Technical Report: On the Relation between Block Diagonal Matrices and Compressive Toeplitz Matrices

Han Lun Yap and Christopher J. Rozell
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332–0250
Email: {yaphanlun,crozell}@gatech.edu

October 28, 2011

Abstract

In a typical communications problem, Toeplitz matrices Φ arise when modeling the task of determining an unknown impulse response a from a given probe signal ϕ. When a is sparse, then whenever Φ formed from the probe signal ϕ satisfy the Restricted Isometry Property (RIP), a can be robustly recovered from its measurements via ℓ_1-minimization. In this paper, we derived the RIP for compressive Toeplitz matrices whose number of rows of the matrices J is much smaller than the number of columns N. We show that J should scale like $J \sim S^2 \log(N)$, where S is the sparsity of the impulse response. While this is marginally worse than the state-of-the-art scaling currently achieved in the literature, the novelty of this work comes from making the relation between the Toeplitz matrix of interest to a block diagonal matrix. The proof of the RIP then follows from using recent results on the concentration of measure inequalities of block diagonal matrices, together with a standard covering-and-counting argument.

I. Introduction

In a typical communications problem, faithful recovery of transmitted data depends on having an accurate estimate of the communications channel. In many settings, the communications channel is modeled by a linear filter and the impulse response is estimated by transmitting a known probe signal. Specifically, denote the unknown impulse response by $a \in \mathbb{R}^N$ and we want to estimate this channel by probing the system with a known signal $\phi \in \mathbb{R}^P$ and examining the system output $\widetilde{y}$:

$$\widetilde{y} = \phi * a = \Phi_{\text{large}} a, \tag{1}$$

where Φ_{large} is an $(N + P - 1) \times N$ Toeplitz matrix given by

$$\Phi_{\text{large}} = \begin{bmatrix} \phi_1 & 0 & 0 \\ \vdots & \ddots & 0 \\ \phi_N & \cdots & \phi_1 \\ \vdots & & \vdots \\ \phi_P & \cdots & \phi_{(P-N+1)} \\ 0 & \ddots & \vdots \\ 0 & 0 & \phi_P \end{bmatrix}. \tag{2}$$

In some applications, the channel can be assumed to be sparse (e.g., when there are only a few multipath reflections), and resource constraints lead to the desire to estimate a using either (1) as short a probe signal ϕ as possible, or (2) as few measurements (i.e., values of $\widetilde{y}$) as possible. In the first case (called the *non-compressive* case), one typically uses a short but random probe signal ϕ and measures all the entries in $\widetilde{y}$ for decoding the sparse channel a. In the second case (called the *compressive* case), one possible strategy is to again use a random (but not necessarily short) probe signal ϕ but now only measure a subset of size $J \leq N+P-1$ of the coefficients in $\widetilde{y}$. In this case, the resulting measurement vector $y \in \mathbb{R}^J$ can be written as $y = \Phi_{\text{small}} a$, where Φ_{small} is a $J \times N$ matrix whose rows are a subset of those from Φ_{large}. In this work, we will focus our study on the compressive case, and in particular, we will focus on the case when $J \leq N$.

Work in the field of compressed sensing (CS) [1] has shown that sparse structure in signals can be leveraged to improve estimation performance by using estimators that explicitly incorporate a sparse signal model. In particular, the CS results make guarantees about robust recoverability of sparse signals x from measurements $y = \Phi x$ when Φ satisfies the Restriced Isometry Property (RIP). We say that a matrix Φ satisfy the RIP of order S (and conditioning ϵ) if *for every* S-sparse signal x, we have

$$(1-\epsilon)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+\epsilon)\|x\|_2^2.$$

While the RIP property is hard to check in general for a specific matrix, much of the recent excitement is due to the fact that $J \times N$ matrices[1] whose entries are i.i.d. random Gaussian variables are known to satisfy the RIP with high probability if the number of measurements J scale linearly in S and logarithmically in N. For these Gaussian random matrices, the proof that they satisfy RIP is particularly easy. It comes from the fact that such matrices satisfy a *uniform concentration of measure inequality*, i.e., for any fixed signal $x \in \mathbb{R}^N$ and any $\epsilon \in (0,1)$, a Gaussian matrix Φ having size $J \times N$ satisfy

$$\mathbb{P}\left\{ \left| \|\Phi x\|_2^2 - \|x\|_2^2 \right| > \epsilon \|x\|_2^2 \right\} \leq 2e^{-J\epsilon^2}. \tag{3}$$

One can then couple the above concentration inequality (3) with some elementary covering arguments and union bounds [2, 3] to show that if $J = O\left(S \log(N/S)\right)$, then Φ satisfies the RIP with high probability. Concentration of measure type results have also been used to prove the RIP for random matrices with subexponential columns [4], and a concentration result of the form (3) has also been used to probabilistically analyze the performance of ℓ_1-minimization [5].

However, good RIP results for Toeplitz matrices (where J scales linearly with S) have been difficult to achieve. The authors in [6] considered a probe ϕ made up of i.i.d. Gaussian random variables (or a probe of i.i.d. Rademacher sequence in [7]), showing that if $P \sim S^2 \log(N)$, then with high probability the non-compressive matrix Φ_{large} satisfies the RIP. This quadratic dependence of the probe length on the sparsity is worse than most cutting edge CS results (where the number of measurements scales linearly with the sparsity). The work in [8] also considers convolution of a random probe $\phi \in \mathbb{R}^P$ with an S-sparse channel $a \in \mathbb{R}^N$ where $P > N$, i.e., there is an added restriction on the probe length being longer than the channel impulse response. The main result there states that the resulting non-compressive Toeplitz matrix Φ_{large} satisfies the RIP with high probability if $S \sim N/\log^5(N)$. There have been several other authors who have proved equivalent results (i.e., quadratic scaling of J with the sparsity S) for contiguous (measurements taken from $\widetilde{y}$ are contiguous) compressive Toeplitz matrices [6, 7, 9, 10] and non-contiguous compressive Toeplitz matrices [11]. The best RIP result so far has been achieved by the authors of [12]. However, they used much heavier technical machinery, essentially bounding the extrema of a random process, and established an RIP bound for partial circulant matrices requiring

[1]The choice of J instead of M to denote the number of row in the matrix Φ is so as to be consistent with our notations of block diagonal matrices as we will see in section II.

only $(S \log N)^{3/2}$ measurements. We remark that other papers, for example in [13], avoids displaying the RIP for (compressive) toeplitz matrices all together but instead gives recoverability guarantees for probabilistically chosen sparse signals.

This paper establishes the RIP for compressive Toeplitz matrices with a number of measurements J proportional to $S^2 \log N$. This result matches the quadratic scaling with sparsity established in most of the literature except for [12]. Despite not achieving state-of-the-art results, we feel that this paper contains 2 novel contributions. First, the tools that we use to derive the RIP is relatively simple. This simplicity is the result of the observation that compressive Toeplitz matrices are in fact related to block diagonal matrices. This relation allows us to use a recent result on the concentration of measure inequality afforded by block diagonal matrices [14] to generate a uniform concentration of measure inequality (for sparse signals) for our compressive Toeplitz matrices. An application of the usual covering-and-counting argument will then get us to the RIP. Second, this paper provides a unified framework for contiguous and non-contiguous sampling of the measurements $\widetilde{y}$. This is possible because of the ambivalence of the concentration of measure result to the different sampling methods.

II. Concentration of Measure for Repeated Block Diagonal Matrices

The result of this paper rely on a recent work on the concentration of measure for block diagonal matrices [14]. The study of block diagonal matrices arises when input data may be divided naturally into discrete subsections (or blocks), with each block acquired via a local measurement operator. In such scenarios, a signal $x \in \mathbb{R}^{NJ}$ is partitioned into J blocks $x_1, x_2, \ldots, x_J \in \mathbb{R}^N$, and for each $j \in \{1, 2, \ldots, J\}$, we 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

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{bmatrix}}_{y:\ (\sum_j M_j)\times 1} = \underbrace{\begin{bmatrix} \Phi_1 & & & \\ & \Phi_2 & & \\ & & \ddots & \\ & & & \Phi_J \end{bmatrix}}_{\Phi:\ (\sum_j M_j)\times NJ} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_J \end{bmatrix}}_{x: NJ\times 1}. \tag{4}$$

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. Our work here will focus on RBD matrices.

Before stating the concentration of measure for RBD matrices, let us 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

$$X := [x_1 \ x_2 \ \cdots \ x_J]^T, \tag{5}$$

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

$$\Lambda_2(x, M) := \frac{M\|\lambda\|_1^2}{\|\lambda\|_2^2} \quad \text{and} \quad \Lambda_\infty(x, M) := \frac{M\|\lambda\|_1}{\|\lambda\|_\infty}. \tag{6}$$

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

Theorem II.1. *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* (4)*, with $\Phi_j = \widetilde{\Phi}$ for all j. Then*

$$\mathbb{P}_{\leq}\left\{\left|\|\Phi x\|_2^2 - \|x\|_2^2\right| > \epsilon\|x\|_2^2\right\} 2\exp\left\{-C_1 \min\left(C_2^2\epsilon^2\Lambda_2(x,M),\ C_2\epsilon\Lambda_\infty(x,M)\right)\right\}, \tag{7}$$

where C_1 and C_2 are absolute constants.

From (7), 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 $0 \leq \epsilon \leq \frac{\Lambda_\infty(x,M)}{C_2\Lambda_2(x,M)}$ and exponentially as a function of $\epsilon\Lambda_\infty(x,M)$ in the case where $\epsilon > \frac{\Lambda_\infty(x,M)}{C_2\Lambda_2(x,M)}$. Thus, the concentration rate depends explicitly on the signal x being measured. The interested reader is referred to [14] for further discussions on this signal-dependent concentration rates.

III. RIP for Toeplitz Matrices

Interestingly, the concentration results of repeated block diagonal matrices can be used as an analytic tool to prove the RIP for sub-sampled compressive and non-compressive Toeplitz measurement matrices that arise in problems such as channel sensing [6, 7, 9–11, 13, 15–17]. While the Toeplitz structure in the convolution matrix shown in (2) does not immediately appear to fall into the block diagonal structure detailed in our concentration results, careful examination reveals that it can be written in such a format. Consider that every output value of $\widetilde{y}$ is the result of multiplying the same probe vector ϕ by a version of the impulse response a that has been shifted by an amount depending on the measurement index. The intuition here is that this computation can be written as an RBD matrix with each block equal to the probe ϕ^T (i.e., $M = 1$), multiplied by a signal x where each block x_k is a time-reversed, shifted, and windowed version of the impulse response a. To make things concrete, suppose $J \leq N$ and let $i_1, i_2, \cdots, i_J$ denote the indices of the measured coefficients of $\widetilde{y}$. For simplicity, we assume that the probe length $P \geq N + J - 1$ and that $i_1, i_2, \cdots, i_J \in [N, P]$, but note that we will not assume that these indices are contiguous.[2] We define block x_k corresponding to the measurement index i_k by appropriately shifting and zero padding the time-reversed impulse response:

$$x_k = \left[\vec{0}_{(i_k - N)}\ \ a_N\ \ \ldots\ \ a_1\ \ \vec{0}_{(P - i_k)}\right]^T, \tag{8}$$

where we denote $\vec{0}_L$ as a row vector of L zeros. With this definition, we now see that y can be written as the multiplication of a $J \times PJ$ RBD matrix Φ (constructed with the $1 \times P$ matrices $\Phi_k = \phi^T$ along the main diagonal) times a length-PJ signal $x = [x_1^T\ \ x_2^T\ \ \ldots\ \ x_J^T]^T$ with J blocks:

$$y = \Phi_{\text{small}}a = \Phi x. \tag{9}$$

The following lemma establishes the concentration of measure result for the subsampled output of the convolution operation, and follows directly from applying Theorem II.1 to the current problem formulation.

Lemma III.1. *Let $a \in \mathbb{R}^N$ be an arbitrary vector, $\phi \in \mathbb{R}^P$ be a random vector with i.i.d. Gaussian entries having mean zero and variance $\sigma^2 = \frac{1}{J}$ with $J \leq N$, and $\widetilde{y}$ be the convolution of a and ϕ as*

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

defined in (2)*. Suppose that* $P \geq J + N - 1$ *and let* $i_1, i_2, \cdots, i_J \in [N, P]$ *denote the indices of the available measurements from this convolution, such that* $y_k = \widetilde{y}_{i_k}$*. Also, define* X *as a concatenation of signal blocks as in* (5)*, with the individual signal blocks* x_k *(depending on the measurement indices* i_k*) defined as in* (8)*. Then, for* $y = \Phi_{\text{small}} a$ *(see* (9)*),* $\mathbb{E}\{\|y\|_2^2\} = \|a\|_2^2$*, and for* C_1, C_2 *appearing in Theorem II.1,*

$$\mathbb{P}\left\{\left|\|y\|_2^2 - \|a\|_2^2\right| > \epsilon\|a\|_2^2\right\} \leq 2\exp\left(-\frac{C_1 J \min(C_2^2\epsilon^2, C_2\epsilon)}{\|\lambda\|_\infty / \|a\|_2^2}\right), \tag{10}$$

where $\{\lambda_i\}$ *are the eigenvalues of* XX^T *and* $\lambda = [\lambda_1 \ \lambda_2 \ \cdots \ \lambda_J]^T$*. If we further suppose that* a *is* S*-sparse, then*

$$\mathbb{P}\left\{\left|\|y\|_2^2 - \|a\|_2^2\right| > \epsilon\|a\|_2^2\right\} \leq 2\exp\left(-\frac{C_1 J \min(C_2^2\epsilon^2, C_2\epsilon)}{S}\right). \tag{11}$$

Proof: See Appendix A.

Lemma III.1 basically states that arbitrary vectors a can have favorable concentration properties when multiplied by compressive Toeplitz matrices. We note that (10) relates the probability of concentration for a vector $a \in \mathbb{R}^N$ to the quantity $\|\lambda\|_\infty$ (which is defined in terms of a). If the vector a is S-sparse, it is possible to derive a useful upper bound for $\|\lambda\|_\infty$ and this is shown in (11). This analysis establishes that when a is a sparse vector measured by a compressive Toeplitz matrix, the concentration exponent can be stated simply in terms of the sparsity S and number of measurements J, making the concentration bound suitable as an analysis tool for establishing results in the CS literature. In particular, we can use this result to prove the RIP for compressive Toeplitz matrices. Using standard covering arguments and following the same steps as in [2], we arrive at the following theorem establishing RIP for the compressive Toeplitz matrices relevant for the channel sensing problem.

Theorem III.1 (RIP)**.** *Suppose* $\Phi_{\text{small}} \in \mathbb{R}^{J \times N}$ *is a compressive Toeplitz matrix as defined in* (9) *(with either contiguous or non-contiguous measurement indices). Then there exist constants* C_3, C_4 *such that if* $J \geq C_3 S^2 \log(N/S)$*,* Φ_{small} *will satisfy the RIP of order* S *with probability at least* $1 - 2\exp(-C_4 J)$*.*

The theorem above establishes the RIP for compressive Toeplitz matrices with a number of measurements J proportional to $S^2 \log N$. As discussed in Section I, this result matches the quadratic scaling with sparsity established in most of the literature except for [12], which had to resort to heavy mathematical machinery for its proof. The novelty of this paper comes from, first, its simplicity, as can be observed from the simple concentration of measure bound in (10) and in (11), and second, its unified framework for both contiguous and non-contiguous measurements, as can be seen in Lemma III.1.

IV. Extensions

Using the same general approach taken in this paper, we can make a similar statement about the RIP of non-compressive Toeplitz matrices. In such cases, one can guarantee the RIP with high probability by taking P proportional to $S^2 \log N$. This result is comparable to the results in [6, 10] (but uses arguably simpler machinery); it is less favorable than the state-of-the-art result of $P \sim S \log^5 N$ implied by [16].

Appendix A
Proof of Lemma III.1

To simplify the proof, it will be beneficial to use operator matrices to define the shifting and windowing operations creating the signal blocks. Specifically, let e_i denote the i^{th} canonical basis vector of $\mathbb{R}^{N+P-1}$

and define the $(N+P-1)\times J$ matrix $R = [e_{i_1} \;\; e_{i_2} \;\; \cdots \;\; e_{i_J}]$ that removes measurements from the convolution operation to isolate just the selected measurements such that $y = R^T\widetilde{y}$. Furthermore, define the windowing matrix $W = \begin{bmatrix} e_N & \dots & e_{(N+P-1)} \end{bmatrix}^T$ to be the $P \times (N+P-1)$ matrix that keeps only the last P coefficients of a length-$(N+P-1)$ vector. Finally, note that we can now write the matrix of concatenated signal blocks $X \in \mathbb{R}^{J\times P}$ defined in (5) as $X^T = WAR$. Here, A is an $(N+P-1)\times(N+P-1)$ circulant matrix whose first column is $\widetilde{a} := [a_N \;\; a_{N-1} \;\; \cdots \;\; a_1 \;\; \vec{0}_{P-1}]^T$ and whose subsequent columns are circularly shifted in the downward direction.

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

$$\mathbb{P}\left\{\left|\|y\|_2^2 - J\|a\|_2^2\right| > \epsilon J\|a\|_2^2)\right\} \leq 2\exp\left(-C_1\min\left(\frac{C_2^2\epsilon^2\|\lambda\|_1^2}{\|\lambda\|_2^2}, \frac{C_2\epsilon\|\lambda\|_1}{\|\lambda\|_\infty}\right)\right). \tag{12}$$

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

By plugging these inequalities into (12), we obtain the probabilities specified in (10). Finally, we note that

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

and so if we suppose the entries of the random probe ϕ actually have variance $\sigma^2 = \frac{1}{J}$, we complete our derivation of (10).

Now, suppose that a has no more than S nonzero components. Then letting $\|D\|_2$ denote the standard operator norm of a matrix D (i.e., the largest singular value of D), we have $\|\lambda\|_\infty = \|XX^T\|_2 = \|X\|_2^2 \leq \|R\|_2^2\|A\|_2^2\|W\|_2^2 = \|A\|_2^2 = \|A^T\|_2^2$, since the largest singular values of both W and R are 1. Because A^T is a circulant matrix, its eigenvalues are equal to the un-normalized discrete Fourier transform (DFT) of its first row $\widetilde{a}^T$. Denoting the un-normalized DFT matrix by $F \in \mathbb{C}^{(N+P-1)\times(N+P-1)}$, we see that $\|A^T\|^2 = \|F\widetilde{a}^T\|_\infty^2$. Expanding on this, we have:

$$\begin{aligned}
\|F\widetilde{a}^T\|_\infty^2 &= \max_{p=1,\cdots,N+P-1}\left|\sum_{k=1,\cdots,N+P-1}\widetilde{a}_k e^{j\frac{2\pi(k-1)(p-1)}{N+P-1}}\right|^2 \\
&\leq \max_{p=1,\cdots,N+P-1}\left(\sum_{k=1,\cdots,N+P-1}|\widetilde{a}_k|\right)^2 \\
&= \|a\|_1^2 \\
&\leq S\|a\|_2^2
\end{aligned}$$

Thus, $\|\lambda\|_\infty \leq S\|a\|_2^2$, which implies that a concentration rate that holds for any S-sparse vector a is

$$\mathbb{P}\left\{\left|\|y\|_2^2 - \|a\|_2^2\right| > \epsilon\|a\|_2^2\right\} \leq 2\exp\left(-\frac{C_1 J\min(C_2^2\epsilon^2, C_2\epsilon)}{S}\right).$$

■

REFERENCES

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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5862–5875, 2010.

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

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

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

[10] 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.

[11] H. Rauhut, "Compressive sensing and structured random matrices," in *Theoretical Foundations and Numerical Methods for Sparse Recovery*, ser. Radon Series Comp. Appl. Math., M. Fornasier, Ed. deGruyter, 2010, vol. 9, pp. 1–92.

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

[13] H. Rauhut, "Circulant and Toeplitz matrices in compressed sensing," in *Proc. Signal Processing with Adaptive Sparse Structured Representations (SPARS)*, vol. 9, 2009.

[14] J. Y. Park, H. L. Yap, C. J. Rozell, and M. B. Wakin, "Concentration of measure for block diagonal matrices with applications to compressive signal processing," *IEEE Trans. Signal Process.*, December 2011.

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

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

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