Geometry of Random Toeplitz-Block Sensing Matrices: Bounds and Implications for Sparse Signal Processing

Waheed U. Bajwa

Department of Electrical and Computer Engineering
Rutgers, The State University of New Jersey
Piscataway, NJ 08854

ABSTRACT

A rich body of literature has emerged during the last decade that seeks to exploit the sparsity of a signal for a reduction in the number of measurements required for various inference tasks. Much of the initial work in this direction has been for the case when the measurements correspond to a projection of the signal of interest onto the column space of (sub)Gaussian and subsampled Fourier matrices. The physics in a number of applications, however, dictates the use of "structured" matrices for measurement purposes. This has led to a recent push in the direction of structured measurement (or sensing) matrices for inference of sparse signals. This paper complements some of the recent work in this direction by studying the geometry of Toeplitz-block sensing matrices. Such matrices are bound to arise in any system that can be modeled as a linear, time-invariant (LTI) system with multiple inputs and single output. The reported results therefore should be of particular benefit to researchers interested in exploiting sparsity in LTI systems with multiple inputs.

Keywords: Compressed sensing, coherence, linear systems, model selection, sparsity, spectral norm, Toeplitz blocks

1. INTRODUCTION

Consider the classical setup in which a signal $x \in \mathbb{R}^N$ is measured according to the linear model $y = Ax + w$. Here, A is an $n \times N$ *measurement* or *sensing matrix*, while $w \in \mathbb{R}^n$ represents additive noise in the measurement system. Despite its simplicity, the linear model suffices to capture the measurement process in a surprisingly large number of engineering applications. The signal processing challenge under this measurement model in an application then is to infer (certain characteristics of) x from the measurements y.

In the absence of any prior (statistical or geometric) knowledge about x, elementary linear algebra dictates that inference of x requires the number of measurements n to be at least equal to N, the extrinsic dimension of the signal. During the last decade, however, it has been successfully argued that the linear dependence of the number of measurements on N is too stringent for signals that can be approximated by a small number of nonzero coefficients in a basis. In particular, it is now a well-known fact that inference of signals that are exactly k-sparse in the canonical basis, $\|x\|_0 := \#\{i : x_i \neq 0\} \leq k$, can be carried out in a computationally efficient manner using only $n \approx \Omega(k \log(N))$ carefully chosen measurements.* The most common inference tasks in this sparse setting include (i) recovery/estimation of x from y, studied under the rubric of "compressed sensing" [1, 2]; (ii) estimation of the locations of nonzero entries of x from y, studied under the monikers of "model selection" [3–5] and "sparsity pattern recovery" [6, 7]; and (iii) testing for the presence of x in noise [8, 9].

Our focus in this paper is on first two of the aforementioned inference tasks. In both these cases, there is a large body of existing literature that characterizes the performance of various optimization-based and greedy methods as a function of certain geometrical properties of the sensing matrix. These properties include the *restricted isometry property* [10], *restricted eigenvalue property* [11], *irrepresentable condition* [3], *incoherence condition* [6], and variants of the *coherence property* [4, 5, 12, 13]. Much of the initial push in the literature has been on leveraging these properties to establish that sensing matrices such as (sub)Gaussian matrices and subsampled Fourier matrices require near-optimal scaling of the number of measurements: $n \approx \Omega(k \log(N))$ [14]. However, the measurement process in a number of applications, such as

E-mail: waheed.bajwa@rutgers.edu. This work was supported in part by the startup funds to W. U. Bajwa from the School of Engineering, Rutgers University.

*Recall the big-O notation: $f(z) = O(g(z))$ if there exist positive C and z_0 such that for all $z > z_0$, $f(z) \leq Cg(z)$. Also, $f(z) = \Omega(g(z))$ if $g(z) = O(f(z))$, and $f(z) = \Theta(g(z))$ if $f(z) = O(g(z))$ and $g(z) = O(f(z))$.

channel estimation [15], radar [16], seismic imaging [17], and optical imaging [18, 19], cannot be adequately described by these "canonical" sensing matrices. This realization has led to a recent push in the direction of studying the geometry of "structured" sensing matrices for inference of sparse signals. In this paper, we complement the existing work on structured sensing matrices by studying the geometry of Toeplitz-block sensing matrices for sparse signal processing. In general, Toeplitz-block sensing matrices are bound to arise in any system that can be modeled as a linear, time-invariant (LTI) system with multiple inputs and single output. The results reported in this paper therefore should be of particular benefit to researchers interested in exploiting sparsity in LTI systems with multiple inputs.

1.1 Our Contribution

The Toeplitz-block sensing matrix A studied in this paper comprises m (partial) Toeplitz matrices with n rows and p columns each. Specifically, we have $A = \begin{bmatrix} T^1 & T^2 & \dots & T^m \end{bmatrix} \in \mathbb{R}^{n \times N}$ with $N = mp$ and

$$T^i = \begin{bmatrix} s^i_p & s^i_{p-1} & \dots & s^i_2 & s^i_1 \\ s^i_{p+1} & s^i_p & \dots & s^i_3 & s^i_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ s^i_{n+p-1} & s^i_{n+p-2} & \dots & s^i_{n+1} & s^i_n \end{bmatrix}, \quad i = 1, \dots, m, \tag{1}$$

where the sequence $\{s^i_j\}_{j=1}^{n+p-1}$ is the *seed* to the $n \times p$ Toeplitz matrix T^i. The $n \times N$ matrix A in this case is completely described by the $m(n+p-1)$ sequence values $\{s^i_j\}_{i,j}$. Note that extensions of this block setup to the case of (partial) Hankel blocks, partial circulant blocks and additional stacking of blocks in the vertical direction are straightforward and therefore will not be discussed in this paper.

The main contribution of this paper is to establish that Toeplitz-block sensing matrices generated from a random seed achieve near-optimal scaling of the number of measurements, $n \approx \Omega(k \log(N))$, for both compressed sensing and model selection problems. The recipe for these results was implicitly provided for Toeplitz-block sensing matrices comprising *full* Toeplitz matrices in our earlier work on multiuser detection [20]. In this paper, we explicitly report these results for the case of *partial* Toeplitz blocks. The key to our results is a characterization of two geometric measures of Toeplitz-block sensing matrices, namely, the *worst-case coherence* and the *spectral norm*. The worst-case coherence of A, defined as

$$\mu_A := \max_{\substack{i,j \in \{1,\dots,N\} \\ i \neq j}} \frac{|\langle a_i, a_j \rangle|}{\|a_i\|_2 \|a_j\|_2}, \tag{2}$$

is a measure of the worst-case similarity between the columns $\{a_i,\, i = 1, \dots, N\}$ of A. The spectral norm of A, on the other hand, can be heuristically regarded as a measure of the similarity between the rows of A, defined as $\|A\|_2 := \sqrt{\lambda_{\max}(A^{\mathrm{T}} A)}$ (i.e., the maximum singular value of A). Specifically, we derive upper bounds on the worst-case coherence and the spectral norm of random Toeplitz-block sensing matrices in Section 2. We then leverage recent results reported in [4, 13, 21] and the results of Section 2 to obtain the scaling relationship $n \approx \Omega(k \log(N))$ for Toeplitz-block sensing matrices in Section 3.

1.2 Relationship to Previous Work

In relation to previous works, our work is a generalization of the earlier work on Toeplitz and circulant matrices in compressed sensing [18, 22–26]. To the best of our knowledge, however, the only works in the sparse signal processing literature that have considered block sensing matrices with Toeplitz or circulant blocks are [27, 28] and [29]. Both [27] and [29] require superlinear scaling (in k) of the number of measurements. The scaling requirements in [28] are linear (modulo a polylogarithmic factor) and its setup is also the one most closely related to our setup. The biggest difference between [28] and this paper is that [28] requires each of its (quasi-circulant) blocks to have full row rank (i.e., $n \geq p$).

2. GEOMETRY OF RANDOM TOEPLITZ-BLOCK SENSING MATRICES

In this section, we specify the column and row geometry of random Toeplitz-block sensing matrices in terms of the worst-case coherence and the spectral norm. We begin with a characterization of μ_A of A in the following.

THEOREM 2.1 (BOUND ON THE WORST-CASE COHERENCE). *Suppose $m \geq 2$ and the $n \times N(= mp)$ Toeplitz-block matrix A is generated from the sequence $\{s_j^i,\ i = 1, \dots, m,\ j = 1, \dots, n+p-1\}$ with each s_j^i drawn independently from a* $\mathrm{Binary}(+1/\sqrt{n}, -1/\sqrt{n})$ *distribution. Then the worst-case coherence of A satisfies*

$$\mu_A \leq \sqrt{\frac{12\log(N)}{n}} \tag{3}$$

with probability exceeding $1 - 4N^{-1}$.

Proof. In order to prove this theorem, notice that the worst-case coherence of A can be expressed in terms of its constituent block matrices as follows:

$$\mu_A = \max\left\{\max_{i=1,\dots,m} \mu_{T^i},\ \max_{\substack{i,j=1,\dots,m\\ i\neq j}} \left\|{T^i}^{\mathrm{T}} T^j\right\|_{\max}\right\}, \tag{4}$$

where μ_{T^i} denotes the worst-case coherence of T^i and $\|\cdot\|_{\max}$ denotes the $\max$ norm of a matrix. We next pick a $\delta \in (0,1)$ and appeal to Theorem 4 in [24], which dictates that the worst-case coherence of a (partial) Toeplitz matrix generated from a random binary sequence satisfies

$$\Pr\left(\mu_{T^i} \geq \delta\right) \leq 2p(p-1)\exp\left(-\frac{n\delta^2}{4}\right). \tag{5}$$

We therefore trivially get from the union bound that

$$\Pr\left(\max_{i=1,\dots,m} \mu_{T^i} \geq \delta\right) \leq 2mp(p-1)\exp\left(-\frac{n\delta^2}{4}\right). \tag{6}$$

In order to provide a similar probabilistic bound on $\max_{\substack{i,j=1,\dots,m\\ i\neq j}} \left\|{T^i}^{\mathrm{T}} T^j\right\|_{\max}$, we define $G^{i,j} = {T^i}^{\mathrm{T}} T^j$ for any fixed i, j with $i \neq j$ and express the (k, ℓ) entry of $G^{i,j}$ as

$$G^{i,j}_{k,\ell} = \sum_{q=1}^{n} s^i_{p+q-k}\, s^j_{p+q-\ell} = \sum_{q=1}^{n} \widetilde{s}_q, \tag{7}$$

where we have suppressed the dependence of (i, j, k, ℓ) on $\widetilde{s}_q = s^i_{p+q-k}\, s^j_{p+q-\ell}$ for ease of notation. Since $i \neq j$, it is easy to argue that the random sequence $\{\widetilde{s}_q\}$ is independent and identically distributed (i.i.d.) as $\mathrm{Binary}(+1/n, -1/n)$. We can therefore apply the Hoeffding inequality [30] to the sum in (7) and obtain

$$\Pr\left(\left|G^{i,j}_{k,\ell}\right| \geq \delta\right) \leq 2\exp\left(-\frac{n\delta^2}{2}\right). \tag{8}$$

We now first apply the union bound for p^2 distinct entries of $G^{i,j}$ to obtain a probabilistic bound for $\left\|{T^i}^{\mathrm{T}} T^j\right\|_{\max}$, followed by the union bound for $m(m-1)/2$ distinct $\left\|{T^i}^{\mathrm{T}} T^j\right\|_{\max}$ to finally obtain

$$\Pr\left(\max_{\substack{i,j=1,\dots,m\\ i\neq j}} \left\|{T^i}^{\mathrm{T}} T^j\right\|_{\max} \geq \delta\right) \leq m(m-1)p^2\exp\left(-\frac{n\delta^2}{2}\right). \tag{9}$$

A final application of the union bound over the two probability events in (6) and (9) along with the mild assumption $m \geq 2$ then results in

$$\Pr\left(\mu_A \geq \delta\right) \leq 4m(m-1)p^2\exp\left(-\frac{n\delta^2}{4}\right). \tag{10}$$

The proof of the theorem now follows by taking $\delta = \sqrt{\frac{12\log(mp)}{n}}$. $\square$

Note that the case of $m = 1$ corresponds to the canonical partial Toeplitz matrix and has been addressed in our earlier work [24]. The normalization factor $1/\sqrt{n}$ in Theorem 2.1 is meant to maintain uniformity with the traditional literature that assumes matrices with unit-norm columns. Any change in this normalization factor, however, does not affect (3) because of the normalization inherent in the definition of the worst-case coherence. Likewise, the choice of binary distribution in this theorem for the seed sequence is not critical. Rather, the result can be readily generalized to sub-Gaussian distributions in much the same way as in [24]. Finally, note that Theorem 2.1 is closely related to Lemma 1 in our earlier work [20], which bounds the worst-case coherence of a random Toeplitz-block matrix with each block being a full Toeplitz matrix with $(n + p - 1)$ rows and p columns.

It is important to point out that Theorem 2.1 alone can be used to provide a crude scaling of the number of measurements of Toeplitz-block matrices. This is because the restricted isometry property (RIP) of a matrix is related to its worst-case coherence through the Geršgorin circle theorem [31]: A sensing matrix A with worst-case coherence μ_A satisfies the RIP of order $k = O\left(\mu_A^{-1}\right)$. This relationship between the RIP and the worst-case coherence, for example, has been exploited in [23,24] and [29] to provide guarantees for Toeplitz-structured sensing matrices. In our case, a similar approach will dictate that a random Toeplitz-block matrix satisfies the RIP of order k as long as $n = \Omega(k^2 \log(N))$. Note that this superlinear scaling of the number of measurements is not due to a loose bound in Theorem 2.1. Indeed, the Welch bound [32] tells us that the bound in Theorem 2.1 is tight up to the $\log(N)$ factor. Rather, the scaling limitation is a direct consequence of approaching the RIP through the worst-case coherence alone. In order to improve upon the $n = \Omega(k^2 \log(N))$ scaling for random Toeplitz-block matrices, we therefore need a handle on another geometric measure of A, namely, the spectral norm. The following result is mainly a consequence of Lemma 2 in our earlier work [20], with changes to account for the partial Toeplitz blocks in our setup.

THEOREM 2.2 (BOUND ON THE SPECTRAL NORM). *Suppose* $n \leq C_1^2 m$ *for some* $C_1 > 0$ *and the* $n \times N(= mp)$ *Toeplitz-block matrix* A *is generated from the sequence* $\{s_j^i,\ i = 1, \dots, m,\ j = 1, \dots, n + p - 1\}$ *with each* s_j^i *drawn independently from a* $\mathrm{Binary}(+1/\sqrt{n}, -1/\sqrt{n})$ *distribution. Then the spectral norm of* A *satisfies*

$$\|A\|_2 \leq 26(1 + C_1)\sqrt{\frac{N}{n}} \tag{11}$$

with probability exceeding $1 - \exp\left(-\frac{\sqrt{nm}}{8} + \log(p)\right)$.

Proof. The spectral norm is invariant under column-interchange operations. We therefore define another $n \times N$ block matrix $\widetilde{A} := \begin{bmatrix}\widetilde{A}^1 & \widetilde{A}^2 & \dots & \widetilde{A}^p\end{bmatrix}$ such that the jth block $\widetilde{A}^j$ is an $n \times m$ matrix comprising the jth column of each of the original Toeplitz blocks $\{T^i\}_{i=1}^m$ of A. Since $\widetilde{A}$ is related to A through column exchanges, we trivially have $\|A\|_2 = \|\widetilde{A}\|_2$. We can now bound $\|\widetilde{A}\|_2$ in terms of the spectral norms of its individual blocks as (see, e.g., Lemma 2 in [20])

$$\|\widetilde{A}\|_2 \leq \sqrt{p} \max_{j=1,\dots,p} \|\widetilde{A}^j\|_2. \tag{12}$$

In order to complete the proof, notice that each matrix $\widetilde{A}^j$ has i.i.d. entries distributed as $\mathrm{Binary}(+1/\sqrt{n}, -1/\sqrt{n})$. It therefore follows from Proposition 2.4 in [33] and the assumption $n \leq C_1^2 m$ that $\|\widetilde{A}^j\|_2 \leq 26(1+C_1)\sqrt{\frac{m}{n}}$ with probability exceeding $1 - \exp\left(-\frac{\sqrt{nm}}{8}\right)$. The proof of the theorem now follows by taking the union bound for the p distinct $\|\widetilde{A}^j\|_2$. $\square$

It is important to point our here that the scaling $\|A\|_2 = O\left(\sqrt{N/n}\right)$ in (2.2) is tight. This can be argued by recalling that the trace of a matrix is equal to the sum of its eigenvalues and noting that

$$\lambda_{\max}\left(A^{\mathrm{T}}A\right) \geq \frac{\sum_{i=1}^n \lambda_i\left(A^{\mathrm{T}}A\right)}{n} = \frac{\mathrm{trace}\left(A^{\mathrm{T}}A\right)}{n} = \frac{N}{n}. \tag{13}$$

The only room of improvement in Theorem 2.2 is the condition $n \leq C_1^2 m$, which may be restrictive for certain applications.

3. TOEPLITZ-BLOCK SENSING MATRICES FOR SPARSE SIGNAL PROCESSING

To the best of our knowledge, the significance of the worst-case coherence and the spectral norm of sensing matrices in the context of recovery of sparse signals was originally highlighted in [13, 21], while their significance in the context of model selection was first highlighted in [4]. In this section, we simply take the bounds obtained in Section 2 and leverage the key results of [4, 13, 21] to evaluate the performance of Toeplitz-block sensing matrices for both (noiseless) recovery of sparse signals and (noisy) model selection. We begin with the result for recovery of sparse signals using Toeplitz-block sensing matrices. The following theorem is due to Tropp and follows from combining results in Section 2 along with the ones reported in [21] and [13].

THEOREM 3.1 (TOEPLITZ-BLOCK SENSING MATRICES FOR RECOVERY OF SPARSE SIGNALS). *Suppose $y = Ax$ for a k-sparse signal $x \in \mathbb{R}^N$ and an estimate of x is obtained from y using the following optimization program:*

$$\widehat{x}_r = \arg\min_{z \in \mathbb{R}^N} \|z\|_1 \quad \textit{subject to} \quad y = Az. \tag{P_1}$$

Then under the assumptions that (i) the nonzero entries of x are independently distributed with zero median, (ii) the support set $\mathcal{S} := \{i : x_i \neq 0\}$ has a uniform distribution over all $\binom{N}{k}$ k-subsets of $\{1, 2, \dots, N\}$, and (iii) A is a random Toeplitz-block sensing matrix defined earlier with $m \geq 2$ and $n \leq C_1^2 m$ for some $C_1 > 0$, the solution of (P_1) satisfies $\widehat{x}_r = x$ with probability exceeding $1 - N^{-2\log(2)} - 5N^{-1} - \exp\left(-\frac{\sqrt{nm}}{8} + \log(p)\right)$ provided

$$n > \max\Big\{c_{r,1} k \log(N), c_{r,2} k \log^2(N), c_{r,3} \log^3(N)\Big\}. \tag{14}$$

Here, $c_{r,1}, c_{r,2}$, and $c_{r,3}$ are positive numerical constants independent of the problem parameters.

The proof of this theorem is omitted here, but it follows from simple but tedious algebraic manipulations of the results of [21] and [13] (see, e.g., Theorem 11 in [34]) along with the union bound involving the results of Section 2. Theorem 3.1 is quite powerful in the sense that it allows linear scaling (in k) of the number of measurements (modulo a logarithmic factor) for recovery of sparse signals using Toeplitz-block sensing matrices. It is also worth pointing out some limitations of this theorem. First, it guarantees recovery of sparse signals in an average sense, rather than the worst-case guarantees implied by the RIP-based analysis. Second, it does not have a straightforward generalization in the presence of noise. The first limitation of course is not that critical in many applications. The second limitation can be overcome if one shifts the focus from signal recovery to exact model selection, as shown below. The following theorem is due to Candès and Plan and follows from combining results in Section 2 along with the ones reported in [4].

THEOREM 3.2 (TOEPLITZ-BLOCK SENSING MATRICES FOR MODEL SELECTION). *Suppose $y = Ax + w$ for a k-sparse signal $x \in \mathbb{R}^N$, the three assumptions stated in Theorem 3.1 hold, and an estimate of x is obtained from y using the following optimization program:*

$$\widehat{x}_m = \arg\min_{z \in \mathbb{R}^N} \frac{1}{2}\|y - Az\|_2 + \tau\|z\|_1. \tag{$P_{1,2}$}$$

Then under the additional assumptions that (i) the additive noise w is distributed as $\mathcal{N}(0, \sigma^2 I)$ and (ii) the smallest (in magnitude) nonzero entry of the sparse signal x obeys $\min_{i \in \mathcal{S}} |x_i| > 8\sqrt{2\sigma^2 \log(N)}$, the solution of $(P_{1,2})$ computed with $\tau = 2\sqrt{2\sigma^2 \log(N)}$ satisfies $\{i : \widehat{x}_{m,i} \neq 0\} = \mathcal{S}$ with probability exceeding $1 - O(N^{-1}) - \exp\left(-\frac{\sqrt{nm}}{8} + \log(p)\right)$ provided

$$n > \max\Big\{c_{m,1} k \log(N), c_{m,2} \log^3(N)\Big\}. \tag{15}$$

Here, $c_{m,1}$ and $c_{m,2}$ are positive numerical constants independent of the problem parameters.

The proof of this theorem is also omitted here, since it is a straightforward application of Theorem 2.1 and Theorem 2.2 to the results of [4]. Similar to the case of Theorem 3.1, we have from Theorem 3.2 that exact model selection using Toeplitz-block sensing matrices in the average case only requires $n \approx \Omega(k \log(N))$ scaling of the number of measurements.

4. CONCLUSIONS

In this paper, we have studied the geometry of random Toeplitz-block sensing matrices in terms of the measures of the worst-case coherence and the spectral norm. The near-tightness of these measures derived in the paper imply that Toeplitz-block sensing matrices require roughly linear scaling (with sparsity) of the number of measurements for both sparse signal recovery and model selection problems. We have also commented on the relationship between the worst-case coherence and the restricted isometry property (RIP), which only results in a quadratic scaling of the number of measurements. The difference in the two approaches of course is that our results report average behavior while the RIP-based results report worst-case behavior. Recently, another geometric measure of *average coherence* has been introduced in [5] that helps eliminate the statistical prior on the nonzero entries of sparse signals. Bounding the average coherence of random Toeplitz-block sensing matrices, however, is a little more involved than bounding the worst-case coherence; this remains a focus of our ongoing research.

REFERENCES

[1] Candès, E. J., Romberg, J., and Tao, T., "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory* **52**, 489–509 (Feb. 2006).

[2] Donoho, D. L., "Compressed sensing," *IEEE Trans. Inform. Theory* **52**, 1289–1306 (Apr. 2006).

[3] Zhao, P. and Yu, B., "On model selection consistency of lasso," *J. Machine Learning Res.* **7**, 2541–2563 (Nov. 2006).

[4] Candès, E. J. and Plan, Y., "Near-ideal model selection by ℓ_1 minimization," *Ann. Statist.* **37**, 2145–2177 (Oct. 2009).

[5] Bajwa, W. U., Calderbank, R., and Jafarpour, S., "Why Gabor frames? Two fundamental measures of coherence and their role in model selection," *J. Commun. Netw.* **12**, 289–307 (Aug. 2010).

[6] Wainwright, M. J., "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (lasso)," *IEEE Trans. Inform. Theory* **55**, 2183–2202 (May 2009).

[7] Fletcher, A. K., Rangan, S., and Goyal, V. K., "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Inform. Theory* **55**, 5758–5772 (Dec. 2009).

[8] Duarte, M. F., Davenport, M. A., Wakin, M. B., and Baraniuk, R. G., "Sparse signal detection from incoherent projections," in [*Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP'06)*], 305–308 (May 2006).

[9] Zahedi, R., Pezeshki, A., and Chong, E. K., "Measurement design for detecting sparse signals," *Phy. Commun.* **5**, 64–75 (June 2012).

[10] Candès, E. J., "The restricted isometry property and its implications for compressed sensing," in [*C. R. Acad. Sci., Ser. I*], **346**, 589–592 (2008).

[11] Bickel, P. J., Ritov, Y., and Tsybakov, A. B., "Simultaneous analysis of lasso and Dantzig selector," *Ann. Statist.* **37**, 1705–1732 (Aug. 2009).

[12] Donoho, D. L., Elad, M., and Temlyakov, V. N., "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inform. Theory* **52**, 6–18 (Jan. 2006).

[13] Tropp, J. A., "On the conditioning of random subdictionaries," *Appl. Comput. Harmon. Anal.* **25**, 1–24 (2008).

[14] Rudelson, M. and Vershynin, R., "On sparse reconstruction from Fourier and Gaussian measurements," *Commun. Pure Appl. Math.* **61**, 1025–1045 (Aug. 2008).

[15] Bajwa, W. U., Haupt, J., Sayeed, A. M., and Nowak, R., "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE* **98**, 1058–1076 (June 2010).

[16] Herman, M. A. and Strohmer, T., "High-resolution radar via compressed sensing," *IEEE Trans. Signal Processing* **57**, 2275–2284 (June 2009).

[17] Neelamani, R., Krohn, C. E., Krebs, J. R., Romberg, J. K., Deffenbaugh, M., and Anderson, J. E., "Efficient seismic forward modeling using simultaneous random sources and sparsity," *Geophysics* **75**, 15–27 (Nov./Dec. 2010).

[18] Romberg, J., "Compressive sensing by random convolution," *SIAM J. Imaging Sci.* **2**(4), 1098–1128 (2009).

[19] Marcia, R. F. and Willett, R. M., "Compressive coded aperture superresolution image reconstruction," in [*Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP'08)*], 833–836 (Apr. 2008).

[20] Applebaum, L., Bajwa, W. U., Duarte, M. F., and Calderbank, R., "Asynchronous code-division random access using convex optimization," *Phy. Commun.* **5**, 129–147 (June 2012).

[21] Tropp, J. A., "Norms of random submatrices and sparse approximation," in [*C. R. Acad. Sci., Ser. I*], **346**, 1271–1274 (2008).

[22] Bajwa, W. U., Haupt, J., Raz, G., Wright, S. J., and Nowak, R., "Toeplitz-structured compressed sensing matrices," in [*Proc. 14th IEEE/SP Workshop Statistical Signal Processing (SSP'07)*], 294–298 (Aug. 2007).

[23] Bajwa, W. U., Haupt, J., Raz, G., and Nowak, R., "Compressed channel sensing," in [*Proc. 42nd Annu. Conf. Information Sciences and Systems (CISS'08)*], 5–10 (Mar. 2008).

[24] Haupt, J., Bajwa, W. U., Raz, G., and Nowak, R., "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inform. Theory* **56**, 5862–5875 (Nov. 2010).

[25] Rauhut, H., "Circulant and Toeplitz matrices in compressed sensing," in [*Proc. Signal Processing with Adaptive Sparse Structured Representations (SPARS'09)*], (Apr. 2009).

[26] Rauhut, H., Romberg, J., and Tropp, J. A., "Restricted isometries for partial random circulant matrices," *Appl. Comput. Harmon. Anal.* **32**, 242–254 (Mar. 2012).

[27] Sebert, F., Zou, Y. M., and Ying, L., "Toeplitz block matrices in compressed sensing and their applications in imaging," in [*Proc. 5th Intl. Conf. Information Technology and Applications in Biomedicine*], 47–50 (May 2008).

[28] Romberg, J. and Neelamani, R., "Sparse channel separation using random probes," *Inverse Problems* **26**(11), 115015 (25 pp.) (2010).

[29] Harmany, Z. T., Marcia, R. F., and Willett, R. M., "Spatio-temporal compressed sensing with coded apertures and keyed exposures," (Nov. 2011). submitted for publication.

[30] Hoeffding, W., "Probability inequalities for sums of bounded random variables," *J. Amer. Statist. Assoc.* **58**, 13–30 (Mar. 1963).

[31] Geršgorin, S. A., "Über die Abgrenzung der Eigenwerte einer Matrix," *Izv. Akad. Nauk SSSR Ser. Fiz.-Mat.* **6**, 749–754 (1931).

[32] Welch, L., "Lower bounds on the maximum cross correlation of signals," *IEEE Trans. Inform. Theory* **20**, 397–399 (May 1974).

[33] Rudelson, M. and Vershynin, R., "Non-asymptotic theory of random matrices: Extreme singular values," in [*Proc. Int. Congr. of Mathematicians*], 1–25 (Aug. 2010).

[34] Bajwa, W. and Pezeshki, A., "Finite frames for sparse signal processing," in [*Finite Frames*], Casazza, P. and Kutyniok, G., eds., Birkhuser Boston, Cambridge, MA (2012).