A Deterministic Sub-linear Time Sparse Fourier Algorithm via Non-adaptive Compressed Sensing Methods*

M. A. Iwen†

Abstract

We study the problem of estimating the best B term Fourier representation for a given frequency-sparse signal (i.e., vector) $\mathbf{A}$ of length $N \gg B$. More precisely, we investigate how to deterministically identify B of the largest magnitude frequencies of $\hat{\mathbf{A}}$, and estimate their coefficients, in polynomial$(B, \log N)$ time. Randomized sub-linear time algorithms, which have a small (controllable) probability of failure for each processed signal, exist for solving this problem. However, for failure intolerant applications such as those involving mission-critical hardware designed to process many signals over a long lifetime, deterministic algorithms with no probability of failure are highly desirable. In this paper we build on the deterministic Compressed Sensing results of Cormode and Muthukrishnan (CM) [26, 6, 7] in order to develop the first known deterministic sub-linear time sparse Fourier Transform algorithm suitable for failure intolerant applications. Furthermore, in the process of developing our new Fourier algorithm, we present a simplified deterministic Compressed Sensing algorithm which improves on CM's algebraic compressibility results while simultaneously maintaining their results concerning exponential decay.

1 Introduction

In many applications only the top few most energetic terms of a signal's Fourier Transform (FT) are of interest. In such applications the Fast Fourier Transform (FFT), which computes all FT terms, is computationally wasteful. To make our point, we next consider a simple application-based example in which the FFT can be replaced by faster approximate Fourier methods.

1.1 Sub-Nyquist Single Frequency Acquisition

Let $f : [0, 2\pi] \to \mathbb{C}$ be a non-identically zero function of the form

$$f(x) = C \cdot e^{\mathbb{i}\omega x}$$

consisting of a single unknown frequency $\omega \in (-N, N]$ (e.g., consider a windowed sinusoidal portion of a wide-band frequency-hopping signal [21]). Sampling at the Nyquist-rate would dictate the need for at least $2N$ equally spaced samples from f in order to discover ω via the FFT without aliasing [3]. Thus, we would have to compute the FFT of the $2N$-length vector

$$\mathbf{A}(j) = f\left(\frac{\pi j}{N}\right), \quad 0 \leq j < 2N.$$

However, if we use aliasing to our advantage, we can correctly determine ω with significantly fewer f-samples as follows:

Let $\mathbf{A}_2$ be a 2-element array of f-samples with

$$\mathbf{A}_2(0) = f(0) = C, \text{ and } \mathbf{A}_2(1) = f(\pi) = C \cdot (-1)^{\omega}.$$

Calculating $\widehat{\mathbf{A}_2}$ we get that

$$\widehat{\mathbf{A}_2}(0) = C \cdot \frac{1 + (-1)^{\omega}}{\sqrt{2}}, \text{ and } \widehat{\mathbf{A}_2}(1) = C \cdot \frac{1 + (-1)^{\omega+1}}{\sqrt{2}}.$$

Note that since ω is an integer, exactly one element of $\widehat{\mathbf{A}_2}$ will be non-zero. If $\widehat{\mathbf{A}_2}(0) \neq 0$ then we know that $\omega \equiv 0$ modulo 2. On the other hand, $\widehat{\mathbf{A}_2}(1) \neq 0$ implies that $\omega \equiv 1$ modulo 2. In this same fashion we may use several potentially aliased Fast Fourier Transforms in parallel to discover ω modulo $3, 5, \ldots$, the $O(\log N)^{\text{th}}$ prime. Once we have collected these moduli we can reconstruct ω via the famous **Chinese Remainder Theorem (CRT)**.

Theorem 1. Chinese Remainder Theorem (CRT): *Any integer x is uniquely specified mod N by its remainders modulo m relatively prime integers $p_1, \ldots, p_m$ as long as $\prod_{l=1}^{m} p_l \geq N$.*

To finish our example, suppose that $N = 500,000$ and that we have used three FFT's with 100, 101, and 103 samples to determine that $\omega \equiv 34$ mod 100, $\omega \equiv 3$ mod 101, and $\omega \equiv 1$ mod 103, respectively. Using that $\omega \equiv 1$ mod 103 we can see that $\omega = 103 \cdot a + 1$ for some integer a. Using this new expression for ω in our second modulus we get

$$(103 \cdot a + 1) \equiv 3 \text{ mod } 101 \Rightarrow a \equiv 1 \text{ mod } 101.$$

Therefore, $a = 101 \cdot b + 1$ for some integer b. Substituting for a we get that $\omega = 10403 \cdot b + 104$. By similar work

*Supported in part by NSF DMS-0510203.

†University of Michigan - Ann Arbor

we can see that $b \equiv 10 \bmod 100$ after considering ω modulo 100. Hence, $\omega = 104,134$ by the CRT. As an added bonus we note that our three FFTs will have also provided us with three different estimates of ω's coefficient C.

The end result is that we have used significantly less than $2N$ samples to determine ω. Using the CRT we required only $100 + 101 + 103 = 304$ samples from f to determine ω since $100 \cdot 101 \cdot 103 > 1,000,000$. In contrast, a million f-samples would be gathered during Nyquist-rate sampling. Besides needing significantly less samples than the FFT, this CRT-based single frequency method dramatically reduces required computational effort. And, it's deterministic. There is no chance of failure. Of course, a single frequency signal is incredibly simple. Signals involving more than 1 nonzero frequency are much more difficult to handle since frequency moduli may begin to collide modulo various numbers. Dealing with the potential difficulties caused by such frequency collisions in a deterministic way comprises the majority of this paper.

1.2 Compressed Sensing and Related Work

Compressed Sensing (CS) methods [4, 28, 26, 6, 7] provide a robust framework for reducing the number of measurements required to summarize sparse signals. For this reason CS methods are useful in areas such as MR imaging [23, 24] and analog-to-digital conversion [21, 20] where measurement costs are high. The general CS setup is as follows: Let $\mathbf{A}$ be an N-length signal/vector with complex valued entries, and Ψ be a full rank $N \times N$ change of basis matrix. Furthermore, suppose that $\Psi \cdot \mathbf{A}$ is sparse (i.e., only $k \ll N$ entries of $\Psi \cdot \mathbf{A}$ are significant/large in magnitude). CS methods deal with generating a $K \times N$ measurement matrix, $\mathcal{M}$, with the smallest number of rows possible (i.e., K minimized) so that the k significant entries of $\Psi \cdot \mathbf{A}$ can be approximately recovered from the K-element result of

$$\mathcal{M} \cdot \Psi \cdot \mathbf{A}. \tag{1.1}$$

Note that CS is inherently algorithmic since a procedure for recovering $\Psi \cdot \mathbf{A}$'s largest k-entries from the result of Equation 1.1 must be specified.

For the remainder of this paper we will consider the special CS case where Ψ is the $N \times N$ Discrete Fourier Transform matrix. Hence, we have

$$\Psi_{i,j} = \frac{e^{\frac{-2\pi \mathrm{i} \cdot i \cdot j}{N}}}{\sqrt{N}}. \tag{1.2}$$

Our problem of interest is to find, and estimate the coefficients of, the k significant entries (i.e., frequencies) of $\hat{\mathbf{A}}$ given a frequency-sparse (i.e., smooth) signal $\mathbf{A}$. In this case the deterministic Fourier CS measurement matrixes, $\mathcal{M} \cdot \Psi$, produced by [28, 26, 6, 7] require super-linear $O(KN)$-time to multiply by $\mathbf{A}$ in Equation 1.1. Similarly, the energetic frequency recovery procedure of [4, 9] requires super-linear time in N. Hence, none of [4, 28, 9, 26, 6, 7] have both sub-linear measurement and reconstruction time.

Existing randomized sub-linear time Fourier algorithms [15, 19, 16] not only show great promise for decreasing measurement costs, but also for speeding up the numerical solution of computationally challenging multi-scale problems [8, 18]. However, these algorithms are not deterministic, and so can produce incorrect results with some small probability on each input signal. Thus, they aren't appropriate for long-lived failure intolerant applications.

In this paper we build on the deterministic Compressed Sensing methods of Cormode and Muthukrishnan (CM) [26, 6, 7] in order to construct the first known deterministic sub-linear time sparse Fourier algorithm. In order to produce our new Fourier algorithm we must modify CM's work in two ways: First, we alter CM's measurement construction in order to allow sub-linear time computation of Fourier measurements via aliasing. Thus, our algorithm can deterministically approximate the result of Equation 1.1 in time $K \cdot \mathrm{polylog}(N)$. Second, CM use a k-strongly selective collection of sets [17] to construct their measurements for algebraically compressible signals. We introduce the notion of a K-majority k-strongly selective collection of sets which leads us to a new reconstruction algorithm with better algebraic compressibility results than CM's algorithm. As a result, our deterministic sub-linear time Fourier algorithm has better then previously possible algebraic compressibility behavior.

The main contributions of this paper are:

1. We present a new deterministic compressed sensing algorithm that both (i) improves on CM's algebraically compressible signal results, and (ii) has comparable measurement/run time requirements to CM's algorithm for exponentially decaying signals.

2. We present the first known deterministic sub-linear time sparse DFT. In the process, we explicitly demonstrate the connection between compressed sensing and sub-linear time Fourier transform methods.

3. We introduce K-majority k-strongly selective collections of sets which have potential applications to streaming algorithms along the lines of [25, 13].

The remainder of this paper is organized as follows:

In section 2 we introduce relevant definitions and terminology. Then, in section 3, we define K-majority k-strongly selective collections of sets and use them to construct our compressed sensing measurements. Section 4 contains our new deterministic compressed sensing algorithm along with analysis of it's accuracy and run time. Finally, we present our deterministic sub-linear time Fourier algorithm in sections 5 and 5.1. Section 6 contains a short conclusion.

2 Preliminaries

Throughout the remainder of this paper we will be interested in complex-valued functions $f : [0, 2\pi] \to \mathbb{C}$ and signals (or arrays) of length N containing f values at various $t \in [0, 2\pi]$. We shall denote such signals by $\mathbf{A}$, where $\mathbf{A}(j) \in \mathbb{C}$ is the signal's j^{th} complex value for all $j \in [0, N-1] \subset \mathbb{N}$. Hereafter we will refer to the process of either calculating, measuring, or retrieving the f value associated any $\mathbf{A}(j) \in \mathbb{C}$ from machine memory as *sampling* from f and/or $\mathbf{A}$. Given a signal $\mathbf{A}$ we define its discrete L^2-norm, or Euclidean norm, to be

$$\|\mathbf{A}\|_2 = \sqrt{\sum_{j=0}^{N-1} |\mathbf{A}(j)|^2}.$$

We will also refer to $\|\mathbf{A}\|_2^2$ as $\mathbf{A}$'s *energy*.

For any signal, $\mathbf{A}$, its Discrete Fourier Transform (DFT), denoted $\hat{\mathbf{A}}$, is another signal of length N defined as follows:

$$\hat{\mathbf{A}}(\omega) = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} e^{\frac{-2\pi i \omega j}{N}} \mathbf{A}(j), \quad \forall \omega \in [0, N).$$

Furthermore, we may recover $\mathbf{A}$ from its DFT via the Inverse Discrete Fourier Transform (IDFT) as follows:

$$\mathbf{A}(j) = \hat{\hat{\mathbf{A}}}^{-1}(j) = \frac{1}{\sqrt{N}} \sum_{\omega=0}^{N-1} e^{\frac{2\pi i \omega j}{N}} \hat{\mathbf{A}}(\omega), \quad \forall j \in [0, N).$$

We will refer to any index, ω, of $\hat{\mathbf{A}}$ as a *frequency*. Furthermore, we will refer to $\hat{\mathbf{A}}(\omega)$ as frequency ω's *coefficient* for each $\omega \in [0, N)$. Parseval's equality tells us that $\|\hat{\mathbf{A}}\|_2 = \|\mathbf{A}\|_2$ for any signal. In other words, the DFT preserves Euclidean norm and energy. Note that any non-zero coefficient frequency will contribute to $\hat{\mathbf{A}}$'s energy. Hence, we will also refer to $|\hat{\mathbf{A}}(\omega)|^2$ as frequency ω's *energy*. If $|\hat{\mathbf{A}}(\omega)|$ is relatively large we'll say that ω is *energetic*.

Our algorithm produces output of the form $(\omega_0, C_0), \ldots, (\omega_{B-1}, C_{B-1})$ where each $(\omega_j, C_j) \in [0, N-1] \times \mathbb{C}$. We will refer to any such set of $B < N$ tuples

$$\left\{ (\omega_j, C_j) \in [0, N-1] \times \mathbb{C} \;\middle|\; 0 \leq j < B \right\}$$

as a **sparse Fourier representation** and denote it with a superscript 's'. Note that if we are given a sparse Fourier representation, $\hat{\mathbf{R}}^{\mathrm{s}}$, we may consider $\hat{\mathbf{R}}^{\mathrm{s}}$ to be a length-N signal. We simply view $\hat{\mathbf{R}}^{\mathrm{s}}$ as the N length signal

$$\hat{\mathbf{R}}(j) = \begin{cases} C_j & \text{if } (j, C_j) \in \widehat{\mathbf{R}}^{\mathrm{s}} \\ 0 & \text{otherwise} \end{cases}$$

for all $j \in [0, N-1]$. Using this idea we may, for example, compute $\mathbf{R}$ from $\hat{\mathbf{R}}^{\mathrm{s}}$ via the IDFT.

A B term/tuple sparse Fourier representation is *B-optimal* for a signal $\mathbf{A}$ if it contains B of the most energetic frequencies of $\hat{\mathbf{A}}$ along with their coefficients. More precisely, we'll say that a sparse Fourier representation

$$\hat{\mathbf{R}}^{\mathrm{s}} = \left\{ (\omega_j, C_j) \in [0, N-1] \times \mathbb{C} \;\middle|\; 0 \leq j < B \right\}$$

is *B-optimal* for $\mathbf{A}$ if there exists a valid ordering of $\hat{\mathbf{A}}$'s coefficients by magnitude

$$|\hat{\mathbf{A}}(\omega_0)| \geq \ldots \geq |\hat{\mathbf{A}}(\omega_j)| \geq \ldots \geq |\hat{\mathbf{A}}(\omega_{N-1})| \tag{2.3}$$

so that $\left\{ (\omega_l, \hat{\mathbf{A}}(\omega_l)) \;\middle|\; l \in [0, B) \right\} = \hat{\mathbf{R}}^{\mathrm{s}}$. Note that a signal may have several B-optimal Fourier representations if its frequency coefficient magnitudes are non-unique. For example, there are two 1-optimal sparse Fourier representations for the signal

$$\mathbf{A}(j) = 5e^{\frac{2\pi i j}{N}} + 5e^{\frac{4\pi i j}{N}}, \; N > 2.$$

However, all B-optimal Fourier representations, $\hat{\mathbf{R}}^{\mathrm{s}}_{\mathrm{opt}}$, for any signal $\mathbf{A}$ will always have both the same unique $\|\mathbf{R}_{\mathrm{opt}}\|_2$ and $\|\mathbf{A} - \mathbf{R}_{\mathrm{opt}}\|_2$ values.

We continue with two final definitions: Let ω_b be a b^{th} most energetic frequency as per Equation 2.3. We will say that a signal $\hat{\mathbf{A}}$ is *(algebraically) p-compressible* for some $p > 1$ if $|\hat{\mathbf{A}}(\omega_b)| = O(b^{-p})$ for all $b \in [1, N)$. If $\mathbf{R}^{\mathrm{s}}_{\mathrm{opt}}$ is a B-optimal Fourier representation we can see that

$$\|\mathbf{A} - \mathbf{R}_{\mathrm{opt}}\|_2^2 = \sum_{b=B}^{N-1} |\hat{\mathbf{A}}(\omega_b)|_2^2 = O\left(\int_B^{\infty} b^{-2p} db \right). \tag{2.4}$$

Hence, any p-compressible signal $\mathbf{A}$ (i.e., any signal with a fixed $c \in \mathbb{R}$ so that $|\mathbf{A}(\omega_b)|_2 \leq c \cdot b^{-p}$ for all $b \in [1, N)$) will have $\|\mathbf{A} - \mathbf{R}_B^{\mathrm{opt}}\|_2^2 \leq \tilde{c}_p \cdot B^{1-2p}$ for some $\tilde{c}_p \in \mathbb{R}$. For any p-compressible signal class (i.e., for any choice of p and c) we will refer to the related optimal $O(B^{1-2p})$-size worst case error value (i.e., Equation 2.4 above) as $\|C_B^{\mathrm{opt}}\|_2^2$. Similarly, we define an *exponentially compressible* (or *exponentially decaying*) signal for a fixed $\alpha \in \mathbb{R}^+$ to be one for which $|\hat{\mathbf{A}}(\omega_b)| = O(2^{-\alpha b})$. The optimal worst case error is then

$$\|C_B^{\mathrm{opt}}\|_2^2 = O\left(\int_B^{\infty} 4^{-\alpha b} db \right) = O(4^{-\alpha B}). \tag{2.5}$$

Fix δ small (e.g., $\delta = 0.1$). Given a compressible input signal, $\mathbf{A}$, our deterministic Fourier algorithm will identify B of the most energetic frequencies from $\hat{\mathbf{A}}$ and approximate their coefficients to produce a Fourier representation $\hat{\mathbf{R}}^{\text{s}}$ with $\|\mathbf{A} - \mathbf{R}\|_2^2 \leq \|\mathbf{A} - \mathbf{R}_{\text{opt}}\|_2^2 + \delta\|C_B^{\text{opt}}\|_2^2$. These are the same types of compressible signal results proven by CM [6, 7].

3 Construction of Measurements

We will use the following types of subset collections to form our measurements:

DEFINITION 3.1. *A collection, $\mathcal{S}$, of s subsets of $[0, N)$ is called K-**majority** k-**strongly selective** if for any $X \subset [0, N)$ with $|X| \leq k$, and for all $x \in X$, the following are true: (i) x belongs to K subsets in $\mathcal{S}$ and, (ii) more than two-thirds of $S_j \in \mathcal{S}$ containing x are such that $S_j \cap X = \{x\}$ (i.e., every member of X occurs separated from all other members of X in more than two-thirds of the K $\mathcal{S}$-subsets it belongs to).*

A K-majority k-strongly selective collection of sets is a more structured version of a **k-strongly selective collection of sets** [17, 26]. Every K-majority k-strongly selective collection of sets not only isolates each $x \in X$, but does so a $\frac{2}{3}^{\text{rd}}$'s majority of the time. Thus, our newly defined K-majority k-strongly selective collections will help us count how many times each significant signal entry is isolated. This added structure allows a new reconstruction algorithm (Algorithm 1) with better algebraic compressibility properties than previous methods.

Next, we will build $O(\log N)$ K-majority k-strongly selective collections of subsets. Each of these $O(\log N)$ collections will ultimately be used to determine energetic frequencies modulo a small prime $< N$. These moduli will then be used along with the Chinese Remainder Theorem to reconstruct each energetic frequency in a manner akin to the introduction's simple example. Our technique is motivated by the method of prime groupings first employed in [25]. To begin, we will denote each of the $O(\log N)$ collections of subsets by $\mathcal{S}_l$ where $0 \leq l \leq O(\log N)$. We construct each of these K-majority k-strongly selective collections as follows:

Define $p_0 = 1$ and let

$$p_1, p_2, \ldots, p_l, \ldots, p_m$$

be the first m primes where m is such that

$$\prod_{l=1}^{m-1} p_l \leq \frac{N}{k} \leq \prod_{l=1}^{m} p_l.$$

Hence, p_l is the l^{th} prime natural number and we have

$$p_0 = 1, p_1 = 2, p_2 = 3, p_3 = 5, \ldots, p_m = O(m \log m).$$

Note that we know $p_m = O(m \log m)$ via the Prime Number Theorem, and so $p_m = O(\log N \log\log N)$. Each p_l will correspond to a different K-majority k-strongly selective collection of subsets of $[0, N) = \{0, \ldots, N-1\}$.

Along these same lines we let q_1 through q_K be the first K (to be specified later) consequitive primes such that

$$\max(p_m, k) \leq q_1 \leq q_2 \leq \ldots \leq q_K.$$

We are now ready to build $\mathcal{S}_0$, our first K-majority k-strongly selective collection of sets. We begin by letting $S_{0,j,h}$ for all $1 \leq j \leq K$ and $0 \leq h \leq q_j - 1$ be

$$S_{0,j,h} = \{n \in [0, N) \mid n \equiv h \bmod q_j\}.$$

Next, we progressively define $S_{0,j}$ to be all integer residues mod q_j, i.e.,

$$S_{0,j} = \{S_{0,j,h} \mid h \in [0, q_j)\},$$

and conclude by setting $\mathcal{S}_0$ equal to all K such q_j-residue groups:

$$\mathcal{S}_0 = \cup_{j=1}^{K} S_{0,j}.$$

More generally, for $0 \leq l \leq m$ we define $\mathcal{S}_l$ to be

$$\cup_{j=1}^{K} \left\{ \{n \in [0, N) \mid n \equiv h \bmod p_l q_j\} \;\middle|\; h \in [0, p_l q_j) \right\}.$$

LEMMA 3.1. *Fix k. If we set $K \geq 3(k-1)\lfloor \log_k N \rfloor + 1$ then $\mathcal{S}_0$ will be a K-majority k-strongly selective collection of sets. Furthermore, if $K = O(k \log_k N)$ then $|\mathcal{S}_0| = O\left(k^2 \log_k^2 N \cdot \max(\log k, \log\log_k N)\right)$.*

Proof. Let $X \subset [0, N)$ be such that $|X| \leq k$. Furthermore, let $x, y \in X$ be such that $x \neq y$. By the Chinese Remainder Theorem we know that x and y may only collide modulo at most $\lfloor \log_k N \rfloor$ of the K q-primes $q_K \geq \ldots \geq q_1 \geq k$. Hence, x may collide with all the other elements of X (i.e., with $X - \{x\}$) modulo at most $(k-1)\lfloor \log_k N \rfloor$ q-primes. We can now see that x will be isolated from all other elements of X modulo at least $K - (k-1)\lfloor \log_k N \rfloor \geq 2(k-1)\lfloor \log_k N \rfloor + 1 > \frac{2K}{3}$ q-primes. This leads us to the conclusion that $\mathcal{S}_0$ is indeed K-majority k-strongly selective.

Finally, we have that

$$|\mathcal{S}_0| \leq \sum_{j=1}^{K} q_j \leq K \cdot q_K.$$

Furthermore, given that $K > \max(k, m)$, the Prime Number Theorem tells us that $q_K = O(K \log K)$. Thus, we can see that $\mathcal{S}_0$ will indeed contain $O\left(k^2 \log_k^2 N \cdot \max(\log k, \log\log_k N)\right)$ sets.

Note that at least $O(k \log_k N)$ primes are required in order to create a (K-majority) k-strongly separating collection of subsets using primes in this fashion. Given any $x \in [0, N)$ a $k-1$ element subset X can be created via the Chinese Remainder Theorem and x moduli so that every element of X collides with x in any desired $O(\log_k N)$ q-primes. We next consider the properties of the other m collections we have defined: $\mathcal{S}_1, \ldots, \mathcal{S}_m$.

LEMMA 3.2. *Let* $S_{l,j,h} = \{n \in [0, N) \mid n \equiv h \bmod p_l q_j\}$, $X \subset [0, N)$ *have* $\leq k$ *elements, and* $x \in X$. *Furthermore, suppose that* $S_{0,j,h} \cap X = \{x\}$. *Then, for all* $l \in [1, m]$, *there exists a unique* $b \in [0, p_l)$ *so that* $S_{l,j,h+b\cdot q_j} \cap X = \{x\}$.

Proof. Fix any $l \in [1, m]$. $S_{0,j,h} \cap X = \{x\}$ implies that $x = h + a \cdot q_j$ for some unique integer a. Using a's unique representation modulo p_l (i.e., $a = b + c \cdot p_l$) we get that $x = h + b \cdot q_j + c \cdot q_j p_l$. Hence, we can see that $x \in S_{l,j,h+bq_j}$. Furthermore, no other element of X is in $S_{l,j,h+t\cdot q_j}$ for any $t \in [0, p_l)$ since it's inclusion therein would imply that it was also an element of $S_{0,j,h}$.

Note that Lemma 3.2 and Lemma 3.1 together imply that each $\mathcal{S}_1, \ldots, \mathcal{S}_m$ is also a K-majority k-strongly separating collection of subsets. Also, we can see that if $x \in S_{l,j,h+b\cdot q_j}$ we can find $x \bmod p_l$ by simply computing $h + bq_j \bmod p_l$. Finally, we form our measurement matrix:

Set $\mathcal{S} = \cup_{l=0}^{m} \mathcal{S}_l$. To form our measurement matrix, $\mathcal{M}$, we simply create one row for each $S_{l,j,h} \in \mathcal{S}$ by computing the N-length characteristic function vector of $S_{l,j,h}$, denoted $\chi_{S_{l,j,h}}$. This leads to $\mathcal{M}$ being a $\tilde{O}(k^2)$ x N measurement matrix. Here we bound the number of rows in $\mathcal{M}$ by noting that: (i) $|\mathcal{S}| < m \cdot K \cdot p_m q_K$, (ii) $m = O(\log N)$, (iii) $p_m = O(\log N \cdot \log\log N)$, (iv) $K = O(k \log N)$, and (v) $q_K = O(K \log K)$.

4 Signal Reconstruction from Measurements

Let $\hat{\mathbf{A}}$ be an N-length signal of complex numbers with it's N entries numbered 0 through $N-1$. Our goal is to identify B of the largest magnitude entries of $\hat{\mathbf{A}}$ (i.e., the first B entries in a valid ordering of $\hat{\mathbf{A}}$ as in Equation 2.3) and then estimate their signal values. Toward this end, set

$$\epsilon = \frac{|\hat{\mathbf{A}}(\omega_B)|}{\sqrt{2}C} \tag{4.6}$$

where $C > 1$ is a constant to be specified later, and let B' be the smallest integer such that

$$\sum_{b=B'}^{N-1} |\hat{\mathbf{A}}(\omega_b)| < \frac{\epsilon}{2}. \tag{4.7}$$

Algorithm 1 SPARSE APPROXIMATE

Input: Signal $\hat{\mathbf{A}}$, integers B, B'
Output: $\hat{\mathbf{R}}^s$, a sparse representation for $\hat{\mathbf{A}}$
Initialize $\hat{\mathbf{R}}^s \leftarrow \emptyset$
Set $K = 3B'\lfloor \log_{B'} N \rfloor$
Form measurement matrix, $\mathcal{M}$, via K-majority B'-strongly selective collections (Section 3)
Compute $\mathcal{M} \cdot \hat{\mathbf{A}}$

IDENTIFICATION

for j from 0 to K **do**
  Sort $\langle \chi_{S_{0,j,0}}, \hat{\mathbf{A}} \rangle, \ldots, \langle \chi_{S_{0,j,q_j-1}}, \hat{\mathbf{A}} \rangle$ by magnitude
  for b from 0 to B' **do**
    $k_{j,b} \leftarrow b^{\text{th}}$ largest magnitude $\langle \chi_{S_{0,j,\cdot}}, \hat{\mathbf{A}} \rangle$
    $r_{0,b} \leftarrow k_{j,b}$'s associated residue mod q_j
    for l from 1 to m **do**
      $t_{\min} \leftarrow \min_{t \in [0,p_l)} |k_{j,b} - \langle \chi_{S_{l,j,t\cdot q_j + r_{0,b}}}, \hat{\mathbf{A}} \rangle|$
      $r_{l,b} \leftarrow r_{0,b} + t_{\min} \cdot q_j \bmod p_l$
    end for
    Construct $\omega_{j,b}$ from $r_{0,b}, \ldots, r_{m,b}$ via the CRT
  end for
end for
Sort $\omega_{j,b}$'s maintaining duplicates and set $C(\omega_{j,b}) =$ the number of times $\omega_{j,b}$ was constructed via line 16

ESTIMATION

for j from 0 to K **do**
  for b from 0 to B' **do**
    if $C(\omega_{j,b}) > \frac{2K}{3}$ **then**
      $C(\omega_{j,b}) \leftarrow 0$
      $x = \text{median}\{\text{real}(k_{j',b'}) | \omega_{j',b'} = \omega_{j,b}\}$
      $y = \text{median}\{\text{imag}(k_{j',b'}) | \omega_{j',b'} = \omega_{j,b}\}$
      $\hat{\mathbf{R}}^s \leftarrow \hat{\mathbf{R}}^s \cup \{(\omega_{j,b}, x + \mathbb{i}y)\}$
    end if
  end for
end for
Output B largest magnitude entries in $\hat{\mathbf{R}}^s$

Note that B' identifies the most energetic *insignificant* frequency (i.e., with energy $<$ a fraction of $|\hat{\mathbf{A}}(\omega_B)|$). We expect to work with sparse/compressible signals so that $B \leq B' \ll N$. Later we will give specific values for C and B' depending on B, the desired approximation error, and $\hat{\mathbf{A}}$'s compressibility characteristics. For now we show that we can identify/approximate B of $\hat{\mathbf{A}}$'s largest magnitude entries each to within ϵ-precision via Algorithm 1.

Algorithm 1 works by using $\mathcal{S}_0$ measurements to separate $\hat{\mathbf{A}}$'s significantly energetic frequencies $\Omega = \{\omega_0, \ldots, \omega_{B'-1}\} \subset [0, N)$. Every measurement which successfully separates an energetic frequency ω_j from all other members of Ω will both (i) provide a good (i.e.,

within $\frac{\epsilon}{2} \leq \frac{|\hat{\mathbf{A}}(\omega_B)|}{2\sqrt{2}}$) coefficient estimate for ω_j, and (ii) yield information about ω_j's identity. Frequency separation occurs because our $\mathcal{S}_0$ measurements can not collide any fixed $\omega_j \in \Omega$ with any other member of Ω modulo more than $(B'-1)\log_{B'} N$ q-primes (see Lemma 3.1). Therefore, more than $\frac{2}{3}^{\text{rds}}$ of $\mathcal{S}_0$'s $3B'\log_{B'} N + 1$ q-primes will isolate any fixed $\omega_j \in \Omega$. This means that our reconstruction algorithm will identify all frequencies at least as energetic as ω_B at least $2B'\log_{B'} N + 1$ times. We can ignore any frequencies that are not recovered this often. On the other hand, for any frequency that is identified more than $2B'\log_{B'} N$ times, at most $B'\log_{B'} N$ of the measurements which lead to this identification can be significantly contaminated via collisions with valid Ω members. Therefore, we can take a median of the more than $2B'\log_{B'} N$ measurements leading to the recovery of each frequency as that frequency's coefficient estimate. Since more than half of these measurements must be accurate, the median will be accurate. The following Theorem is proved in the appendix.

THEOREM 2. *Let $\hat{\boldsymbol{R}}_{\text{opt}}$ be a B-optimal Fourier representation for our input signal $\boldsymbol{A}$. Then, the B term representation, $\hat{\boldsymbol{R}}^{\text{s}}$, returned from Algorithm 1 is such that $\|\boldsymbol{A}-\boldsymbol{R}\|_2^2 \leq \|\boldsymbol{A}-\boldsymbol{R}_{\text{opt}}\|_2^2 + \frac{6B\cdot|\hat{\boldsymbol{A}}(\omega_B)|^2}{C}$. Furthermore, Algorithm 1's Identification and Estimation (lines 7 - 30) run time is $O(B'^2\log^4 N)$. The number of measurements used is $O(B'^2\log^6 N)$.*

Theorem 2 immediately indicates that Algorithm 1 gives us a deterministic $O(m^2\log^6 N)$-measurement, $O(m^2\log^4 N)$-reconstruction time method for exactly recovering vectors with m non-zero entries. If $\hat{\mathbf{A}}$ has exactly m non-zero entries then setting $B' = B = m$ and $C = 1$ will be sufficient to guarantee that both $|\hat{\mathbf{A}}(\omega_B)|^2 = 0$ and $\sum_{b=B'}^{N-1}|\hat{\mathbf{A}}(\omega_b)| = 0$ are true. Hence, we may apply Theorem 2 with $B' = B = m$ and $C = 1$ to obtain a perfect reconstruction via Algorithm 1. However, we are mainly interested in the more realistic cases where $\hat{\mathbf{A}}$ is either algebraically or exponentially compressible. The following theorem presents itself.

THEOREM 3. *Let $\hat{\boldsymbol{A}}$ be p-compressible. Then, Algorithm 1 can return a B term sparse representation, $\hat{\boldsymbol{R}}^{\text{s}}$, with $\|\boldsymbol{A}-\boldsymbol{R}\|_2^2 \leq \|\boldsymbol{A}-\boldsymbol{R}_{\text{opt}}\|_2^2 + \delta\|C_B^{\text{opt}}\|_2^2$ using $O\left(B^{\frac{2p}{p-1}}\delta^{\frac{2}{1-p}}\log^4 N\right)$ total identification/estimation time and $O\left(B^{\frac{2p}{p-1}}\delta^{\frac{2}{1-p}}\log^6 N\right)$ measurements. If $\hat{\boldsymbol{A}}$ decays exponentially, Algorithm 1 can return a B term sparse representation, $\hat{\boldsymbol{R}}^{\text{s}}$, with $\|\boldsymbol{A}-\boldsymbol{R}\|_2^2 \leq \|\boldsymbol{A}-\boldsymbol{R}_{\text{opt}}\|_2^2 + \delta\|C_B^{\text{opt}}\|_2^2$ using both $\left(B^2+\log^2\delta^{\frac{-1}{\alpha}}\right)\cdot$ polylog(N) measurements and identification/estimation time.*

Algorithm 2 FOURIER MEASURE

1: **Input:** f**-samples, integers** m, K
2: **Output:** $<\chi_{S_{l,j,h}}, \hat{f}>$**-measurements**
3: Zero a $O(q_K p_m)$-element array, $\mathbf{A}$
4: **for** j from 1 to K **do**
5: **for** l from 1 to m **do**
6: $\mathbf{A} \leftarrow f(0), f\left(\frac{2\pi}{q_j p_l}\right), \ldots, f\left(\frac{2\pi(q_j p_l - 1)}{q_j p_l}\right)$
7: Calculate $\hat{\mathbf{A}}$ via Chirp z-Transform [27, 2]
8: $<\chi_{S_{l,j,h}}, \hat{f}> \leftarrow \hat{\mathbf{A}}(h)$ for each $h \in [0, q_j p_l)$
9: **end for**
10: **end for**
11: Output $<\chi_{S_{l,j,h}}, \hat{f}>$-measurements

For p-compressible signals, $p > 2$, CM's algorithm [6, 7] takes $O\left(B^{\frac{6p}{p-2}}\delta^{\frac{6}{2-p}}\log^6 N\right)$- identification/estimation time and $O\left(B^{\frac{4p}{p-2}}\delta^{\frac{4}{2-p}}\log^4 N\right)$-measurements to achieve the same error bound. As a concrete comparison, CM's algorithm requires $O(B^{18}\delta^{-6}\log^6 N)$- identification/estimation time and $O(B^{12}\delta^{-4}\log^4 N)$-measurements for 3-compressible signals. Algorithm 1, on the other hand, requires only $O(B^3\delta^{-1}\log^4 N)$- identification/estimation time and $O(B^3\delta^{-1}\log^6 N)$-measurements. Hence, we have improved on CM's algebraic compressibility results. All that's left to do in order to develop a deterministic sub-linear time Fourier algorithm is to compute our CS Fourier measurements (Algorithm 1 lines 1 - 6) in sub-linear time.

5 Fast Fourier Measurement Acquisition

Our goal in this section is to demonstrate how to use Algorithm 1 as means to approximate the Fourier transform of a signal/function $f : [0, 2\pi] \to \mathbb{C}$, where (i) f has an integrable p^{th} derivative, and (ii) $f(0) = f(2\pi), f'(0) = f'(2\pi), \ldots, f^{(p-2)}(0) = f^{(p-2)}(2\pi)$. In this case we know the Fourier coefficients for f to be p-compressible [3, 12]. Hence, for $N = q_1 \cdot p_1 \cdots p_m$ sufficiently large, if we can collect the necessary Algorithm 1 (lines 5 and 6) measurements in sub-linear time we will indeed be able to use Algorithm 1 as a sub-linear time Fourier algorithm for f.

Note that in order to validate the use of Algorithm 1 (or any other sparse approximate Fourier Transform method [15, 16]) we must assume that f exhibits some multiscale behavior. If $\hat{f}$ contains no unpredictably energetic large (relative to the number of desired Fourier coefficients) frequencies then it is more computationally

efficient to simply use standard FFT/USFFT methods [5, 22, 1, 10, 11]. The responsible user, therefore, is not entirely released from the obligation to consider $\hat{f}$'s likely characteristics before proceeding with computations.

Choose any Section 3 q-prime q_j, $j \in [1, K]$, and any p-prime p_l with $l \in [0, m]$. Furthermore, pick $h \in [0, q_j p_l)$. Throughout the rest of this discussion we will consider f to be accessible to sampling at any desired predetermined positions $t \in [0, 2\pi]$. Given this assumption, we may sample f at $t = 0, \frac{2\pi}{q_j p_l}, \ldots, \frac{2\pi(q_j p_l - 1)}{q_j p_l}$ in order to perform the following DFT computation:

$$< \chi_{S_{l,j,h}}, \hat{f} >= \frac{1}{q_j p_l} \sum_{k=0}^{q_j p_l - 1} f\left(\frac{2\pi k}{q_j p_l}\right) e^{\frac{-2\pi \mathrm{i} h k}{q_j p_l}}.$$

Using the Fourier expansion for f yields

$$< \chi_{S_{l,j,h}}, \hat{f} >= \frac{1}{q_j p_l} \sum_{k=0}^{q_j p_l - 1} \left(\sum_{\omega=-\infty}^{\infty} \hat{f}(\omega) e^{\frac{2\pi \mathrm{i} \omega k}{q_j p_l}} \right) e^{\frac{-2\pi \mathrm{i} h k}{q_j p_l}}.$$

Finally, exchanging the order of summation above, we see that $< \chi_{S_{l,j,h}}, \hat{f} >$ reduces to

$$\frac{1}{q_j p_l} \sum_{\omega=-\infty}^{\infty} \hat{f}(\omega) \sum_{k=0}^{q_j p_l - 1} e^{\frac{2\pi \mathrm{i}(\omega - h)k}{q_j p_l}} = \sum_{\omega \equiv h \bmod q_j p_l} \hat{f}(\omega)$$

via aliasing [3].

Using Sections 3 and 4 we can see that these measurements are exactly what we need in order to determine B of the most energetic frequencies of $\hat{f}$ modulo $N = q_1 \cdot p_1 \cdots p_m$ (i.e., B of the most energetic frequencies of f's band-limited interpolant's DFT). We are now in the position to modify Algorithm 1 in order to find a sparse Fourier representation for $\hat{f}$. To do so we proceed as follows: First, remove lines 5 and 6 and replace them with Algorithm 2 for computing all the necessary $< \chi_{S_{l,j,h}}, \hat{f} >$-measurements. Second, replace each "$< \chi_{S_{l,j,h}}, \hat{\mathbf{A}} >$" by "$< \chi_{S_{l,j,h}}, \hat{f} >$" in Algorithm 1's IDENTIFICATION section. It remains to show that these Algorithm 1 modifications indeed yield a sub-linear time approximate Fourier transform. The following theorem presents itself:

THEOREM 4. *Let $f : [0, 2\pi] \to \mathbb{C}$ have (i) an integrable p^{th} derivative, and (ii) $f(0) = f(2\pi), f'(0) = f'(2\pi), \ldots, f^{(p-2)}(0) = f^{(p-2)}(2\pi)$ for some $p > 1$. Furthermore, assume that $\hat{f}$'s $B' = O\left(B^{\frac{p}{p-1}} \delta^{\frac{1}{1-p}}\right)$ largest magnitude frequencies all belong to $\left(-\left\lceil \frac{N}{2} \right\rceil, \left\lfloor \frac{N}{2} \right\rfloor\right]$. Then, we may use Algorithm 1 to return a B term sparse Fourier representation, $\hat{\boldsymbol{R}}^{\text{s}}$, for $\hat{f}$ such that $\|\hat{f} - \hat{\boldsymbol{R}}\|_2^2 \leq \|\hat{f} - \hat{\boldsymbol{R}}_{\text{opt}}\|_2^2 + \delta \|C_B^{\text{opt}}\|_2^2$ using $O\left(B^{\frac{2p}{p-1}} \delta^{\frac{2}{1-p}} \log^7 N\right)$-time and $O\left(B^{\frac{2p}{p-1}} \delta^{\frac{2}{1-p}} \log^6 N\right)$-measurements from f.*

If $f : [0, 2\pi] \to \mathbb{C}$ is smooth (i.e., has infinitely many continuous derivatives on the unit circle where 0 is identified with 2π) it follows from Theorem 4 that Algorithm 1 can be used to find an δ-accurate, with $\delta = O\left(\frac{1}{N}\right)$, sparse B-term Fourier representation for $\hat{f}$ using $\tilde{O}(B^2)$-time/measurements. This result differs from previous sub-linear time Fourier algorithms [15, 16] in that both the algorithm and the measurements/samples it requires are deterministic. Recall that the deterministic nature of the algorithm's required samples is potentially beneficial for failure intolerant hardware. In signal processing applications the sub-Nyquist sampling required to compute Algorithm 1's $< \chi_{S_{l,j,h}}, \hat{f} >$-measurements could be accomplished via $\tilde{O}(B)$ parallel low-rate analog-to-digital converters.

5.1 DFT from Inaccessible Signal Samples

Throughout the remainder of this section we will consider our N-length compressible vector $\hat{\mathbf{A}}$ to be the product of the N x N DFT matrix, Ψ, and a non-sparse N-length vector $\mathbf{A}$. Thus,

$$\hat{\mathbf{A}} = \Psi \mathbf{A}.$$

Furthermore, we will assume that $\mathbf{A}$ contains equally spaced samples from some unknown smooth function $f : [0, 2\pi] \to \mathbb{C}$ (i.e., $\mathbf{A}$'s band-limited interpolent). Hence,

$$\mathbf{A}(j) = f\left(\frac{2\pi j}{N}\right), \; j \in [0, N).$$

We would like to use our modified Algorithm 1 along with Algorithm 2 to find a sparse Fourier representation for $\hat{\mathbf{A}}$. However, unless $\frac{N}{q_j p_l} \in \mathbb{N}$ for all $q_j p_l$-pairs (which would imply f had been grossly oversampled), $\mathbf{A}$ won't contain all the f-samples required by Algorithm 2. Not having access to f directly, and restricting ourselves to sub-linear time approaches only, we have little recourse but to locally interpolate f around our required samples.

For each required Algorithm 2 f-sample at $t = \frac{2\pi h}{q_j p_l}, h \in [0, q_j p_l)$, we may approximate $f(t)$ to within $O(N^{-2\kappa})$-error by constructing 2 local interpolents (one real, one imaginary) around t using $\mathbf{A}$'s nearest 2κ entries [14]. These errors in f-samples can lead to errors of size $O(N^{-2\kappa} \cdot p_m q_K \log p_m q_K)$ in our $< \chi_{S_{l,j,h}}, \hat{f} >$ calculations. However, as long as the $< \chi_{S_{l,j,h}}, \hat{f} >$-measurement errors are small enough (e.g., $O(\delta \cdot B^{-p})$ in the p-compressible case) Theorem 4 and all related Section 4 results and will still hold. After some scratch

work we can see that using $2\kappa = O(\log \delta^{-1} + p)$ interpolation points per f-sample ensures all our $< \chi_{S_{l,j,h}}, \hat{f} >$-measurement errors are $O(\delta \cdot B^{-p})$. We have the following result:

THEOREM 5. *Let $\hat{\boldsymbol{A}} = \Psi \boldsymbol{A}$ be p-compressible. Then, we may use Algorithms 1 and 2 to return a B term sparse representation, $\hat{\boldsymbol{R}}^{\text{s}}$, for $\hat{\boldsymbol{A}}$ such that $\|\boldsymbol{A} - \boldsymbol{R}\|_2^2 \leq \|\boldsymbol{A} - \boldsymbol{R}_{\text{opt}}\|_2^2 + \delta \|C_B^{\text{opt}}\|_2^2$ using $\tilde{O}\left(B^{\frac{2p}{p-1}} \delta^{\frac{2}{1-p}} (\log \delta^{-1} + p)^2\right)$-time and $\tilde{O}\left(B^{\frac{2p}{p-1}} \delta^{\frac{2}{1-p}} (\log \delta^{-1} + p)\right)$-samples from $\boldsymbol{A}$.*

Notice that Theorem 5 no longer guarantees an $\delta = O(\frac{1}{N})$-accurate $\tilde{O}(B^2)$-time DFT algorithm for smooth data (i.e., $\mathbf{A}$'s containing samples from a smooth function f). This is because as $p \to \infty$ we require an increasingly large number of interpolation points per f-sample in order to guarantee our $< \chi_{S_{l,j,h}}, \hat{f} >$-measurements remain $O(\delta \cdot B^{-p})$-accurate. However, for $\delta = O(\log^{-1} N)$, we can still consider smooth data $\mathbf{A}$ to be $O(\log N)$-compressible and so achieve a $\tilde{O}(B^2)$-time DFT algorithm.

6 Conclusion

Compressed Sensing (CS) methods provide algorithms for approximating the result of any large matrix multiplication as long as it is known in advance that the result will be sparse/compressible. Hence, CS is potentially valuable for many numerical applications such as those involving multiscale aspects [8, 18]. In this paper we used CS methods to develop the first known deterministic sub-linear time sparse Fourier transform algorithm. In the process, we introduced a new deterministic Compressed Sensing algorithm along the lines of Cormode and Muthukrishnan (CM) [6, 7]. Our new deterministic CS algorithm improves on CM's algebraic compressibility results while simultaneously maintaining their results concerning exponential compressibility.

Compressed Sensing is closely related to hashing methods, combinatorial group testing, and many other algorithmic problems [25, 13]. Thus, K-majority k-strongly selective collections of sets and Algorithm 1 may help improve deterministic results concerning stream hashing/heavy-hitter identification. Further development of these/other algorithmic applications is left as future work.

It is also worthwhile to note that Monte Carlo Fourier results similar to those of [16] may be obtained by altering our measurement construction in Section 3. If we construct our $\mathcal{S}_l$ collections by using only a small subset of $O(\log B')$ randomly chosen q_j's, we will still locate all sufficiently energetic entries of $\hat{\mathbf{A}}$ with high probability. The discovered entries' coefficients can then be approximated by using either (i) standard USFFT techniques [16, 10, 11, 22], or (ii) another $O(\log B')$ randomly chosen q_j-measurement groups. In either case, the end result will be a $O(B' \cdot \text{polylog}(N))$-time/measurement Fourier algorithm that produces the same results (e.g., Theorem 4) as above with high probability.

7 Acknowledgments

We would like to thank Graham Cormode and S. Muthukrishnan for answering questions about their work. We would also like to thank Martin Strauss, Anna Gilbert, Joel Lepak, and Hualong Feng for helpful discussions, advice, and comments.

References

[1] C. Anderson and M. D. Dahleh, *Rapid computation of the discrete Fourier transform*, SIAM J. Sci. Comput., 17:913–919, 1996.

[2] L. I. Bluestein. *A Linear Filtering Approach to the Computation of Discrete Fourier Transform*, IEEE Transactions on Audio and Electroacoustics, 18:451–455, 1970.

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

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

[5] J. Cooley and J. Tukey, *An algorithm for the machine calculation of complex Fourier series*, Math. Comput., 19:297–301, 1965.

[6] G. Cormode and S. Muthukrishnan, *Combinatorial Algorithms for Compressed Sensing*, Technical Report DIMACS TR 2005-40, 2005.

[7] G. Cormode and S. Muthukrishnan, *Combinatorial Algorithms for Compressed Sensing*, Conference on Information Sciences and Systems, March 2006.

[8] I. Daubechies, O. Runborg, and J. Zou, *A sparse spectral method for homogenization multiscale problems*, Multiscale Model. Sim., 2007.

[9] R. A. DeVore, *Deterministic constructions of compressed sensing matrices*, Preprint at *http://www.ima.umn.edu/2006-2007/ND6.4-15.07/activities/DeVore-Ronald/Henrykfinal.pdf*, 2007.

[10] A. Dutt and V. Rokhlin, *Fast Fourier transforms for nonequispaced data*, SIAM J. Sci. Comput., 14:1368–1383, 1993.

[11] J. A. Fessler and B. P. Sutton, *Nonuniform Fast fourier transforms using min-max interpolation*, IEEE Trans. Signal Proc., 51:560–574, 2003.

[12] G. B. Folland, *Fourier Analysis and Its Applications*, Brooks/Cole Publishing Company, 1992.

[13] S. Ganguly and A. Majumder, *CR-precis: A deterministic summary structure for update data streams*, ArXiv Computer Science e-prints, Sept. 2006.

[14] C. F. Gerald and P. O. Wheatley, *Applied Numerical Analysis*, Addison-Wesley Publishing Company, 1994.

[15] A. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, and M. Strauss, *Near-optimal sparse Fourier estimation via sampling*, ACM STOC, pages 152–161, 2002.

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

[17] P. Indyk. *Explicit constructions of selectors and related combinatorial structures, with applications*, SODA '02, pages 697–704, Philadelphia, PA, USA, 2002.

[18] M. A. Iwen, Unpublished Results, http://www-personal.umich.edu/ markiwen/.

[19] M. A. Iwen, A. C. Gilbert, and M. J. Strauss, *Empirical evaluation of a sub-linear time sparse DFT algorithm*, Submitted for Publication, 2007.

[20] S. Kirolos, J. Laska, M. Wakin, M. Duarte, D. Baron, T. Ragheb, Y. Massoud, and R. Baraniuk, *Analog-to-information conversion via random demodulation. in Proc. IEEE Dallas Circuits and Systems Conference*, 2006.

[21] J. Laska, S. Kirolos, Y. Massoud, R. Baraniuk, A. Gilbert, M. Iwen, and M. Strauss, *Random sampling for analog-to-information conversion of wideband signals*, Proc. IEEE Dallas Circuits and Systems Conference, 2006.

[22] J.-Y. Lee and L. Greengard, *The type 3 nonuniform FFT and its applications*, J Comput. Phys., 206:1–5, 2005.

[23] M. Lustig, D. Donoho, and J. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Submitted for publication, 2007.

[24] R. Maleh, A. C. Gilbert, and M. J. Strauss, *Signal recovery from partial information via orthogonal matching pursuit*, IEEE Int. Conf. on Image Processing, 2007.

[25] S. Muthukrishnan, *Data Streams: Algorithms and Applications*, Foundations and Trends in Theoretical Computer Science, 1, 2005.

[26] S. Muthukrishnan, *Some Algorithmic Problems and Results in Compressed Sensing*, Allerton Conference, 2006.

[27] L. Rabiner, R. Schafer, and C. Rader, *The Chirp z-Transform Algorithm*, IEEE Transactions on Audio and Electroacoustics, AU-17(2):86–92, June 1969.

[28] J. Tropp and A. Gilbert, *Signal recovery from partial information via orthogonal matching pursuit*, Submitted for Publication, 2005.

A Proof of Theorem 2

We begin by proving two lemmas.

LEMMA A.1. *IDENTIFICATION: Lines 7 through 19 of Algorithm 1 are guaranteed to recover all valid $\omega_0, \ldots, \omega_{B-1}$ (i.e., all ω with $|\hat{\boldsymbol{A}}(\omega)| \geq |\hat{\boldsymbol{A}}(\omega_B)|$ - there may be $> B$ such entries) more then $\frac{2K}{3}$ times. Hence, despite line 22, an entry for all such $\omega_b, 0 \leq b < B$, will be added to $\hat{\boldsymbol{R}}^s$ in line 26.*

Proof. Because of the construction of $\mathcal{S}_0$ (i.e., proof of Lemma 3.1) we know that for each $b \in [0, B)$ there exist more then $\frac{2K}{3}$ subsets $S \in \mathcal{S}_0$ such that $S \cap \{\omega_{b'} \mid b' \in [0, B')\} = \{\omega_b\}$. Choose any $b \in [0, B)$. Denote the q-primes that isolate ω_b from all of $\omega_0, \ldots, \omega_{b-1}, \omega_{b+1}, \ldots, \omega_{B'-1}$ by

$$q_{j_1}, q_{j_2}, \ldots, q_{j_{K'}}, \quad \frac{2K}{3} < K' \leq K.$$

We next show that, for each $k' \in [1, K']$, we get $< \chi_{S_{0,j_{k'},\omega_b} \bmod q_{j_{k'}}}, \mathbf{A} >$ as one of the $B'+1$ largest magnitude $< \chi_{S_{0,j_{k'},\cdot}}, \hat{\mathbf{A}} >$-measurements identified in line 10.

Choose any $k' \in [1, K']$. We know that

$$\begin{aligned} \frac{\epsilon}{2} \;<\; \frac{\epsilon}{\sqrt{2}} \;&<\; |\hat{\mathbf{A}}(\omega_B)| - \sqrt{2} \sum_{b'=B'}^{N-1} |\hat{\mathbf{A}}(\omega_{b'})| \\ &\leq\; |\hat{\mathbf{A}}(\omega_b)| - \left| \sum_{b' \in [B', N),\ \omega_{b'} \equiv \omega_b} \hat{\mathbf{A}}(\omega_{b'}) \right| \\ &\leq\; \left| < \chi_{S_{0,j_{k'},\omega_b} \bmod q_{j_{k'}}}, \hat{\mathbf{A}} > \right|. \end{aligned}$$

We also know that the $(B'+1)^{\text{st}}$ largest measurement L^2-magnitude must be $< \frac{\epsilon}{2}$. Hence, we are guaranteed to execute lines 12-15 with an $r_{0,\cdot} = \omega_b \bmod q_{j_{k'}}$.

Choose any $l \in [1, m]$ and set $\tilde{\Omega}$ to be

$$\left\{ \omega_{b'} \middle| b' \in [B', N), \omega_{b'} \equiv \omega_b \mathrm{mod} q_{j_{k'}}, \omega_{b'} \neq \omega_b \mathrm{mod} q_{j_{k'}} p_l \right\}.$$

Using Lemma 3.2 we can see that line 13 inspects all the necessary residues of $\omega_b \bmod p_l q_{j_{k'}}$. To see that t_{min} will be chosen correctly we note first that

$$\begin{aligned} &\left| < \chi_{S_{0,j_{k'},\omega_b} \mathrm{mod}_{q_{j_{k'}}}}, \hat{\mathbf{A}} > - < \chi_{S_{0,j_{k'},\omega_b} \mathrm{mod}_{p_l q_{j_{k'}}}}, \hat{\mathbf{A}} > \right| \\ &\quad = \left| \sum_{\omega_{b'} \in \tilde{\Omega}} \hat{\mathbf{A}}(\omega_{b'}) \right| \;\leq\; \sqrt{2} \sum_{\omega_{b'} \in \tilde{\Omega}} |\hat{\mathbf{A}}(\omega_{b'})|. \end{aligned}$$

Furthermore, setting $r_{0,\cdot} = \omega_b \bmod q_{j_{k'}}$ and Ω' to be

$$\left\{ \omega_{b'} \;\middle|\; b' \in [B', N),\ \omega_{b'} \equiv \omega_b \bmod q_{j_{k'}}, \omega_{b'} \neq (r_{0,\cdot} + t q_{j_{k'}}) \right.$$
$$\left. \bmod q_{j_{k'}} p_l \text{ for } t \text{ with } (r_{0,\cdot} + t q_{j_{k'}}) \neq \omega_b \bmod q_{j_{k'}} p_l \right\},$$

we have

$$\sqrt{2} \sum_{\omega_{b'} \in \tilde{\Omega}} |\hat{\mathbf{A}}(\omega_{b'})| < \frac{\epsilon}{\sqrt{2}} \;<\; |\hat{\mathbf{A}}(\omega_B)| - \sqrt{2} \sum_{b'=B'}^{N-1} |\hat{\mathbf{A}}(\omega_{b'})|$$

$$\leq |\hat{\mathbf{A}}(\omega_b)| - \left|\sum_{\omega_{b'}\in\Omega'} \hat{\mathbf{A}}(\omega_{b'})\right|.$$

Finally we can see that

$$|\hat{\mathbf{A}}(\omega_b)| - \left|\sum_{\omega_{b'}\in\Omega'} \hat{\mathbf{A}}(\omega_{b'})\right| \leq \Big| < \chi_{S_{0,j_{k'},\omega_b \bmod q_{j_{k'}}}}, \hat{\mathbf{A}} > - < \chi_{S_{0,j_{k'},(r_{0,\cdot}+tq_{j_{k'}})\neq\omega_b \bmod p_l q_{j_{k'}}}}, \hat{\mathbf{A}} > \Big|.$$

Hence, lines 13 and 14 will indeed select the correct residue for ω_b modulo p_l. Therefore, line 16 will correctly reconstruct ω_b at least $K' > \frac{2K}{3}$ times.

LEMMA A.2. *ESTIMATION: Every* $(\omega, \tilde{\hat{\mathrm{A}}}_\omega)$ *stored in* $\hat{\boldsymbol{R}}^{\mathrm{s}}$ *in line 26 is such that* $|\hat{\boldsymbol{A}}(\omega) - \tilde{\hat{\mathrm{A}}}_\omega|_2 < \epsilon$.

Proof. Suppose that $(\omega, \tilde{\hat{\mathrm{A}}}_\omega)$ is stored in $\hat{\mathbf{R}}^{\mathrm{s}}$. This only happens if $\hat{\mathbf{A}}(\omega)$ has been estimated by

$$< \chi_{S_{0,j,\omega} \bmod q_j}, \hat{\mathbf{A}} > \; = \sum_{\tilde{\omega}\equiv\omega \bmod q_j} \hat{\mathbf{A}}(\tilde{\omega})$$

for more then $\frac{2K}{3}$ q_j-primes. The only way that any such estimate can have $|\hat{\mathbf{A}}(\omega) - < \chi_{S_{0,j,\omega} \bmod q_j}, \hat{\mathbf{A}} > | \geq \frac{\epsilon}{2}$ is if ω collides with one of $\omega_0, \ldots, \omega_{B'-1}$ modulo q_j (this is due to the definition of B' in Equation 4.7). By the proof of Lemma 3.1 we know this can happen at most $B'\lfloor \log_{B'} N \rfloor < \frac{K}{3}$ times. Hence, more then half of the $\frac{2K}{3}$ estimates, $\tilde{\hat{\mathrm{A}}}'_\omega$, must be such that $|\hat{\mathbf{A}}(\omega) - \tilde{\hat{\mathrm{A}}}'_\omega| < \frac{\epsilon}{2}$. It follows that taking medians as per lines 24 and 25 will result in the desired ϵ-accurate estimate for $\hat{\mathbf{A}}(\omega)$.

We are now ready to prove Theorem 2.

Theorem 2 *Let* $\hat{\boldsymbol{R}}_{\mathrm{opt}}$ *be a* B*-optimal Fourier representation for our input signal* $\hat{\boldsymbol{A}}$*. Then, the* B *term representation* $\hat{\boldsymbol{R}}^{\mathrm{s}}$ *returned from Algorithm 1 is such that* $\|\boldsymbol{A} - \boldsymbol{R}\|_2^2 \leq \|\boldsymbol{A} - \boldsymbol{R}_{\mathrm{opt}}\|_2^2 + \frac{6B\cdot|\hat{\boldsymbol{A}}(\omega_B)|^2}{C}$*. Furthermore, Algorithm 1's Identification and Estimation (lines 7 - 30) run time is* $O(B'^2 \log^4 N)$*. The number of measurements used is* $O(B'^2 \log^6 N)$.

Proof. Choose any $b \in [0, B)$. Using Lemmas A.1 and A.2 we can see that only way some $\omega_b \notin \hat{\mathbf{R}}^{\mathrm{s}}_B$ is if there exists some associated $b' \in [B, N)$ so that $\omega_{b'} \in \hat{\mathbf{R}}^{\mathrm{s}}$ and

$$|\hat{\mathbf{A}}(\omega_B)| + \epsilon \geq |\hat{\mathbf{A}}(\omega_{b'})| + \epsilon > |\tilde{\hat{\mathrm{A}}}_{\omega_{b'}}| \geq |\tilde{\hat{\mathrm{A}}}_{\omega_b}| > |\hat{\mathbf{A}}(\omega_b)| - \epsilon \geq |\hat{\mathbf{A}}(\omega_B)| - \epsilon.$$

In this case we get $2\epsilon > |\hat{\mathbf{A}}(\omega_b)| - |\hat{\mathbf{A}}(\omega_{b'})| \geq 0$ so that

$$|\hat{\mathbf{A}}(\omega_{b'})|^2 + 4\epsilon\left(\epsilon + |\hat{\mathbf{A}}(\omega_B)|\right) \geq |\hat{\mathbf{A}}(\omega_{b'})|^2 + 4\epsilon\left(\epsilon + |\hat{\mathbf{A}}(\omega_{b'})|\right) > |\hat{\mathbf{A}}(\omega_b)|^2. \tag{1.8}$$

Now using Lemma A.2 we can see that

$$\|\hat{\mathbf{A}} - \hat{\mathbf{R}}\|^2 = \sum_{(\omega,\cdot)\notin\hat{\mathbf{R}}^{\mathrm{s}}} |\hat{\mathbf{A}}(\omega)|^2 + \sum_{(\omega,\tilde{\hat{\mathrm{A}}}_\omega)\in\hat{\mathbf{R}}^{\mathrm{s}}} |\hat{\mathbf{A}}(\omega) - \tilde{\hat{\mathrm{A}}}_\omega|^2 < \sum_{(\omega,\cdot)\notin\hat{\mathbf{R}}^{\mathrm{s}}} |\hat{\mathbf{A}}(\omega)|^2 + B\cdot\epsilon^2.$$

Furthermore, we have

$$B\cdot\epsilon^2 + \sum_{(\omega,\cdot)\notin\hat{\mathbf{R}}^{\mathrm{s}}} |\hat{\mathbf{A}}(\omega)|^2 = B\cdot\epsilon^2 + \sum_{b\in[0,B),\ \omega_b\notin\hat{\mathbf{R}}^{\mathrm{s}}} |\hat{\mathbf{A}}(\omega_b)|^2 + \sum_{b'\in[B,N),\ \omega_{b'}\notin\hat{\mathbf{R}}^{\mathrm{s}}} |\hat{\mathbf{A}}(\omega_{b'})|^2.$$

Using observation 1.8 above we can see that this last expression is bounded above by

$$B\cdot(5\epsilon^2 + 4\epsilon|\hat{\mathbf{A}}(\omega_B)|) + \sum_{b'\in[B,N),\ \omega_{b'}\in\hat{\mathbf{R}}^{\mathrm{s}}} |\hat{\mathbf{A}}(\omega_{b'})|^2 + \sum_{b'\in[B,N),\ \omega_{b'}\notin\hat{\mathbf{R}}^{\mathrm{s}}} |\hat{\mathbf{A}}(\omega_{b'})|^2 \leq \|\hat{\mathbf{A}} - \hat{\mathbf{R}}_{\mathrm{opt}}\|_2^2 + B\cdot(5\epsilon^2 + 4\epsilon|\hat{\mathbf{A}}(\omega_B)|).$$

Substituting for ϵ (see Equation 4.6) gives us our result.

We next focus on run time. Algorithm 1's Identification (i.e., lines 7 through 19) run time is dominated by the $O(KB'm)$ executions of line 13. And, each execution of line 13 takes time $O(p_m)$. Hence, given that $m = O(\log N)$, $p_m = O(\log N \cdot \log\log N)$, and $K = O(B' \log_{B'} N)$, we can see that Identification requires $O(B'^2 \cdot \log_{B'} N \cdot \log^2 N \cdot \log\log N)$-time.

Continuing, Algorithm 1's Estimation (i.e., lines 20 through 30) run time is ultimately determined by line 22's IF-statement. Although line 22 is executed $O(KB') = O(B'^2 \log_{B'} N)$ times, it can only evaluate to true $O(B')$ times. Hence, each line 24/25 $O(B' \log_{B'} N \log B')$-time median operation will be evaluated at most $O(B')$ times. The resulting Estimation runtime is therefore $O(B'^2 \log_{B'} N \log B')$.

To bound the number of measurements we recall that: (i) the number of measurements is $< m \cdot K \cdot p_m q_K$, (ii) $m = O(\log N)$, (iii) $p_m = O(\log N \cdot \log\log N)$, (iv) $K = O(B' \log N)$, and (v) $q_K = O(K \log K)$. Hence, the number of measurements is $O\left(K^2 \log K \log^2 N \log\log N\right)$. Substituting for K gives us the desired bound.