TOEPLITZ-STRUCTURED COMPRESSED SENSING MATRICES

Waheed U. Bajwa, Jarvis D. Haupt*, Gil M. Raz**, Stephen J. Wright***, and Robert D. Nowak**

*Department of Electrical and Computer Engineering, University of Wisconsin-Madison
**GMR Research and Technology, Concord, MA 01742-3819
***Department of Computer Sciences, University of Wisconsin-Madison

ABSTRACT

The problem of recovering a sparse signal $x \in \mathbb{R}^n$ from a relatively small number of its observations of the form $y = Ax \in \mathbb{R}^k$, where A is a known matrix and $k \ll n$, has recently received a lot of attention under the rubric of *compressed sensing* (CS) and has applications in many areas of signal processing such as data compression, image processing, dimensionality reduction, etc. Recent work has established that if A is a random matrix with entries drawn independently from certain probability distributions then exact recovery of x from these observations can be guaranteed with high probability. In this paper, we show that Toeplitz-structured matrices with entries drawn independently from the same distributions are also sufficient to recover x from y with high probability, and we compare the performance of such matrices with that of fully independent and identically distributed ones. The use of Toeplitz matrices in CS applications has several potential advantages: (i) they require the generation of only $O(n)$ independent random variables; (ii) multiplication with Toeplitz matrices can be efficiently implemented using fast Fourier transform, resulting in faster acquisition and reconstruction algorithms; and (iii) Toeplitz-structured matrices arise naturally in certain application areas such as system identification.

Index Terms— Compressed sensing, restricted isometry property, system identification, Toeplitz matrices, underdetermined systems of linear equations

1. INTRODUCTION

1.1. Background

We begin by revisiting the problem of recovering a signal $x \in \mathbb{R}^n$ from linear observations of the form

$$y \ = \ Ax \quad : \quad \|x\|_0 \leq m, \tag{1}$$

where $\|\cdot\|_0$ counts the number of non-zero entries in a vector, and $A \in \mathbb{R}^{k\times n}$ is a known matrix. Of particular interest is the special case of highly underdetermined system, $k \ll n$, that has applications in many areas of signal processing such as data compression, image processing, dimensionality reduction etc. and has recently received a lot of attention under the rubric of *compressed sensing* (CS) – starting in particular with some of the earlier works of Candes, Romberg and Tao [1, 2, 3] and Donoho [4].

One of the fundamental problems in CS is to identify the observation matrices that are sufficient to ensure exact recovery of x from y; we term such matrices as the CS matrices. Independently, Donoho [4], and Candes and Tao [1, 3] have provided sufficient conditions for CS matrices. In particular, it was established in [3] (and refined in [1]) that for a $k \times n$ observation matrix A to be a CS matrix, it is sufficient that it satisfies *restricted isometry property* (RIP) of order $3m$ in the following sense: let $T \subset \{1, 2, \ldots, n\}$ and A_T be the $k \times |T|$ submatrix obtained by retaining the columns of A corresponding to the indices in T; then, there exists a constant $\delta_{3m} \in (0, 1/3)$ such that

$$\forall\, z \in \mathbb{R}^{|T|}, \quad (1 - \delta_{3m})\|z\|_2^2 \leq \|A_T z\|_2^2 \leq (1 + \delta_{3m})\|z\|_2^2 \tag{2}$$

holds for all subsets T with $|T| \leq 3m$.[1] Moreover, it was also shown in [1] that x can be exactly recovered in that case by the convex program

$$x = \arg\left(\min_{z\in\mathbb{R}^n} \|z\|_1 \quad \text{subject to} \quad y = Az\right), \tag{3}$$

which is attractive because it can be solved in a computationally tractable manner using linear programming and convex optimization techniques – see, e.g., [1, 4, 5]. Note that the RIP of order $3m$ is equivalent to saying that the singular values of all $k \times 3m$ submatrices of A lie in the interval $\left(\sqrt{2/3}, \sqrt{4/3}\right)$. And while the definition of RIP does not guarantee the existence of CS matrices, recent work has shown that (appropriately scaled) random matrices with entries drawn independently from certain probability distributions satisfy RIP of order $3m$ with high probability for every $\delta_{3m} \in (0, 1/3)$ provided $k \geq const \cdot m \ln(n/m)$ – see, e.g., [1, 3, 4, 6]; we refer to such matrices as independent and identically distributed (IID) CS matrices.

1.2. Main Result

In this paper, we show that if a probability distribution $P(a)$ yields an IID CS matrix (having unit-norm columns in expectation) then a $k \times n$ (partial) Toeplitz matrix A (also having unit-norm columns in expectation) of the form

$$A \ = \begin{bmatrix} a_n & a_{n-1} & \ldots & a_2 & a_1 \\ a_{n+1} & a_n & \ldots & a_3 & a_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_{n+k-1} & a_{n+k-2} & \ldots\ldots & & a_k \end{bmatrix}, \tag{4}$$

This research was supported in part by the National Science Foundation under grants CCF-0353079 and ECS-0529381, and by the DARPA Analog-to-Information Program. E-mails: bajwa@cae.wisc.edu, jdhaupt@wisc.edu, raz@gmrtech.com, swright@cs.wisc.edu, nowak@engr.wisc.edu.

[1] This is a slightly weaker version of the sufficient condition originally given by Candes and Tao; for the sake of brevity, however, and because it suffices to illustrate the principles, we limit ourselves to this condition and refer the reader to [1, 3] for further details.

where the entries $\{a_i\}_{i=1}^{n+k-1}$ have been drawn independently from $P(a)$, is also a CS matrix in the sense that it satisfies RIP of order $3m$ with high probability for every $\delta_{3m} \in (0, 1/3)$ provided $k \geq const \cdot m^3 \ln(n/m)$. Essentially, the reduction in the number of degrees of freedom (DoF) of a Toeplitz random matrix seems to result in an increase in the required number of observations. Note, however, that the result established in this paper is a sufficient condition for exact recovery of *all* m-sparse signals, and simulation results show that actual performance of Toeplitz CS matrices tends to be comparable to that of IID CS matrices for *many*, if not all, such signals. The proof technique used for obtaining this sufficient condition is a novel combination of existing results on IID CS matrices and equitable coloring of graphs, and should be of particular interest to people working in the area of compressed sensing.

The use of Toeplitz CS matrices is a desirable alternative for a number of application areas because (i) IID CS matrices require generation of $O(kn)$ independent random variables, which could be particularly troublesome for large-scale applications, whereas Toeplitz CS matrices require generation of only $O(n)$ independent random variables; (ii) multiplication with IID CS matrices requires $O(kn)$ operations resulting in longer data acquisition and reconstruction times, while multiplication with a Toeplitz CS matrix can be efficiently implemented using fast Fourier transform (FFT) and consequently requires only $O(n \log_2(n))$ operations; and (iii) Toeplitz-structured matrices arise naturally in certain application areas such as identification of a linear time-invariant (LTI) system and consequently, IID CS matrix results are not applicable in such cases.

1.3. Organization

The rest of this paper is organized as follows. In Section 2, we prove that a Toeplitz matrix of the form given in (4) satisfies RIP of order $3m$ with high probability. In Section 3, we discuss extensions of the result of Section 2 to circulant matrices, *left-shifted* Toeplitz-structured matrices, identification of LTI systems having sparse impulse responses and recovery of signals that are sparse in some transform domain. In Section 4, we numerically compare the performance of Toeplitz and circulant CS matrices to that of IID ones and finally, in Section 5, we present some concluding remarks.

2. PROOF OF MAIN RESULT

To establish that Toeplitz-structured matrices with entries drawn independently from "good" probability distributions are also sufficient to recover x from y with high probability, we first observe that if (2) holds for any T then it also holds for all $T' \subset T$. Consequently, it suffices to show that Toeplitz submatrices satisfy (2) with high probability for all $T \subset \{1, 2, \dots, n\}$ such that $|T| = 3m$. Next, we lower bound the probability that Toeplitz submatrices satisfy (2) for any fixed subset T with $|T| = 3m$.

Lemma 1. *Suppose that n, m are given, and let $P(a)$ be a probability distribution that generates a $k \times n$ IID matrix having unit-norm columns in expectation such that, for every $\delta_{3m} \in (0, 1/3)$ and every $T \subset \{1, 2, \dots, n\}$ with $|T| = 3m$, the $k \times |T|$ IID submatrix obtained by retaining the columns corresponding to the indices in T satisfies* (2) *with probability at least*

$$1 - e^{-f(k,m,\delta_{3m})}, \tag{5}$$

where $f(k, m, \delta_{3m})$ is some real-valued function of k, m and δ_{3m}. Let $\{a_i\}_{i=1}^{n+k-1}$ be a sequence of random variables drawn independently from the same distribution, and A be a $k \times n$ Toeplitz matrix of the form given in (4)*. Then, for every $\delta_{3m} \in (0, 1/3)$ and every $T \subset \{1, 2, \dots, n\}$ with $|T| = 3m$, the Toeplitz submatrix A_T satisfies* (2) *with probability at least*

$$1 - e^{-f(\lfloor k/q \rfloor, m, \delta_{3m}) + \ln(q)}, \tag{6}$$

where $q = 3m(3m - 1) + 1$.

Remark 1. As an illustration, let the probability distribution $P(a)$ be given by

$$\mathcal{N}\left(0, \frac{1}{k}\right), \quad \begin{cases} +\sqrt{\frac{1}{k}} & \text{with probability } \frac{1}{2}, \\ -\sqrt{\frac{1}{k}} & \text{with probability } \frac{1}{2}, \end{cases} \quad \text{or}$$

$$\begin{cases} +\sqrt{\frac{3}{k}} & \text{with probability } \frac{1}{6}, \\ 0 & \text{with probability } \frac{2}{3}, \\ -\sqrt{\frac{3}{k}} & \text{with probability } \frac{1}{6}. \end{cases} \tag{7}$$

Then,

$$f(k, m, \delta_{3m}) \;=\; c_0 k - 3m \ln(12/\delta_{3m}) - \ln(2), \tag{8}$$

where $c_0 = c_0(\delta_{3m}) = \delta_{3m}^2/16 - \delta_{3m}^3/48$ (see, e.g., [6]).

Proof. Fix $\delta_{3m} \in (0, 1/3)$ and $T \subset \{1, 2, \dots, n\}$ with $|T| = 3m$. Let $A_{T,i}$ denote the i-th row of A_T and construct an undirected (dependency) graph $G = (V, E)$ such that $V = \{1, 2, \dots, k\}$ and

$$E = \left\{(i, i') \in V \times V : i \neq i', A_{T,i} \text{ and } A_{T,i'} \text{ are dependent}\right\}.$$

Notice that because of the Toeplitz nature of A, $A_{T,i}$ can at most be dependent with

$$2 \cdot (1 + 2 + \dots + (|T| - 1)) \;=\; 2 \cdot |T|(|T| - 1)/2 \;=\; q - 1$$

other rows of A_T. This implies that the maximum degree Δ of G (defined as the maximum number of edges originating from any vertex) is given by $\Delta \leq (q - 1)$.[2] Consequently, using the well-known Hajnal-Szemerédi theorem on equitable coloring of graphs [8], we can always partition G using q (or more) colors such that

$$\lfloor k/q \rfloor \;\leq\; \min_{j \in \{1,2,\dots,q\}} |C_j| \;\leq\; \max_{j \in \{1,2,\dots,q\}} |C_j| \;\leq\; \lceil k/q \rceil, \tag{9}$$

where $\{C_j\}_{j=1}^{q}$ correspond to the different color classes.[3] Next, let A_T^j be the $|C_j| \times |T|$ partition submatrix obtained by retaining the rows of A_T corresponding to the indices in C_j and note that

$$\forall\, z \in \mathbb{R}^{|T|}, \quad \|A_T z\|_2^2 \;=\; \sum_{j=1}^{q} \|A_T^j z\|_2^2 \;=\; \sum_{j=1}^{q} \frac{|C_j|}{k} \|\tilde{A}_T^j z\|_2^2, \tag{10}$$

where $\tilde{A}_T^j$ is defined to be $\tilde{A}_T^j = \sqrt{\frac{k}{|C_j|}}\, A_T^j$ (to ensure unit-norm columns in expectation). Then, from the definition of A_T^j's, we have that each $\tilde{A}_T^j$ is a $|C_j| \times |T|$ submatrix with IID entries from the distribution $P(a)$ and hence, satisfies (2) with probability at least

$$1 - e^{-f(|C_j|, m, \delta_{3m})} \;\geq\; 1 - e^{-f(\lfloor k/q \rfloor, m, \delta_{3m})}. \tag{11}$$

[2] We refer the reader to [7] for a review of basic terminology in graph theory.

[3] Recall that coloring of a graph means that each $C_j \subset V$, $\cup_j C_j = V$, $C_j \cap C_{j'} = \phi$ and no two vertices in a color class share an edge.

Also, note that $\sum_{j=1}^{q} \frac{|C_j|}{k} = 1$ and therefore, from (10), we have that occurrence of the event

$$\left\{ \bigcap_{j=1}^{q} \tilde{A}_T^j \text{ satisfies (2)} \right\}$$

implies that $\forall\, z \in \mathbb{R}^{|T|}$

$$\sum_{j=1}^{q} \frac{|C_j|}{k}(1-\delta_{3m})\|z\|_2^2 \;\le\; \sum_{j=1}^{q} \frac{|C_j|}{k}\, \|\tilde{A}_T^j z\|_2^2 \;\le\; \sum_{j=1}^{q} \frac{|C_j|}{k}(1+\delta_{3m})\|z\|_2^2$$

$$\implies (1-\delta_{3m})\|z\|_2^2 \;\le\; \|A_T z\|_2^2 \;\le\; (1+\delta_{3m})\|z\|_2^2, \quad (12)$$

that is,

$$\left\{ \bigcap_{j=1}^{q} \tilde{A}_T^j \text{ satisfies (2)} \right\} \subset \{A_T \text{ satisfies (2)}\}. \qquad (13)$$

Consequently, we have that

$$\begin{aligned}
&\Pr\left(\{A_T \text{ satisfies (2)}\}\right) \\
&\quad = 1 - \Pr\left(\{A_T \text{ does not satisfy (2)}\}\right) \\
&\quad \overset{(a)}{\ge} 1 - \Pr\left(\left\{ \bigcup_{j=1}^{q} \tilde{A}_T^j \text{ does not satisfy (2)} \right\}\right) \\
&\quad \overset{(b)}{\ge} 1 - \sum_{j=1}^{q} \Pr\left(\left\{ \tilde{A}_T^j \text{ does not satisfy (2)} \right\}\right) \\
&\quad \overset{(c)}{\ge} 1 - \sum_{j=1}^{q} e^{-f(\lfloor k/q \rfloor, m, \delta_{3m})} \\
&\quad = 1 - e^{-f(\lfloor k/q \rfloor, m, \delta_{3m}) + \ln(q)}, \qquad (14)
\end{aligned}$$

where (a) follows from (13), (b) follows from union bounding the event $\left\{ \bigcup_{j=1}^{q} \tilde{A}_T^j \text{ does not satisfy (2)} \right\}$ and (c) follows from (11); this completes the proof of the lemma. □

Remark 2. Note that the idea of using equitable coloring of graphs to partition a set of dependent random variables into disjoint sets having approximately equal number of independent random variables is not new and has been previously used by researchers to derive deviation bounds for sums of dependent random variables that exhibit limited dependence – see, e.g., [9].

Loosely speaking (and for k sufficiently large), Lemma 1 says that for any subset T of cardinality $3m$, if an IID submatrix satisfies (2) with probability at least $1 - e^{-O(k)}$, then a Toeplitz submatrix with entries drawn independently from the same distribution also satisfies (2) with probability at least $1 - e^{-O(k/m^2)}$. The next step of lower bounding the probability that Toeplitz submatrices satisfy (2) for *all* subsets T of cardinality $3m$ follows trivially from this result by union bounding over the choice of $\binom{n}{3m}$ such subsets. Below, we specifically state the implications of this union bound for the distributions given in (7).

Theorem 1. *Suppose that n, m are given, and let A be a $k \times n$ Toeplitz matrix of the form given in* (4)*, where the entries $\{a_i\}_{i=1}^{n+k-1}$ are drawn independently from one of the probability distributions given in* (7)*. Then, there exist constants $c_1, c_2 > 0$ depending only on δ_{3m} such that for any $k \ge c_1\, m^3 \ln(n/m)$, A satisfies RIP of order $3m$ for every $\delta_{3m} \in (0, 1/3)$ with probability at least*

$$1 - e^{-c_2 k/m^2}. \qquad (15)$$

Proof. Fix $\delta_{3m} \in (0, 1/3)$. From Lemma 1 and (8), A satisfies (2) for any $T \subset \{1, 2, \ldots, n\}$ of cardinality $3m$ with probability at least

$$1 - e^{-c_0 \lfloor k/q \rfloor + 3m \ln(12/\delta_{3m}) + \ln(2) + \ln(q)} \;\ge\; 1 - e^{-c_0 k/9m^2 + 3m \ln(12/\delta_{3m}) + \ln(9m^2) + \ln(2) + c_0},$$

and there are $\binom{n}{3m} \le (en/3m)^{3m}$ such subsets. Consequently, union bounding over these subsets yields that A satisfies RIP of order $3m$ with probability at least

$$1 - e^{-c_0 k/9m^2 + 3m[\ln(n/3m) + \ln(12/\delta_{3m}) + 1] + \ln(9m^2) + \ln(2) + c_0}. \qquad (16)$$

Next, fix $c_2 > 0$ and pick $c_1 > 27c_3/(c_0 - 9c_2)$, where $c_3 = \ln(12/\delta_{3m}) + \ln(2) + c_0 + 4$. Then, for any $k \ge c_1\, m^3 \ln(n/m)$, the exponent in the exponential in (16) is upper bounded by $-c_2 k/m^2$ and this completes the proof of the theorem. □

Before discussing natural extensions to Toeplitz CS matrices, it is instructional to compare the result of Theorem 1 with that for IID CS matrices. Specifically, previous work has shown that IID CS matrices generated from the distributions given in (7) satisfy RIP of order $3m$ for every $\delta_{3m} \in (0, 1/3)$ with probability $\ge 1 - e^{-c_2' k}$ provided $k \ge c_1'\, m \ln(n/m)$, where $c_1', c_2' > 0$ are constants depending only on δ_{3m} – see, e.g., [3, 6]. It might be tempting, therefore, to conclude that reduction in the number of DoFs of a Toeplitz matrix from $O(kn)$ to $O(n)$ results in a factor of $O(m^2)$ increase in the required number of observations. One needs to apply caution, however, as Theorem 1 bounds the worst case performance of Toeplitz CS matrices for *all* m-sparse signals and it might very well be that this oversampling is not required for *most* signals in the class. Extensive simulations carried out for a number of m-sparse signals using IID and Toeplitz matrices of equal dimensions, in fact, support this intuition. It is also interesting to note that somewhat similar numerical results (without any performance guarantees) have been reported in [10] in the context of *random filters*.

3. EXTENSIONS

In this section, we discuss natural extensions of the result of Section 2 to circulant and left-shifted Toeplitz-structured matrices. Further, we also describe how the results for Toeplitz-structured CS matrices lend themselves to (i) identification of LTI systems having sparse impulse responses; and (ii) recovery of signals that are either piecewise constant (PWC) or sparse in the Haar wavelet domain.

3.1. Circulant CS Matrices

Theorem 2. *Suppose that n, m are given, and let A be a $k \times n$ (partial) circulant matrix of the form*

$$A = \begin{bmatrix} a_n & a_{n-1} & \ldots & a_2 & a_1 \\ a_1 & a_n & \ldots & a_3 & a_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_{k-1} & a_{k-2} & \multicolumn{2}{c}{\ldots\ldots\ldots} & a_k \end{bmatrix}, \qquad (17)$$

where the entries $\{a_i\}_{i=1}^n$ are drawn independently from one of the distributions given in (7). Then, there exist constants $c_1'', c_2'' > 0$ depending only on δ_{3m} such that for any $k \geq c_1'' m^3 \ln(n/m)$, A satisfies RIP of order $3m$ for every $\delta_{3m} \in (0, 1/3)$ with probability at least

$$1 - e^{-c_2'' k/m^2}. \tag{18}$$

Sketch of Proof. Using the notational convention of Lemma 1, note that for any fixed $T \subset \{1, 2, \ldots, n\}$ with $|T| = 3m$, the i-th row $A_{T,i}$ of a circulant submatrix A_T can at most be dependent with $2|T|(|T| - 1) \leq 6m(3m - 1)$ other rows of A_T because of the circulant nature of A, i.e., the maximum degree Δ of the dependency graph (as defined in Lemma 1) of A_T is given by $\Delta \leq 6m(3m-1)$. The rest of the proof follows along exactly the same lines as for the Toeplitz case. □

3.2. Left-shifted Toeplitz and Circulant CS Matrices

The results of Theorem 1 and 2 apply equally well to left-shifted Toeplitz and circulant matrices of the form

$$\begin{bmatrix} a_1 & a_2 & \ldots & a_{n-1} & a_n \\ a_2 & a_3 & \ldots & a_n & a_{n+1} \\ \vdots & / & / & \vdots & \vdots \\ a_k & \ldots & \ldots & a_{n+k-2} & a_{n+k-1} \end{bmatrix}, \tag{19}$$

and

$$\begin{bmatrix} a_1 & a_2 & \ldots & a_{n-1} & a_n \\ a_2 & a_3 & \ldots & a_n & a_1 \\ \vdots & / & / & \vdots & \vdots \\ a_k & \ldots & \ldots & a_{k-2} & a_{k-1} \end{bmatrix}, \tag{20}$$

because the submatrices of such matrices also give rise to dependency graphs with maximum degrees upper bounded by $3m(3m-1)$ and $6m(3m-1)$, respectively.

3.3. System Identification

The area of estimation of the impulse response of an LTI system from the knowledge of its input and output signals, commonly termed as system identification, is of considerable importance in signal processing because of its applicability to a wide range of problems – see, e.g., [11, 12]. In the case of a finite impulse response (FIR) LTI system, this typically involves probing the system with a (known) white noise sequence of duration orders of magnitude greater than that of the impulse response [13], which may be prohibitive because of the delay incurred in solving for the impulse response and the difficulty of generating a truly white noise sequence. For the purposes of deconvolving an LTI system having a sparse impulse response, however, a more promising alternative is to appeal to the results of Section 2.

As an illustration, let $x[\ell]$ be an m-sparse impulse response of an LTI system (of duration n) and $a[\ell]$ be an IID sequence of duration $(n + k - 1)$ that has been drawn from one of the probability distributions given in (7). Then, probing the given system with $a[\ell]$ yields $y[\ell] = a[\ell] * x[\ell]$ and the theory of CS along with Theorem 1 guarantees that, with high probability, $x[\ell]$ can be exactly recovered by solving the convex program

$$x[\ell] \;=\; \arg\left(\min_{z \in \mathbb{R}^n} \|z\|_1 \quad \text{subject to} \quad y = Az\right), \tag{21}$$

where, in this case, $y = \begin{bmatrix} y[n-1] \\ y[n] \\ \vdots \\ y[n+k-2] \end{bmatrix}$, and

$$A = \begin{bmatrix} a[n-1] & a[n-2] & \ldots & a[1] & a[0] \\ a[n] & a[n-1] & \ldots & a[2] & a[1] \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a[n+k-2] & a[n+k-3] & \ldots & \ldots & a[k-1] \end{bmatrix}.$$

3.4. Beyond Sparse Signals

We have proven above that Toeplitz (and circulant) matrices, having entries drawn independently from probability distributions that yield IID CS matrices, satisfy RIP of order $3m$ with high probability. Often, we are interested in signals that are sparse in some transform domain $\Psi \neq I$, i.e., $x = \Psi\theta$ and $\theta \in \mathbb{R}^n$ is m-sparse, in which case it is required that the product matrix $A\Psi$ satisfies RIP of order $3m$ for successful recovery of θ (and hence x). This is indeed the case when A happens to be an IID CS matrix and Ψ is any orthonormal basis [6]. Toeplitz matrices, however, seem to lack this *universality* property because of their highly structured nature. Nevertheless, the results of Section 2 can still be leveraged to design CS matrices for *fixed* transformations to retain some of the benefits of Toeplitz-structured CS matrices such as generation of only $O(n)$ independent random variables, and faster acquisition and reconstruction algorithms.

As an illustration, let x be an m-piece PWC signal; such a signal can be written as $x = L\theta$, where $\theta \in \mathbb{R}^n$ is m-sparse and $L \in \mathbb{R}^{n \times n}$ – the discrete integral transform – is given by

$$L \;=\; \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 1 & 1 & \ddots & 0 \\ 1 & \ddots & \ddots & 0 \\ 1 & \ldots & 1 & 1 \end{bmatrix}. \tag{22}$$

Further, let $\{a_i\}_{i=1}^{n+k-1}$ be a sequence of random variables drawn independently from a distribution that yields an IID CS matrix and $A_L \in \mathbb{R}^{k \times n}$ be the cascade of a $k \times n$ Toeplitz matrix A and the $n \times n$ differencing operator

$$D \;=\; \begin{bmatrix} 1 & 0 & & \\ -1 & 1 & \ddots & \\ & \ddots & \ddots & 0 \\ & & -1 & 1 \end{bmatrix}, \tag{23}$$

that is,

$$A_L \;=\; \begin{bmatrix} (a_n - a_{n-1}) & \ldots & (a_2 - a_1) & a_1 \\ (a_{n+1} - a_n) & \ldots & (a_3 - a_2) & a_2 \\ \vdots & \ddots & \vdots & \vdots \\ (a_{n+k-1} - a_{n+k-2}) & \ldots & (a_{k+1} - a_k) & a_k \end{bmatrix}. \tag{24}$$

Then, by construction, (i) A_L has only $(n + k - 1)$ DoFs; (ii) multiplication with $A_L = AD$ requires only $O(n \log_2(n))$ operations; and (iii) the product matrix $A_L L = ADL = A$ is a Toeplitz CS matrix and consequently, satisfies RIP with high probability. Likewise, if x happened to be m-sparse in the Haar wavelet domain, i.e., $\Psi = W^{-1}$ (the inverse Haar wavelet transform matrix), then a CS matrix of the form $A_W = AW$ would also have these three properties.

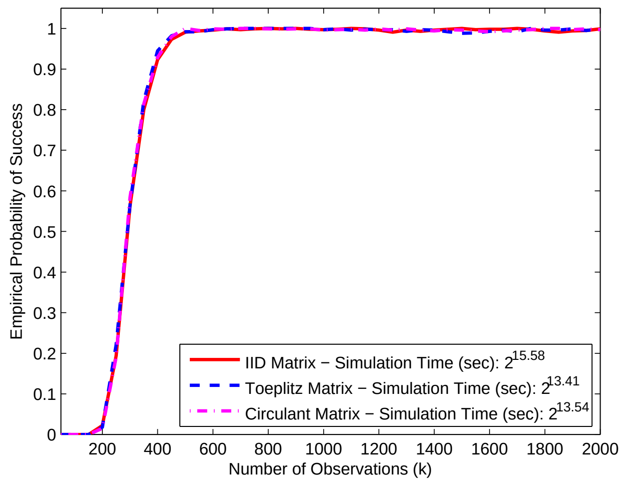

Fig. 1. Empirical probability of success as a function of number of observations k $(n = 2048, m = 20)$.

4. NUMERICAL RESULTS

In this section, we numerically compare the performance of Toeplitz and circulant CS matrices to that of IID ones. The experimental setup involves generating a length $n = 2048$ signal with randomly placed $m = 20$ non-zero entries drawn independently from $\mathcal{N}(0, 1)$. Each such generated signal is sampled using $k \times n$ IID, Toeplitz and circulant matrices with entries drawn independently from the Bernoulli $= \left\{+\sqrt{\frac{1}{k}} \text{ with probability } \frac{1}{2}, -\sqrt{\frac{1}{k}} \text{ with probability } \frac{1}{2}\right\}$ distribution and reconstructed using the gradient projection algorithm described in [5], where matrix multiplications are carried out using FFT in the case of Toeplitz and circulant observation matrices. *Success* is declared if the algorithm exactly recovers the signal (taking into account machine precision errors), and the empirical probability of success for each value of k is determined by repeating this process 1000 times and calculating the fraction of successes. While running this experiment for all $x \in \mathbb{R}^n$ or even all $\binom{2048}{20}$ unique sparsity patterns does not seem possible, simulation results show that for a large number of synthesized signals (and for the reasons described earlier), Toeplitz and circulant matrices perform as well as IID ones in terms of the empirical probability of success. We plot the empirical probability of success versus number of observations k for one such signal in Fig. 1.

5. CONCLUSIONS

In this paper, we have shown that Toeplitz-structured matrices with entries drawn independently from probability distributions that yield IID CS matrices are also sufficient to recover undersampled sparse signals. The use of such matrices is a desirable alternative for a number of application areas because it greatly reduces the computational and storage complexity in large-dimensional problems.[4] Our proof technique uses the celebrated Hajnal-Szemerédi theorem on equitable coloring of graphs to partition a $k \times |T|$ Toeplitz-structured submatrix A_T into roughly $O(m^2)$ IID submatrices having dimensions approximately equal to $O(k/m^2) \times |T|$. It is interesting to note that certain special types of graphs can be equitably colored using far fewer colors – see, e.g., [9]. This implies interesting extensions to the work presented here. First, to the best of our knowledge, the minimum number of colors needed to equitably color the structured dependency graphs considered here is still unknown and sharper coloring results could provide a reduction in the required number of observations in the above results. In addition, it might be possible to obtain other structured CS matrices by starting with dependency graphs for which sharp equitable coloring results are known and working backwards.

[4]We refer the reader to [14] for a different take on solving the problem of computational and storage complexity in CS applications.

6. REFERENCES

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

[2] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

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

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

[5] M. Figueiredo, R. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," pre-print. [Online]. Available: `http://www.lx.it.pt/~mtf/GPSR/`

[6] R. Baraniuk, M. Davenport, R. A. DeVore, and M. B. Wakin, "A simple proof of the restricted isometry property for random matrices," pre-print. [Online]. Available: `http://www.dsp.ece.rice.edu/cs/`

[7] D. B. West, *Introduction to Graph Theory*, Prentice Hall, Upper Saddle River, NJ, 2000.

[8] A. Hajnal and E. Szemerédi, "Proof of a conjecture of P. Erdős," in *Combinatorial Theory and its Application*, P. Erdős, A. Rényi, and V. T. Sós, Eds., pp. 601–623. North-Holland, Amsterdam, 1970.

[9] S. Pemmaraju, "Equitable coloring extends Chernoff-Hoeffding bounds," in *Proc. RANDOM-APPROX 2001*, Berkeley, CA, Aug. 2001, pp. 285–296.

[10] J. A. Tropp, M. B. Wakin, M. F. Duarte, D. Baron, and R. G. Baraniuk, "Random filters for compressive sampling and reconstruction," in *Proc. IEEE Intl. Conf. Acoustics, Speech and Sig. Proc. (ICASSP'06)*, Toulouse, France, May 2006, pp. 872–875.

[11] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, Englewood Cliffs, NJ, first edition, 1987.

[12] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, NJ, fourth edition, 2001.

[13] L. R. Rabiner, R. E. Crochiere, and J. B. Allen, "FIR system modeling and identification in the presence of noise and with band-limited inputs," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 26, no. 4, pp. 319–333, Aug. 1978.

[14] E. J. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," pre-print. [Online]. Available: `http://www.acm.caltech.edu/~emmanuel/publications.html`