On the Restricted Isometry of Deterministically Subsampled Fourier Matrices

Jarvis Haupt, Lorne Applebaum, and Robert Nowak

***Abstract*—Matrices satisfying the Restricted Isometry Property (RIP) are central to the emerging theory of compressive sensing (CS). Initial results in CS established that the recovery of sparse vectors x from a relatively small number of linear observations of the form $y = Ax$ can be achieved, using a tractable convex optimization, whenever A is a matrix that satisfies the RIP; similar results also hold when x is nearly sparse or the observations are corrupted by noise. In contrast to random constructions prevalent in many prior works in CS, this paper establishes a collection of deterministic matrices, formed by deterministic selection of rows of Fourier matrices, which satisfy the RIP. Implications of this result for the recovery of signals having sparse spectral content over a large bandwidth are discussed.**

I. Introduction

The emerging theory of compressive sensing (CS) establishes that sparse vectors $x \in \mathbb{R}^p$ can be recovered exactly from a relatively small number of non-adaptive linear measurements of the form $y = Ax$, where the number of rows of the measurement matrix A, denoted by m, can be far fewer than the ambient signal dimension p. Further, and perhaps most remarkably, this recovery can be accomplished by solving a tractable convex optimization [1]–[3].

A particularly concise way of describing for which measurement matrices such exact recovery is possible, is using the Restricted Isometry Property (RIP), proposed in [4].

Definition 1 (Restricted Isometry Property)**.** *The matrix A satisfies the Restricted Isometry Property of order s with parameter $\delta_s \in [0, 1)$ if*

$$(1 - \delta_s)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s)\|x\|_2^2 \tag{1}$$

holds simultaneously for all sparse vectors x having no more than s nonzero entries.

Remark 1. *For vectors $z \in \mathbb{C}^p$, the function $\|z\|_2^2$ denotes the ℓ_2 norm, given by $\|z\|_2^2 = \sum_{i=1}^p z_i^2$. In the sequel, we will also make use of the ℓ_1 norm, which is given by $\|z\|_1 = \sum_{i=1}^p |z_i|$, and we will use the ℓ_0 quasi-norm, $\|z\|_0$, to denote the number of nonzero entries of the vector z.*

In essence, matrices A satisfying the RIP of order s with parameter δ_s—denoted RIP(s,δ_s) for shorthand—are those for which all column submatrices composed of no more than s distinct columns of A behave like a near-isometry, in the sense that their singular values lie in the range $(\sqrt{1-\delta_s}, \sqrt{1+\delta_s})$. For matrices that satisfy the RIP, the following recovery result, established in [4] and refined in [5], is representative.

Lemma 1 (Exact Recovery Using RIP Matrices)**.** *Let A be a matrix satisfying RIP($2s$,δ_{2s}) with $\delta_{2s} < \sqrt{2} - 1$, and let $y = Ax$ be a vector of observations of any sparse signal $x \in \mathbb{R}^p$ having no more than s nonzero entries. Then, the estimate*

$$\widehat{x} = \arg\min_z \ \|z\|_1 \ \ \textit{subject to} \ \ y = Az, \tag{2}$$

is unique and equal to x.

Remark 2. *The above recovery results can be extended to settings where x is a complex vector; see, for example, [6]. In that case if A satisfies RIP($2s$,δ_{2s}) with $\delta_{2s} < 0.3$, then any $x \in \mathbb{C}^p$ having no more than s nonzero entries can be recovered using the same constrained optimization.*

A variety of additional CS recovery results that leverage the RIP, including those that guarantee approximate recovery of nearly-sparse signals and stable recovery when the observations are corrupted by noise, can be found in the CS literature. See, for example, [4], [5], [7]–[10].

While there is currently no known polynomial-time algorithm to test whether a given matrix satisfies the RIP, certain randomly-constructed matrices have been shown to satisfy the RIP with high probability. For example, $m \times p$ random matrices whose entries are independent and identically distributed (iid) realizations of certain zero-mean random variables have been shown to satisfy RIP(s,δ_s) with high probability for any integer s satisfying

$$s \leq c(\delta_s) \cdot \frac{m}{\log p}, \tag{3}$$

where $c(\delta_s)$ is a constant that depends on δ_s but not on m or p [2]–[4], [11]. In such cases, the results of Lemma 1 hold with high probability.

Random matrices with more structure have also been shown to satisfy the RIP. Define the $p \times p$ Discrete Fourier Transform (DFT) matrix, denoted here as $\mathcal{F}^p$, to be the complex-valued matrix whose (j, k)th entry is given by

$$\{\mathcal{F}^p\}_{j,k} = \exp\left(\frac{2\pi i j k}{p}\right), \tag{4}$$

where $i = \sqrt{-1}$ is the complex element and $j, k = 0, 1, \ldots, p-1$. We denote by $\mathcal{F}_T^p$ the $|T| \times p$ submatrix formed by selecting only the rows of $\mathcal{F}^p$ indexed by elements in a set T whose entries come from the set $\{0, 1, \ldots p-1\}$. Recent

This work was supported by the AFOSR, grant no. FA9550-09-1-0140, and by the DARPA Analog-to-Information Program.

J.H. is with the Department of Electrical and Computer Engineering at Rice University in Houston, TX; L.A. is with the Department of Electrical Engineering at Princeton University in Princeton, NJ; and R.N. is with the Department of Electrical and Computer Engineering at the University of Wisconsin–Madison, Madison, WI. Email: jdhaupt@rice.edu, lappleba@princeton.edu, nowak@engr.wisc.edu.

results in [12] improve upon initial results in [1], establishing that when $T \subset \{0, 1, \ldots, p-1\}$ is a subset of size $|T| = m$, selected uniformly at random without replacement, the (scaled) submatrix $m^{-1/2}\mathcal{F}_T^p$ satisfies RIP(s,δ_s) with high probability when

$$s \leq c'(\delta_s) \cdot \frac{m}{\log^4 p}, \tag{5}$$

where $c'(\delta_s)$ is a constant that does not depend on m or p. In this paper we identify a collection of matrices, formed by specific selections of rows of the Fourier matrix, that satisfy the RIP *deterministically*. Our results leverage ideas from number theory, specifically the analysis of certain structured exponential sums, and a result from classical eigenanalysis known as Geršgorin's Disc Theorem.

The remainder of the paper is organized as follows. Our main results, that matrices formed by certain deterministic selections of rows of Fourier matrices satisfy the RIP, are stated in the next section. Applications of these results, in the context of recovering time-varying signals whose spectral representations are sparse over a large bandwidth, are discussed in Section III, and the approach is examined via simulation in Section IV. The method described here is discussed in the broader context of sub-Nyquist sampling in Section V. Finally, some helpful lemmata, as well as the proofs of the main results, are relegated to the appendix.

II. Main Results

Our main results concern submatrices formed by deterministically selecting a set of rows of $\mathcal{F}^p$. When the indices of the rows selected correspond to the (integer) outputs of certain polynomial functions mod p, the resulting submatrices satisfy the restricted isometry property (RIP). Analogous results also hold for submatrices formed by selecting rows of the Fourier inverse matrix $(\mathcal{F}^p)^{-1}$, which is the $p \times p$ matrix whose (j,k)th element is given by

$$\left\{(\mathcal{F}^p)^{-1}\right\}_{j,k} = \frac{1}{p}\ \exp\left(\frac{-2\pi i j k}{p}\right), \tag{6}$$

$j, k = 0, 1, \ldots, p-1$. The first main result of this paper is stated below as a theorem.

Theorem 1. *Let $p > 2$ be a prime integer and let $f(n)$ be any polynomial of degree $d \geq 2$ of the form $f(n) = a_1 n + \cdots + a_d n^d$, with real integer coefficients $a_j \in \{0, 1, \ldots, p-1\}$ for $j = 1, 2, \ldots, d-1$, and $a_d \in \{1, 2, \ldots, p-1\}$. For any $\epsilon_1 \in (0,1)$, choose m to be an integer satisfying*

$$p^{1/(d-\epsilon_1)} \leq m \leq p. \tag{7}$$

Let $T = \{f(n) \mod p : n = 1, 2, \ldots, m\}$ (note that $|T| = m$ and T may contain duplicate entries). Then, for any $\delta_s \in (0,1)$ and $\epsilon_2 \in (0, \epsilon_1)$, the matrix $m^{-1/2}\mathcal{F}_T^p$ satisfies RIP(s,δ_s) whenever

$$s \leq \delta_s \cdot C(d, \epsilon_2) \cdot m^{(\epsilon_1 - \epsilon_2)/2^{d-1}}, \tag{8}$$

where $C(d, \epsilon_2)$ is a constant that does not depend on m or p. The same result also holds for the subsampled Fourier inverse matrix, $pm^{-1/2}\,(\mathcal{F}^p)_T^{-1}$, for the same choice of T.

When the degree of the polynomial is large, the following result provides a slightly stronger statement with respect to the exponent on m.

Theorem 2. *Let $p > 2$ be a prime integer and let $f(n)$ be any polynomial of degree $d > 2$ of the form $f(n) = a_1 n + \cdots + a_d n^d$, with real integer coefficients $a_j \in \{0, 1, \ldots, p-1\}$ for $j = 1, 2, \ldots, d-1$, and $a_d \in \{1, 2, \ldots, p-1\}$. Choose m to be an integer satisfying*

$$p^{1/(d-1)} \leq m \leq p. \tag{9}$$

Let $T = \{f(n) \mod p : n = 1, 2, \ldots, m\}$. Then, for any $\delta_s \in (0,1)$, the matrix $m^{-1/2}\mathcal{F}_T^p$ satisfies RIP(s, δ_s) whenever

$$s \leq \delta_s \cdot C'(d) \cdot m^{1/(9d^2 \log d)}, \tag{10}$$

where $C'(d)$ is a constant that does not depend on m or p. The same result also holds for the subsampled Fourier inverse matrix, $pm^{-1/2}\,(\mathcal{F}^p)_T^{-1}$, for the same choice of T.

We briefly compare these results with existing results for similar random constructions. As stated in the introduction, matrices $m^{-1/2}\mathcal{F}_T^p$ formed by selecting a set of m rows of the p-dimensional Fourier matrix uniformly at random satisfy the RIP of order s with parameter δ_s with high probability provided

$$s \leq c'(\delta_s) \cdot \frac{m}{\log^4 p}. \tag{11}$$

In contrast, here we identify an analogous class of *deterministic* matrices that satisfy the RIP of order s with parameter δ_s for

$$s \leq \max\left\{\delta_s\ C(d, \epsilon_2) \cdot m^{(\epsilon_1 - \epsilon_2)/2^{d-1}}\ ,\right.$$
$$\left.\delta_s\ C'(d) \cdot m^{1/(9d^2 \log d)}\right\}, \tag{12}$$

where $0 < \epsilon_2 < \epsilon_1 < 1$, d is the degree of the polynomial used to generate the set of rows to be selected, and the constants C and C' do not depend on the signal dimension p or the number of measurements m.

Relative to the results for random constructions, the deterministic results presented here exhibit slightly less-favorable scaling behavior with respect to the parameter m. Specifically, while a randomly constructed matrix with m rows allows for recovery of signals with $O(m/\log^4 p)$ nonzero entries, our deterministic results only guarantee recovery if the number of nonzero entries is, at most, $O(\sqrt{m})$. An open question remains as to whether there exists a deterministic sampling strategy from which any signal having $O(m)$ nonzero entries can be recovered from $O(m)$ observations. In addition, the results we present here also require that the number of measurements m exceeds some fractional power of the signal dimension p. This minimum sampling requirement is more restrictive than what exists in the CS literature for random sampling methods, where the dependence of the minimum number of samples on the signal dimension is typically some power of $\log p$. The minimum sampling conditions obtained here are artifacts of the analysis techniques, and another open question remains as to the minimum number of samples required for deterministic sampling strategies.

III. Applications

In this section we describe how the results established above can be utilized in the analysis of radio frequency (RF) receiver systems that acquire time-varying signals having sparse frequency-domain representations using structured, deterministic, non-uniform sampling. The approach we consider here entails collecting samples at times that are related to the outputs of certain polynomial functions, and is motivated in part by the **Ny**quist **F**olding **R**eceiver (NyFR) architecture proposed in [13]. Upon casting the sampling processes in the canonical CS framework, where the target signal is the (sparse) frequency-domain representation of the time-varying signal of interest, the results of the previous section guarantee that the "effective" observation matrix satisfies the restricted isometry property (RIP). Consequently, any of a number of CS recovery algorithms can then be utilized to recover the target signal.

Consider signals $x(t)$ defined on $t \in (0, t_d]$, for some time duration $t_d > 0$. Let p be a large prime, and for $\tau = t_d/p$, define the *sampled signal vector* $x \in \mathbb{R}^p$ to be the vector whose kth entry is given by $x_k = x(k\tau)$, for $k = 1, 2, \ldots, p$. The class of signals we will be interested in here are those having sparse frequency-domain representations—specifically, signals for which $\|\mathcal{F}^p x\|_0 \leq s \ll p$. Motivated by the results in the previous section, we suggest two methods for acquiring samples of the signals of interest according to the outputs of certain polynomial functions. The first method corresponds to a "streaming" sampling process, where the output of the polynomial function determines a set of actual sampling times. The second method is a process where the sample times are determined a priori, and can be collected over a relatively small number of periods of the signal. Each of the methods is briefly described below.

First, in the case where the polynomial function determines actual sampling times, we let $f(n)$ be a degree-d polynomial of the form $f(n) = a_1 n + \cdots + a_d n^d$, where $a_i \in \{0, 1, \ldots, p-1\}$ for $i = 1, \ldots, d-1$, and $a_d \in \{1, 2, \ldots, p-1\}$. A total of m samples of $x(t)$ are collected, at sampling times described by the entries of the set $\{\tau \cdot f(n) : n = 1, 2, \ldots, m\}$. Notice that, because of the periodicity of $x(t)$ (and its corresponding sampled signal vector), the acquisition scheme can be modeled by a matrix-vector product corresponding to the action of a *sampling matrix* S on the periodic extension of the sampled signal vector,

$$\overbrace{\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}}_{m\times 1}}^{y} = \overbrace{\underbrace{\begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}}_{m\times rp}}^{S} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_1 \\ \vdots \\ x_p \end{bmatrix}}_{rp\times 1}. \tag{13}$$

Here, each row of the sampling matrix contains only one nonzero (unit) entry corresponding to the time at which the sample is collected, and $r \geq 1$ is an integer determined by the maximum sampling time $\tau \cdot f(m)$.

Now, to simplify this expression, we note that

$$\underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_1 \\ \vdots \\ x_p \end{bmatrix}}_{rp\times 1} = \frac{1}{r} \cdot \underbrace{\begin{bmatrix} (\mathcal{F}^p)^{-1} \\ \vdots \\ (\mathcal{F}^p)^{-1} \end{bmatrix}}_{rp\times p} \underbrace{\begin{bmatrix} \mathcal{F}^p & \ldots & \mathcal{F}^p \end{bmatrix}}_{p\times rp} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_1 \\ \vdots \\ x_p \end{bmatrix}}_{rp\times 1}. \tag{14}$$

Substituting this into the above, we can rewrite the observation process as

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \underbrace{S \begin{bmatrix} (\mathcal{F}^p)^{-1} \\ \vdots \\ (\mathcal{F}^p)^{-1} \end{bmatrix}}_{m\times p} \cdot \underbrace{\mathcal{F}^p \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}}_{p\times 1}. \tag{15}$$

Now, the problem has been recast into the canonical CS framework, where the observation model comprises m samples of a length-p sparse (complex) vector. In addition, the effective observation matrix is formed by selecting a set of rows of the Fourier inverse matrix, which is the setting examined theoretically in the previous section. In other words, the results of Theorems 1 and 2 specify the number of samples m (as a function of the degree of the polynomial) that must be collected in order for the corresponding effective observation matrix to satisfy the RIP. Any of a number of recovery techniques whose success is conditional on the RIP can then be employed to recover the spectral representation of the signal being observed.

Notice that in the above discussion, the total observation time required to obtain m samples is no less than $\tau \cdot f(m) = t_d \cdot f(m)/p$. In some applications where the available observation time window is limited, or where the signal of interest is from a broader class of signals which are periodic but only for a short time duration (as is the case with frequency-hopping transmissions), this time requirement might be prohibitive. To overcome this, the second method we propose allows for the collection of the same number of samples over fewer periods of the signal (and consequently, less time) by designing a collection of sampling points in time a priori, based on the output of the polynomial function. Specifically, the results of Theorems 1 and 2 also apply when the sampling points are given by the entries of the set $T = \{\tau \cdot (f(n) \mod p) : n = 1, 2, \ldots, m\}$.

Note that in this case, T may contain duplicate entries, corresponding to indices $n_1 \neq n_2$ for which $f(n_1) = f(n_2) \mod p$. We suggest two possible ways to collect the required samples in this setting. First, for points in time where duplicate observations are required, a single observation can be collected and simply replicated as often as necessary in the overall vector of observations, prior to the implementation of a CS recovery technique. Alternatively, the observations can be collected at distinct times over the periodic extension of the signal, similar in spirit to the above setting. The upshot here

is that the observations can be collected over fewer periods of the signal, and consequently, a shorter actual time.

Indeed, in order to collect distinct samples for each time prescribed in T, the number of periods of the signal that must be observed needs to be no greater than the largest multiplicity of any entry of T. By the Fundamental Theorem of Algebra, the number of solutions of the polynomial congruence $f(n) = 0 \mod p$ is at most the degree of the polynomial when p is a prime and the polynomial has integer coefficients. It follows that the number of points n_i for which the sample time indices $f(n_i)$ are equal (mod p) is no more than d (the degree of the polynomial that generates the set of sampling times), provided the number of measurements satisfies $m < p$. Consequently, all of the prescribed sampling points in time can be acquired by unique observations that occur over an *actual* observation time that need not exceed $t_d \cdot d$, which is much smaller than $t_d \cdot f(m)/p$ when m satisfies the conditions of the theorems.

IV. Simulations

In this section we demonstrate the effectiveness of the proposed sampling and recovery technique via simulation. Three settings are examined, corresponding to two different choices of polynomials used to generate deterministic sampling times as described above, and one setting where the sampling times are chosen at random. In each setting, the signal length p is chosen to be the largest prime less than 2^{16} (which is 65521), and we consider collecting a total of $m = \lfloor p/10 \rfloor = 6552$ observations, which corresponds to downsampling by a factor of about 10. The deterministic sampling times are generated by the polynomials $f_2(n) = 10n+n^2$ and $f_3(n) = 10n+n^2+n^3$ (mod p) for $n = 1, \ldots, m$, and in the case of random sampling, a total of m sample times are chosen uniformly at random (without replacement) from the set $\{1, \ldots, p\}$. For simplicity we consider only noise-free settings.

For each sparsity level, we generate the signals of interest randomly as follows. First, we select a set of locations for the nonzero spectral components uniformly at random from the set $\{1, \ldots, p\}$. The real and imaginary parts of the amplitude of each nonzero component are obtained as independent draws from the unit-norm Gaussian distribution. Finally, the sampled signal vector (the discretized version of the real-valued time-varying signal) is taken to be the real part of the inverse Fourier transform of this complex vector. For each sparsity level and each sampling approach we perform 50 trials, each using a unique randomly-generated signal, and we tally the number of trials for which exact recovery was obtained by solving (2). We perform the optimization using the complex Basis Pursuit solver in the SPGL1 software suite; see [14], [15].

The results are depicted in Figure 1, which shows how the average rate of successful recovery varies as a function of sparsity level for the three settings examined. The solid line corresponds to sampling times chosen randomly, while the dashed and dash-dot lines correspond to sampling times generated using $f_2(n)$ and $f_3(n)$, respectively. The results suggest that samples obtained at times generated by the second-order polynomial can be as effective for recovery as random samples, while samples obtained at times generated by a third-order polynomial may be slightly less effective.

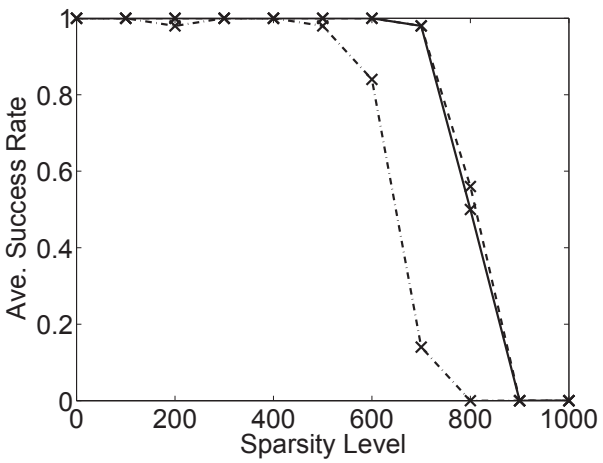

Fig. 1. Average rate of successful signal recovery as a function of sparsity level, for two sample-time generating polynomials of differing orders. In each case, the signal dimension $p = 65521$, and the number of observations $m = 6552$. The sparsity level quantifies the number of distinct sinusoids present in the signal (the actual complex vector being recovered has twice that number of nonzero components). The solid line corresponds to sample times generated at random, while the dashed and dash-dot lines correspond to sample times generated by a second-order and third-order polynomial, respectively.

It is worth noting that, because of the way the signals are constructed, the sparsity level stated on the figures is the number of distinct sinusoids present in the time-varying signal, so that the overall complex vector being recovered actually consists of twice as many nonzero entries. Strictly speaking, the optimization in (2) does not take into account the symmetry of the frequency-domain representation of real-valued periodic time-varying signals, though in practice such conditions should, of course, be leveraged. Also, as mentioned in the previous sections, the fact that the effective measurement matrices in these settings satisfies the RIP allows for the potential utilization of any of a number of recovery techniques (including efficient *greedy* methods), and ensures that reliable recovery can be achieved in noisy settings as well.

V. Comparison with Other Works

Sub-Nyquist sampling methods for analog RF signals comprise a well-studied field. Initial works in sub-Nyquist sampling considered the case of multiband (rather than strictly bandlimited) signals, establishing that recovery was possible provided that the average sampling rate exceeds the so-called Nyquist-Landau rate, which is defined to be the total occupied Fourier bandwidth of the signal [16], [17]. In these works, the proposed recovery approach consists of bandpass filtering and demodulation of each individual signal to baseband, followed by uniform sampling. Such approaches, of course, require prior knowledge of the carrier frequencies for each band.

Alternative sampling and recovery techniques for multiband signals have also been examined in the literature. For example, in [18], the authors consider samples obtained by interleaving the outputs of several channels, each of which performs uniform downsampling on the unknown multiband signal. In this approach, however, the recovery process requires knowledge of the locations of the occupied bands. Similar approaches using interleaved analog-to-digital converters

(ADCs) have been examined—see, for example, [19]–[21]. Interleaving techniques all suffer from inherent technological limitations, such as a limited allowable input bandwidth and timing jitter/synchronization issues.

Recently, with the emergence of compressive sensing, a powerful new set of theoretical results and techniques are being leveraged and utilized to address the multiband sampling and recovery problem. For example, random demodulator approaches [22], [23] employ single-channel receivers in which the incoming signal is mixed with a high-rate pseudo-random signal, integrated, and downsampled. Another concurrent path of research entails multi-channel systems, employing either a form of multi-coset sampling, or a setting in which each channel performs an analog pre-mixing of the incoming signal with a periodic waveform followed by filtering and downsampling [24], [25]. Each of these approaches enjoy certain theoretical performance guarantees for recovering sparse, multi-band signals from a large RF bandwidth.

The structured non-uniform sampling technique proposed here differs markedly from these two approaches. First, our approach utilizes a single receiver channel, in contrast to multi-coset and multi-channel sampling techniques described above. In practice, fewer channels corresponds to reduced power requirements. Further, the structured non-uniform sampling approach proposed here can be implemented in practice by mixing the incoming signal with an appropriate analog pulse sequence, low-pass filtering, and then sampling the baseband representation, thus overcoming the bandwidth limitations of traditional ADCs. See [13] for details.

In contrast to the random demodulator approach, the method examined here does not require any analog front-end components that oscillate at the true Nyquist rate of the target signal. In addition, while recovery from random demodulator observations requires storage of and operation with a large (pseudo-)random matrix, the approach described here can instead leverage fast Fourier transform techniques and simple indexing operations to speed the recovery. Finally, as stated earlier, compared with the random demodulator approach the structured non-uniform sampling approach examined here has the distinct benefit of deterministic recovery guarantees.

VI. Appendix

A. Helpful Lemmata

The proofs will utilize a standard result in classical eigenanalysis known as *Geršgorin's Disc Theorem*, stated here as a lemma without proof (for more details see, for example, [26]).

Lemma 2 (Geršgorin)**.** *The eigenvalues of a $p \times p$ complex matrix M all lie in the union of p discs $d_j = d_j(c_j, r_j)$, $j = 1, 2, \ldots, p$, centered at $c_j = M_{j,j}$, and with radius*

$$r_j = \sum_{\substack{i=1 \\ i \neq j}}^{p} |M_{j,i}|. \tag{16}$$

When applying Geršgorin's Theorem, we will control the sums of off-diagonal elements using bounds on certain exponential sums. The following result is credited to H. Weyl [27], and appears in the present form in [28].

Lemma 3 (Weyl)**.** *Let $d \geq 2$, and let $g(n) = a_1 n + \cdots + a_d n^d$, where $a_d = \alpha/p + \theta/p^2$, $|\theta| \leq 1$, and $\gcd(\alpha, p) = 1$. If, for $0 < \epsilon_2 < \epsilon_1 < 1$, the condition $m^{\epsilon_1} \leq p \leq m^{d-\epsilon_1}$ holds for some integer m, then*

$$\frac{1}{m} \left| \sum_{n=1}^{m} \exp\left(2\pi i g(n)\right) \right| \leq c(d, \epsilon_2) \cdot m^{(\epsilon_2 - \epsilon_1)/2^{d-1}}. \tag{17}$$

For completeness, we note the constant $c(d, \epsilon_2)$ is given by

$$c(d, \epsilon_2) = 2 \left[\left(\frac{64d}{\epsilon_2} \right) \left(\frac{d^2}{\epsilon_2 \log 2} \right)^{\exp\left(d^2/\epsilon_2\right)} d! \right]^{1/2^{d-1}}. \tag{18}$$

Theorem 2 utilizes an improvement of the above for large values of d, credited to I. Vinogradov [29], and appearing in its present form in [28].

Lemma 4 (Vinogradov)**.** *Let $d > 2$, and let $g(n) = a_1 n + \cdots + a_d n^d$, where $a_d = \alpha/p + \theta/p^2$, $|\theta| \leq 1$, and $\gcd(\alpha, p) = 1$. If $m \leq p \leq m^{d-1}$, then*

$$\frac{1}{m} \left| \sum_{n=1}^{m} \exp\left(2\pi i g(n)\right) \right| \leq c(d) \cdot m^{-1/(9d^2 \log d)}. \tag{19}$$

For completeness, we note the constant $c(d) = \exp(3d)$.

B. Proof of Theorem 1

Recall that the matrix of interest is the $|T| \times p$ matrix $m^{-1/2}\mathcal{F}_T^p$, where the set T indicates which rows of the DFT matrix $\mathcal{F}^p$ are present. We begin by analyzing the Gram matrix $G = m^{-1}(\mathcal{F}_T^p)^H \mathcal{F}_T^p$, where the superscript H denotes the complex conjugate transpose (Hermitian) operator. Let $f(n)$ be any polynomial of degree $d > 2$ of the form $f(n) = a_1 n + \cdots + a_d n^d$, with real integer coefficients $a_j \in \{0, 1, \ldots, p-1\}$ for $j = 1, 2, \ldots, d-1$, and $a_d \in \{1, 2, \ldots, p-1\}$, as specified in the theorem. The entries of the Gram matrix G correspond to inner products between the columns of the matrix of interest, whose entries are given by

$$\begin{aligned} \left\{ m^{-1/2}\mathcal{F}_T^p \right\}_{j,n} &= m^{-1/2} \exp\left(\frac{2\pi i j f(n) \mod p}{p} \right) \\ &= m^{-1/2} \exp\left(\frac{2\pi i j f(n)}{p} \right), \end{aligned} \tag{20}$$

for $j = 0, 1, \ldots, p-1$ and $n = 1, 2, \ldots, m$. Using this fact, we have that the entry of $G = m^{-1}(\mathcal{F}_T^p)^H \mathcal{F}_T^p$ in row y and column z is given by

$$\{G\}_{y,z} = m^{-1} \sum_{n=1}^{m} \exp\left(\frac{2\pi i (z-y) f(n)}{p} \right), \tag{21}$$

for $y, z = 0, 1, \ldots, p-1$ and $n = 1, \ldots, m$.

The first point to note is that the diagonal elements of G (where $y = z$) are all equal to one since the argument of the exponential function is identically zero. For the off-diagonal elements, we can make the substitution $g(n) = (z-y)f(n)/p$ to obtain

$$\{G\}_{y,z} = m^{-1} \sum_{n=1}^{m} \exp\left(2\pi i g(n)\right), \tag{22}$$

for $y, z = 0, 1, \ldots, p-1$, $y \neq z$, and $n = 1, \ldots, m$.

Note that the coefficient of the highest order term in $g(n)$ is given by $(z-y)a_d/p$. For $z > y$, it is easy to see that the polynomial $g(n)$ satisfies the conditions of Lemma 3 because p is prime and $(z-y)$ is a positive integer less than p. For the case $z < y$, we can instead consider the polynomial $g'(n) = g(n) + n^d$, since in this case the periodicity of the complex exponential implies

$$\begin{aligned} \{G\}_{y,z} &= m^{-1} \sum_{n=1}^{m} \exp\left(2\pi i g(n)\right) \\ &= m^{-1} \sum_{n=1}^{m} \exp\left(2\pi i g'(n)\right). \end{aligned} \tag{23}$$

If $z < y$, then $(z-y)$ is negative but $(z-y+p)$ is positive (since $y \leq p-1$). Thus, the leading coefficient of $g'(n)$ is of the form of a positive integer (less than p) divided by p, and satisfies the conditions of Lemma 3 in this case as well. If, in addition, $0 < \epsilon_2 < \epsilon_1 < 1$ and m satisfies the condition $p^{1/(d-\epsilon_1)} \leq m \leq p$, we can apply Lemma 3 to bound the magnitude of the off-diagonal elements of G by

$$\left|\{G\}_{y,z}\right| \leq c(d,\epsilon_2) \cdot m^{(\epsilon_2-\epsilon_1)/2^{d-1}}, \tag{24}$$

for $y, z = 0, 1, \ldots, p-1$ and $y \neq z$.

Now, let S be a subset of $\{0, 1, \ldots, p-1\}$ satisfying $|S| \leq s$, and denote by $\mathcal{F}^p_{T,S}$ the $|T| \times |S|$ submatrix of $m^{-1/2}\mathcal{F}^p_T$ formed by retaining the columns indexed by the elements of S. Notice that if for each unique choice of S, the eigenvalues of the Gram matrix $G_S = m^{-1}(\mathcal{F}^p_{T,S})^H \mathcal{F}^p_{T,S}$ are all within the range $(1-\delta_s, 1+\delta_s)$ then the RIP is satisfied. But, since each Gram matrix G_S is a proper submatrix of G, Geršgorin's Theorem guarantees that the eigenvalue bound is satisfied simultaneously for all choices of S whenever the sum of any $s-1$ off-diagonal elements of G does not exceed δ_s. In other words, the RIP is satisfied whenever

$$\begin{aligned} s &\leq \frac{\delta_s}{c(d,\epsilon_2)} m^{(\epsilon_1-\epsilon_2)/2^{d-1}} \\ &= \delta_s \cdot C(d,\epsilon_2) \cdot m^{(\epsilon_1-\epsilon_2)/2^{d-1}}, \end{aligned} \tag{25}$$

as claimed. The same method of proof also establishes the RIP for the matrix $pm^{-1/2}\left(\mathcal{F}^p\right)^{-1}_T$.

C. Proof of Theorem 2

The proof proceeds along the same lines as above, with references to Lemma 4 instead of Lemma 3 where appropriate.

Acknowledgments

The authors thank Jerry Fudge at L-3 Communications, whose structured non-uniform sampling architecture motivated this research direction, and Robert Calderbank for discussing his own related work and for collaborating on this project.

References

[1] E. J. Candès, 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.

[2] E. J. Candès 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.

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

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

[5] E. Candès, "The restricted isometry property and its implications for compressed sensing," in *C. R. Acad. Sci., Ser. I*, vol. 346, pp. 589–592. Paris, 2008.

[6] W. U. Bajwa, *New Information Processing Theory and Methods for Exploting Sparsity in Wireless Systems*, Ph.D. thesis, University of Wisconsin—Madison, 2009.

[7] E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, no. 6, pp. 2313–2351, Dec. 2007.

[8] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of Computational Mathematics*, vol. 9, no. 3, pp. 317–334, 2007.

[9] D. Needell and R. Vershynin, "Signal recovery from inaccurate and incomplete measurements via regularized orthogonal matching pursuit," *IEEE J. Sel. Topics in Sig. Proc.*, 2008, to appear.

[10] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comp. Harmonic Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.

[11] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," in *Constructive Approximation*. Springer, New York, 2008.

[12] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Commun. Pure Appl. Math.*, vol. 61, no. 8, pp. 1025–1045, Aug. 2008.

[13] G. Fudge, M. Chivers, S. Ravindran, R. Bland, P. Pace, and J. Haupt, "A Nyquist folding analog-to-information receiver," in *Proc. 42nd Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 2008.

[14] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comp.*, vol. 31, no. 2, pp. 890–912, 2008.

[15] E. van den Berg and M. P. Friedlander, "SPGL1: A solver for large-scale sparse reconstruction," June 2007, Online: `http://www.cs.ubc.ca/labs/scl/spgl1`.

[16] H. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Math.*, vol. 117, no. 1, pp. 37–52, Feb. 1967.

[17] H. Landau, "Sampling, data transmission, and the Nyquist rate," *Proc. IEEE*, vol. 55, no. 10, pp. 1701–1706, Oct. 1967.

[18] Y.-P. Lin and P. Vaidyanathan, "Periodically nonuniform sampling of bandpass signals," *IEEE Trans. Circuits Syst. II*, vol. 45, no. 3, pp. 340–351, Mar. 1998.

[19] W. Black and D. Hodges, "Time interleaved converter arrays," *IEEE J. Solid-State Circuits*, vol. 15, no. 6, pp. 1022–1029, Dec. 1980.

[20] Y.-C. Jenq, "Digital spectra of nonuniformly sampled signals: A robust sampling time offset estimation algorithm for ultra high-speed waveform digitizers using interleaving," *IEEE Trans. Instrum. Meas*, vol. 39, no. 1, pp. 71–75, Feb. 1990.

[21] R. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[22] J. Laska, S. Kirolos, M. Duarte, T. Ragheb, R. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *Proc. IEEE Int. Symp. Circuits and Systems*, May 2007, pp. 1959–1962.

[23] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," Jan. 2009, Online: `http://arxiv.org/abs/0902.0026`.

[24] M. Mishali and Y. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.

[25] M. Mishali and Y. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. of Sel. Topics Signal Proc.: Special Issue on Compressed Sensing*, to appear.

[26] R. Varga, *Geršgorin and His Circles*, Number 36 in Springer Series in Computational Mathematics. Springer-Verlag, Berlin, Germany, 2004.

[27] H. Weyl, "Über die Gleichverteilung von Zahlen mod. Eins," *Math. Ann.*, vol. 77, pp. 312–352, 1916.

[28] N. Korobov, *Exponential sums and their applications*, Kluwer Academic Publishers, 1992.

[29] I. Vinogradov, *Selected Works*, Springer-Verlag, Berlin-New York, 1985.