Simple Deterministically Constructible RIP Matrices with Sublinear Fourier Sampling Requirements

M. A. Iwen
Institute for Mathematics and its Applications (IMA)
University of Minnesota, 114 Lind Hall
Minneapolis, MN 55455-0134
Email: iwen@ima.umn.edu

***Abstract*—We present a deterministic number theoretic construction for matrices with the Restricted Isometry Property (RIP). Furthermore, we show that the number theoretic properties of our RIP matrices allow their products with Discrete Fourier Transform (DFT) matrices to be well approximated via a few highly sparse matrix multiplications. Hence, our RIP matrices may be approximately multiplied by the DFT of any input vector in sublinear-time by reading only a small fraction of its entries. As a consequence, we obtain small deterministic sample sets which are guaranteed to allow the recovery of near-optimal sparse Fourier representations for all periodic functions having an integrable second derivative over a single period. Explicit bounds are provided for the sizes of our RIP matrices, the sizes of their associated sublinear Fourier sampling sets, and the errors incurred by quickly approximating their products with DFT matrices. The Fourier sampling requirements obtained herein improve on previous deterministic Fourier sampling results in [1], [2].**

***Index Terms*—Fourier transforms, Discrete Fourier transforms, Algorithms, Number theory, Signal processing**

I. Introduction

There has been interest in deterministic Compressed Sensing (CS) RIP matrix constructions ever since the introduction of universal near-optimal randomized constructions [3]. Since then deterministic constructions have been discovered [4], [5], but they do not lead to small sampling sets in the important Fourier CS case. In this paper we present a simple deterministic construction for RIP matrices which leads to small deterministic Fourier sampling set constructions. As a consequence, we obtain a deterministic sparse Fourier transform method which is guaranteed to recover a near-optimal sparse Fourier representation (if one exists) for any input signal by reading only a small deterministic subset of its entries.

Suppose we want to find a near-optimal k-term Fourier approximation for a signal (i.e., periodic function) with bandwidth N. Randomized RIP matrix construction results tell us that $O(k \cdot \log^4 N)$ random signal samples suffice for the near-optimal k-term approximation of all N-bandwidth signals with high probability (see [6], and Theorem 1 below, proven in [7]). However, verifying that any such randomly selected set of samples actually works for all signals is computationally infeasible. In addition, random sampling can be difficult to incorporate into some circuit designs. In this paper we show that $O\left(k^2 \cdot \frac{\log^3 N}{\log^2 k}\right)$ deterministic samples suffice for guaranteed near-optimal Fourier approximation. Furthermore, the deterministic sampling entailed is simple.

Coherence arguments show that all N-bandwidth signals consisting of *exactly* k nonzero modes can be recovered with high probability using $O(k^2 \cdot \log N)$ random sampling positions [8]. Furthermore, the associated coherence of any such random sampling set can be checked in time $O(k^2 \cdot N^2 \cdot \log N)$. The end result is a polynomial-time Las Vegas algorithm for generating sets of samples which are guaranteed to allow the recovery of all k-sparse Fourier signals. However, these results do not consider noise (i.e., signals which are only approximately sparse). Furthermore, if k is $\Omega(N^\alpha)$ for a fixed $\alpha \in (0, 1)$, we achieve sampling sets of comparable size explicitly and deterministically. We don't need to check coherence properties for new k and N-valued sample sets.

Finally, we note that the deterministic Fourier sampling results herein improve on previous deterministic sparse Fourier sampling results concerning periodic functions, f, whose Fourier transforms decay algebraically [1], [2]. The best such result states that if f's b^{th}-largest magnitude Fourier coefficient is $O(b^{-p})$ for some $p > 1$, then a near optimal k-term Fourier representation for f can be calculated using $O\left(k^{\frac{2p}{p-1}} \cdot \frac{\log^3 N}{\log^2 k}\right)$ deterministic samples [2]. Hence, the number of samples required for guaranteed approximation depends on the signal's Fourier compressibility. We improve on these deterministic sampling bounds for all fixed $p \in (1, \infty)$.

The remainder of this paper is organized as follows: In Section II we introduce required definitions and briefly review compressed sensing results. In Section III we introduce our deterministic RIP matrix construction method, and give explicit row bounds. Doing so allows us to compare it to previous RIP matrix constructions and conclude that it is near-optimal for its class (binary entries, L^2-RIP). Finally, we note that our number theoretic matrices are easy to build and cheap to store.

In Section IV we demonstrate how our measurement matrices' number-theoretic structure allows them to produce small deterministic sparse Fourier sampling sets. In doing so we provide fast algorithms for approximating the product of our measurement matrices with the discrete Fourier transform of an array of samples from an arbitrary Fourier-compressible signal. Explicit error bounds for the approximation methods are derived. In Section IV-A these results are extended to

include standard discrete Fourier transform results (where only given equally-spaced signal samples can be utilized).

Finally, in Section V we conclude with a short discussion of reconstruction methods. In doing so we note that our RIP matrices lead to efficient linear-time (in the bandwidth) sparse Fourier transforms. Better yet, we briefly discuss the possibility of sublinearizing the runtime of these methods by using a small number of additional samples.

II. Preliminaries

We are interested in signals (i.e., arrays), $\mathbf{A}$, of N complex values with an associated $N \times N$ complex valued matrix Ψ under which $\mathbf{A}$ is sparse. Thus, we assume that only $k < N$ entries of $\Psi \cdot \mathbf{A}$ are significant (or large) in magnitude. In this case standard compressed sensing (CS) methods [3], [5], [9], [7] deal with recovering the k most significant entries of $\Psi \cdot \mathbf{A}$ using only a small (i.e., compressed) set of measurements given by $\mathcal{M} \cdot \Psi \cdot \mathbf{A}$, where $\mathcal{M}$ is a special type of rectangular $K \times N$ matrix with the *Restricted Isometry Property (RIP)*.

Let the discrete L^q-norm of any array $\mathbf{A}$ be

$$\|\mathbf{A}\|_q = \left(\sum_{j=0}^{N-1} |\mathbf{A}(j)|^q \right)^{\frac{1}{q}}.$$

A matrix $\mathcal{M}$ has the RIP(N,k,ϵ) if

$$(1-\epsilon)\|x\|_2^2 \leq \|\mathcal{M}x\|_2^2 \leq (1+\epsilon)\|x\|_2^2$$

for all $x \in \mathbb{C}^N$ containing at most k non-zero coordinates. Given a RIP(N,$O(k)$,ϵ) matrix $\mathcal{M}$ with $\epsilon \in \left(0, \frac{1}{2}\right)$ sufficiently small, a near optimal $O(k)$-term approximation can be obtained for $\Psi \cdot \mathbf{A}$ using linear programming [3], [5] or Regularized Orthogonal Matching Pursuit (ROMP) variants [9], [7]. In particular, we will later utilize the following theorem proven in [7].

Theorem 1. *Suppose that $\mathcal{M}$ is a $K \times N$ measurement matrix with RIP(N,$2k$,ϵ), where ϵ is sufficiently small. Fix precision parameter $\eta \in \mathbb{R}$ and let $\mathbf{U} = \mathcal{M}\Psi\mathbf{A} + \mathbf{E}$ be measurements for a given $\mathbf{A} \in \mathbb{C}^N$, contaminated with arbitrary noise $\mathbf{E}$. Then, we may use a ROMP variant with input $\mathbf{U}$ to produce a $2k$-sparse approximation $\mathbf{R}$ satisfying*

$$\|\Psi \cdot \mathbf{A} - \mathbf{R}\|_2 \leq \text{Const} \cdot \max\left\{\eta, \frac{1}{\sqrt{k}} \cdot \|\Psi \cdot \mathbf{A} - \mathbf{R}_{\text{opt}}\|_1 + \|\mathbf{E}\|_2\right\}$$

where $\mathbf{R}_{\text{opt}}$ is a best k-sparse approximation to $\Psi \cdot \mathbf{A}$. The runtime is $O(C \cdot \log(\|\Psi \cdot \mathbf{A}\|_2/\eta))$, where C bounds the cost of a matrix-vector multiply with $\mathcal{M}$ or $\mathcal{M}^*$.

In this paper we are primarily interested the special CS case where Ψ is the $N \times N$ Discrete Fourier Transform (DFT) matrix

$$\Psi_{i,j} = \frac{2\pi}{N} \cdot e^{\frac{2\pi \mathbf{i} \cdot i \cdot j}{N}}.$$

Thus, $\mathbf{A}$'s DFT, denoted $\widehat{\mathbf{A}}$, is a sparse/compressible signal of length N defined as follows:

$$\widehat{\mathbf{A}}(\omega) = \Psi \cdot \mathbf{A} = \frac{2\pi}{N} \cdot \sum_{j=0}^{N-1} e^{\frac{-2\pi \mathbf{i} \omega j}{N}} \mathbf{A}(j), \quad \forall \omega \in \left(-\left\lceil \frac{N}{2} \right\rceil, \left\lfloor \frac{N}{2} \right\rfloor\right].$$

Later we will assume that $\mathbf{A}$ consists of point samples from a function $f : [0, 2\pi] \mapsto \mathbb{C}$ so that $\mathbf{A}(j) = f\left(j \cdot \frac{2\pi}{N}\right)$ for all $j \in [0, N) \cap \mathbb{Z}$. We note that this is always the case since $\mathbf{A}$ always has the bandlimited interpolant

$$f(x) = \frac{1}{2\pi} \sum_{\omega=1-\lceil \frac{N}{2} \rceil}^{\lfloor \frac{N}{2} \rfloor} \widehat{\mathbf{A}}(\omega) \cdot e^{\mathbf{i}\omega \cdot x}, \quad x \in [0, 2\pi].$$

However, our results easily extend to the case where f is known to not be bandlimited, but instead exhibits sufficiently fast decay of its large Fourier modes. We formalize this notion by defining (c, γ, N)-decay for any function $f : [0, 2\pi] \mapsto \mathbb{C}$. A function f has (c, γ, N)-decay for $c, \gamma \in \mathbb{R}^+$, $\gamma > 1$, if $|\hat{f}(\omega)| \leq c \cdot |\omega|^{-\gamma}$ for all $\omega \notin \left(-\left\lceil \frac{N}{2} \right\rceil, \left\lfloor \frac{N}{2} \right\rfloor\right]$. Note that any 2π-periodic function f with an integrable second derivative over $[0, 2\pi]$ has (c, γ, N)-decay with $\gamma > 1$ for some constant $c \in \mathbb{R}^+$ (see [10]).

III. Deterministic RIP Construction

Define $p_0 = 1$ and let p_l be the l^{th} prime natural number. Thus, we have

$$p_0 = 1, p_1 = 2, p_2 = 3, p_3 = 5, p_4 = 7, \ldots \tag{1}$$

Choose $q \in \mathbb{N}$ so that

$$p_{q-1} < k \leq p_q. \tag{2}$$

We will use the first $K \in \mathbb{N}$ (to be specified later) primes no smaller than k

$$k \leq p_q < p_{q+1} < \cdots < p_{q+K-1} \tag{3}$$

to create a measurement matrix

$$\mathcal{M} \in \left\{0, \frac{1}{\sqrt{K}}\right\}^{\left(\sum_{j=0}^{K-1} p_{q+j}\right) \times N}.$$

We create our measurement matrix $\mathcal{M}$ as follows: We produce a row $r_{j,h}$, $j \in [0, K) \cap \mathbb{Z}$ and $h \in [0, p_{q+j}) \cap \mathbb{Z}$, in $\mathcal{M}$ for each possible residue of each of our p_{q+j} primes. Each $r_{j,h}$ row's n^{th} entry, $n \in [0, N) \cap \mathbb{Z}$, is given by

$$\begin{aligned}(r_{j,h})_n &= \delta\left((n-h) \bmod p_{q+j}\right) \\ &= \begin{cases} 1 & \text{if } n \equiv h \bmod p_{q+j} \\ 0 & \text{otherwise} \end{cases}.\end{aligned} \tag{4}$$

We then set

$$\mathcal{M} = \frac{1}{\sqrt{K}} \begin{pmatrix} r_{0,0} \\ r_{0,1} \\ \vdots \\ r_{0,p_q-1} \\ r_{1,0} \\ \vdots \\ r_{1,p_{q+1}-1} \\ \vdots \\ r_{K-1,p_{q+K-1}-1} \end{pmatrix}. \tag{5}$$

For an example measurement matrix see Figure 1.

$\mathbf{n} \in [\mathbf{0}, \mathbf{N})$	**0**	**1**	**2**	**3**	**4**	**5**	**6**	...
$\mathbf{n} \equiv \mathbf{0} \bmod \mathbf{2}$	1	0	1	0	1	0	1	...
$\mathbf{n} \equiv \mathbf{1} \bmod \mathbf{2}$	0	1	0	1	0	1	0	...
$\mathbf{n} \equiv \mathbf{0} \bmod \mathbf{3}$	1	0	0	1	0	0	1	...
$\mathbf{n} \equiv \mathbf{1} \bmod 3$	0	1	0	0	1	0	0	...
$\mathbf{n} \equiv \mathbf{2} \bmod \mathbf{3}$	0	0	1	0	0	1	0	...
$\vdots$				$\vdots$				
$\mathbf{n} \equiv \mathbf{1} \bmod \mathbf{5}$	0	1	0	0	0	0	1	...
$\vdots$				$\vdots$				

Fig. 1. Measurement Matrix, $\sqrt{K} \cdot \mathcal{M}$, Using $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, ...

Given the first $k + K$ primes we can see that it is easy to form our measurement matrix $\mathcal{M}$. Thus, we quickly mention that efficient algorithms for generating the first n primes exist (see [11]) and have been widely implemented (e.g., MATLAB's PRIMES function). In addition, previously computed lists of primes can be found online (e.g., the first 50 million primes can be downloaded at [12]). We now concentrate on demonstrating that $\mathcal{M}$ has the Restricted Isometry Property.

Theorem 2. *Our measurement matrix $\mathcal{M}$ has RIP(N,k,ϵ) for all $K \geq \frac{(k-1)\cdot\lfloor \log_k N \rfloor}{\epsilon}$.*

Proof:

We prove the result via an argument similar to one employed by DeVore in [4]. Let $S = \{n_1, \ldots, n_k\} \subset [0, N)$. Given any such S we define $\mathcal{M}_S$ to be the $\left(\sum_{j=0}^{K-1} p_{q+j}\right) \times k$ matrix consisting of $\mathcal{M}$'s S columns. We now consider the $k \times k$ Grammian (and therefore symmetric and non-negative definite) matrix

$$\mathcal{M}_S^* \mathcal{M}_S = \mathcal{I} + \mathcal{D}_S.$$

If we can show that both $\| \mathcal{D}_S \|_1 \leq \epsilon$ and $\| \mathcal{D}_S \|_\infty \leq \epsilon$ are true, we will have our result.

Each off diagonal entry $(\mathcal{D}_S)_{l,m}, l \neq m$, is the inner product of $\mathcal{M}$'s associated n_l and n_m columns. Thus, we have

$$(\mathcal{D}_S)_{l,m} = \frac{1}{K} \cdot \sum_{j=0}^{K-1} \delta\left((n_l - n_m) \bmod p_{q+j}\right) \leq \frac{\lfloor \log_k N \rfloor}{K}$$

by the Chinese Remainder Theorem (see [11]). The end result is that both $\| \mathcal{D}_S \|_1$ and $\| \mathcal{D}_S \|_\infty$ are $\leq \frac{(k-1)\cdot\lfloor \log_k N \rfloor}{K}$. Theorem 2 follows. □

We will now study the number of rows, $\sum_{j=0}^{K-1} p_{q+j}$, in our measurement matrix. It follows from results in [13] that

$$\sum_{j=0}^{K-1} p_{q+j} = \frac{p_{q+K}^2}{2 \ln p_{q+K}} \cdot \left(1 + O\left(\frac{1}{\ln p_{q+K}}\right)\right) - \frac{p_q^2}{2 \ln p_q} \cdot \left(1 + O\left(\frac{1}{\ln p_q}\right)\right). \quad (6)$$

Furthermore, the Prime Number Theorem (see [11]) tells us that

$$q = \frac{k}{\ln k}\left(1 + O\left(\frac{1}{\ln k}\right)\right).$$

Thus, if we use the smallest possible value for K in order to obtain a measurement matrix $\mathcal{M}$ with RIP(N,k,ϵ) we will have

$$q + K = \frac{k \cdot \lfloor \log_k N \rfloor}{\epsilon}\left(1 + O\left(\frac{\epsilon}{\ln N}\right)\right).$$

Here we have assumed that K is an natural number (e.g., that $\epsilon^{-1} \in \mathbb{N}$) and that $k + K$ is less than N.

Applying the Prime Number Theorem once more we have that

$$p_q = k\left(1 + O\left(\frac{\ln \ln k}{\ln k}\right)\right)$$

and

$$p_{q+K} = \frac{k \cdot \lfloor \log_k N \rfloor \cdot \ln\left(\frac{k \cdot \ln N}{\epsilon}\right)}{\epsilon}\left(1 + O\left(\frac{\ln \ln\left(\frac{k \ln N}{\epsilon}\right)}{\ln\left(\frac{k \ln N}{\epsilon}\right)}\right)\right). \quad (7)$$

Utilizing Equation 6 now yields

$$\sum_{j=0}^{K-1} p_{q+j} = \frac{k^2 \cdot \lfloor \log_k N \rfloor^2 \cdot \ln\left(\frac{k \cdot \ln N}{\epsilon}\right)}{2\epsilon^2}\left(1 + O\left(\frac{\ln \ln\left(\frac{k \ln N}{\epsilon}\right)}{\ln\left(\frac{k \ln N}{\epsilon}\right)}\right)\right). \quad (8)$$

Hence, we have an asymptotic for the number of rows in $\mathcal{M}$. In the next theorem we provide a concrete bound.

Theorem 3. *Suppose that $\epsilon^{-1} \in \mathbb{N} - \{1\}$ and $N > k \geq 2$. If we use the smallest $K \geq 5$ possible so that $\mathcal{M}$ has RIP(N,k,ϵ), the number of rows in $\mathcal{M}$ will be bounded above by*

$$\frac{7k^2 \cdot \lfloor \log_k N \rfloor^2}{\epsilon^2} \ln\left(\frac{3.05 \cdot k \cdot \lfloor \log_k N \rfloor}{\epsilon}\right).$$

Tighter (and asymptotically correct) upper row bounds may be explicitly calculated using Equations 9 – 12 below.

Proof:

Let $\pi(n)$ be the number of primes no greater than n. In [14] it is shown that

$$\frac{n}{\ln n}\left(1 + \frac{0.992}{\ln n}\right) \leq \pi(n) \leq \frac{n}{\ln n}\left(1 + \frac{1.2762}{\ln n}\right)$$

for all $n \geq 599$. Using this result (in combination with numerical tests for $n < 600$) we obtain the following bounds

for $q+K$ and q.

$$q+K \leq \pi(k)+K+1 \leq \frac{k\cdot\lfloor\log_k N\rfloor}{\epsilon}\left(1+\frac{\epsilon}{\ln k\cdot\lfloor\log_k N\rfloor}+\frac{1.2762\cdot\epsilon}{\ln^2 k\cdot\lfloor\log_k N\rfloor}\right). \quad (9)$$

and

$$q \geq \pi(k) \geq \max\left\{\frac{k}{\ln k}\left(1+\frac{0.992}{\ln k}-\frac{8.85}{k}\right), 1\right\}. \quad (10)$$

Continuing, we can bound the sums of our utilized primes by noting that

$$\sum_{j=1}^{q-1} p_j \geq \sum_{j=1}^{q-1} j\cdot\ln(j) \quad \text{(see [14])} \geq \int_1^{q-1} x\cdot\ln x\, dx \geq \frac{(q-1)^2}{2}\left(\ln(q-1)-\frac{1}{2}\right) \quad (11)$$

and

$$\begin{aligned}\sum_{j=1}^{q+K-1} p_j &\leq 10+\sum_{j=4}^{q+K-1} j\cdot\ln(p_j) \quad \text{(see [14])}\\ &\leq 10+\ln(p_{q+K})\cdot\left(\sum_{j=4}^{q+K-1} j\right) \quad \text{(see [14])}\\ &\leq \frac{(q+K-1)(q+K)}{2}\\ &\quad\cdot\ln\left((q+K)\cdot(\ln(q+K)+\ln\ln(q+K))\right) \quad (12)\\ &\leq \frac{3}{4}(q+K)^2\cdot\ln(q+K). \quad (13)\end{aligned}$$

Using Equation 9 together with Equation 13 finishes the proof. □

Theorem 3 establishes our explicit measurement matrix construction as being near-optimal (i.e., within logarithmic factors) with respect the number of rows for explicit RIP matrices with binary entries [15]. Unfortunately, RIP(N,k,ϵ) matrices consisting of binary entries are suboptimal with respect to row number [4], [15]. It is known that $\Omega(k\cdot\log(N/k))$ rows are necessary for fixed ϵ, and randomized constructions [6], [16] can generate RIP matrices containing numbers of rows within constant factors of this lower bound. However, the structure of explicitly constructed RIP matrices can endow them with beneficial properties that randomly constructed RIP matrices lack. Examples of such properties include (*i*) guaranteed recovery of all sufficiently sparse signals, (*ii*) small storage requirements, and (*iii*) fast matrix multiplication procedures for important change-of-basis matrices Ψ.

Clearly, any explicit RIP matrix construction will have recovery guarantees (property (*i*)). Verifying that a randomly constructed matrix has the RIP property is computationally intractable. Furthermore, randomly constructed RIP matrices generally require $\Omega(N\cdot k\cdot\log(N/k))$ bits of storage (e.g., in the Gaussian and Bernoulli cases [6], [16]). On the other hand, our explicit number theoretic construction requires only

$$O\left(\sum_{j=0}^{K-1}\log p_{q+j}\right) = O\left((q+K)\cdot\ln(q+K)\right) = O\left(\frac{k\cdot\lfloor\log_k N\rfloor}{\epsilon}\cdot\ln\left(\frac{k\cdot\lfloor\log_k N\rfloor}{\epsilon}\right)\right)$$

bits of storage (see [14] and Equation 9). Finally, our RIP matrix is relatively sparse. Its easy to see that it (as well as its adjoint) can be multiplied by a vector in $O(K\cdot N) = O\left(\frac{k\cdot\lfloor\log_k N\rfloor}{\epsilon}\cdot N\right)$ time. Of greater interest, however, is that our RIP matrices have fast multiplication algorithms which utilize a small number of function samples in the Fourier CS case (i.e., when Ψ is a DFT matrix).

IV. The Fourier Case

Let P be the least common multiple of $\{N, p_q, \ldots, p_{q+K-1}\}$ and imagine that we form the new Fourier CS problem

$$\tilde{\mathbf{A}}(h) = f\left(h\cdot\frac{2\pi}{P}\right) \text{ for } h\in[0,P)\cap\mathbb{Z}, \quad x = \tilde{\Psi}\tilde{\mathbf{A}}$$

where $f:[0,2\pi]\mapsto\mathbb{C}$ is such that

$$\mathbf{A}(j) = f\left(j\cdot\frac{2\pi}{N}\right) = \tilde{\mathbf{A}}\left(j\cdot\frac{P}{N}\right)$$

and $\tilde{\Psi}$ is the $P\times P$ DFT matrix defined by $\tilde{\Psi}_{\omega,j} = \frac{2\pi}{P}\cdot e^{\frac{2\pi\mathrm{i}\cdot\omega\cdot j}{P}}$. Recall that there is always such a function f (e.g., we can always let f be $\mathbf{A}$'s bandlimited interpolant). For the time being we will assume we have direct access to $\tilde{\mathbf{A}}$. In section IV-A we will again restrict our sample usage to values from $\mathbf{A}$.

To solve these new problems we will use an extended version of our number theoretic RIP(N,k,ϵ) matrix $\mathcal{M}$. This extended RIP matrix, $\mathcal{M}_P$, is the $\left(\sum_{j=0}^{K-1} p_{q+j}\right)\times P$ matrix formed by extending each row $r_{j,h}$ of $\mathcal{M}$ as per Equation 4 for all $n\in[0,P)$. We now consider the product of $\mathcal{M}_P$ and $\tilde{\Psi}$. For each row $r_{j,h}$ of $\mathcal{M}_P$ and column ω of $\tilde{\Psi}$ we have

$$\begin{aligned}\left(\mathcal{M}_P\cdot\tilde{\Psi}\right)_{r_{j,h},\omega} &= \frac{2\pi}{P\cdot\sqrt{K}}\sum_{l=0}^{\frac{P}{p_{q+j}}-1} e^{\frac{2\pi\mathrm{i}\cdot\omega\cdot\left(h+l\cdot p_{q+j}\right)}{P}}\\ &= \frac{2\pi\cdot e^{\frac{2\pi\mathrm{i}\cdot\omega\cdot h}{P}}}{P\cdot\sqrt{K}}\sum_{l=0}^{\frac{P}{p_{q+j}}-1} e^{\frac{2\pi\mathrm{i}\cdot\omega\cdot l}{P/p_{q+j}}} \quad (14)\\ &= \begin{cases}\frac{2\pi\cdot e^{\frac{2\pi\mathrm{i}\cdot\omega\cdot h}{P}}}{p_{q+j}\cdot\sqrt{K}} & \text{if } \omega\equiv 0 \bmod \frac{P}{p_{q+j}}\\ 0 & \text{otherwise}\end{cases}.\end{aligned}$$

Thus the product $\mathcal{M}_P\cdot\tilde{\Psi}$ is highly sparse. In fact, we can see that each $r_{j,h}$ row contains only p_{q+j} non-zero entries. Better still, all the rows associated with a given p_{q+j} will access all the same $\tilde{\mathbf{A}}$ entries in a pattern consistent with a small FFT. This aliasing phenomena results in a fast algorithm for computing $\mathcal{M}_P\cdot\tilde{\Psi}\cdot\tilde{\mathbf{A}}$ (see Algorithm 1). Theorem 1 shows that $\mathcal{M}_P\tilde{\Psi}\tilde{\mathbf{A}}$ is a good approximation to $\mathcal{M}\Psi\mathbf{A}$ for all signals, f, whose Fourier transforms decay sufficiently fast.

Algorithm 1 FAST MULTIPLY

Input: Signal pointer f, integers $k < K < N$, primes $p_q, \ldots, p_{q+K-1}$
Output: $\mathcal{M}_P \cdot \tilde{\Psi} \cdot \tilde{\mathbf{A}}$
for j from 0 to $K-1$ **do**
  $\mathbf{A}_{p_{q+j}} \leftarrow f(0), f\left(\frac{2\pi}{p_{q+j}}\right), \ldots, f\left(\frac{2\pi(p_{q+j}-1)}{p_{q+j}}\right)$
  $\widehat{\mathbf{A}_{p_{q+j}}} \leftarrow \frac{1}{\sqrt{K}} \cdot \mathbf{FFT}[\mathbf{A}_{p_{q+j}}]$
end for
Output $\left(\widehat{\mathbf{A}_{p_q}}, \widehat{\mathbf{A}_{p_{q+1}}}, \ldots, \widehat{\mathbf{A}_{p_{q+K-1}}}\right)^{\mathrm{T}}$

Lemma 1. *Suppose that $\mathcal{M}$ has RIP(N,k,ϵ). Furthermore, suppose that $N \geq 3$, $p_{q+K} \leq \frac{N}{2}$, and that $\hat{f}$ has (c, γ, N)-decay. Then, $\| \mathcal{M}\Psi\boldsymbol{A} - \mathcal{M}_P\tilde{\Psi}\tilde{\boldsymbol{A}} \|_2$ is $O\left(\frac{c}{\sqrt{k}\cdot(\gamma-1)}\left(\frac{N}{4}\right)^{1-\gamma}\right)$.*

Proof:

Suppose N is odd (the even case is analogous). Then, for all $r_{j,h}$ we have that

$$\begin{aligned}
&\left|\mathcal{M}\Psi\mathbf{A}(r_{j,h}) - \mathcal{M}_P\tilde{\Psi}\tilde{\mathbf{A}}(r_{j,h})\right| = \\
&\frac{1}{\sqrt{K}} \cdot \left| \sum_{l,\ |h+l\cdot p_{q+j}| \leq \frac{N-1}{2}} \widehat{\mathbf{A}}(h + l \cdot p_{q+j}) - \sum_{\omega \equiv h \bmod p_{q+j}} \hat{f}(\omega) \right| = \\
&\frac{1}{\sqrt{K}} \cdot \left| \sum_{l,\ |h+l\cdot p_{q+j}| \leq \frac{N-1}{2}} \ \sum_{\tilde{\omega} \equiv (h+l\cdot p_{q+j}) \bmod N} \hat{f}(\tilde{\omega}) - \sum_{\omega \equiv h \bmod p_{q+j}} \hat{f}(\omega) \right| = \\
&\frac{1}{\sqrt{K}} \cdot \Bigg| \sum_{l,\ |h+l\cdot p_{q+j}| \leq \frac{N-1}{2}} \ \sum_{\tilde{\omega} \equiv (h+l\cdot p_{q+j}) \bmod N,\ |\tilde{\omega}| \geq \frac{N+1}{2}} \hat{f}(\tilde{\omega}) \\
&\qquad - \sum_{\omega \equiv h \bmod p_{q+j},\ |\omega| \geq \frac{N+1}{2}} \hat{f}(\omega) \Bigg| \\
&\leq \frac{2c}{\sqrt{K}} \cdot \left[\left(\frac{N}{p_{q+j}} + 2 \right) \cdot \sum_{l=0}^{\infty} \left(\frac{N+1}{2} + l \cdot N \right)^{-\gamma} \right. \\
&\qquad \left. + \sum_{l=0}^{\infty} \left(\frac{N+1}{2} + l \cdot p_{q+j} \right)^{-\gamma} \right] \\
&\leq \frac{2c}{p_{q+j} \cdot \sqrt{K} \cdot (\gamma - 1)} \cdot \left(8\left(\frac{N}{4}\right)^{1-\gamma} + 2\left(\frac{N - p_{q+K}}{2}\right)^{1-\gamma} \right).
\end{aligned}$$

Thus, we can see that

$$\begin{aligned}
&\| \mathcal{M}\Psi\mathbf{A} - \mathcal{M}_P\tilde{\Psi}\tilde{\mathbf{A}} \|_2^2 \\
&\leq \frac{4c^2}{p_q \cdot (\gamma-1)^2} \cdot \left(8\left(\frac{N}{4}\right)^{1-\gamma} + 2\left(\frac{N - p_{q+K}}{2}\right)^{1-\gamma} \right)^2 \\
&\leq \frac{400 \cdot c^2}{k \cdot (\gamma-1)^2} \left(\frac{N}{4}\right)^{2-2\gamma}.
\end{aligned}$$

The result follows. □

Looking at Section III's Theorem 3 we can see than Algorithm 1 utilizes only $O\left(\frac{k^2 \cdot \lfloor \log_k N \rfloor^2 \cdot \ln\left(\frac{k \cdot \ln N}{\epsilon}\right)}{\epsilon^2}\right)$ f-samples/$\tilde{\mathbf{A}}$-entries. Similarly, we can see that Algorithm 1 runs in time $O\left(\sum_{j=0}^{K-1} p_{q+j} \log p_{q+j}\right)$ if we calculate the FFTs using a chirp z-transform [17]. Thus, the runtime is

$$\begin{aligned}
O\left(\sum_{j=0}^{K-1} p_{q+j} \log p_{q+j}\right) &= O\left(p_{q+K}^2\right) \quad \text{(see [13])} \\
&= O\left(\frac{k^2 \cdot \lfloor \log_k N \rfloor^2 \cdot \ln^2\left(\frac{k \cdot \ln N}{\epsilon}\right)}{\epsilon^2}\right)
\end{aligned}$$

using Equation 7. We next consider the case where we only have access to entries of $\mathbf{A}$ and can not freely sample from f (or $\tilde{\mathbf{A}}$).

A. Restricted Sampling

To restrict our sampling to $\mathbf{A}$ but still use a small number of samples we will utilize a $P \times N$ interpolation matrix based on polynomials in Lagrangian form [18]. We begin by defining n_p as

$$n_p = \max\left\{ n \in [0, N) \cap \mathbb{Z} \;\middle|\; n \leq \frac{p \cdot N}{P} \right\}.$$

Furthermore, we define the polynomial $l_{b,m}^{\kappa} : \mathbb{R} \mapsto \mathbb{R}$ for given $b, \kappa \in [0, N) \cap \mathbb{Z}$ and $m \in (b, b + 2\kappa]$ as

$$l_{b,m}^{\kappa}(x) = \prod_{h=1, b+h \neq m}^{2\kappa} \frac{\frac{x \cdot N}{P} - b - h}{m - b - h}.$$

Now we define the $P \times N$ matrix $\mathcal{L}^{2\kappa}$ by

$$\mathcal{L}_{p,n}^{2\kappa} = \begin{cases} l_{n_p - \kappa, m(n)}^{\kappa}(p) & \text{if } n \in (n_p - \kappa, n_p + \kappa] \bmod N \\ 0 & \text{otherwise} \end{cases}, \quad (15)$$

where $m(n) \in (n_p - \kappa, n_p + \kappa] \cap \mathbb{Z}$ and $m(n) \equiv n \bmod N$.

We will use $\mathcal{L}^{2\kappa}$ to locally interpolate $\mathbf{A}$ to a finer grid containing P equally spaced points. Put another way, we will sample Section IV's $\tilde{\mathbf{A}}$ as needed by calculating $\mathcal{L}^{2\kappa}\mathbf{A} \approx \tilde{\mathbf{A}}$. We can then use Section IV's techniques to approximate the bandlimited interpolant, f, of $\mathbf{A}$. We have the following theorem.

Lemma 2. $\| \mathcal{M}\Psi\boldsymbol{A} - \mathcal{M}_P\tilde{\Psi}\mathcal{L}^{2\kappa}\boldsymbol{A} \|_2$ *is*

$$O\left(\sqrt{\frac{k \cdot \lfloor \log_k N \rfloor \cdot \ln\left(\frac{k \cdot \ln N}{\epsilon}\right)}{\epsilon}} \cdot \frac{\| \hat{\boldsymbol{A}} \|_1}{8^{\kappa}} \right).$$

Proof:

We take $f : [0, 2\pi] \mapsto \mathbb{C}$ in Lemma 1 to be the bandlimited interpolant of $\mathbf{A}$. Thus, if we set

$$\beta = \mathcal{L}^{2\kappa}\mathbf{A} - \tilde{\mathbf{A}}$$

we can see that

$$\| \mathcal{M}\Psi\mathbf{A} - \mathcal{M}_P\tilde{\Psi}\mathcal{L}^{2\kappa}\mathbf{A} \|_2 = \| \mathcal{M}_P\tilde{\Psi}\beta \|_2$$

since $\hat{f}$ (i.e., $\hat{\mathbf{A}}$) has $(0, \gamma, N)$-decay for any $\gamma > 0$. In order to bound this last expression it suffices to bound $\| \beta \|_{\infty}$.

Using standard results concerning polynomial interpolation error (see [18]) we can see that

$$\| \beta \|_\infty \;\leq\; \frac{2 \| \hat{\mathbf{A}} \|_1}{(2\kappa)!} \left(\frac{N}{2}\right)^{2\kappa} \prod_{j=1}^{\kappa} \left(\frac{j}{N}\right)^2 \;\leq\; \frac{2 \| \hat{\mathbf{A}} \|_1}{8^\kappa}.$$

Therefore, using Equation 14, we have that

$$\left|\mathcal{M}_P \tilde{\Psi} \beta(r_{j,h})\right| \;\leq\; \frac{4\pi}{\sqrt{K}} \frac{\| \hat{\mathbf{A}} \|_1}{8^\kappa}.$$

Finishing, we have that

$$\| \mathcal{M}_P \tilde{\Psi} \beta \|_2^2 \;\leq\; 16\pi^2 \cdot p_{q+K} \cdot \frac{\| \hat{\mathbf{A}} \|_1^2}{8^{2\kappa}}.$$

The result follows from Equation 7. □

Theorem 3 together with Algorithm 1 and Equation 15 for $\mathcal{L}^{2\kappa}$ imply that we can approximately calculate $\mathcal{M}\Psi\mathbf{A}$ by $\mathcal{M}_P \tilde{\Psi} \mathcal{L}^{2\kappa} \mathbf{A}$ using only $O\left(\kappa \cdot \frac{k^2 \cdot \lfloor \log_k N \rfloor^2 \cdot \ln\left(\frac{k \cdot \ln N}{\epsilon}\right)}{\epsilon^2}\right)$ $\mathbf{A}$-entries. The required runtime is $O\left(\left(\kappa + \ln\left(\frac{k \cdot \ln N}{\epsilon}\right)\right) \cdot \frac{k^2 \cdot \lfloor \log_k N \rfloor^2 \cdot \ln\left(\frac{k \cdot \ln N}{\epsilon}\right)}{\epsilon^2}\right)$. We now conclude by briefly considering CS recovery.

V. Conclusion

By combining our work herein with Theorem 1 we obtain the following deterministic Fourier approximation result.

Corollary 1. *Suppose that $f : [0, 2\pi] \mapsto \mathbb{C}$ has (c, γ, N)-decay. Fix precision parameter $\eta \in \mathbb{R}$. We can produce a $2k$-term N-bandlimited Fourier approximation, $\widehat{\mathbf{R}}$, so that $\|\hat{f} - \widehat{\mathbf{R}}\|_2$ is*

$$O\left(\frac{1}{\sqrt{k}} \cdot \|\hat{f} - \widehat{\mathbf{R}}_{\text{opt}}\|_1 + \frac{c \cdot (N/4)^{1-\gamma}}{\gamma - 1} + \eta\right).$$

Here $\widehat{\mathbf{R}}_{\text{opt}}$ is a best k-term N-bandlimited Fourier approximation to $\hat{f}$. The required number of samples is $O\left(k^2 \cdot \lfloor \log_k N \rfloor^2 \cdot \ln(k \ln N)\right)$. The required runtime is $O\left(N \cdot k \cdot \lfloor \log_k N \rfloor \cdot \log\left(\frac{1}{\eta} \cdot \left[\|\hat{f}\|_2 + \frac{c}{(\gamma-1)\sqrt{N}}\right]\right)\right)$.

Proof Sketch:

We begin by applying Theorem 1 in combination with Theorem 2, Theorem 3, and Lemma 1. This tells us that we can deterministically approximate $\widehat{\mathbf{A}} = \Psi\mathbf{A}$ well with a $2k$-term Fourier representation using $\tilde{O}(k^2)$ samples. We continue by restating the result in terms of $\hat{f}$. A calculation similar to the one performed for the proof of Lemma 1 reveals that both $\| \chi_{(-\lceil \frac{N}{2} \rceil, \lfloor \frac{N}{2} \rfloor]} \cdot \hat{f} - \Psi\mathbf{A} \|_2$ and $\| \hat{f} - \Psi\mathbf{A} \|_2$ are $O\left(\frac{c}{\gamma-1}\left(\frac{N}{4}\right)^{\frac{1}{2}-\gamma}\right)$. Similarly, $\| \hat{f} - \Psi\mathbf{A} \|_1$ is $O\left(\frac{c}{\gamma-1}\left(\frac{N}{4}\right)^{1-\gamma}\right)$. Finally, one can see that the optimal $2k$-term Fourier representation for $\Psi\mathbf{A}$ must also be a near-optimal representation for $2k$ of f's near-largest magnitude low frequency Fourier coefficients. The result follows. □

Corollary 1 improves on previous deterministic linear-time in bandwidth results [2] in terms of both sampling and runtime requirements. Unlike these previous results, Corollary 1 exhibits no runtime or sampling dependence on the Fourier compressibility of our input signal f. Following the logic of [2] one immediately considers the possibility of sublinearizing Corollary 1's runtime, thereby improving the sublinear-time algorithms developed therein. Indeed, the reconstruction algorithm [7] as used in Corollary 1 is sublinear-time except that it requires multiplications with $\mathcal{M}^*$. However, these multiplications are only used to identify energetic frequencies in $\hat{f}$. In principle they can be replaced by sublinear-time energetic frequency identification procedures from [19] at the expense of additional sampling. Although these identification procedures fail with some small probability, our guaranteed RIP properties allow us to check identified frequencies for energy. This approach bears fruit in the noiseless exact k-sparse signal case. Handling noise is less straightforward.

Acknowledgments

I would like to thank Martin Strauss, Anna Gilbert, Holger Rauhut, Joel Tropp, Felix Krahmer, and Tsvetanka Sendova for helpful comments and advice.

References

[1] M. A. Iwen, "A deterministic sub-linear time sparse fourier algorithm via non-adaptive compressed sensing methods," in *Proc. of ACM-SIAM symposium on Discrete algorithms (SODA'08)*, 2008.

[2] M. A. Iwen, "Combinatorial sublinear-time fourier algorithms," *Submitted. Available at http://www.ima.umn.edu/~iwen/index.html*, 2008.

[3] Emmanuel Candes and Terence Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. on Information Theory*, 2006.

[4] R. DeVore, "Deterministic Constructions of Compressed Sensing Matrices," http://www.dsp.ece.rice.edu/cs/Henryk.pdf, Preprint 2007.

[5] R. Berinde, A. C. Gilbert, P. Indyk, H. Karloff, and M.J. Strauss, "Combining geometry and combinatorics: A unified approach to sparse signal recovery," *preprint*, 2008.

[6] Mark Rudelson and Roman Vershynin, "Sparse reconstruction by convex relaxation: Fourier and gaussian measurements," in *40th Annual Conference on Information Sciences and Systems (CISS)*, 2006.

[7] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *preprint*, 2008.

[8] Stefan Kunis and Holger Rauhut, "Random Sampling of Sparse Trigonometric Polynomials II - Orthogonal Matching Pursuit versus Basis Pursuit," *Foundations of Computational Mathematics*, to appear.

[9] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *submitted*, 2007.

[10] J. P. Boyd, *Chebyshev and Fourier Spectral Methods*, Dover Publications, Inc., 2001.

[11] Ivan Niven, Herbert S. Zuckerman, and Hugh L. Montgomery, *An Introduction to The Theory of Numbers*, John Wiley & Sons, Inc., 1991.

[12] "The Prime Pages," http://primes.utm.edu/.

[13] M. A. Iwen and C. V. Spencer, "Improved bounds for a deterministic sublinear-time sparse fourier algorithm," in *Conference on Information Sciences and Systems (CISS)*, 2008.

[14] Pierre Dusart, "The k^{th} prime is greater than $k(\ln k + \ln \ln k - 1)$ for $k \geq 2$," *Mathematics of Computation*, vol. 68, no. 225, 1999.

[15] Venkat Chandar, "A negative result concerning explicit matrices with the restricted isometry property," *Preprint*, 2008.

[16] David L. Donoho and Jared Tanner, "Thresholds for the recovery of sparse solutions via l1 minimization," in *40th Annual Conference on Information Sciences and Systems (CISS)*, 2006.

[17] L.R. Rabiner, R.W. Schafer, and C.M. Rader, "The Chirp z-Transform Algorithm," *IEEE Transactions on Audio and Electroacoustics*, vol. AU-17, no. 2, pp. 86–92, June 1969.

[18] David Kincaid and Ward Cheney, *Numerical Analysis: Mathematics of Scientific Computing*, China Machine Press, 2003.

[19] A. Gilbert, S. Muthukrishnan, and M. Strauss, "Improved time bounds for near-optimal sparse Fourier representations," *SPIE*, 2005.