Concentration of Measure Inequalities for Compressive Toeplitz Matrices with Applications to Detection and System Identification

Borhan M. Sanandaji, Tyrone L. Vincent, and Michael B. Wakin

Abstract— In this paper, we derive concentration of measure inequalities for compressive Toeplitz matrices (having fewer rows than columns) with entries drawn from an independent and identically distributed (i.i.d.) Gaussian random sequence. These inequalities show that the norm of a vector mapped by a Toeplitz matrix to a lower dimensional space concentrates around its mean with a tail probability bound that decays exponentially in the dimension of the range space divided by a factor that is a function of the sample covariance of the vector.

Motivated by the emerging field of Compressive Sensing (CS), we apply these inequalities to problems involving the analysis of high-dimensional systems from convolution-based compressive measurements. We discuss applications such as system identification, namely the estimation of the impulse response of a system, in cases where one can assume that the impulse response is high-dimensional, but sparse. We also consider the problem of detecting a change in the dynamic behavior of a system, where the change itself can be modeled by a system with a sparse impulse response.

I. INTRODUCTION

We live in an era that one might call "*the century of data explosion*"—indeed, we have now reached a point where all of the data generated by intelligent sensors, digital cameras, and so on exceeds the world's total data storage capacity [1]. Motivated to reduce the burdens of acquiring, transmitting, storing, and analyzing such vast quantities of data, signal processing researchers have over the last few decades developed a variety of techniques for data compression and dimensionality reduction. Unfortunately, many of these techniques require a raw, high-dimensional data set to be acquired before its essential low-dimensional structure can be identified, extracted, and exploited. In contrast, what would be truly desirable are sensors that require fewer raw measurements yet still capture the essential information in a data set. One positive outcome of this past work, however, has been the development of sparse and compressible representations as concise models for high-dimensional data sets.

Building on these principles, Compressive Sensing (CS) has very recently emerged as a powerful paradigm for combining measurement with compression. First introduced by Candès, Romberg and Tao [2]–[4], and Donoho [5], the CS problem can be viewed as recovery of an s-sparse signal $a \in \mathbb{R}^n$ from just $d < n$ (or even $d \ll n$) observations $y = Xa \in \mathbb{R}^d$, where $X \in \mathbb{R}^{d \times n}$ is a matrix representing

the linear measurement process. An s-sparse signal $a \in \mathbb{R}^n$ is a signal of length n with $s < n$ nonzero (significant) entries. The notation $s := \|a\|_0$ denotes the sparsity level of a. Since the null space of X is non-trivial, exact recovery of the signal depends on whether the solution set for $y = Xa$ contains only one solution of the required sparseness. It has been shown that if X is a fully populated random matrix with entries drawn from an independent and identically distributed (i.i.d.) Gaussian random sequence, then exact recovery of a can be guaranteed with high probability from $d = \mathcal{O}\left(s \log\left(\frac{n}{s}\right)\right)$ measurements [6]. In addition, recovery is possible by solving a convex optimization problem and is robust to measurement noise [7]. As an extension of CS, there is also recent work on the detection of sparse signals from compressive measurements [8].

Concentration of measure inequalities are one of the leading techniques used in the theoretical analysis of randomized compressive operators [9]. A typical concentration inequality takes the following form [10]. For any fixed signal $a \in \mathbb{R}^n$, and a suitable random matrix X, the norm of the projected signal by X will be highly concentrated around the norm of the original signal with high probability. In other words, there exist constants c_1 and c_2, such that for any fixed $a \in \mathbb{R}^n$,

$$\mathbf{P}\left\{ \left| \|Xa\|_2^2 - \mathbf{E}\left[\|Xa\|_2^2 \right] \right| \geq \epsilon \mathbf{E}\left[\|Xa\|_2^2 \right] \right\} \leq c_1 e^{-c_2 d c_0(\epsilon)},$$

where $c_0(\epsilon)$ is a function of $\epsilon \in (0,1)$. In this paper, we derive concentration of measure inequalities for random, compressive Toeplitz matrices and discuss applications to detection and system identification.

A. Compressive Toeplitz Matrices

While the recovery results discussed in the previous section are for a measurement matrix X with independent elements, many applications will require X to have a particular structure. In particular, when dynamical systems are involved, the measurement process may involve the convolution of an (unknown) system impulse response with a (known) input. Consider identifying the impulse response of a linear time-invariant (LTI) system from its input-output observations. Let $\{x_k\}_{k=1}^{n+d-1}$ be the applied input sequence to an LTI system characterized by its finite impulse response $\{a_k\}_{k=1}^n$. Then the corresponding output y is calculated from the time-domain convolution $y = x * a$. Considering the x_k and a_k sequences to be zero-padded from both sides, each

output sample y_i can be written as

$$y_i = \sum_{j=1}^{n} a_j x_{i-j}. \tag{1}$$

If we only keep d observations of the system, $\{y_k\}_{k=n+1}^{n+d}$, then (1) can be written in a matrix-vector multiplication format as

$$y = Xa, \tag{2}$$

where

$$X = \begin{bmatrix} x_n & x_{n-1} & \cdots & x_1 \\ x_{n+1} & x_n & \cdots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{n+d-1} & x_{n+d-2} & \cdots & x_d \end{bmatrix} \tag{3}$$

is a $d \times n$ Toeplitz matrix. Supposing the system has an s-sparse impulse response, we are interested in efficiently acquiring and retrieving a via its multiplication by the $d \times n$ compressive Toeplitz matrix X. We call this a compressive Toeplitz matrix because $d < n$. In addition to the application to convolution measurement problems, it is worth noting that while fully populated random matrices require nd elements to be generated, Toeplitz matrices require only $n + d - 1$ distinct random entries, which may provide an advantage in other CS applications.

B. Related Results

Compressive Toeplitz matrices and convolution have been previously studied in the context of compressive sensing in [11]–[17], with applications including channel estimation and synthetic aperture radar. The work in this paper is most closely related to [17], which also derives a concentration of measure bound for compressive Toeplitz matrices and utilizes this bound to study recovery. A comparison between our work and this work is given in the next section, with further discussion of related work postponed to later sections.

C. Main Result

In this paper, we derive a concentration of measure bound for compressive Toeplitz matrices as given in (3) with entries drawn from an i.i.d. Gaussian random sequence. Our main result, detailed in Theorem 1, states that the upper and lower probability tail bounds depend on the number of measurements d and on the eigenvalues of the measurement covariance matrix $P = P(a)$ defined as:

$$P = \begin{bmatrix} \mathcal{R}_a(0) & \mathcal{R}_a(1) & \cdots & \mathcal{R}_a(d-1) \\ \mathcal{R}_a(1) & \mathcal{R}_a(0) & \cdots & \mathcal{R}_a(d-2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{R}_a(d-1) & \mathcal{R}_a(d-2) & \cdots & \mathcal{R}_a(0) \end{bmatrix}, \tag{4}$$

where

$$\mathcal{R}_a(\tau) := \sum_{i=1}^{n-\tau} a_i a_{i+\tau} \tag{}$$

denotes the un-normalized sample autocorrelation function of $a \in \mathbb{R}^n$.

Theorem 1: Let $a \in \mathbb{R}^n$ be fixed. Define two quantities $\rho(a)$ and $\mu(a)$ associated with the eigenvalues of the measurement covariance matrix $P = P(a)$ as:

$$\rho(a) = \frac{\max_i \lambda_i}{\|a\|_2^2} \tag{5}$$

and

$$\mu(a) = \frac{\sum_{i=1}^{d} \lambda_i^2}{d\|a\|_2^4} \tag{6}$$

where $\{\lambda_i\}$ are the eigenvalues of P. Suppose X is a random compressive Toeplitz matrix with i.i.d. Gaussian entries having zero mean and unit variance. Then for any $\epsilon \in (0,1)$, the upper tail probability bound is

$$\mathbf{P}\left\{\|y\|_2^2 > d\|a\|_2^2 (1+\epsilon)\right\} \le \left((1+\epsilon)e^{-\epsilon}\right)^{\frac{d}{2\rho(a)}} \tag{7}$$

and the lower tail probability bound is

$$\mathbf{P}\left\{\|y\|_2^2 < d\|a\|_2^2 (1-\epsilon)]\right\} \le \left(e^{\frac{-2\epsilon^3 - \epsilon^2}{2(1+\epsilon)^2}}\right)^{\frac{d}{2\mu(a)}}. \tag{8}$$

In [17], an upper and lower bound essentially identical to (7) is given, but only for a range of ϵ bounded away from 0. Our result (8) and proof technique appear to be new.

D. Organization

The rest of the paper is organized as follows. In Section II, we state the proof of the main result. In Section III, we provide bounds on the spectral norm of the covariance matrix P. A system identification application is considered in Section IV. In Section V, we explore the problem of detection with compressive sensing and present experimental results. We conclude in Section VI.

II. PROOF OF MAIN RESULT

Our analysis begins with the observation that, for fixed a and random Gaussian X, $y = Xa$ will be a Gaussian random vector.

Lemma 1: If $\{x_k\}_{k=1}^{n+d-1}$ is a zero mean, unit variance i.i.d. Gaussian random sequence, then $y = Xa$ is a $d \times 1$ Gaussian random vector with zero mean and $d \times d$ covariance matrix $P = P(a)$ given in (4).

Proof: See Appendix A. ∎

From Lemma 1, it quickly follows that $\mathbf{E}\left[\|y\|_2^2\right] = d\|a\|_2^2$. It is also easy to see that P is a $d \times d$ symmetric Toeplitz matrix with $\mathrm{tr}(P) = d\|a\|_2^2$ where $\{\lambda_i(P)\}_{i=1}^{d}$ denote the eigenvalues of P. Note that $\lambda_i(P) \ge 0$ for all $i = 1, 2, \ldots, d$.

The proof of the main theorem utilizes Markov's inequality along with a suitable bound on the moment generating function of $\|y\|_2^2$. This bound depends on the following lemma.

Lemma 2: If $y \in \mathbb{R}^d$ is a Gaussian random vector with covariance matrix P, then

$$\mathbf{E}\left[e^{\pm t y' y}\right] = \frac{1}{\sqrt{\det(I \mp 2tP)}}, \tag{9}$$

where we require $t \in (0, \frac{1}{2(\max_i \lambda_i)})$ for the case involving $\det(I - 2tP)$, and y' denotes the transpose of y.

Proof:

$$\mathbf{E}\left[e^{\pm ty'y}\right] = \int \frac{1}{(2\pi \det(P))^{\frac{1}{2}}} e^{\pm ty'y} e^{-\frac{1}{2}y'P^{-1}y} dy$$

$$= \int \frac{1}{(2\pi \det(P))^{\frac{1}{2}}} e^{-\frac{1}{2}y'\left(P^{-1}\mp 2tI\right)y} dy$$

$$= \frac{\det^{\frac{1}{2}}\left(\left(P^{-1}\mp 2tI\right)^{-1}\right)}{\det^{\frac{1}{2}}(P)}$$

$$= \frac{1}{\left(\det\left(P^{-1}\mp 2tI\right)\det P\right)^{\frac{1}{2}}}$$

$$= \frac{1}{\sqrt{\det\left(I\mp 2tP\right)}}.$$

∎

Remark 1: As a special case of Lemma 2, if $y \in \mathbb{R}$ is a scalar Gaussian random variable of unit variance, then we obtain the well known result

$$\mathbf{E}\left[e^{\pm ty^2}\right] = \frac{1}{\sqrt{1\mp 2t}}. \tag{10}$$

The main objective of this paper is to find bounds on

$$\mathbf{P}\left\{\|y\|_2^2 > d\|a\|_2^2(1+\epsilon)\right\} \tag{11a}$$

and

$$\mathbf{P}\left\{\|y\|_2^2 < d\|a\|_2^2(1-\epsilon)\right\}. \tag{11b}$$

We use Markov's inequality and Chernoff's bounding method for computing the upper and lower probability tail bounds (11a) and (11b). Specifically, for a random variable Z, and all $t > 0$,

$$\mathbf{P}\left\{Z > \epsilon\right\} = \mathbf{P}\left\{e^{tZ} > e^{t\epsilon}\right\} \le \frac{\mathbf{E}\left[e^{tZ}\right]}{e^{t\epsilon}} \tag{12}$$

(see e.g. [18]). Applying (12) to (11a) yields

$$\mathbf{P}\left\{\|y\|_2^2 > d\|a\|_2^2(1+\epsilon)\right\} \le \frac{\mathbf{E}\left[e^{ty'y}\right]}{e^{d\|a\|_2^2(1+\epsilon)t}}. \tag{13}$$

Using Lemma 2, we rewrite the right hand side of (13) as

$$\frac{\mathbf{E}\left[e^{ty'y}\right]}{e^{d\|a\|_2^2(1+\epsilon)t}} = \left(\det\left(I - 2tP\right)\right)^{-\frac{1}{2}} e^{-d\|a\|_2^2(1+\epsilon)t}. \tag{14}$$

In (14), $t > 0$ is a free variable which can be varied to find the tightest possible bound, resulting in the Chernoff bound approximation of the inequality.

Claim 1: The value of t that minimizes (14) satisfies

$$\sum_{i=1}^{d} \frac{\lambda_i}{1 - 2t\lambda_i} = d\|a\|_2^2(1+\epsilon). \tag{15}$$

Proof: See Appendix B. ∎

Unfortunately, (15) is difficult to use for further analysis. However, since (14) is valid for all t, we are free to choose a suboptimal, but still useful value. We propose to use

$$t = \frac{\epsilon}{2(1+\epsilon)f(a)\|a\|_2^2},$$

where f is a function of a that will be chosen later in a way that leads to tighter bounds. Now we are ready to prove the

main theorem. We state the upper tail probability bound in Lemma 3 and the lower tail probability bound in Lemma 4.

Lemma 3: Let $y = Xa \in \mathbb{R}^d$ be a zero mean Gaussian random vector with $d \times d$ covariance matrix P, where X is a $d \times n$ compressive Toeplitz matrix populated with i.i.d. zero mean, unit variance Gaussian random variables, and a is an s-sparse vector of length n. Then, for any $\epsilon \in (0,1)$,

$$\mathbf{P}\left\{\|y\|_2^2 > d\|a\|_2^2(1+\epsilon)\right\} \le \left((1+\epsilon)e^{-\epsilon}\right)^{\frac{d}{2\rho(a)}}. \tag{16}$$

Proof: From (13) and (14), we get

$$\mathbf{P}\left\{\|y\|_2^2 > d\|a\|_2^2(1+\epsilon)\right\} \le$$
$$\left(\left(\det\left(I - 2tP\right)\right)^{-\frac{1}{d}} e^{-2\|a\|_2^2(1+\epsilon)t}\right)^{\frac{d}{2}}. \tag{17}$$

Choosing t as

$$t = \frac{\epsilon}{2(1+\epsilon)\rho(a)\|a\|_2^2}, \tag{18}$$

the right hand side of (17) can be written as

$$\left(\left(\det\left(I - \frac{\epsilon}{(1+\epsilon)}\frac{P}{\rho(a)\|a\|_2^2}\right)\right)^{-\frac{1}{d}} e^{-\frac{\epsilon}{\rho(a)}}\right)^{\frac{d}{2}}. \tag{19}$$

This expression can be simplified by applying the previous lemmas. Note that

$$\det\left(I - \frac{\epsilon}{(1+\epsilon)}\frac{P}{\rho(a)\|a\|_2^2}\right) = \prod_{i=1}^{d} 1 - \frac{\epsilon}{(1+\epsilon)}\frac{\lambda_i}{\rho(a)\|a\|_2^2}$$
$$= e^{\sum_{i=1}^{d} \log\left(1 - \frac{\epsilon}{(1+\epsilon)}\frac{\lambda_i}{\rho(a)\|a\|_2^2}\right)}.$$

Using the fact that $\log(1 - c_1 c_2) \ge c_2 \log(1 - c_1)$ for any $c_1, c_2 \in [0,1]$ and $\text{tr}(P) = d\|a\|_2^2$, we have

$$e^{\sum_{i=1}^{d} \log\left(1 - \frac{\epsilon}{(1+\epsilon)}\frac{\lambda_i}{\rho(a)\|a\|_2^2}\right)} \ge e^{\sum_{i=1}^{d} \frac{\lambda_i}{\rho(a)\|a\|_2^2} \log\left(1 - \frac{\epsilon}{1+\epsilon}\right)}$$
$$= e^{\frac{\text{tr}(P)}{\rho(a)\|a\|_2^2} \log\left(\frac{1}{1+\epsilon}\right)}$$
$$= e^{\frac{d}{\rho(a)} \log\left(\frac{1}{1+\epsilon}\right)}$$
$$= \left(\frac{1}{1+\epsilon}\right)^{\frac{d}{\rho(a)}}. \tag{20}$$

Combining (17), (19), and (20) gives us

$$\mathbf{P}\left\{\|y\|_2^2 > d\|a\|_2^2(1+\epsilon)\right\} \le \left(\left(\frac{1}{1+\epsilon}\right)^{-\frac{1}{\rho(a)}} e^{-\frac{\epsilon}{\rho(a)}}\right)^{\frac{d}{2}}$$
$$= \left((1+\epsilon)e^{-\epsilon}\right)^{\frac{d}{2\rho(a)}}.$$

∎

Lemma 4: Using the same assumptions as in Lemma 3, for any $\epsilon \in (0,1)$,

$$\mathbf{P}\left\{\|y\|_2^2 < d\|a\|_2^2(1-\epsilon)\right\} \le \left(e^{\frac{-2\epsilon^3-\epsilon^2}{2(1+\epsilon)^2}}\right)^{\frac{d}{2\mu(a)}}.$$

Proof: Applying Markov's inequality to (11b), we obtain

$$\mathbf{P}\left\{\|y\|_2^2 < d\|a\|_2^2(1-\epsilon)\right\} = \mathbf{P}\left\{-\|y\|_2^2 > -d\|a\|_2^2(1-\epsilon)\right\}$$

$$\leq \frac{\mathbf{E}\left[e^{-ty'y}\right]}{e^{-d\|a\|_2^2(1-\epsilon)t}}. \qquad (21)$$

Using Lemma 2, this implies

$$\mathbf{P}\left\{\|y\|_2^2 < d\|a\|_2^2(1-\epsilon)\right\} \leq$$

$$\left(\left(\det\left(I+2tP\right)\right)^{-\frac{1}{d}} e^{2\|a\|_2^2(1-\epsilon)t}\right)^{\frac{d}{2}}. \quad (22)$$

In this case, we choose

$$t = \frac{\epsilon}{2(1+\epsilon)\mu(a)\|a\|_2^2}.$$

Plugging t into (22) and following similar steps as for the upper tail bound, we get

$$\det(I+2tP) = \det\left(I + \frac{\epsilon}{(1+\epsilon)}\frac{P}{\mu(a)\|a\|_2^2}\right)$$

$$= \prod_{i=1}^{d} 1 + \frac{\epsilon}{(1+\epsilon)}\frac{\lambda_i}{\mu(a)\|a\|_2^2}$$

$$= e^{\sum_{i=1}^{d} \log\left(1+\frac{\epsilon}{(1+\epsilon)}\frac{\lambda_i}{\mu(a)\|a\|_2^2}\right)}. \qquad (23)$$

Since $\log(1+c) \geq c - \frac{c^2}{2}$ for $c > 0$,

$$\log\left(1 + \frac{\epsilon}{(1+\epsilon)}\frac{\lambda_i}{\mu(a)\|a\|_2^2}\right) \geq$$

$$\frac{\epsilon}{(1+\epsilon)}\frac{\lambda_i}{\mu(a)\|a\|_2^2} - \frac{1}{2}\left(\frac{\epsilon}{(1+\epsilon)}\frac{\lambda_i}{\mu(a)\|a\|_2^2}\right)^2.$$

Thus,

$$\sum_{i=1}^{d} \log\left(1 + \frac{\epsilon}{(1+\epsilon)}\frac{\lambda_i}{\mu(a)\|a\|_2^2}\right)$$

$$\geq \sum_{i=1}^{d} \frac{\epsilon}{(1+\epsilon)}\frac{\lambda_i}{\mu(a)\|a\|_2^2} - \frac{1}{2}\left(\frac{\epsilon}{(1+\epsilon)}\frac{\lambda_i}{\mu(a)\|a\|_2^2}\right)^2$$

$$= \frac{\epsilon}{(1+\epsilon)}\frac{\sum_{i=1}^{d}\lambda_i}{\mu(a)\|a\|_2^2} - \frac{1}{2}\left(\frac{\epsilon}{(1+\epsilon)\mu(a)\|a\|_2^2}\right)^2 \sum_{i=1}^{d}\lambda_i^2$$

$$= \frac{\epsilon}{(1+\epsilon)}\frac{d}{\mu(a)} - \frac{1}{2}\left(\frac{\epsilon}{1+\epsilon}\right)^2 \frac{d}{\mu(a)}$$

$$= \frac{d}{\mu(a)}\left(\frac{\epsilon^2 + 2\epsilon}{2(1+\epsilon)^2}\right).$$

Using this in (23) gives the bound

$$\det(I+2tP) \geq e^{\frac{d}{\mu(a)}\left(\frac{\epsilon^2+2\epsilon}{2(1+\epsilon)^2}\right)}$$

$$= \left(e^{\frac{\epsilon^2+2\epsilon}{2(1+\epsilon)^2}}\right)^{\frac{d}{\mu(a)}}. \qquad (24)$$
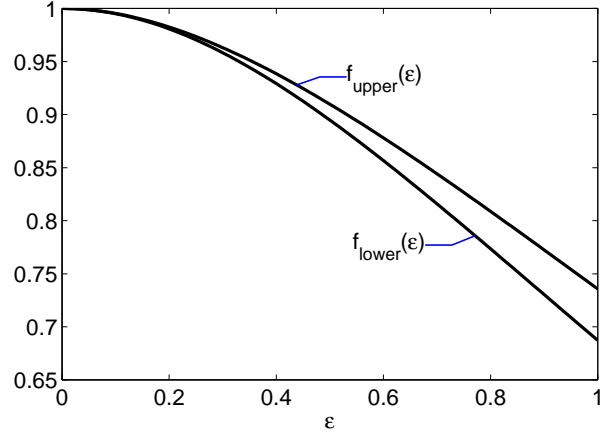

Fig. 1.   Comparison between upper and lower tail bound functions.

By substituting (24) in (22), we obtain

$$\mathbf{P}\left\{\|y\|_2^2 < d\|a\|_2^2(1-\epsilon)\right\} \leq \left(e^{\frac{-\epsilon^2-2\epsilon}{2(1+\epsilon)^2}} e^{\frac{\epsilon(1-\epsilon)}{1+\epsilon}}\right)^{\frac{d}{2\mu(a)}}$$

$$= \left(e^{\frac{-2\epsilon^3-\epsilon^2}{2(1+\epsilon)^2}}\right)^{\frac{d}{2\mu(a)}}.$$

■

III. CONCENTRATION BOUNDS FOR SPARSE SIGNALS

As can be seen the statement of Theorem 1, the upper and lower tail probability bounds are functions of the quantities $\rho(a)$ and $\mu(a)$. For any $\epsilon \in (0,1)$, we have the upper tail probability bound

$$\mathbf{P}\left\{\|y\|_2^2 > d\|a\|_2^2(1+\epsilon)\right\} \leq f_{\text{upper}}(\epsilon)^{\frac{d}{2\rho(a)}}$$

and the lower tail probability bound

$$\mathbf{P}\left\{\|y\|_2^2 < d\|a\|_2^2(1-\epsilon)\right\} \leq f_{\text{lower}}(\epsilon)^{\frac{d}{2\mu(a)}},$$

where

$$f_{\text{upper}}(\epsilon) = (1+\epsilon)e^{-\epsilon} \quad \text{and} \quad f_{\text{lower}}(\epsilon) = e^{\frac{-2\epsilon^3-\epsilon^2}{2(1+\epsilon)^2}}.$$

Figure 1 shows a comparison between $f_{\text{upper}}(\epsilon)$ and $f_{\text{lower}}(\epsilon)$ for $\epsilon \in (0,1)$. These functions are similar, and behave as $e^{-\frac{\epsilon^2}{4}}$ for small ϵ. Typically, we are seeking to find values of d for which the bound is small for a particular class of signals, which in our case will be s-sparse signals. Consequently, it is necessary to obtain an upper bound for the quantities $\rho(a)$ and $\mu(a)$ for the class of signals of interest. It is easy to show that for all a, $\mu(a) \leq \rho(a)$. Therefore, we limit our problem to finding the sharpest bound for $\rho(a)$.

Note that P is a $d \times d$ symmetric Toeplitz matrix which can be decomposed as:

$$P = A'A,$$

where A is an $(n + d - 1) \times d$ Toeplitz matrix formed by a,

$$A = \begin{bmatrix} a_1 & 0 & \ldots & & 0 \\ \vdots & \ddots & & (0) & \vdots \\ a_n & \ldots & a_1 & & \\ 0 & \ddots & & \ddots & 0 \\ \vdots & & a_n & & a_1 \\ & (0) & & \ddots & \vdots \\ 0 & & \ldots & 0 & a_n \end{bmatrix},$$

and A' is the transpose of A. By the Cauchy Interlacing Theorem [19],

$$\max \lambda_i(P) = \max \lambda_i(A'A) \le \max \lambda_i(\widetilde{A}'\widetilde{A}),$$

where $\widetilde{A}$ is an $(n + d - 1) \times (n + d - 1)$ circulant matrix with

$$\widetilde{a} = [a_1, a_2, \ldots, a_n, 0, \ldots, 0],$$

as its first row. Since $\lambda_i(\widetilde{A}'\widetilde{A}) = |\lambda_i(\widetilde{A})|^2$, an upper bound for $\rho(a)$ is provided by the maximum eigenvalue of $\widetilde{A}$. Since $\widetilde{A}$ is circulant, $\lambda_i(\widetilde{A})$ simply equals the un-normalized $(n + d - 1)$-length Discrete Fourier Transform (DFT) of the first row of $\widetilde{A}$. As a result,

$$\lambda_i(\widetilde{A}) = \sum_{k=1}^{n} a_k e^{-j2\pi(i-1)k/(n+d-1)}. \tag{25}$$

When a is an s-sparse signal, only $s < (n + d - 1)$ terms in this summation are nonzero, therefore $\lambda_i(\widetilde{A}) \le \sqrt{s}\|a\|_2$. This results in the bound

$$\rho(a) \le s.$$

Remark 2: Although $\mu(a) \le \rho(a) \le s$ for all s-sparse signals, this appears to be a highly pessimistic bound for most signals. Numerical experiments have shown that for Gaussian random s-sparse signals, $\mathbf{E}[\rho(a)] \sim \log(s)$. However, it remains to be seen if this non-uniformity can be exploited.

IV. SYSTEM IDENTIFICATION / TOEPLITZ RECOVERY

In this section, we address problem of recovering a sparse signal after convolution. We seek to find a number of measurements d of an s-sparse signal a using a random, compressive Toeplitz measurement matrix X that are sufficient to ensure that a can be recovered with high probability.

A. Related Work

Considering structure in the measurement matrix X was addressed by Tropp et al., as one of the early papers which considered reconstruction of a signal from its convolution with a fixed Finite Impulse Response (FIR) filter having random taps [11]. Bajwa et al. [12] studied Toeplitz-structured compressed sensing matrices with applications to sparse channel estimation. They allowed the entries of the measurement matrix be drawn from a symmetric Bernoulli distribution. Later they extended this to random matrices

whose entries are bounded or Gaussian-distributed. It is shown that $d = \mathcal{O}\left(s^2 \log(n)\right)$ measurements are sufficient in order to have exact signal recovery [14], [15], which appears to be the best result to date. This compares to $d = \mathcal{O}\left(s \log(n/s)\right)$ measurements when X is unstructured, see e.g. [20].

Rauhut considered circulant and Toeplitz matrices whose entries are independent Bernoulli ± 1 random variables. Furthermore, they constrained the signs of the non-zero entries of the s-sparse signal to be drawn from a Bernoulli distribution as well. Considering this configuration, they proposed that the required number of measurements for exact recovery scales linearly with sparsity up to a log-factor of the signal length [16]. Xiang et al. proposed convolution with a white noise waveform followed by deterministic subsampling as a framework for CS. Imposing a uniformly random sign pattern on the non-zero entries of the signal, they required $d = \mathcal{O}\left(s \log^{1.5}(n/\delta)\right)$ measurements to have exact recovery with probability exceeding $(1 - \delta)$ [17].

B. Recovery Condition: The Restricted Isometry Property

Introduced by Candès and Tao [21], the Restricted Isometry Property (RIP) is an isometry condition on measurement matrices used in CS.

Definition 1: X satisfies the RIP of order s if there exists a $\delta_s \in (0, 1)$ such that

$$(1 - \delta_s)\|a\|_2^2 \le \|Xa\|_2^2 \le (1 + \delta_s)\|a\|_2^2 \tag{26}$$

holds for all s-sparse signals a.

$\square$

Given a matrix $X \in \mathbb{R}^{d \times n}$ and any set T of column indices, we denote by X_T the $d \times |T|$ submatrix of X composed of these columns. Equation (26) can be interpreted as a requirement that all of the eigenvalues of the Grammian matrix $X_T'X_T$ lie in the range $[1 - \delta_s, 1 + \delta_s]$ for all sets T with $|T| \le s$. When the RIP is satisfied of order $2s$ with $\delta_{2s} < \sqrt{2} - 1$, then correct recovery of an s-sparse signal can be obtained by finding the minimizer of the following optimization problem [7]:

$$\min_{a \in \mathbb{R}^n} \|a\|_1 \quad \text{subject to} \quad y = Aa.$$

In [7], it is also shown that robust recovery is possible in the presence of measurement noise.

Verifying the RIP is generally a difficult task. This property requires bounded condition number for all submatrices built by selecting $s = |T|$ arbitrary columns. Therefore, there are $\binom{n}{s}$ such submatrices to check, and computing the spectral norm of a matrix is not generally an easy task. However, for certain random constructions of X, the RIP follows in a simple way from concentration of measure inequalities.

An important result in this vein concerns the $d \times n$ matrix X whose entries $x_{i,j}$ are independent realizations of Gaussian random variables, $x_{i,j} \sim \mathcal{N}\left(0, \frac{1}{d}\right)$. (All Gaussian random variables mentioned in this section will have this same mean and variance.) Note that in this realization of X, there are $d \times n$ fully independent Gaussian random entries.
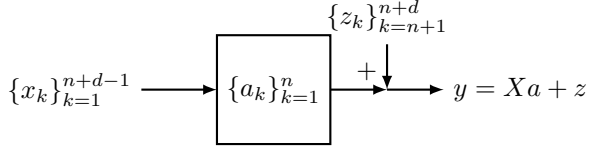
Fig. 2.   FIR filter with impulse response a_k.

We call such realizations simply *Unstructured X* comparing to *Toeplitz X* which have the form of (3). In previous work, a concentration of measure inequality has been used to show that *Unstructured X* satisfy the RIP with high probability.

The relevant concentration of measure inequality for *Unstructured X* is as follows (see e.g. [10]). For any $\epsilon \in (0, 1)$

$$\mathbf{P}\left\{\left(\left|\|Xa\|_2^2 - \|a\|_2^2\right| \geq \epsilon\|a\|_2^2\right)\right\} \leq 2e^{-dc_0(\epsilon)} \quad (27)$$

with $c_0(\epsilon) = \frac{\epsilon^2}{4} - \frac{\epsilon^3}{6}$. Based on this this concentration inequality, the following result was proven in [20].

Theorem 2: An *Unstructured X* with i.i.d. Gaussian entries satisfies the RIP of order s with high probability for $d = \mathcal{O}\left(s \log\left(\frac{n}{s}\right)\right)$

An approach identical to the one taken in [20] can be used to establish the RIP for *Toeplitz X* based on the concentration of measure inequalities given in Theorem 1. In particular, some basic bounding gives us the following result for *Toeplitz X*. For any $\epsilon \in (0, 1)$,

$$\mathbf{P}\left\{\left(\left|\|Xa\|_2^2 - \|a\|_2^2\right| \geq \epsilon\|a\|_2^2\right)\right\} \leq 2e^{-\frac{d}{\rho(P)}c_0(\epsilon)}. \quad (28)$$

Comparing inequality (28) with (27) indicates that these concentration inequalities differ only by the factor $\rho(a)$. Since for sparse signals, $\rho(a)$ is bounded by s, we have the following result, whose proof is omitted for space.

Theorem 3: A *Toeplitz X* with i.i.d. Gaussian entries satisfies the RIP of order s with high probability if $d = \mathcal{O}\left(s^2 \log\left(\frac{n}{s}\right)\right)$.

Note that this result is essentially identical to the bounds given previously in [14], [15]. However, given the extremely non-uniform distribution of $\rho(a)$ over the set of all s-sparse signals a, it may be that a modified approach to the RIP condition will allow the recovery results to be tightened. In particular, we note that if the sign pattern for the non-zero entries of the signal a is chosen to be random, $\rho(a) \approx \log(s)$. This remains an area of current research.

V. DETECTION WITH COMPRESSIVE SENSING

A. Problem Setup

In another application, we consider solving a CS detection problem. Consider an FIR filter with impulse response $\{a_k\}_{k=1}^n$. The response of this filter to a test signal x_k is as described in (1). Moreover, suppose the observations are corrupted by a random additive measurement noise z. Figure 2 shows the schematic of this filter.

Now a detection problem can be considered as follows: the impulse response a_k is known, but we wish to detect whether the dynamics of the system change to b_k, where the difference $c_k = b_k - a_k$ is known. The response $\{y_k\}_{k=n+1}^{n+d}$

to a known input x_k is monitored and the goal is to detect when the impulse response of the system changes using $d < n$ measurements. Since the the nominal impulse response is known, the expected response Xa can be subtracted off, and thus without loss of generality, we can consider $a = 0$.

The detection problem can be formulated as follows [8]. Define two events $\mathcal{E}_0$ and $\mathcal{E}_1$ as:

$$\mathcal{E}_0 \triangleq y = z$$
$$\mathcal{E}_1 \triangleq y = Xc + z$$

where z is a vector of i.i.d. Gaussian noise. However, due to the presence of random noise, the algorithm can result in false detection. This leads to defining the probabilities P_{FA} and P_D as:

$$P_{FA} = \mathbf{P}\left\{(\mathcal{E}_1 \text{ chosen when } \mathcal{E}_0)\right\}$$
$$P_D = \mathbf{P}\left\{(\mathcal{E}_1 \text{ chosen when } \mathcal{E}_1)\right\}$$

where P_{FA} denotes the false-alarm probability and P_D denotes the detection probability. A Receiver Operating Curve (ROC) is a plot of P_D as a function of P_{FA}. A Neyman-Pearson (NP) detector maximizes P_D for a given limit on failure probability, $P_{FA} \leq \alpha$. The NP test is based on the *likelihood ratio* which in our case can be written as

$$y'Xc \underset{\mathcal{E}_0}{\overset{\mathcal{E}_1}{\gtrless}} \gamma \quad (29)$$

where the threshold γ is chosen to meet the constraint $P_{FA} \leq \alpha$. Consequently, we consider the compressive detector t as

$$t := y'Xc. \quad (30)$$

By evaluating t and considering the threshold γ, we are now able to distinguish between the occurrence of two events, $\mathcal{E}_0$ and $\mathcal{E}_1$. To fix the failure limit, we set $P_{FA} = \alpha$ which leads to:

$$P_D(\alpha) = Q\left(Q^{-1}(\alpha) - \frac{\|Xc\|_2}{\sigma}\right) \quad (31)$$

where

$$Q(q) = \frac{1}{\sqrt{2\pi}} \int_q^\infty e^{-\frac{u^2}{2}} \, du. \quad (32)$$

Since $P_D(\alpha)$ is a function of $\|Xc\|_2$, we postulate that the ultimate performance of the detector will depend on $\mu(c)$, $\rho(c)$ and consequently on P which is the sample covariance matrix for c. Furthermore, this dependence would not occur if the measurement utilized an *Unstructured* measurement matrix (which, of course, would not apply to the FIR filter measurement considered here but is a useful comparison to the structured measurement imposed by convolution.)

B. Experiments and ROCs

In these experiments, we test different signals c with different $\rho(c)$ values. Several signal classes are designed based on the location, sign and the value of their non-zero elements. For a given fixed signal c, 1000 *Unstructured* and *Toeplitz X* matrices are generated. For each X, a curve of P_D over P_{FA} is computed using (31). Figure 3 shows the
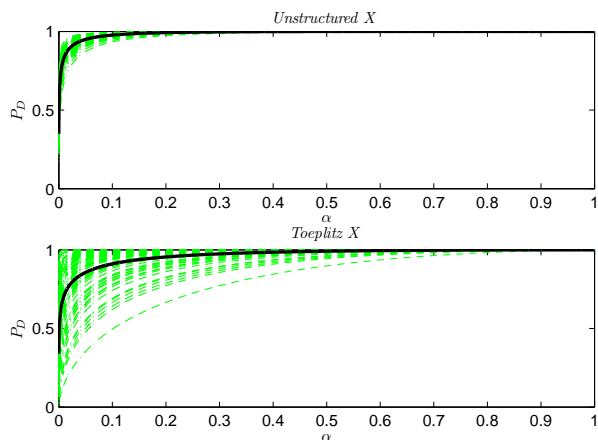
Fig. 3. ROCs for 1000 X Unstructured and Toeplitz matrices for a fixed signal c ($\rho(c) = 50$). The solid black curve is the average of 1000 curves.

result for a special class of signals whose non-zero elements are put in a block position in front with same value and sign. In this experiment signal length is $n = 250$ and number of measurements $d = 125$.

As can be seen, the ROCs associated with *Toeplitz X* are more scattered comparing to the ROCs when we apply *Unstructured X*. This result is due to the weaker concentration of $\|Xc\|$ for *Toeplitz X*. In order to compare the ROCs of different signals with different $\rho(c)$ values when projected by random $d \times n$ Toeplitz matrices, we design an experiment with 6 different signals. Figures 4 and 5 show how these signals get treated by *Unstructured* and *Toeplitz X* matrices.
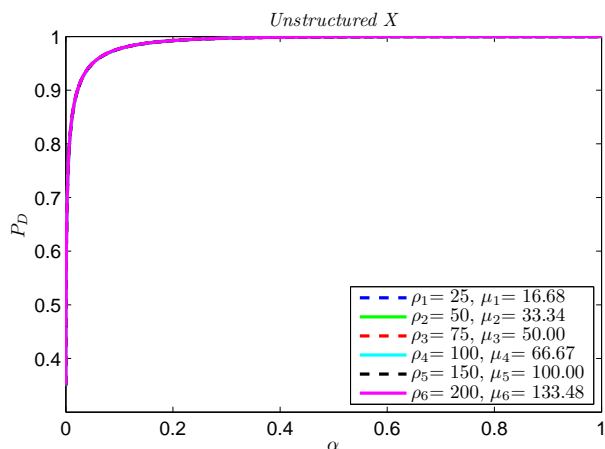

Fig. 4. Average ROCs with 1000 Unstructured X for 6 different signals c. Note that all curves are overlapping.

As can be seen, the average ROCs do not change with respect to different signals when we apply *Unstructured X*. However, they change when applying *Toeplitz X*. More importantly, it can be concluded that $\rho(c)$ has a direct influence on these curves. Generally signals with higher spectral norm $\rho(c)$ have lower ROCs, although this relation is not firm. In all of these experiments, we fix $\|c\|_2 = 1$ and let the noise standard deviation $\sigma = 0.3$.
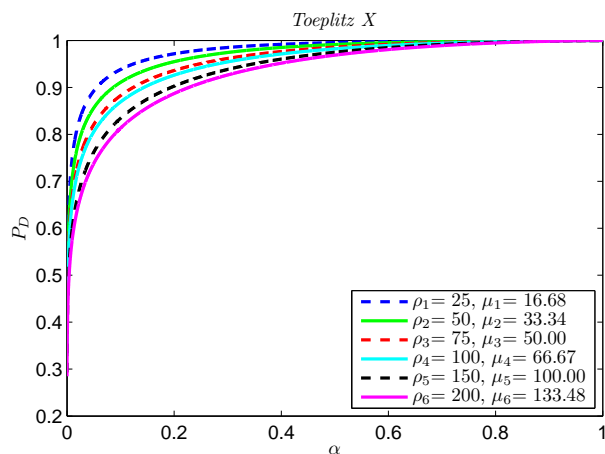

Fig. 5. Average ROCs with 1000 Toeplitz X for 6 different signals c. The curves descend the same order they appear in the legend box.

VI. CONCLUSION

Concentration of measure inequalities for a compressive Toeplitz matrix are derived in this work. We have considered two important quantities $\rho(a)$ and $\mu(a)$ associated with the eigenvalues of the covariance matrix P. Two applications were addressed for these types of matrices. A Toeplitz recovery application was considered where we want to recover an s-sparse signal a from its compressive measurements via $y = Xa$, where X is a random, compressive Toeplitz matrix. We showed that for Toeplitz X, the required number of measurements should scale with $s\rho(a)$ for X to satisfy RIP of order s for all s-sparse signals a. In another application, a CS detection problem is considered. Experimental results show that signals with different $\rho(a)$ values have different ROCs when measured by a Toeplitz matrix.

APPENDIX

A. Proof of Lemma 1

Proof: It is easy to show that y is a zero mean vector, since

$$\mathbf{E}[y] = \mathbf{E}[Xa] = \mathbf{E}[X]\,a = 0.$$

The $d \times d$ covariance matrix P associated with y can be calculated as follows:

$$P = \mathbf{E}[yy'] = \begin{bmatrix} \mathbf{E}[y_1 y_1] & \mathbf{E}[y_1 y_2] & \cdots & \mathbf{E}[y_1 y_d] \\ \mathbf{E}[y_2 y_1] & \mathbf{E}[y_2 y_2] & \cdots & \mathbf{E}[y_2 y_d] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{E}[y_d y_1] & \mathbf{E}[y_d y_2] & \cdots & \mathbf{E}[y_d y_d] \end{bmatrix}.$$

Now consider one of the elements e.g. $\mathbf{E}\left[y_\ell y_m\right]$. Without loss of generality, assume $m < \ell$. Then

$$\mathbf{E}\left[y_\ell y_m\right] =$$

$$a'\mathbf{E}\left[\begin{bmatrix} x_{n+\ell-1} \\ x_{n+\ell-2} \\ \cdots \\ x_\ell \end{bmatrix}\begin{bmatrix} x_{n+m-1} & x_{n+m-2} & \cdots & x_m \end{bmatrix}\right]a$$

$$= a'\begin{bmatrix} 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & & & \vdots \\ 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \end{bmatrix}a = a'\begin{bmatrix} 0 \\ \vdots \\ 0 \\ a_1 \\ a_2 \\ \vdots \\ a_{n-(\ell-m)} \end{bmatrix}$$

$$= \sum_{i=1}^{n-(\ell-m)} a_i a_{i+\ell-m} = \mathcal{R}_a(\ell-m),$$

where the ones start at row $\ell - m + 1$. Clearly, with $\mathbf{E}\left[y_\ell y_m\right] = \mathcal{R}_a(\ell-m)$, the covariance matrix P has the form as in (4) which completes the proof. Note that the covariance matrix is a $d \times d$ symmetric Toeplitz matrix. ∎

B. Proof of Claim 1

Proof: We start by taking the derivative of (14) with respect to t and then putting it equal to zero. Using Jacobi's formula for an invertible matrix A,

$$\frac{d\det(A)}{d\alpha} = \det(A)\mathrm{tr}\left(A^{-1}\frac{dA}{d\alpha}\right),$$

we have

$$\frac{d}{dt}\left(\left(\det\left(I - 2tP\right)\right)^{-\frac{1}{d}} e^{-2\|a\|_2^2(1+\epsilon)t}\right)$$

$$= -\frac{1}{d}\left(\det\left(I - 2tP\right)\right)^{-\frac{1}{d}-1} e^{-2\|a\|_2^2(1+\epsilon)t} \times$$

$$\left(\det\left(I - 2tP\right) \times \mathrm{tr}\left(\left(I - 2tP\right)^{-1}\left(-2P\right)\right)\right) -$$

$$2\|a\|_2^2\left(1+\epsilon\right)e^{-2\|a\|_2^2(1+\epsilon)t}\left(\det\left(I - 2tP\right)\right)^{-\frac{1}{d}}$$

$$= \left(\det\left(I - 2tP\right)\right)^{-\frac{1}{d}} e^{-2\|a\|_2^2(1+\epsilon)t} \times$$

$$\left(-\frac{1}{d} \times \mathrm{tr}\left(\left(I - 2tP\right)^{-1}\left(-2P\right)\right) - 2\|a\|_2^2\left(1+\epsilon\right)\right).$$

Now setting the above equation equal to zero, we get

$$\mathrm{tr}\left(\left(I - 2tP\right)^{-1}\left(P\right)\right) = d\|a\|_2^2\left(1+\epsilon\right).$$

Knowing that P is a symmetric matrix, we can decompose it using the Singular Value Decomposition (SVD) $P = U\Sigma U^T$, where U is an orthogonal matrix and Σ is a diagonal matrix. Now for $0 < t < 0.5$, we get

$$\mathrm{tr}\left(\left(I - 2tP\right)^{-1}\left(P\right)\right) = \mathrm{tr}\left(\left(UU^T - 2tU\Sigma U^T\right)^{-1}\left(U\Sigma U^T\right)\right)$$

$$= \mathrm{tr}\left(\left(U\left(I - 2t\Sigma\right)U^T\right)^{-1}\left(U\Sigma U^T\right)\right)$$

$$= \mathrm{tr}\left(\left(I - 2t\Sigma\right)^{-1}\Sigma\right).$$

From here, we get the following polynomial equation:

$$\sum_{i=1}^{d}\frac{\lambda_i}{1 - 2t\lambda_i} = d\|a\|_2^2(1+\epsilon)$$

where λ_i's are the eigenvalues of the P matrix which lie along the diagonal of Σ. ∎

REFERENCES

[1] "Data, data everywhere," *The Economist*, 25 Feb. 2010.
[2] E. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, vol. 1, no. 3. Citeseer, 2006.
[3] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
[4] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.
[5] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
[6] E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
[7] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes rendus-Mathématique*, vol. 346, no. 9-10, pp. 589–592, 2008.
[8] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, "Signal processing with compressive measurements," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 445–460, 2010.
[9] M. Ledoux, *The concentration of measure phenomenon.* Amer Mathematical Society, 2001.
[10] D. Achlioptas, "Database-friendly random projections: Johnson-lindenstrauss with binary coins," *Journal of Computer and System Sciences*, vol. 66, no. 4, pp. 671–687, 2003.
[11] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk, "Random filters for compressive sampling and reconstruction," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2006.
[12] W. Bajwa, J. Haupt, G. Raz, S. Wright, and R. Nowak, "Toeplitz-structured compressed sensing matrices," in *IEEE/SP 14th Workshop on Statistical Signal Processing, 2007. SSP'07*, 2007, pp. 294–298.
[13] J. Romberg, "Compressive sensing by random convolution," *SIAM Journal on Imaging Sciences*, vol. 2, no. 4, pp. 1098–1128, 2009.
[14] W. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Proc. 42nd Annu. Conf. Information Sciences and Systems (CISS08)*, 2008, pp. 5–10.
[15] J. Haupt, W. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inform. Theory*, 2008.
[16] H. Rauhut, "Circulant and Toeplitz matrices in compressed sensing," *Proc. SPARS*, vol. 9, 2009.
[17] Y. Xiang, L. Li, and F. Li, "Compressive sensing by white random convolution," 2009. [Online]. Available: http://www.citebase.org/abstract?id=oai:arXiv.org:0909.2737
[18] G. Lugosi, "Concentration-of-measure inequalities," *Lecture notes*, 2004.
[19] S. Hwang, "Cauchy's Interlace Theorem for Eigenvalues of Hermitian Matrices," *The American Mathematical Monthly*, vol. 111, no. 2, pp. 157–159, 2004.
[20] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
[21] E. Candès and T. Tao, "Decoding via linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.