Near-Optimal Sparse Fourier Representations via Sampling

[Extended Abstract]

A. C. Gilbert* S. Guha† P. Indyk‡ S. Muthukrishnan* M. Strauss*

ABSTRACT

We give an algorithm for finding a Fourier representation $\mathbf{R}$ of B terms for a given discrete signal $\mathbf{A}$ of length N, such that $\|\mathbf{A}-\mathbf{R}\|_2^2$ is within the factor $(1+\epsilon)$ of best possible $\|\mathbf{A}-\mathbf{R}_{\rm opt}\|_2^2$. Our algorithm can access $\mathbf{A}$ by reading its values on a sample set $T \subseteq [0,N)$, chosen randomly from a (non-product) distribution of our choice, independent of $\mathbf{A}$. That is, we sample non-adaptively. The total time cost of the algorithm is polynomial in $B\log(N)\log(M)/\epsilon$ (where M is the ratio of largest to smallest numerical quantity encountered), which implies a similar bound for the number of samples.

1. INTRODUCTION

A Discrete Fourier Transform decomposes a signal into its trigonometric components, which vibrate at various frequencies. Formally, let $\mathbf{A} = (\mathbf{A}(0), \ldots, \mathbf{A}(N-1))$ be a discrete signal of length N. We let $\psi_\omega(t)$ denote the ω'th Fourier basis function $\frac{1}{\sqrt{N}}e^{2\pi i\omega t/N}$. Then the ω'th Fourier coefficient $\widehat{\mathbf{A}}(\omega)$ of $\mathbf{A}$ is the inner product $\langle \mathbf{A}, \psi_\omega\rangle$ of $\mathbf{A}$ with ψ_ω, $\widehat{\mathbf{A}}(\omega) = \frac{1}{\sqrt{N}}\sum_t \mathbf{A}(t)e^{-2\pi i\omega t/N}$. Fourier analysis, which consists of studying the Fourier coefficients of functions, is a prolific area of mathematical research by itself, with applications to lossy compression, signal and speech processing, fast computation of convolutions, solving partial differential equations, and others too numerous to list. The Fast Fourier Transform that computes *all* the N Fourier coefficients in $O(N\log N)$ operations is a cornerstone: it is used for computing convolutions and elsewhere in computational algebra. Often, however, Fourier analysis involves filtering out noise (the Fourier coefficients corresponding to unwanted frequencies can be eliminated) or detecting an underlying signal (Fourier coefficients that are smaller—in absolute value—than some specified tolerance can be discarded). This filtering is done because the large coefficients capture the major time-invariant wave-like features of the signal. The large coefficients are useful in data compression, feature extraction, finding approximate periods and other data mining tasks. Thus the problem of finding a small set of the largest Fourier coefficients of a signal that capture most of the signal trends is a fundamental task in Fourier analysis and its applications. We address the question of how these coefficients can be estimated fast and accurately.

Suppose we are given an upper bound M on the ratio of largest to smallest numerical quantity encountered. Our *main result* in this paper is an algorithm that samples at most $\mathrm{poly}(B\log(N)\log(M)/\epsilon)$ positions of the signal and outputs a Fourier representation $\mathbf{R}$ of B terms such that $\|\mathbf{A}-\mathbf{R}\|_2^2$ is within the factor $(1+\epsilon)$ of the best possible error $\|\mathbf{A}-\mathbf{R}_{\rm opt}\|_2^2$. In fact, the sampled positions are independent of the signal. The total computation time is also $\mathrm{poly}(B\log(N)\log(M)/\epsilon)$. Even though the number of sampled positions is significantly sublinear in N, the outcome is nearly the best choice of B Fourier coefficients. This is the first known result that accurately estimates Fourier representations compactly (in polylogarithmic "space" or samples) and rapidly (in polylogarithmic time without even having to read the entire input). This is possible because the basis functions have wide support. Our result should be compared with what is *not* possible in this model. No algorithm can even estimate the *average* of the signal values, or, therefore, any nontrivial norm in similar bounds using $o(N)$ samples. In particular, no algorithm can estimate $\|\mathbf{A}\|_2$ accurately within the sampling complexity above. This in turn means we can not estimate $\|\mathbf{A}-\mathbf{R}\|_2^2$, the quality of our output representation. Nevertheless, our analysis proves that our algorithm based on samples is provably accurate to within $1+\epsilon$ in providing the best B-term Fourier coefficient representation of a signal. We expect our algorithm to find many uses in computational Fourier analysis.

Mathematically speaking, we are recovering a signal from sample values of that function but using a particular type of signal recovery. We ask to recover $\mathbf{A}$ from the class of B-term trigonometric polynomials. This recovery $\mathbf{R}$ is not the optimal representation $\mathbf{R}_{\rm opt}$ from among all possible but we can guarantee that if the optimal ℓ^2 error is $\delta \neq 0$, then the error of our recovery is within a factor $(1+\epsilon)$; *i.e.*, $\|\mathbf{A}-\mathbf{R}\|_2^2 \le (1+\epsilon)\delta$. (Up to precision issues, we also recover $\mathbf{A}$ exactly if $\mathbf{A}$ consists exactly of B frequencies.) Our work is related to mathematical ideas that have been around for decades, as well as recent results in theoretical computer sci-

*AT&T Labs—Research, {agilbert, mstrauss, muthu}@research.att.com
†University of Pennsylvania, sudipto@cis.upenn.edu
‡MIT, indyk@theory.lcs.mit.edu

STOC'02, May 19-21, 2002, Montreal, Quebec, Canada.

ence. In what follows, we will place our work in the context of topics from signal processing, mathematical uncertainty principles and learning theory, among other areas.

Signal Processing. There are many signal processing results pertaining to signal recovery from "sampled" values of a function. We restrict ourselves to the "classical" result: the Shannon sampling theorem [14]. It says that you can exactly reconstruct a continuous-time function $\mathbf{F}$ from its sampled values, provided the nonzero Fourier coefficients all lie within the interval $[-\pi/T, \pi/T]$ and the samples are taken at regular spaced points in time, each separated by T. If the largest frequency that $\mathbf{F}$ contains is $1/2T$, then we are sampling $\mathbf{F}$ at the Nyquist frequency (and hence fewer samples would be insufficient to reconstruct $\mathbf{F}$ exactly). The reconstruction procedure in the Shannon sampling theorem is very different from our setting. The representation is a linear combination of translates of one basis function with sample values as weights and it is exact, as opposed to approximate reconstruction using trigonometric polynomials that we do here. There are many other interpolation and/or signal recovery algorithms that use sample values or simple functions of sample values as coefficients in a linear reconstruction procedure with different building block functions.

Mathematics and Classical Uncertainty Principles. Let $\mathbf{A}$ be a sequence of length N and let $\widehat{\mathbf{A}}$ be its discrete Fourier transform. Let N_t and $N_w = B$ be the number of nonzero values in $\mathbf{A}$ and $\widehat{\mathbf{A}}$ respectively. The discrete time uncertainty principle says

THEOREM 1. *(Donoho and Stark [5]) For all* $\mathbf{A}$, $N_t N_w \geq \Omega(N)$.

This implies that any approximation to a signal cannot be simultaneously sparse in both time and frequency. Thus we cannot hope to construct a representation for $\mathbf{A}$ that consists of B frequencies and B spikes if $B^2 < N$. This is in contrast to other types of representations (*e.g.*, histograms and singular value decompositions), in which one can produce compact representations that are sparse in time as well.

Algorithmic Version of Uncertainty Principle. Suppose that we wish to reconstruct a signal $\mathbf{A}$ but we observe $\mathbf{A}(t)$ only on some index set T. In addition, we know that $\mathbf{A}$ is synthesized using only B frequencies. The (contrapositive of the) uncertainty principle tells us [5]:

> For all B and *all* sample sets T, if $\mathbf{A}$ has only B frequencies and $|T| \geq N - O(N/B)$, then $\mathbf{A}$ can be reconstructed from samples on T.

This result is tight over the entire range of $|T|$ and B. For any B that divides N, there are two different functions, each limited to B frequencies, that agree on some T of size just less than $(N - N/B)$—these functions cannot be distinguished. By contrast, one of our results says (where each sample in T is chosen uniformly and independently from $[0, N)$ and $\widetilde{\Omega}$ suppresses factors of $\log(N)\log(M)$),

> For all B and *most* sample sets T, if $\mathbf{A}$ has only B frequencies and $|T| \geq \widetilde{\Omega}(B)$, then $\mathbf{A}$ can be reconstructed from samples on T.

In some applications, the sample set T is adversarially forced. When T can be chosen even randomly, however, and $B \ll N$, our results give reconstruction from a sample set of size perhaps exponentially smaller than what is guaranteed by the uncertainty principle. Thus our results here imply that the uncertainty bound is loose in this context, which may be of independent interest.

The preceding discussion of the context of our work concerns only the number of samples needed to reconstruct $\mathbf{A}$, and applies even when the reconstruction time is $\Omega(N)$. We note that a significant contribution of our work is bounding the *time* to construct a representation. (Indeed, in order to bound the reconstruction time, our algorithm uses a sample set T that is larger than above and drawn from a distribution more complicated than the uniform distribution.) Thus we regard our work as a (computational) time-bounded strengthening of the uncertainty principle.

Theoretical Computer Science. There are several relevant areas, and we will distinguish our work from related work in each.

First, a relationship was established in [11] between Fourier spectra and learnability. For the class of Boolean functions, the discrete Boolean Fourier basis was defined based on the parity of subsets of the input variables. This representation was used to demonstrate learnability of functions. In [10], authors presented a polynomial time algorithm to find all the large coefficients for this Fourier basis for given Boolean functions. Our work here is related to [10]. Clearly their Fourier basis is different from ours (parity of subsets vs trigonometric polynomials). Our basis is the classical one used in convolutions and other well-known applications, while the one in [11, 10] proves useful in learning theory context (see [13] for a nice overview of the relationship between their Fourier transforms and complexity theory). Our technique is related to [10] at the high level involving the test of groups of coefficients in order to isolate the large ones, but the technical details are vastly different.

Our work can be regarded as a recovery of coefficients in the complex polynomial $p(z) = \sum_\omega \widehat{\mathbf{A}}(\omega) z^\omega$, from samples of p among those equally spaced along the unit circle, $\mathbf{A}(t) = p\left(e^{2\pi i t/N}\right)$. Indeed, if only B coefficients are nonzero, we recover p exactly. In general, however, our result also provides some incomplete and approximate recovery of p, good in the ℓ_2 sense, provided p is close in the ℓ_2 sense to a sparse polynomial. While there has been work in sparse polynomial recovery over finite fields, especially for polynomials of bounded degree [3] (in which, if the polynomial is noisy, the noise bound is the Hamming norm), there has been work in exact polynomial recovery over characteristic zero (especially in the multivariate case) [9], and there has been work [2] in approximate recovery, over characteristic zero under ℓ_1 norm, of individual coefficients (but not recovery of representations) we are not aware of previous work in approximate polynomial reconstruction over the complex numbers with the bounds we give.

Next, there has been recent work on approximating matrices by low rank ones. In particular, in [6], the authors find a matrix D^* from a given matrix A in time polynomial in k such that $\|A - D^*\|_F \leq \max_{\mathrm{rank}(D) \leq k} \|A - D\|_F + \epsilon \|A\|_F$; here $\|\cdot\|_F$ is the Frobenius norm which is the sum of the squares of matrix entries. This work is related to ours since the authors find the k significant singular values; however, the two contexts and techniques are very different. In particular, their work involves weighted sampling of entries in A which can be implemented by scanning A once in linear time. They do not face the challenge of estimating norms or inner products or group testing to find the largest coefficients as we do here. See [1] for further work, but no direct relevance to our work exists.

Finally, our work here is related to our earlier work on finding B significant wavelet coefficients or B-bucket histograms that are piecewise constant representation of the signal. Both seek a $(1+\epsilon)$ approximation to the best size-B representation (B histogram buckets, B wavelet terms, or B Fourier terms). Both use the same high-level strategy: Identify big basis functions by randomized group testing, estimate coefficient values, and subtract from the residual under an overall classic greedy pursuit algorithm. However, there the results themselves differ greatly on most technical issues. The algorithm of [7] works in the streaming model and looks at the entire signal as it necessarily must because sampling will not provide the result there as the basis functions have narrow support. It uses very large-support randomized linear projections a la Johnson and Lindenstrauss [8]. By contrast, we use only samples. The main obstacle in this paper is that we cannot even reliably estimate norms by sampling; by contrast, in [7] norm estimation is easy. As a result, our work here requires significantly different technical machinery as will be evident from the rest of this paper.

Organization. The paper is organized as follows. In Section 3, we give an ideal version of our algorithm, that is infeasible in the sampling model. We thereby give the overall structure of the algorithm that we later introduce and we give names for the various parts. In Section 4, we revisit each element of the ideal algorithm, modifying it if necessary, then implementing it in the sampling model. In Section 5, we show how to estimate $\|\mathbf{A}-\mathbf{R}\|_2^2$, given $\|\mathbf{A}\|_2^2$ by oracle.

2. NOTATION AND PRELIMINARIES

Let $\mathbf{A} = (\mathbf{A}(0), \ldots, \mathbf{A}(N-1))$ be a signal indexed by t, regarded as an integer mod N. We denote by $\psi_\omega(t)$ the ω'th Fourier basis function $\frac{1}{\sqrt{N}} e^{2\pi i\omega t/N}$. Then $\widehat{\mathbf{A}}(\omega) = \langle \mathbf{A}, \psi_\omega \rangle = \frac{1}{\sqrt{N}} \sum_t \mathbf{A}(t) e^{-2\pi i\omega t/N}$ is the ω'th Fourier coefficient of $\mathbf{A}$. (Note that $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_t \mathbf{A}(t)\overline{b(t)}$, where $\overline{b(t)}$ is the complex conjugate of $b(t)$.) The vector $\widehat{\mathbf{A}}$ is the *spectrum* of $\mathbf{A}$. The basis functions in the time domain are denoted δ_t, so $\delta_t(s)$ is 1 if $s = t$ and 0 otherwise. We say that $\widehat{\mathbf{A}}(\omega)$ is *bigger* than $\widehat{\mathbf{A}}(\omega')$ if $|\widehat{\mathbf{A}}(\omega)| > |\widehat{\mathbf{A}}(\omega')|$. The *energy* of $\mathbf{A}$ is $\|\mathbf{A}\|_2^2$. We also refer to $|\widehat{\mathbf{A}}(\omega)|^2$ as the energy of the Fourier coefficient $\widehat{\mathbf{A}}(\omega)$ (or the energy of ω) and, similarly, the energy of a set of Fourier coefficients is the sum of the squares of their magnitudes.

We write $\mathbf{F} \star \mathbf{G}$ to denote the convolution, $(\mathbf{F} \star \mathbf{G})(t) = \sum_s \mathbf{F}(s)\mathbf{G}(t-s)$. It follows that $\widehat{\mathbf{F} \star \mathbf{G}} = \widehat{\mathbf{F}}\widehat{\mathbf{G}}$. Parseval's equality says that $\sum_t |\mathbf{F}(t)|^2 = \sum_\omega |\widehat{\mathbf{F}}(\omega)|^2$. We have $\cos(x) = (e^{ix} + e^{-ix})/2$. We denote by χ_S the signal that equals 1 on S and zero elsewhere. The index to χ_S may be time or frequency; this is made clear from context. For more background on Fourier analysis, see [15].

For small real $\epsilon > 0$, we will write $\widetilde{x} = (1 \pm \epsilon)x$ to mean $|\widetilde{x} - x| \le \epsilon|x|$, where $|x|$ is the complex absolute value. It is not necessary that the complex-valued $\widetilde{x}$ be a real multiple of the complex-valued x.

We have the following access to $\mathbf{A}$. We toss coins, then, based on the coins, compute a set $T \subseteq [0, N)$ of indices, and learn $\mathbf{A}(t)$ for $t \in T$. Thus, for example, we could pick t at random and then learn $\mathbf{A}(t), \mathbf{A}(t+1)$, and $\mathbf{A}(t+2)$, so that the samples are dependent, but we cannot adapt our choice of sample to the values seen. Our goal is to bound the time used by the algorithm, which implies a bound on the number of samples made.

If $\mathbf{A}$ contains large or precise entries, we will need extra time and space (polynomial in the input size) just to read and process input samples. Similarly, if $\mathbf{A}$ is exactly or very nearly representable by B coefficients, we will need time and space to perform computations precise enough to capture $\mathbf{A}$ as well as we claim or as well as desired (in the case of an exact superposition).

Let ι denote a small number such that, if $\|\mathbf{A}-\mathbf{R}\|_2^2 \le \iota$, then we consider the signal to have been recovered. (Henceforth, assume $\iota = 1$.) Let M be a crude upper bound on $\|\mathbf{A}\|_2^2$. (Note that $\|\mathbf{A}\|_2 \ge \|\mathbf{A}\|_\infty$, the maximum size of input values.) We assume that we know M. Note that most arithmetic on integers up to M require $\log(M)$ bit operations; but, even if we assume that common arithmetic on numbers can be peformed in constant time (as we do here), some of our algorithms will have cost depending on M, since some of our algorithms perform $\log(M)$ arithmetic operations.

Let $c(N) \le 2^{N^{o(1)}}$ be any increasing function. Then any sampling algorithm that, given a vector $\mathbf{F}$ of length N, distinguishes $\|\mathbf{F}\|_2^2 \le 1$ from $\|\mathbf{F}\|_2^2 \ge c(N)$, makes $\log(M)N^{1-o(1)}$ samples. This is because suppose $\mathbf{F} = c(N)\delta_t$, for randomly chosen t. It is easy to see that an algorithm that distinguishes $\mathbf{F}$ from 0 makes $\Omega(N) \ge \log(M)N^{1-o(1)}$ samples.

Our algorithms are randomized. That is, for all inputs $\mathbf{A}$ and 3/4 of the random choices of our algorithm, the algorithm succeeds. The success probability 3/4 ("significant") can be boosted to as close to 1 as desired ("high") using standard inexpensive techniques, not discussed further.

3. AN IDEAL ALGORITHM

In this section we present an idealized greedy pursuit (IGP) algorithm for finding a near-optimal Fourier representation. (See, *e.g.*, [12] for similar algorithms.) While IGP is not feasible in the sampling model, it can be modified appropriately. These modifications form the bulk of our technical contribution and will be discussed in the next section. The main components of IGP are, however, crucial for the feasible algorithm and it is illustrative to specify them precisely. The three main components of IGP are:

- *Identification* of all frequencies ω that contribute significantly to the signal's energy ($|\widehat{\mathbf{A}}(\omega)|^2$ is large);
- *Estimation* of $\widehat{\mathbf{A}}(\omega)$ when $|\widehat{\mathbf{A}}(\omega)|$ is large (call this estimate $\widetilde{\widehat{\mathbf{A}}}(\omega)$);
- *Iteration* on the residual signal $\mathbf{A} - \widetilde{\widehat{a}}(\omega)\psi_\omega$ (*i.e.*, we add the representation for $\mathbf{A} - \widetilde{\widehat{\mathbf{A}}}(\omega)\psi_\omega$ to $\widetilde{\widehat{\mathbf{A}}}(\omega)\psi_\omega$).

In the ideal setting, this algorithm is easily seen to find, in B steps, the representation consisting of the B terms with largest-magnitude coefficients, which is optimal.

3.1 Ideal Greedy Pursuit

In the ideal setting, the *estimation* step is exact and the *iteration* step is simply greedy pursuit. That is, we assume that, once we have identified a significant frequency ω_0, we have some technique to compute $\widehat{\mathbf{A}}(\omega_0)$ exactly. Because the Fourier basis $\{\psi_\omega\}$ is an orthonormal one, the best B-term representation for $\mathbf{A}$ is the best 1-term representation $\widehat{\mathbf{A}}(\omega_0)\psi_{\omega_0}$, plus the best $(B-1)$-term representation for $\mathbf{A} - \widehat{\mathbf{A}}(\omega_0)\psi_{\omega_0}$ (in which ω_0 is absent). We greedily add $\widehat{\mathbf{A}}(\omega_0)\psi_{\omega_0}$ to a growing representation $\mathbf{R}$ and iterate on $\mathbf{A} - \widehat{\mathbf{A}}(\omega_0)\psi_{\omega_0}$.

While the *estimation* and *iteration* steps have direct analogs in the sampling model, the *identification* step does not. We now consider more closely two components of *identification*, namely *group testing* and *isolation*, that have direct analogs.

Group Testing. We assume initially that $\mathbf{A}$ consists of one pure frequency ω_0 plus orthogonal terms of insignificant energy; that is, $\mathbf{A} = \widehat{\mathbf{A}}(\omega_0)\psi_{\omega_0} + \rho$ with $\langle\psi_{\omega_0}, \rho\rangle = 0$ and $\|\rho\|_2^2 \leq (0.25)\|\mathbf{A}\|_2^2$. Our goal is to identify ω_0. We will make progress towards this goal by repeatedly halving a list of candidates that always contains ω_0.

Consider the vector $\widehat{\mathbf{A}}\chi_{[0,k)}$ consisting of the first k Fourier coefficients of $\mathbf{A}$, where $k \approx N/2$. Observe that, if $\omega_0 \in [0,k)$, then $\left\|\widehat{\mathbf{A}}\chi_{[0,k)}\right\|_2^2 \geq |\widehat{\mathbf{A}}(\omega_0)|^2 \geq (0.75)\|\widehat{\mathbf{A}}\|_2^2$, and

$$\left\|\widehat{\mathbf{A}}\chi_{[k,N)}\right\|_2^2 \leq \|\rho\|_2^2 \leq (0.25)\|\widehat{\mathbf{A}}\|_2^2 < \left\|\widehat{\mathbf{A}}\chi_{[0,k)}\right\|_2^2. \qquad (1)$$

Thus, by *grouping* the first and second halves of the spectrum of $\mathbf{A}$ with two filters $\chi_0 = \chi_{[0,k)}$ and $\chi_1 = \chi_{[k,N)}$ and by *testing* whether the energy of the first half is greater than or less than the energy of the second half, we cut roughly in half the number of possibilities for ω_0. Note that we can compute $\|\widehat{\mathbf{A}}\chi_{[0,k)}\|_2^2$ in the time domain as $\|\mathbf{A} \star \chi\check{}_{[0,k)}\|_2^2$, where $\chi\check{}_{[0,k)}$ is the inverse Fourier transform of $\chi_{[0,k)}$. (Even in this infeasible ideal algorithm, we don't want to compute $\widehat{\mathbf{A}}$ directly.)

Suppose N is an odd prime and suppose, without loss of generality, that $0 \leq \omega_0 < N/2$. We now adaptively cut roughly in half the remaining possibilities for ω_0. To do this, define $\mathbf{F}(t) = \mathbf{A}(t/2)$. It follows that the sequence $\widehat{\mathbf{A}}(0), \widehat{\mathbf{A}}(1), \widehat{\mathbf{A}}(2), \ldots, \widehat{\mathbf{A}}(N/2-1)$ is mapped into

$$\{\widehat{\mathbf{F}}(0), \widehat{\mathbf{F}}(2), \widehat{\mathbf{F}}(4), \ldots, \widehat{\mathbf{F}}(\approx N)\},$$

and the other (low-energy) frequencies of $\widehat{\mathbf{A}}$ in $[N/2, N)$ are mapped to the complementary set. We repeat the $[0, N/2)$ versus $[N/2, N)$ test for the big frequencies of $\mathbf{F}$, thus eliminating roughly half the positions for ω_0. Continuing this way, we learn all of ω_0.

Isolation. The group-testing procedure is successful if $\mathbf{A}$ consists predominantly of one pure tone. On the other hand, if, for example, $\mathbf{A}$ comprises two pure frequencies, each with 1/3 the total energy, plus orthogonal noise at 1/3 the energy, group testing will fail to find directly either frequency. From a test that tells us that $S_1 \subseteq [0, N)$ contains some significant frequency and another test that tells us that $S_2 \subseteq [0, N)$ contains a significant frequency, we cannot conclude that $S_1 \cap S_2$ contains a big frequency. This is in contrast to the situation of a single overwhelmingly energetic tone. We proceed by isolating big Fourier coefficients first then using group testing to identify them precisely.

Assume that $\mathbf{A}$ consists of one or more pure tones plus orthogonal noise but that the energy of the pure tones is no longer overwhelming. That is, $\mathbf{A} = \psi + \rho$ with $\langle\psi, \rho\rangle = 0$ and $\|\psi\|_2^2 \geq \eta\|\mathbf{A}\|_2^2$, but η is considerably less than 1/2. We seek a frequency ω_0 such that $|\widehat{\mathbf{A}}(\omega_0)|^2 \geq \eta\|\mathbf{A}\|_2^2$.

From $\mathbf{A}$, first construct a vector $R\mathbf{A}$ such that the spectrum of $R\mathbf{A}$ is a random permutation of the spectrum of $\mathbf{A}$. Now, consider the family of functions $\mathbf{F}_j$, $j = 0, \ldots, 1/\eta^2$, obtained by filtering $R\mathbf{A}$ with the filters $\chi_j = \chi_{[j\eta^2 N, (j+1)\eta^2 N)}$ in the frequency domain. Note that each $\widehat{\mathbf{F}}_j$ consists of (at most) $\eta^2 N$ non-zero coefficients that form a random subset of $\eta^2 N$ coefficients of $\widehat{\mathbf{A}}$. One can show that, with significant probability, each of the frequencies ω_i of $\mathbf{A}$ that contains energy at least $\eta\|\mathbf{A}\|_2^2$ is *isolated* to one $\mathbf{F}_{j_i}$, and all frequencies are isolated this way. Furthermore, we show that each such coefficient satisfies $|\widehat{\mathbf{F}}_{j_i}(\omega_i)|^2 \geq (0.75)\|\mathbf{F}_{j_i}\|_2^2$, so we can appeal to the previous group testing procedure to identify ω_i precisely.

3.2 Infeasible Aspects of IGP

There are several aspects of IGP that are infeasible in the sampling model. In particular, a sampling algorithm cannot estimate $\|\mathbf{A}\|_2^2$ or other norms well. For many aspects of IGP, estimation of norms is part of a straightforward implementation; fortunately, it is not necessary for all implementations. In addition, although a sampling algorithm can produce a representation $\mathbf{R}$ that is near optimal, the algorithm has no way to assess the error $\|\mathbf{A} - \mathbf{R}\|_2^2$.

The infeasible aspects are:

- We cannot compute $\widehat{\mathbf{A}}(\omega)$ exactly, nor even with good relative error. Besides estimation itself, this obstacle also affects *identification* (we can't always resolve which coefficient is largest) and *iteration* (frequency ω may appear in $\mathbf{A} - \widetilde{\widehat{a}}(\omega)\psi_\omega$; ω may even dominate).
- We cannot compare $\left\|\widehat{\mathbf{A}}\chi_{[0,\frac{N}{2})}\right\|_2^2$ with $\left\|\widehat{\mathbf{A}}\chi_{[\frac{N}{2},N)}\right\|_2^2$ directly, since a sampling algorithm cannot, in general, estimate norms, even to within huge relative or additive error.
- The vector $\mathbf{A}\star\chi\check{}$ is too expensive to compute by sampling, even approximately, since each point of $\mathbf{A}\star\chi\check{}$ depends significantly on most points of $\mathbf{A}$. (Our goal is to sample just a few points of $\mathbf{A}$.)[1]
- Constructing $R\mathbf{A}$ to have spectrum that is a truly random permutation of $\mathbf{A}$'s spectrum is too expensive for a sampling algorithm.

4. A FEASIBLE ALGORITHM

In the previous section, we listed four aspects of IGP that are infeasible in the sampling model. In this section we address those problems with specific technical modifications. In Section 4.1, we show how to estimate $\widehat{\mathbf{A}}(\omega)$ as $\widetilde{\widehat{\mathbf{A}}}(\omega)$, with $|\widehat{\mathbf{A}}(\omega) - \widetilde{\widehat{\mathbf{A}}}(\omega)|^2 \leq \epsilon\|\mathbf{A}\|_2^2$. In Section 4.2, we show that a sampling algorithm can get upper and lower bounds on L^2 norms, which are good enough to perform the analog of the comparison of norms in (1). In Sections 4.3 and 4.4, we

[1] If N is even, then one can sample from $\mathbf{A} \star \chi_{\mathrm{even}}\check{}$ by making two samples from $\mathbf{A}$—this is the basis of the FFT algorithm. This construction, however, does not generalize.

show how to perform *identification* and *isolation* using time-limited filters suitable for a sampling algorithm (as opposed to $\chi^{\check{}}$). There we use a pairwise-independent permutation of the spectrum of $\mathbf{A}$. Finally, in Section 4.5, we analyze the resulting greedy pursuit algorithm.

4.1 Estimating Individual Fourier Coefficients

LEMMA 2. *There exists a sampling algorithm A with cost $O(\log(M)/\epsilon)$ such that, for each frequency ω and each vector $\mathbf{A}$, and each $\epsilon > 0$, with high probability, the output of A on input $(\mathbf{A}, \omega, \epsilon)$ is $\widehat{\widetilde{\mathbf{A}}}(\omega)$ with $|\widehat{\widetilde{\mathbf{A}}}(\omega) - \widehat{\mathbf{A}}(\omega)|^2 \le \epsilon \|\mathbf{A}\|_2^2$.*

PROOF. Pick t uniformly at random and let $\mathbf{A}' = N\mathbf{A}(t)\delta_t$, where N is the length of the signal. Then we have $E[\widehat{\mathbf{A}'}(\omega)] = \widehat{\mathbf{A}}(\omega)$ and $E[|\widehat{\mathbf{A}'}(\omega)|^2] \le \|\mathbf{A}\|_2^2$. Thus, if we take an average $\mathbf{A}''$ of $O(1/\epsilon)$ independent copies of $\mathbf{A}'$ and compute $\widehat{\mathbf{A}''}(\omega)$, the result, X, has $E[X] = \widehat{\mathbf{A}}(\omega)$ and $E[|X - \widehat{\mathbf{A}}(\omega)|^2] \le O\left(\epsilon\|\mathbf{A}\|_2^2\right)$. The result follows by the Chebychev inequality. We omit the details. □

4.2 Estimating Norms

We turn now to estimating norms by sampling.

LEMMA 3. *There's a sampling algorithm, A, that takes a bound M on $\|\mathbf{F}\|_2$, makes $O(\log\log(M))$ samples and runs in time polynomial in $\log(M)$, such that, on input $\mathbf{F}$ (of length N) with biggest Fourier coefficient $\widehat{\mathbf{F}}(\omega)$, A produces, with high probability, a random output X such that*

1. *$X \le \|\mathbf{F}\|_2^2$ (for any $\mathbf{F}$).*
2. *If $|\widehat{\mathbf{F}}(\omega)|^2 \ge 0.95\|\mathbf{F}\|_2^2$, then $X \ge 0.5\|\mathbf{F}\|_2^2$.*

Intuitively, the proof is as follows. For random $t \in [0, N)$, let $Y = N|\mathbf{F}(t)|^2$. Then $E[Y] = \|\mathbf{F}\|_2^2$, and the average of independent copies of Y would be a good estimate if the variance of Y were small. The variance comes from spikes, so let Z be a random variable that equals Y except that we set $Z = 0$ if $\mathbf{F}$ has a large spike at t. Then $0 \le E[Z] \le E[Y] = \|\mathbf{F}\|_2^2$, and Z has small variance. Thus we have the first statement. Note that if $\mathbf{F}$ is a pure frequency ψ_ω (which has no spikes—in fact, $|\psi_\omega|$ is constant), then Y is always exactly equal to $\|\mathbf{F}\|_2^2$. If 95% of $\mathbf{F}$ is concentrated in a pure frequency, then the energy of $\mathbf{F}$ contained in any spike is small, so $E[Z] \approx E[Y] = \|\mathbf{F}\|_2^2$, and, since the variance of Z is small, Z is a good approximation with high probability.

The algorithm sets a ceiling above which all values of the function are clipped as if they were spikes. The algorithm then estimates the energy of the clipped function. The algorithm gradually lowers the clipping ceiling until the energy estimate is consistent with the ceiling value.

PROOF. For each number c, define $K_c\mathbf{F}$ ("$\mathbf{F}$ clipped at c") by

$$(K_c f)(t) = \begin{cases} \mathbf{F}(t), & |\mathbf{F}(t)| \le c/\sqrt{N} \\ 0 & \text{otherwise.} \end{cases}$$

Let $\beta > 0$ be a small number (independent of N and M) that we will determine later. Perform the algorithm of Figure 1.

Figure 1: Approximation algorithm for $\|\mathbf{F}\|_2^2$. The parameters $\beta > 0$ and $k = k(\beta)$ are described in the text.

```
c ← 2M/β
Do
{
    Pick k random independent samples t_j from [0, N)
    X ← avg_j N|K_c F(t_j)|^2
    c ← c/(1 + β)
} while X < β^2 c^2 and c ≥ 1
Output X/(1 + β).
```

First, the algorithm halts in $O(\log(M))$ iterations, since $c = 1$ is reached from an initial value of $c = 2M/\beta$, by constant-fraction reductions. If the algorithm halts because $c < 1$, then $K_c\mathbf{F}$ is considered to be zero under our specified precision, so $X = 0$, and the Lemma holds. So, henceforth, assume that, when the algorithm halts, $X \ge \beta^2 c^2$.

It is easy to see that $E[X] = \|K_c\mathbf{F}\|_2^2$ immediately after X is set, as a loop invariant. Next,

$$\begin{aligned} E[|X|^2] &= \frac{1}{k}\frac{1}{N}\sum_{t=0}^{N} N^2|K_c\mathbf{F}(t)|^4 = \frac{N}{k}\sum_{t=0}^{N}|K_c\mathbf{F}(t)|^4 \\ &\le \frac{N}{k}\|K_c f\|_\infty^2 \sum_{t=0}^{N}|K_c\mathbf{F}(t)|^2 \le \frac{1}{k}c^2\|K_c\mathbf{F}\|^2. \end{aligned}$$

Thus $X = \|K_c\mathbf{F}\|_2^2 \pm O\left(\frac{1}{\sqrt{k}}c\|K_c\mathbf{F}\|_2\right)$ with significant probability.

As long as $\beta^2c^2 \ge 2\|K_c\mathbf{F}\|_2^2$, and $k \le O(1/\beta^2)$ is sufficiently large, $X < \beta^2c^2$ with significant probability, so, if $c \ge 1$, the algorithm will not halt. On the other hand, if $\beta^2c^2 \le 4\|K_c\mathbf{F}\|_2^2$, the variance of X is at most

$$\frac{1}{k}c^2\|K_c\mathbf{F}\|_2^2 \le \frac{4}{\beta^2 k}\|K_c\mathbf{F}\|_2^4 \le \beta^2\|K_c\mathbf{F}\|_2^4,$$

provided $k \ge 4/\beta^4$. It follows that

$$X = (1 \pm \beta)\|K_c\mathbf{F}\|_2^2 \qquad (2)$$

is a good approximation to $\|K_c\mathbf{F}\|_2^2$. Thus, when the algorithm halts, $X/(1+\beta) \le \|K_c\mathbf{F}\|_2^2 \le \|\mathbf{F}\|_2^2$. This gives the first statement (and will also be used below).

Now, suppose $\mathbf{F} = \psi + \rho$, where $\psi = b\psi_\omega$ is a pure frequency, $\langle\psi, \rho\rangle = 0$ and $\|\rho\|_2^2 \le 0.05\|\mathbf{F}\|_2^2$.

Consider a ceiling value c with $\|\psi\|_2^2 < \beta c^2$ and let $S = \{t : |\mathbf{F}(t)| \ge c/\sqrt{N}\}$ be the set of indices at which $\mathbf{F}$ is clipped. We will now estimate the energy of ψ and ρ each restricted to $S^C = [0, N) \setminus S$. Observe that, by Markov's inequality, $|S| \cdot c^2/N \le \|\mathbf{F}\|_2^2$, or $|S| \le N\|\mathbf{F}\|_2^2/c^2 \le \frac{N}{0.95}\|\psi\|_2^2/c^2 < N(1.06)\beta$. Thus, since $|\psi(t)|$ is constant for all t,

$$\|\psi\chi_{S^C}\|_2^2 = (1 - |S|/N)\|\psi\|_2^2 \ge (1 - 1.06\beta)\|\psi\|_2^2.$$

Also, $\|\rho\chi_{S^C}\|_2 \le \|\rho\|_2$. It follows, using the triangle inequality, that

$$\begin{aligned} \|K_c\mathbf{F}\|_2 &= \|\mathbf{F}\chi_{S^C}\|_2 \ge \|\psi\chi_{S^C}\|_2 - \|\rho\chi_{S^C}\|_2 \\ &\ge \sqrt{1 - 1.06\beta}\|\psi\|_2 - \|\rho\|_2 \\ &\ge \sqrt{1 - 1.06\beta}\sqrt{0.95}\|\mathbf{F}\|_2 - \sqrt{0.05}\|\mathbf{F}\|_2 \\ &\ge \sqrt{1 - 1.06\beta}(0.75)\|\mathbf{F}\|_2. \end{aligned}$$

Thus

$$\begin{aligned} \|K_c\mathbf{F}\|_2^2 &\geq (1-1.06\beta)(0.75)^2\|\mathbf{F}\|_2^2 \\ &\geq 0.52\|\mathbf{F}\|_2^2, \end{aligned} \quad (3)$$

for sufficiently small β.

Above we showed that, when the algorithm halts, X is a good approximation to $\|K_c\mathbf{F}\|_2^2$. That is, $X/(1+\beta) \geq \frac{1-\beta}{1+\beta}\|K_c\mathbf{F}\|_2^2$. Thus, if the algorithm halts for $\beta c^2 > \|\psi\|_2^2$,

$$X/(1+\beta) \geq \frac{1-\beta}{1+\beta}\|K_c\mathbf{F}\|_2^2 \geq 0.52\frac{1-\beta}{1+\beta}\|\mathbf{F}\|_2^2 \geq 0.5\|\mathbf{F}\|_2^2,$$

for sufficiently small β.

Finally, suppose $\|\psi\|_2^2 < \beta c^2 \leq 4\|\psi\|_2^2$. Since

$$\beta c^2 \leq 4\|\psi\|_2^2 \leq 4\|\mathbf{F}\|_2^2 \leq 8\|K_c\mathbf{F}\|_2^2 \leq \frac{4}{\beta}\|K_c\mathbf{F}\|_2^2,$$

it follows by (2) that $X = (1\pm\beta)\|K_c\mathbf{F}\|_2^2$. By (3), $0.52\|\mathbf{F}\|_2^2 \leq \|K_c\mathbf{F}\|_2^2$, and, as above, it follows that $X \geq 0.5\|\mathbf{F}\|_2^2$. Thus

$$X \geq 0.5\|\mathbf{F}\|_2^2 \geq 0.5\|\psi\|_2^2 \geq 0.5\beta c^2/4 \geq \beta^2 c^2,$$

for sufficiently small β, and the algorithm terminates for c satisfying $\|\psi\|_2^2 < \beta c^2 \leq 4\|\psi\|_2^2$, if it hasn't already terminated for larger c. (Note that, in the first iteration, $\|\psi\|_2^2 \leq \|\mathbf{F}\|_2^2 \leq M^2 = \beta^2c^2/4 < \beta c^2$, so there are such c.) In any case, the output satisfies $X/(1+\beta) \geq 0.5\|\mathbf{F}\|_2^2$.

The number of iterations is at most $\log_{1+\beta}(2M/\beta)$, to reduce c from the initial value of $2M/\beta$ to 1, through reductions by the factor $1/(1+\beta)$. The number k of samples at each step depends only on $\beta > 0$, which is independent of N and M. Finally, note that we can reuse the same sample for each iteration. If we want the overall probability of success to be at least 3/4, then the probability of failure at each iteration needs to be at most $O(\beta/\log(M/\beta))$, *i.e.*, 1/4 divided by the number of iterations. To achieve this, we need $O(\log\log(M))$ samples. $\square$

The probability of success can be boosted from "significant" (*i.e.*, 3/4) to "high" (*i.e.*, $1-p$, for any desired p), by taking a median of $O(\log(1/p))$ repetitions and appealing to the Chernoff bound, as usual.

4.3 Group Testing

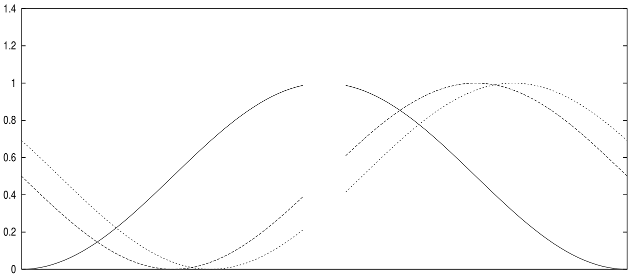
Recall that, in the ideal algorithm, for *identification*, we used filter functions defined in the frequency domain by $\chi_{[0,\frac{N}{2})}$ and $\chi_{[\frac{N}{2},N)}$. In this section, we first define imperfect filter functions based on cosines that will be easier for a sampling algorithm to use, then show how to use them.

Lemma 4. *For* $0 \leq k < 16$, *let* $\widehat{\mathbf{G}}_k(\omega) = (1+\cos(2\pi\omega/N - 2\pi k/16))/2$. *Then the inverse Fourier transform of* $\widehat{\mathbf{G}}_k$ *is given by* $\mathbf{G}_k(t) = \frac{\sqrt{N}}{2}e^{2\pi ikt/16}$ *for* $t = \pm 1$, $\mathbf{G}_k(0) = \sqrt{N}$, *and* $\mathbf{G}_k(t) = 0$ *otherwise.*

Thus $\mathbf{G}_k$ is a cosine of period N, shifted up and scaled to take values between 0 and 1, and shifted to the right by $k/16$ of a period. (See Figure 2.) For all k, $(\mathbf{F} \star \mathbf{G}_k)(t)$ is a linear combination of three values of $\mathbf{F}$, and can be sampled efficiently by an algorithm that samples $\mathbf{F}$.

Definition 1. Denote by pass_k the set $\{\omega : |2\pi\omega/N - 2\pi k/16| \leq 2\pi/32\}$.

Figure 2: Filter functions $\widehat{\mathbf{G}}_k$, for $k = 0, 4$, and 5. The region pass_0 is indicated by the gap. Each filter has period N and takes values between 0 and 1.



Thus pass_k consists of the ω's within 1/32 of a period to the maximum of $\widehat{\mathbf{G}}_k$. Note that, for $\omega \in \text{pass}_k$, $\widehat{\mathbf{G}}_k(\omega) \geq (\cos(2\pi/32)+1)/2 \geq 0.99$.

We turn now to group testing. In this section, we assume $\mathbf{F}$ consists mainly of a single frequency:

Definition 2. Let $\mathcal{F}$ denote the class of complex-valued vectors $\mathbf{F}$ of length N, such that, for some ω and some complex number b, $\mathbf{F} = \psi + \rho$, where $\langle\psi, \rho\rangle = 0$, $\psi = b\psi_\omega$ is a pure frequency, and $\|\psi\|_2^2 \geq (0.98)\|\mathbf{F}\|_2^2$. For $\mathbf{F} \in \mathcal{F}$, we say that $\mathbf{F}$ is "98% pure."

Finally, we also need the following transform on $\mathbf{F}$, that scales and translates $\widehat{\mathbf{F}}$.

Definition 3. Define $R_{\theta,\sigma}\mathbf{F}$ by $R_{\theta,\sigma}\mathbf{F}(t) = e^{2\pi i\theta t/N}\mathbf{F}(\sigma t)$. Thus $(\widehat{R_{\theta,\sigma}\mathbf{F}})(\sigma\omega + \theta) = \widehat{\mathbf{F}}(\omega)$.

Lemma 5. *There is a sampling algorithm, A, with time and sample complexity* $\log(M)\log^{O(1)}(N)$, *such that, on input a 98% pure vector* $\mathbf{F}$ *of length N, A identifies the frequency* ω' *of the biggest Fourier coefficient of* $\mathbf{F}$.

Proof. Observe that, for some k, $0 \leq k < 16$, $\omega' \in \text{pass}_k$; without loss of generality, assume $k = 0$. Now, consider the function $\mathbf{F} \star \mathbf{G}_0$. Since $\omega' \in \text{pass}_0$, we have $|\widehat{\mathbf{F} \star \mathbf{G}_0}(\omega')| \geq 0.99|\widehat{\mathbf{F}}(\omega')|$. Thus the pure frequency $\widehat{\mathbf{F} \star \mathbf{G}_0}(\omega')\psi_{\omega'}$ satisfies

$$\begin{aligned} |\widehat{\mathbf{F} \star \mathbf{G}_0}(\omega')|^2 &\geq (0.99)^2\|\psi\|_2^2 \geq (0.99)^2(0.98)\|\mathbf{F}\|_2^2 \\ &\geq (0.95)\|\widehat{\mathbf{F}}\|_2^2 \geq (0.95)\|\widehat{\mathbf{F}}\widehat{\mathbf{G}}_0\|_2^2 \\ &= (0.95)\|\mathbf{F} \star \mathbf{G}_0\|_2^2. \end{aligned}$$

It follows that, if we use Lemma 3 to estimate $\|\mathbf{F} \star \mathbf{G}_0\|_2^2$, the estimate will be at least

$$\begin{aligned} (0.5)\|\mathbf{F} \star \mathbf{G}_0\|_2^2 &\geq (0.5)|\widehat{\mathbf{F} \star \mathbf{G}_0}(\omega')|^2 \\ &\geq (0.5)(0.99)^2(0.98)\|\mathbf{F}\|_2^2 \geq (0.48)\|\mathbf{F}\|_2^2. \end{aligned}$$

On the other hand, now consider $f \star \mathbf{G}_4$. Note that

$$\begin{aligned} |\widehat{\mathbf{F} \star \mathbf{G}_4}(\omega')| &= |\widehat{\mathbf{F}}(\omega')|\widehat{\mathbf{G}_4}(\omega') \\ &\leq |\widehat{\mathbf{F}}(\omega')|(1+\cos(2\pi(1/32 - 4/16)))/2 \\ &\leq 0.6|\widehat{\mathbf{F}}(\omega')| \leq 0.6\|\mathbf{F}\|_2, \end{aligned}$$

since $\cos(2\pi(\omega/N - 4/16)))/2$ increases on $\text{pass}_0 \ni \omega'$. Also,

$$\|\widehat{\mathbf{F} \star \mathbf{G}_4}\|_2^2 - |\widehat{\mathbf{F} \star \mathbf{G}_4}(\omega')|^2 \le \|\rho\|_2^2 \le 0.02\|\mathbf{F}\|_2^2.$$

Thus $\|\mathbf{F} \star \mathbf{G}_4\|_2^2 \le (0.6)^2\|\mathbf{F}\|_2^2 + 0.02\|\mathbf{F}\|_2^2 = 0.38\|\mathbf{F}\|_2^2$. It follows that if we use Lemma 3 to estimate $\|\mathbf{F} \star \mathbf{G}_4\|_2^2$, the result will be at most $(0.38)\|\mathbf{F}\|_2^2 \le (0.48)\|\mathbf{F}\|_2^2$, which is a lower bound for the estimate of $\|\mathbf{F} \star \mathbf{G}_0\|_2^2$. Similarly, for all $4 \le k \le 12$, if we use Lemma 3 to estimate $\|\mathbf{F} \star \mathbf{G}_k\|_2^2$, then, since $\mathbf{G}_k(\omega') \le (1 + \cos(2\pi(1/32 - 4/16)))/2$, the result will be less than the estimate of $\|\mathbf{F} \star \mathbf{G}_0\|_2^2$. In general, if $\omega' \in \text{pass}_{k'}$ for k' not necessarily 0, we can reliably determine that $\omega \notin \text{pass}_k$, for $|k - k'| \ge 4$, *i.e.*, we can rule out 9/16 of the possibilities opposite the region $\text{pass}_{k'}$ where ω' lies. Thus, for $0 \le k' < 16$ and $|k - k'| \ge 4$, we compare $\|\mathbf{F} \star \mathbf{G}_{k'}\|_2^2$ with $\|\mathbf{F} \star \mathbf{G}_k\|_2^2$. If there is some such k with $\|\mathbf{F} \star \mathbf{G}_k\|_2^2$ apparently larger than $\|\mathbf{F} \star \mathbf{G}_{k'}\|_2^2$, then we conclude $\omega' \notin \text{pass}_{k'}$; otherwise, possibly, $\omega' \in \text{pass}_{k'}$. By the above argument, we can always eliminate 9 consecutive pass regions out of the 16, leaving a cyclic interval of length at most $7N/16$.

In the sequel, it will be convenient to express the situation as follows. Let P denote a cyclic interval of odd size at most $7N/16 + 1$ that includes all possibilities for ω'. Let b_1 denote the center of P. Then the spectrum of $\mathbf{F}_1(t) = e^{-2\pi i b_1 t/N}\mathbf{F}(t)$ is a shift of the spectrum of $\mathbf{F}$ by $-b_1$. Thus we know b_1, we know that $\omega' - b_1$ is the biggest frequency of $e^{2\pi i b_1 t/N}\mathbf{F}(t)$, and we know $\omega' - b_1$ is in the range $-(7N/32 + 1)$ to $+(7N/32 + 1)$. We will now seek $\omega' - b_1$.

Intuitively, we dilate the spectrum of $\mathbf{F}_1$ by 2, which can be accomplished in the time domain by dilating $\mathbf{F}_1$ by 1/2. Thus the interval of length just less than $N/2$ known to contain $\omega' - b_1$ is dilated to the alternate positions in an interval of length just less than N. We then rule out a cyclic interval of length at least $N/2$ in the dilated spectrum, leaving at most an interval of length $N/2$ in the dilated spectrum. We now undo the dilation, getting an interval of length just less than $N/4$, centered at some known b_2, as the possibilities for $\omega' - b_1$. That is, we've learned the equivalent of the second most significant bit of ω'. We repeat this to learn the other bits of ω'. A formal proof is technical, to do arithmetic simultaneously over the integers and over the integers mod N. We defer the details for the final version of this paper.

We proceed as in the group testing procedure of IGP, except we use cosines $\widehat{\mathbf{G}}_k$ instead of characteristic functions. To learn the most significant bit of ω' we use $\widehat{\mathbf{G}}_k$ instead of $\chi_{[0, \frac{N}{2})}$. We will use "band-pass" regions pass_k of $\mathbf{G}_k$ with size $N/16$ and antipodal "band-reject" regions with size $9N/16$, so we have to use all 16 filters $\widehat{\mathbf{G}}$ together to complete this test. To learn the second most significant bit of ω' after learning the most significant (and assuming, without loss of generality, $-N/4 < \omega' < +N/4$), we dilate $\mathbf{F}$ by 2 and apply the 16 filters again. Continuing this way, we learn ω'. $\square$

Note that our algorithm adapts to the signal values (for example, the b_j's depend on the signal values). But the choices of samples to the original signal do not depend on the signal values. For example, to sample $\mathbf{F}_1(t) = e^{-2\pi i b_1 t/N}\mathbf{F}(t)$ at t, just sample $\mathbf{F}$ at t and then multiply by $e^{-2\pi i b_1 t/N}$.

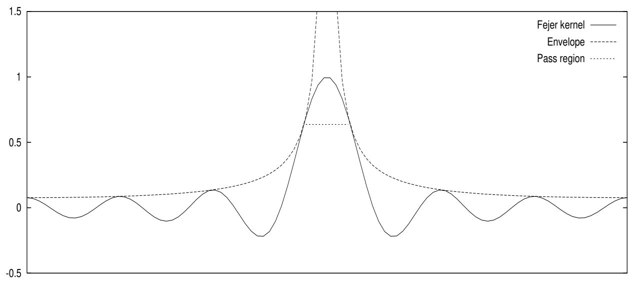
Above we assumed that N is prime, so that 2 is relatively prime to N. We then used the transform $R_{0,2^j}$. In general, we just need some number p to take the role of 2, such that p is relatively prime to N and $p \le \log^{O(1)}(N)$. In particular, if N is a power of 2, we could use $p = 3$. The necessary changes to the above algorithm are straightforward, and omitted. Finally, the prime number theorem states that the $\log(N)$'th prime is approximately $\log(N)\log\log(N)$. Since N is less than the product of the first $\log(N)$ primes, some prime less than $\log(N)\log\log(N)$ does not divide N. The least prime p not dividing N can be found in time $\log^{O(1)}(N)$ by linear search. It follows that the above techniques hold for all N.

4.4 Isolation

Finally, we consider *isolation* of the significant coefficients, the last part of *identification*. That is, we are given a signal $\mathbf{A} = \psi + \rho$, where ψ is a pure frequency $\psi = b\psi_{\omega'}$, $\langle\psi, \rho\rangle = 0$, and $\|\psi\|_2^2 \ge \eta\|\mathbf{A}\|_2^2$, where $\eta > 0$ may be much less than 0.98—even much less than 1/2. Our goal is to output a short list of frequencies that contains ω'. To that end, we construct a short sequence $\mathbf{F}_0, \mathbf{F}_1, \ldots$ of signals, such that, for some j, $|\widehat{\mathbf{F}}_j(\omega')|^2 \ge 0.98\|\mathbf{F}_j\|_2^2$. We can then use the techniques of previous sections on each $\mathbf{F}_j$ to identify a candidate for ω'. The following filter function will be used for isolation. (See Figure 3.)

Definition 4. (Fejér kernel; see, *e.g.*, [15].) For integer k, define $\mathbf{H}_k$ by $\mathbf{H}_k(t) = \frac{\sqrt{N}}{2k+1}\chi_{[-k,k]}$.

Figure 3: Filter function $\widehat{\mathbf{H}}_k$ based on Fejér kernel (wavy curve), together with envelope (monotone curves) and pass region (flat segment).



The proofs of the following lemmas are immediate.

LEMMA 6. *We have the following.*

1. *For all k,*

$$\widehat{\mathbf{H}}_k(\omega) = \frac{1}{2k+1}\sum_{t=-k}^{k} e^{-2\pi i\omega t/N} = \begin{cases} \frac{\sin(\pi(2k+1)\omega/N)}{(2k+1)\sin(\pi\omega/N)}, & \omega \ne 0 \\ 1, & \omega = 0. \end{cases}$$

2. *For all k and all ω, $|\widehat{\mathbf{H}}_k(\omega)| \le 1$.*

3. *For all k and ω such that $|\omega| \le \frac{N}{2(2k+1)}$, we have $\widehat{\mathbf{H}}_k(\omega) \ge 2/\pi$.*

4. *For all k, $\|\widehat{\mathbf{H}}_k\|_2^2 = \frac{N}{2k+1}$.*

We now show how to use $\mathbf{H}_k$, for appropriate k, to isolate frequencies. Again we assume that N is prime, though the techniques of this section can be modified to work for all N.

LEMMA 7. *Suppose θ and σ are chosen randomly mod N, with σ invertible. Then $\omega \mapsto \sigma\omega + \theta$ is a pairwise-independent permutation. That is, if $\omega_1 \neq \omega_2$ and $\omega_3 \neq \omega_4$, then $\omega_1 \mapsto \omega_3$ and $\omega_2 \mapsto \omega_4$ with probability $\frac{1}{N} \cdot \frac{1}{N-1}$.*

Note that $\widehat{R_{\theta,\sigma}\mathbf{A}}(\omega) = \widehat{\mathbf{A}}(\sigma\omega + \theta)$.

LEMMA 8. *Given signal $\mathbf{A}$ and number η we can create $2k+1$ new signals, $\mathbf{F}_0, \ldots, \mathbf{F}_{2k}$, such that:*

- *For each ω' such that $|\widehat{\mathbf{A}}(\omega')|^2 \geq \eta\|\mathbf{A}\|_2^2$, there exists a j such that $|\widehat{\mathbf{F}}_j(\omega')|^2 \geq 0.98\|\mathbf{F}_j\|_2^2$.*
- *Each $\mathbf{F}_j$ can be sampled by sampling non-adaptively from $\mathbf{A}$ in $O(k)$ places.*
- *k is small,* i.e., *$k \leq O(1/\eta)$.*

PROOF. Pick θ and σ at random, with σ invertible mod N. Let $k \leq O(1/\eta)$ be sufficiently large. Then consider the signal $\mathbf{F}_j = (e^{2\pi ijt/(2k+1)}\mathbf{H}_k) \star R_{\theta,\sigma}\mathbf{A}$, for $0 \leq j < 2k+1$. Semantically, $\widehat{\mathbf{F}}_j$ is formed by starting with $\widehat{\mathbf{A}}$, performing a pairwise-independent permutation (in the frequency domain), then multiplying pointwise by a phase shift of $\widehat{\mathbf{H}}_k$ by $jN/(2k+1)$.

One can sample a chosen point of $R_{\theta,\sigma}\mathbf{A}$ by sampling a single chosen point from $\mathbf{A}$. It follows that one can sample a chosen point from $\mathbf{F}_j = (e^{2\pi ijt/(2k+1)}\mathbf{H}_k) \star R_{\theta,\sigma}\mathbf{A}$ by sampling $|\mathrm{supp}(\mathbf{H}_k)| = 2k+1$ chosen points from $\mathbf{A}$.

Fix j such that $(j - 1/2)N/(2k+1) \leq \sigma\omega' + \theta < (j + 1/2)N/(2k+1)$; wlog, $j = 0$. We show that $|\widehat{\mathbf{F}_0}(\sigma\omega' + \theta)|^2 \geq 0.98\|\mathbf{F}_0\|_2^2$. First, observe that

$$|\widehat{\mathbf{F}_0}(\sigma\omega' + \theta)|^2 \geq (2/\pi)^2|\widehat{\mathbf{A}}(\omega')|^2 \geq (\eta/4)\|\widehat{\mathbf{A}}\|_2^2,$$

since $|\widehat{\mathbf{H}}_k(\sigma\omega' + \theta)| \geq (2/\pi)$. We now show that

$$\|\widehat{\mathbf{F}}_0\chi_{\{\sigma\omega'+\theta\}^C}\|_2^2 \leq (\eta/200)\|\widehat{\mathbf{A}}\|_2^2.$$

Fix some $\omega_0 \neq \omega'$, and consider the contribution of ω_0 to $\left\|\widehat{\mathbf{F}}_0\chi_{\{\sigma\omega'+\theta\}^C}\right\|_2^2$. By pairwise independence, the contribution $|\widehat{\mathbf{A}}(\omega_0)|^2$, which appears as $|\widehat{R\mathbf{A}}(\sigma\omega_0 + \theta)|^2$, appears at a position $\sigma\omega_0 + \theta$ chosen uniformly at random except that it avoids $\sigma\omega' + \theta$. It then gets attenuated by the factor $|\mathbf{H}_k(\sigma\omega_0 + \theta)|^2$. By hypothesis, $\sigma\omega' + \theta$ is in the "pass" region of $\mathbf{H}_k$, *i.e.*, $|\mathbf{H}_k(\sigma\omega' + \theta)|^2$ is greater than the average value, so the expected conditional attenutation $\frac{1}{N-1}\sum_{\omega\neq\omega'}|\mathbf{H}_k(\sigma\omega+\theta)|^2$ applied to ω_0 is more severe than the unconditional expected attenuation, *i.e.*,

$$\frac{1}{N-1}\sum_{\omega\neq\omega'}|\mathbf{H}_k(\sigma\omega + \theta)|^2 \leq \frac{\|\mathbf{H}_k\|_2^2}{N} = \frac{1}{2k+1}.$$

The expected attenuation hits all $\omega \neq \omega'$ equally, and, since $\sum_{\omega\neq\omega'}|\mathbf{A}(\omega)|^2 \leq \|\mathbf{A}\|_2^2$, we have

$$E_{\theta,\sigma}\left[\left\|\widehat{\mathbf{F}}_0\chi_{\{\sigma\omega'+\theta\}^C}\right\|_2^2 \;\middle|\; -\tfrac{N}{2(2k+1)} \leq \sigma\omega' + \theta < \tfrac{N}{2(2k+1)}\right] \leq \tfrac{\|\mathbf{A}\|_2^2}{2k+1}.$$

Thus, By Markov's inequality with $2k+1 = 800/\eta$,

$$\Pr_{\theta,\sigma}\left[\left\|\widehat{\mathbf{F}}_j\chi_{\{\sigma\omega'+\theta\}^C}\right\|_2^2 > (\eta/200)\|\widehat{\mathbf{A}}\|_2^2 \;\middle|\; -\tfrac{N}{2(2k+1)} \leq \sigma\omega' + \theta < \tfrac{N}{2(2k+1)}\right] \leq 1/4.$$

As usual, the success probability 3/4 can be boosted to probability $1-p$ with cost factor $O(\log(1/p))$. In this case, repeat the experiment $O(\log(1/p))$ times with independent randomness, and take the union of all $O(k\log(1/p))$ candidates for ω'. The probability that ω is not among them is the product of the probabilities, $(1/4)^{\log(1/p)} = p$. We can therefore make the probability that ω' is among the candidates at least $1 - \eta/4$. Since there are at most $1/\eta$ possible ω''s, the probability that *all* ω''s are among the candidates is at least 3/4. In turn, this probability can also be boosted, if desired. □

In case N is not prime, the above proof can be modified to work. First, choose σ at random but relatively prime to N, so that σ is still invertible. The permutation $\omega \mapsto \pi(\omega) = \sigma\omega + \theta$ is not pairwise independent; for example, if N is a power of 2, then $0 \mapsto \theta$ and $N/4 \mapsto N/4 + \theta$ or $N/4 \mapsto 3N/4 + \theta$. For each q dividing N, we need to sum the contribution to $E_{\theta,\sigma}\left[\left\|\widehat{\mathbf{F}}_j\chi_{\{\sigma\omega'+\theta\}^C}\right\|_2^2\right]$ due to ω with $\gcd(\omega' - \omega, N) = q$. As above, one can show that the expected contribution due to all such $\omega \neq \omega'$ is $\frac{1}{2k+1}$ of the total. By summing over q, the result follows. We omit the details.

4.5 Adaptive Greedy Pursuit

Above we showed, for any ϵ and B, given an $\mathbf{A}$, we can construct a short list of frequencies that contains all ω' such that $|\widehat{\mathbf{A}}(\omega')|^2 \geq \eta\|\mathbf{A}\|_2^2$, for appropriate η. Using Lemma 2, we can estimate the value of each corresponding coefficient to within $\eta\|\mathbf{A}\|_2$, additively. In this section, we show that this suffices for an adaptive greedy algorithm. This is our main result.

THEOREM 9. *Fix a signal $\mathbf{A}$. Let $\mathbf{R}'$ denote a representation for $\mathbf{A}$ with at most B terms and set $\mathbf{E} = \mathbf{A} - \mathbf{R}'$. Suppose, for all representations $\mathbf{R}'$ of at most B terms, we can, with cost $1/\eta^{O(1)}$,*

- *estimate $\widehat{\mathbf{E}}(\omega)$ as $\widetilde{\widehat{\mathbf{E}}}(\omega)$ with $|\widehat{\mathbf{E}}(\omega) - \widetilde{\widehat{\mathbf{E}}}(\omega)|^2 \leq \eta\|\mathbf{E}\|_2^2$*
- *output a list of at most $1/\eta^{O(1)}$ frequencies that contains all frequencies ω such that $|\widehat{\mathbf{E}}(\omega)|^2 \geq \eta\|\mathbf{E}\|_2^2$.*

Then we can, on input B and ϵ, with cost $(B\log(N)\log(M)/\epsilon)^{O(1)}$, output a B-term representation $\mathbf{R}$ with sum-square-error $\|\mathbf{A} - \mathbf{R}\|_2^2 \leq (1+\epsilon)\|\mathbf{A} - \mathbf{R}_{\mathrm{opt}}\|_2^2$, where $\mathbf{R}_{\mathrm{opt}}$ is the B-term representation for $\mathbf{A}$ with the least sum-square-error.

PROOF. We use the greedy algorithm of Figure 4. This is the classic greedy pursuit algorithm, except:

- When searching for a new coefficient for our representation, we cannot find the biggest coefficient, only a near-biggest. Since (in this variant of greedy pursuit) we have no facility to eject a wrongly-chosen coefficient, we need to argue that each coefficient we take is acceptable.

Figure 4: Greedy algorithm for Fourier approximation to signal A

$S' \leftarrow \emptyset$ // set of $\leq B$ basis vectors
$\mathbf{R}' \leftarrow 0$ // representation
$J \leftarrow (B \log(N) \log(M/\delta)/\epsilon)^{O(1)}$
for($j \leftarrow 0$; $j < J$; $j \leftarrow j+1$)
{
  $\Lambda \leftarrow$ identify $\left(\mathbf{A} - \mathbf{R}', \frac{\epsilon}{2B}\right)$
  if $|S'| == B$
    $\Lambda \leftarrow \Lambda \cap S'$
  if $\Lambda == \emptyset$
    Output $\mathbf{R}'$ and halt
LOOPING:
  for($\omega \in \Lambda$)
    $\widetilde{c}_\omega \leftarrow$ estimate $\left(\mathbf{A} - \mathbf{R}', \omega, \frac{\epsilon}{3 \cdot 24 j^2 B^2}\right)$
  $\omega' \leftarrow \mathrm{argmax}_{\omega \in \Lambda}(\widetilde{c}_\omega)$
  $S' \leftarrow S' \cup \{\omega'\}$
  $\mathbf{R}' \leftarrow \mathbf{R}' + c_{\omega'} \psi_{\omega'}$
}

- Because our estimates of coefficients are not perfect, our algorithm must revisit frequencies. So it cannot be expected to halt after B steps. Furthermore, unlike a classic algorithm that can measure its progress and can halt when no further improvement is possible, our algorithm in general continues even after no further progress is possible. We need to show that potential extraneous rounds of greedy pursuit performed by our algorithm don't worsen the error by much.

Each of these is straightforward to check. For completeness, we present the entire greedy pursuit algorithm.

The algorithm clearly halts, quickly, and outputs a representation of at most B terms. We now argue that the representation has sufficiently small error. Let δ denote the error of the best B-term representation (with optimal coefficients).

We show that the following invariant is maintained:

- At line LOOPING, $\mathbf{A}$ has an approximate representation of B terms, with error at most $\left(1 + \frac{\epsilon|S'|}{2B}\right)\delta$, extending S'.

By "extension" of S', we mean a representation $\mathbf{R}''$ over frequencies in $S'' \supseteq S'$. An extension of S' might be the current representation $\mathbf{R}'$ on S' itself, might involve new terms, or might have better approximations to the terms in $\mathbf{R}'$.

By hypothesis, the invariant is true the first time line LOOPING is executed. Suppose execution is at Line LOOPING at an arbitrary time, with the invariant holding. We now show that if, at Line LOOPING, S' is extendible, then $S' \cup \{\omega'\}$ is extendible. This is clearly true if $\omega' \in S'$, so we assume that $\omega' \notin S'$, so that $\Lambda \leftarrow \Lambda \cap S'$ had not been executed, and ω' is the argmax of the originally identified set.

First, suppose $\delta < \frac{1}{6B}\|\mathbf{A} - \mathbf{R}'\|_2^2$. Then, since some B-term extension of $\mathbf{R}'$ has error at most $(1+\epsilon)\delta$, it follows that the largest coefficient has energy at least

$$\begin{aligned}(\|\mathbf{A} - \mathbf{R}'\|_2^2 - (1+\epsilon)\delta)/B &\geq (\|\mathbf{A} - \mathbf{R}'\|_2^2 - 2B\delta)/B \\ &\geq 2\|\mathbf{A} - \mathbf{R}'\|_2^2/(3B),\end{aligned}$$

so appears on our list of candidates. Thus the frequency ω' whose coefficient appears largest by our estimation has energy at least

$$\begin{aligned}&\frac{2\|\mathbf{A} - \mathbf{R}'\|_2^2}{3B} - \frac{2\epsilon}{24j^2B^2}\|\mathbf{A} - \mathbf{R}'\|_2^2 \\ &\quad\geq \frac{2\|\mathbf{A} - \mathbf{R}'\|_2^2}{3B} - \frac{1}{3B}\|\mathbf{A} - \mathbf{R}'\|_2^2 \\ &\quad\geq \frac{\|\mathbf{A} - \mathbf{R}'\|_2^2}{3B} > 2\delta \geq (1+\epsilon)\delta,\end{aligned}$$

for $\epsilon < 1$, so ω' is in any extension with error at most $(1+\epsilon)\delta$.

Now, suppose $\delta \geq \frac{1}{6B}\|\mathbf{A} - \mathbf{R}'\|_2^2$. Let ω^* be the frequency whose coefficient is actually largest, so ω^* can go into any optimal extension. We have

$$\begin{aligned}|c_{\omega'}|^2 &\geq |\widetilde{c}_{\omega'}|^2 - \frac{\epsilon}{24j^2B^2}\|\mathbf{A} - \mathbf{R}'\|_2^2 \\ &\geq |\widetilde{c}_{\omega^*}|^2 - \frac{\epsilon}{24j^2B^2}\|\mathbf{A} - \mathbf{R}'\|_2^2 \\ &\geq |c_{\omega^*}|^2 - \frac{\epsilon}{12j^2B^2}\|\mathbf{A} - \mathbf{R}'\|_2^2 \\ &\geq |c_{\omega^*}|^2 - \frac{\epsilon}{2B}\delta.\end{aligned}$$

Thus, if we choose ω' instead of ω^*, at worst, we displace ω^* by ω' in an optimal extension. It follows that the error of the best extension to $S' \cup \{\omega'\}$ is at most the error of the best extension of S' plus $|c_{\omega^*}|^2 - |c_{\omega'}|^2$, *i.e.*, at most

$$\begin{aligned}\delta\left(1 + \frac{\epsilon|S'|}{2B}\right) + |c_{\omega^*}|^2 - |c_{\omega'}|^2 &\leq \delta\left(1 + \frac{\epsilon|S'|}{2B}\right) + \frac{\epsilon}{2B}\delta \\ &\leq \delta\left(1 + \frac{\epsilon|S' \cup \{j\}|}{2B}\right).\end{aligned}$$

It follows that the invariant is preserved. We now consider the correctness of the output. First we show that, for some iteration $j' < J$, the representation is good enough; next we argue that subsequent iterations cannot worsen the representation by much. Thus, however we exit the loop, the final representation has acceptable error.

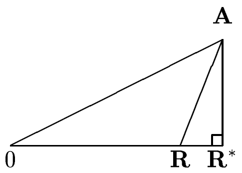
Suppose $\|\mathbf{A} - \mathbf{R}'\|_2^2 > (1 + 7\epsilon/8)\delta$. Then, since an extension has error at most $(1+\epsilon/2)\delta$, it follows that some optimal coefficient c_{ω^*} with $|S' \cup \{\omega^*\}| \leq B$ has energy at least $(3\epsilon/(8B))\|\mathbf{A} - \mathbf{R}'\|_2^2 \geq (\epsilon/(4B))\|\mathbf{A} - \mathbf{R}'\|_2^2$, so is identified. Thus

$$\begin{aligned}|c_{\omega'}|^2 &\geq |\widetilde{c}_{\omega'}|^2 - (\epsilon/(24j^2B^2))\|\mathbf{A} - \mathbf{R}'\|_2^2 \\ &\geq |\widetilde{c}_{\omega^*}|^2 - (\epsilon/(24j^2B^2))\|\mathbf{A} - \mathbf{R}'\|_2^2 \\ &\geq |c_{\omega^*}|^2 - (\epsilon/(12j^2B^2))\|\mathbf{A} - \mathbf{R}'\|_2^2 \\ &\geq (3\epsilon/(8B))\|\mathbf{A} - \mathbf{R}'\|_2^2 - (\epsilon/(12j^2B^2))\|\mathbf{A} - \mathbf{R}'\|_2^2 \\ &\geq (7\epsilon/(24B))\|\mathbf{A} - \mathbf{R}'\|_2^2.\end{aligned}$$

It follows that the new representation, $\mathbf{R}' + \widetilde{c}_{\omega'}\psi_{\omega'}$, has error

$$\begin{aligned}&\|\mathbf{A} - (\mathbf{R}' + \widetilde{c}_{\omega'}\psi_{\omega'})\|_2^2 \\ &\quad= \|\mathbf{A} - \mathbf{R}'\|_2^2 - |c_{\omega'}|^2 + |\widetilde{c}_{\omega'} - c_{\omega'}|^2 \\ &\quad\leq \|\mathbf{A} - \mathbf{R}'\|_2^2 - (7\epsilon/(24B))\|\mathbf{A} - \mathbf{R}'\|_2^2 \\ &\qquad+(\epsilon/(24j^2B^2))\|\mathbf{A} - \mathbf{R}'\|_2^2 \\ &\quad\leq (1 - \epsilon/(4B))\|\mathbf{A} - \mathbf{R}'\|_2^2.\end{aligned}$$

Figure 5: Assessing the quality of R.



Thus, after $O(B \log(M)/\epsilon)$ iterations, the error of $\mathbf{R}'$ has been reduced from $\|\mathbf{A}\|_2^2$ by the factor

$$(1 - \epsilon/(4B))^{O(B \log(M)/\epsilon)} \le 1/M,$$

so we regard $\mathbf{R}'$ as an exact recovery and every subsequent sample will be zero to our specified precision. It follows that, for some $j' < J$, after j' iterations, $\|\mathbf{A} - \mathbf{R}'\|_2^2 \le (1 + 7\epsilon/8)\delta$.

Now consider the representation after the j''th iteration. The list of identified coefficients will still be short, but is otherwise arbitrary. Thus we choose some arbitrary frequency ω', estimate $c_{\omega'}$ up to $(\epsilon/(24j^2B^2))\|\mathbf{A} - \mathbf{R}'\|_2^2$ as $\widetilde{c}_{\omega'}$, and add $\widetilde{c}_{\omega'}\psi_{\omega'}$ to our growing representation. It follows that the new representation has error at most $(1 + \epsilon/(24j^2B^2))\|\mathbf{A} - \mathbf{R}'\|_2^2$. Multiplying over all $j < J$, it follows that the final representation has error at most $(1 + \epsilon)\delta$. $\square$

5. ASSESSING THE ERROR

Unfortunately, a sampling-only algorithm has no way to assess the error $\|\mathbf{A} - \mathbf{R}\|_2^2$ of the output representation $\mathbf{R}$. For example, if $\mathbf{A} = b\delta_s$ is a spike in time at s, a sampling algorithm will miss s altogether, and cannot estimate b, the error of the output representation $\mathbf{R} = 0$ (which is near-best in this case). We now show that, if we happen to know a good approximation to $\|\mathbf{A}\|_2^2$, then we *can* assess the quality, at least roughly.

Lemma 10. *Suppose* $\mathbf{R}$ *is a representation for* $\mathbf{A}$ *such that* $\|\mathbf{A} - \mathbf{R}\|_2^2 \le (1 + \epsilon)\|\mathbf{A} - \mathbf{R}_{\mathrm{opt}}\|_2^2$. *Then* $\|\mathbf{A} - \mathbf{R}\|_2^2 = \|\mathbf{A}\|_2^2 - \|\mathbf{R}\|_2^2 \pm 4\sqrt{\epsilon}\|\mathbf{A}\|_2^2$.

Proof. Let $\mathbf{R}^*$ denote the multiple of $\mathbf{R}$ that minimizes $\|\mathbf{A} - \mathbf{R}^*\|_2^2$. Thus the $\mathbf{A}$-$\mathbf{R}^*$-0 angle is a right angle. (See Figure 5.) The statement follows easily from near-optimality of $\mathbf{R}$ and the fact that $\mathbf{R}^*$ is a B-term representation, using the Pythagorean theorem. We omit the details. $\square$

6. CONCLUSION

We provide a sampling algorithm that yields, with high probability, a B-term Fourier representation $\mathbf{R}$ for any input signal $\mathbf{A}$ of length N, with the guarantee that $\|\mathbf{A} - \mathbf{R}\|_2^2$ is within a factor $(1 + \epsilon)$ of the best possible B-term Fourier representation. The time used is at most $\mathrm{poly}(B \log(N)\|\mathbf{A}\|/\epsilon)$ and these sample positions are independent of the signal. The three main components of this algorithm are the *identification* of significant frequencies, the *estimation* of the Fourier coefficients corresponding to the significant frequencies, and *iteration* on the residual representation.

This algorithm may have applications in several major areas. In mathematics, our algorithm may have implications for Uncertainty Principles, function approximation, and solution to differential equations. In datamining and signal processing, our algorithm can be used to find approximate periods and for feature detection, particularly for speech signals, *e.g.*, to track formants. In computer science, our algorithm has potential uses for approximate convolution and noisy polynomial interpolation.

Acknowledgments

For helpful discussions, we thank: Ingrid Daubechies, Ron DeVore, Sinan Güntürk, and Henry Landau.

7. REFERENCES

[1] D. Achlioptas, F. McSherry: Fast computation of low rank matrix. STOC 2001: 611-618.

[2] N. Alon and Y. Mansour. ϵ-discrepancy sets and their application for interpolation of sparse polynomials. Information Processing Letters, (54)6, 337–342, 1995.

[3] S. Ar, R. Lipton, R. Rubinfeld, and M. Sudan. Reconstructing Algebraic Functions from Mixed Data. SIAM J. Comput. 28(2): 487-510 (1998)

[4] N. Alon, Y. Matias, and M. Szegedy. The Space Complexity of Approximating the Frequency Moments. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing* (STOC), pp. 20–29, May 1996.

[5] D. Donoho and P. Stark. Uncertainty principles and signal recovery. SIAM J. Appl. Math., vol. 49, no. 3, pp. 906–931, June 1989.

[6] A. Frieze, R. Kannan and S. Vempala. Fast Monte-Carlo Algorithms for Finding Low-Rank Approximations. Proc. of the 39th Foundations of Computer Science (FOCS '98), Palo Alto, 1998.

[7] A. C. Gilbert, S. Guha, P. Indyk, Y. Kotidis, S. Muthukrishnan, and M. J. Strauss. Fast, Small-Space Algorithms for Approximate Histogram Maintenance. 2001, these proceedings.

[8] W. B. Johnson and J. Lindenstrauss. 'Extensions of Lipshitz mapping into Hilbert space. *Contemporary Mathematics*, **26** 189–206, 1984.

[9] E. Kaltofen and L. Yagati. Improved Sparse Multivariate Polynomial Interpolation Algorithms. International Symposium on Symbolic and Algebraic Computation. Proc. ISSAC 1988, Springer-Verlag LNCS 358, 1988, pp. 467–474.

[10] E. Kushilevitz, Y. Mansour: Learning Decision Trees Using the Fourier Sprectrum. STOC 1991: 455-464.

[11] N. Linial, Y. Mansour, N. Nisan. Constant Depth Circuits, Fourier Transform, and Learnability. JACM 40(3): 607-620 (1993).

[12] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. IEEE Transactions on Signal Processing, vol. 41, no. 12, pp. 3397–3415, 1993.

[13] Y. Mansour. Learning Boolean Functions via the Fourier Transform. In *Theoretical Advances in Neural Computation and Learning*, (V.P. Roychodhury and K-Y. Siu and A. Orlitsky, ed.), 391–424 (1994).

[14] C. E. Shannon. Communications in the presence of noise. *Proc. of the IRE* (37), 10–21, January 1949.

[15] H. J. Weaver. *Applications of Discrete and Continuous Fourier Analysis.* Wiley, 1983.