (c) Non-orthogonal signal components, fully dense 64×1024 Φ. (d) Non-orthogonal signal components, block diagonal 64×1024 Φ with repeated blocks. (e) Non-orthogonal signal components, extended block diagonal 192×1024 Φ with repeated blocks.*

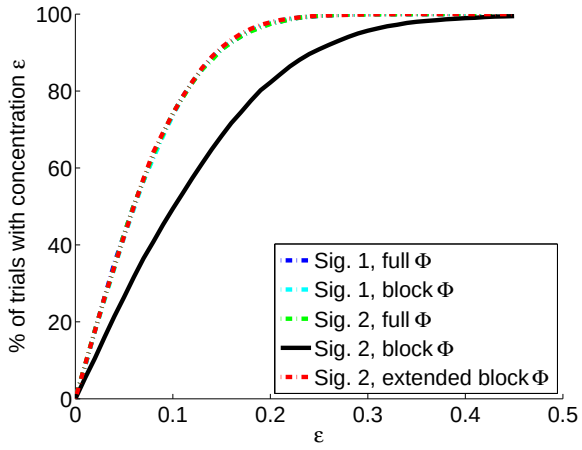

Fig. 4. *The percent of trials for which $(1-\epsilon) \leq \|\Phi x\|_2/\|x\|_2 \leq (1+\epsilon)$. Note that all curves overlap except for the signal having non-orthogonal signal components with block diagonal Φ with repeated blocks.*

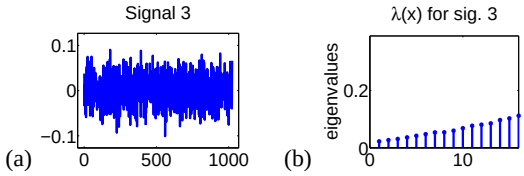

Fig. 5. *(a) Random signal (signal 3). (b) $\lambda(x)$ for random signal.*

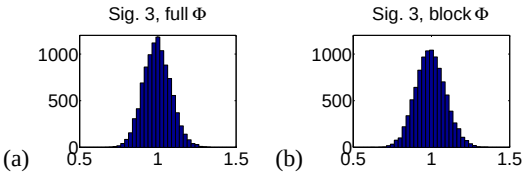

Fig. 6. *Histogram of $\|\Phi x\|_2/\|x\|_2$ for fixed x across 10000 randomly generated matrices Φ. (a) random signal, fully dense 64×1024 Φ. (b) random signal, block diagonal 64×1024 Φ with repeated blocks.*

mean Gaussian random variables of variance $\frac{1}{4}$. As $\Lambda(x) = 5.3341 < J = 16$, Theorem II.2 suggests that the effectiveness of using such matrices is diminished. This is seen in Figure 3(d) and Figure 4.

Since in this case we had access to $\Lambda(x)$ which we know to be 3 times less than the optimal value of $J = 16$, we could consider block diagonal matrices with additional rows that would compensate for this fact. In particular, consider extended block diagonal matrices Φ with $J = 16$ repeated blocks, but each block now has dimensions $M' = 12$ and $N = 64$. Applying Theorem II.2, we see that $M'\Lambda(x) \approx MJ = 64$, and thus we would obtain the same concentration exponent as for fully dense Φ of size 64×1024. Decidedly from Figure 3(e) and Figure 4, the concentration of measurement behavior for these $M'J \times NJ = 192 \times 1024$ block diagonal matrices on this signal is indistinguishable from the earlier favorable cases. This suggests that the factor $\Lambda(x)$ plays an important and precise role in dictating the concentration of measure phenomena for these block diagonal matrices.

Lastly, we shall demonstrate that block diagonal matrices with repeated blocks work almost as well as fully dense matrices for random signals. Using the same parameters as above, we construct a vector x of length $NJ = 1024$ whose entries are i.i.d. zero mean Gaussian random variables with equal variance. This signal is shown in Figure 5 and has $\Lambda(x) = 13.4846 \sim J = 16$. We plot in Figure 6 the histograms of $\frac{\|\Phi x\|_2}{\|x\|_2}$ for fully dense and block diagonal measurement matrices respectively. In Figure 7, we plot as a function of ϵ the percentage of trials for which $(1-\epsilon) \leq \frac{\|\Phi x\|_2}{\|x\|_2} \leq (1+\epsilon)$ for all 3 types of signals when using block diagonal matrices of size $MJ \times NJ = 64 \times 1024$ with repeated blocks. Here we see that using block diagonal matrices on a random signal gives almost the same concentration of measure as using fully dense matrices since for our choice of J and N, we have $J \ll N$.

IV. Conclusion

We have derived a concentration of measure bound for block diagonal matrices with repeated blocks. These matrices model constrained data acquisition systems where the data is viewed repeatedly with the same measurement block. Our main result, Theorem II.2, shows that the concentration exponent may scale as that for a fully dense matrix despite the block diagonal matrices having many fewer nonzeros and much lower complexity than the dense matrices. We have also identified the role that the signal diversity (i.e., $\Lambda(x)$) plays in this concentration exponent. The main intuition provided by this result is that repeated applications of the same measurement operator can leverage the diversity in the signal

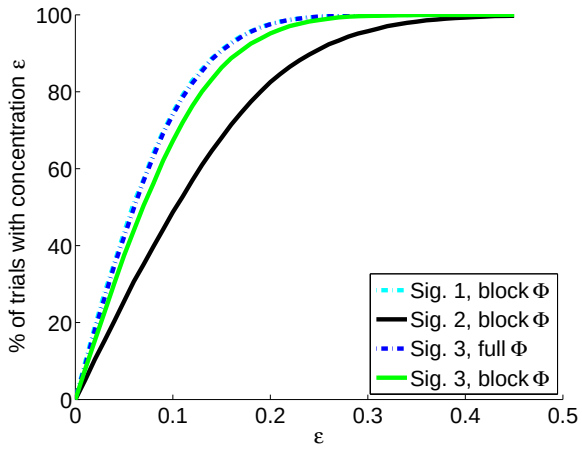

Fig. 7. *The percent of trials for which* $(1-\epsilon) \leq \|\Phi x\|_2/\|x\|_2 \leq (1+\epsilon)$. *Here we see that using block diagonal matrices with repeated blocks on a random signal gives almost the same concentration of measure as using fully dense matrices.*

itself to provide concentration results comparable to those of an unconstrained dense measurement matrix. In the other extreme, when the signal is essentially unchanging (providing no diversity), these restricted measurement operators provide concentration results as if they were only getting data from one measurement block applied to one signal component (which is effectively all that can happen in this case).

This bound may provide the foundation for proving results such as the JL lemma or RIP for sensing systems that require using the same randomized measurement operator repeatedly. As these results will vary depending on the signal structure, it will be important to characterize the types of signals that can provide the diversity necessary to achieve favorable concentration rates. This characterization of the signal diversity structure for various signal classes is the subject of our ongoing research.

V. Appendix

Proof of Lemma II.2:

We first rewrite the block measurement equations as $y' = X'\phi'$, with

$$y' = \begin{pmatrix} y_1(1) \\ \vdots \\ y_J(1) \\ y_1(2) \\ \vdots \\ y_J(M) \end{pmatrix}, \; X' = \begin{pmatrix} X & & \\ & \ddots & \\ & & X \end{pmatrix},$$

$$\text{and } \phi' = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_M \end{pmatrix},$$

where $y_j(i)$ denotes the ith component of the measurement vector y_j and $y' \in \mathbb{R}^{MJ}, X' \in \mathbb{R}^{MJ \times MN}$, and $\phi' \in \mathbb{R}^{MN}$.

We note that y' is just a re-arrangement of y and thus $\|y'\|_2 = \|y\|_2$. Therefore $\|y\|_2^2 = \|y'\|_2^2 = \phi'^T X'^T X' \phi' = \sum_{i=1}^M \phi_i^T A \phi_i$.

■

Proof of Lemma II.3:

Because A is symmetric, it has an eigen-decomposition $A = VDV^T$, 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.

■

References

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

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

[3] M. Ledoux, "The concentration of measure phenomenon," *Mathematical Surveys and Monographs, American Math. Society*, 2001.

[4] R. DeVore, G. Petrova, and P. Wojtaszczyk, "Instance-optimality in probability with an ℓ_1-minimization decoder," 2008, to appear in *Appl. Comp. Harmonic Anal.*

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

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

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

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

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

[10] 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, Signal Proc. (ICASSP)*, March 2010.

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

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