

What’s the Frequency, Kenneth?: Sublinear Fourier Sampling Off the Grid

Petros Boufounos¹, Volkan Cevher², Anna C. Gilbert³, Yi Li⁴, and Martin J. Strauss⁵

¹ Mitsubishi Electric Research Labs, 201 Broadway, Cambridge, MA 02139
`petrosb@merl.com*`

² EPFL, Laboratory for Information and Inference Systems, Lausanne, Switzerland
`volkan.cevher@epfl.ch**`

³ Department of Mathematics, University of Michigan, Ann Arbor
`annacg@umich.edu***`

⁴ Department of EECS, University of Michigan, Ann Arbor `leeyi@umich.edu`

⁵ Departments of Mathematics and EECS, University of Michigan, Ann Arbor
`martinjs@umich.edu†`

Abstract. We design a sublinear Fourier sampling algorithm for a case of sparse *off-grid* frequency recovery. These are signals with the form $f(t) = \sum_{j=1}^k a_j e^{i\omega_j t + \hat{\nu}}$, $t \in \mathbb{Z}$; i.e., exponential polynomials with a noise term. The frequencies $\{\omega_j\}$ satisfy $\omega_j \in [\eta, 2\pi - \eta]$ and $\min_{i \neq j} |\omega_i - \omega_j| \geq \eta$ for some $\eta > 0$. We design a sublinear time randomized algorithm, which takes $O(k \log k \log(1/\eta)(\log k + \log(\|a\|_1/\|\nu\|_1)))$ samples of $f(t)$ and runs in time proportional to number of samples, recovering $\{\omega_j\}$ and $\{a_j\}$ such that, with probability $\Omega(1)$, the approximation error satisfies $|\omega'_j - \omega_j| \leq \eta/k$ and $|a_j - a'_j| \leq \|\nu\|_1/k$ for all j with $|a_j| \geq \|\nu\|_1/k$.

1 Introduction

Many natural and man-made signals can be described as having a few degrees of freedom relative to their size due to natural parameterizations or constraints; examples include AM, FM, and other communication signals and per-flow traffic measurements of the Internet. Sparse models capture the inherent structure of such signals via concise linear representations: A signal $y \in \mathbb{R}^N$ has a sparse representation as $y = \Psi x$ in a basis $\Psi \in \mathbb{R}^{N \times N}$ when $k \ll N$ coefficients x can exactly represent the signal y . Sparse models guide the way we acquire signals (e.g., sampling or sketching) and how we efficiently recover them from limited observations (e.g., sublinear recovery algorithms).

* exclusively supported by Mitsubishi Electric Research Laboratories.

** supported by a Rice Faculty Fellowship, MIRG-268398, ERC Future Proof, SNSF 200021-132620, and DARPA KeCoM program #11-DARPA-1055.

*** supported in part by NSF DMS 0354600 and partially supported by DARPA ONR N66001-06-1-2011

† supported in part by NSF DMS 0354600 and NSF DMS 0510203 and partially supported by DARPA ONR N66001-06-1-2011

There has been considerable effort to develop sublinear algorithms within the theoretical computer science community for recovering signals with a few significant discrete Fourier components, beginning with Kushilevitz and Mansour [1], including [2–4], and culminating in the recent work of Hassanieh, et al. [5, 6]. All of these algorithms are predicated upon treating the vector y as periodic and the discrete Fourier transform of a vector x being approximately k -sparse.

Unfortunately, these assumptions are too strong for many practical applications where the discrete Fourier transform coefficients are only approximation of an underlying continuous Fourier transform. For example, if we want to measure the approaching speed (the “doppler”) of an object via the Doppler effect, we transmit a sinusoid wave $e^{i\omega_0 t}$ (where t is time in this example) and receive a sinusoid wave whose frequency offset from ω_0 depends on the unknown doppler, v . Since v can be essentially any continuous value, so can be the received frequency. If there are two or more speeding objects in view, the received signal is of the form $f(t) = a_1 e^{i\omega_1 t} + a_2 e^{i\omega_2 t}$, where ω_1/ω_2 is not necessarily a rational number, so that $f(t)$ is not periodic. This practical and common example does not directly fit the discrete Fourier transform setting of [1–6].

To illustrate why we cannot simply reduce the continuous problem to the previous discrete Fourier sampling techniques, consider $f(t) = a_1 e^{i\omega_1 t} + a_2 e^{i\omega_2 t}$ and simply sample it on N equally-spaced points t . The Discrete Fourier Transform (DFT) of these samples produces a set of N coefficients at corresponding frequencies $2\pi\ell/N$, $\ell = 0, \dots, N-1$, uniformly spaced in the interval $[0, 2\pi]$. It is also possible to compute the oversampled DFT, producing a larger set of coefficients $N' > N$, also corresponding to frequencies $2\pi\ell/N'$, $\ell = 0, \dots, N'-1$, uniformly spaced in $[0, 2\pi]$. In this setting, the existing DFT-based methods often fail to capture the true sparsity of the signal and may blow up the sparsity in an unacceptable fashion. Indeed, even a 1-sparse original signal $f(t) = e^{i\omega_1 t}$ for, say, $\omega_1 = 5.3 \cdot 2\pi/N$, will lead to a discretized signal whose Fourier transform is concentrated around $5 \cdot 2\pi/N$ and $6 \cdot 2\pi/N$, but is significant in $\Omega(N)$ places. This phenomenon arises even with an oversampled DFT, no matter how finely we discretize the frequency grid; *i.e.*, no matter how large N' is.

To this end, our approach lets ω range continuously while keeping t discrete. In Fourier analysis on (locally compact abelian) groups, the variables t and ω must be members of dual groups, which include the pairings $\mathbb{Z}_N \leftrightarrow \mathbb{Z}_N$, $\mathbb{Z} \leftrightarrow \mathbb{S}^1$ (where \mathbb{S}^1 denotes a continuous circle, \mathbb{R}/\mathbb{Z}), and $\mathbb{R} \leftrightarrow \mathbb{R}$. We take $t \in \mathbb{Z}$ and $\omega \in \mathbb{S}^1$. The generalization benefits over $\mathbb{Z}_N \leftrightarrow \mathbb{Z}_N$ are as follows.

- In $\mathbb{Z}_N \leftrightarrow \mathbb{Z}_N$, the data $\{f(n)\}$ are completely specified by N consecutive samples; we can treat double-ended infinite sequences $\{f(n)\}_{n \in \mathbb{Z}}$ provided $\sum_{n \in \mathbb{Z}} |f(n)| < \infty$.
- In $\mathbb{Z}_N \leftrightarrow \mathbb{Z}_N$, the frequencies ω_j must lie on a discrete grid; we can treat frequencies in the continuous space \mathbb{S}^1 .

A concrete application of our approach (studied in the extended version of this paper) is the bearing (or angular direction) estimation of sources transmitting at fixed frequencies, a canonical array signal processing problem with applications

to radar, sonar, and remote sensing. Other applications also include the finite rate of innovation problems [7].

The organization of the paper is as follows. In Sect. 2, we define our model and problem formulation. In Sect. 3, we present our algorithm and its analysis. An application of this result to bearing estimation problems can be found in the full version of this paper.

2 Preliminaries

In this section, we define the problem of sublinear recovery of sparse off-grid frequencies, set the stage notationally, and then detail our results.

2.1 The problem

We define a spectrally sparse function f with *off-grid* frequencies as a function $f : \mathbb{Z} \rightarrow \mathbb{C}$ with k frequencies $\omega_1, \dots, \omega_k \in \mathbb{S}^1$, and we allow for noise ν in the spectrum that is supported on a set $I_\nu \subset \mathbb{S}^1$. We fix a minimum frequency resolution η and assume that $\{[\omega_j - \eta/2, \omega_j + \eta/2]\}_{j=1}^k$ and $[I_\nu - \eta/2, I_\nu + \eta/2]$ are all mutually disjoint. That is, the frequencies are not on a fixed, discrete grid but they are separated from each other and from the noise by a minimum frequency resolution. In our analysis below, we assume that $|\omega_j| > \eta$ without loss of generality. Specifically, we assume f is of the form

$$f(t) = \sum_{j=1}^k a_j e^{i\omega_j t} + \int_{I_\nu} \nu(\omega) e^{i\omega t} d\omega, \quad t \in \mathbb{R},$$

with $\nu \in L^1(I_\nu)$. Without loss of generality, we assume that $a_j \neq 0$ for all j .⁶

Our goal is to find all (a_j, ω_j) with

$$|a_j| \geq \frac{1}{k} \int_{I_\nu} |\nu(\omega)| d\omega \quad (1)$$

making as few samples on \mathbb{Z} as possible (and with the smallest support) from f and for the shortest duration and to produce such a list in time comparable to the number of samples. The number of samples and the size of the support set of the samples should be proportional to a polynomial in k and $\log(1/\eta)$, the number of desired frequencies and precision. We call the frequencies ω_j whose associated amplitude a_j meet the threshold condition (1) *significant*.

⁶ Strictly speaking these functions are not well-defined as, in the current definition, f is not in $L^1(\mathbb{Z})$ and does not have a mathematically well-defined Fourier transform (without resorting to more sophisticated mathematical techniques, such as tempered distributions). To be mathematically correct, we define f as above and then multiply it by a Gaussian window of width η^{100} . Call this mollified function \tilde{f} . The spectrum of \tilde{f} is thus the convolution of \hat{f} with a Gaussian of width η^{-100} . Up to the precision factor η/k , the spectra of \tilde{f} and f are indistinguishable. Henceforth, we consider f with the understanding that \tilde{f} is the well-defined version.

We observe that if we dilate the frequency domain \mathbb{S}^1 by a factor $1/d \in \mathbb{R}$ (i.e., map ω to ω/d), we produce an equivalent sequence of samples $f(t)$, at regularly spaced real-valued points $t = nd$, $n \in \mathbb{Z}$. While the points are indexed by the integers, the values themselves $t = nd$ are in \mathbb{R} . The dilation factor d determines the “rate” at which we sample the underlying signal and the total number of samples times the sampling rate is the duration over which we sample. Both the rate and the total number of samples are resources for our algorithm.

2.2 Notation

Let Ω be a domain (which can be either continuous or discrete). Roughly speaking, we call a function $K : \Omega \rightarrow \mathbb{R}$ a *filter* if K is or approximates the characteristic function χ_E of some set $E \subset \Omega$, which will be called the pass region of K . The resulting signal of applying filter K to signal f (viewed as a function on Ω) is the pointwise product $K \cdot f$.

Let K_m be a kernel defined on \mathbb{S}^1 (identified with $(-\pi, \pi]$) that satisfies the following properties:

- it is continuous on \mathbb{S}^1 ,
- its Fourier transform $\widehat{K}_m : \mathbb{Z} \rightarrow \mathbb{C}$ has finite support: $|\text{supp } \widehat{K}_m| = O(\frac{m}{\alpha} \log \frac{1}{\epsilon})$,
- it approximates $\chi_{[-\frac{\pi}{m}, \frac{\pi}{m}]}$ (so K_m is a filter): $|K_m(x)| \leq \epsilon$ for $|x| \geq \frac{\pi}{m}$, $|K_m(x) - 1| \leq \epsilon$ for $|x| \leq (1 - \alpha)\frac{\pi}{m}$ and $K_m(x) \in [-\epsilon, 1 + \epsilon]$ elsewhere.

A Dolph-Chebyshev filter convolved with the characteristic function of an interval meets these criteria. We call the region $[-(1 - \alpha)\frac{\pi}{m}, (1 - \alpha)\frac{\pi}{m}]$ the plateau of K_m . The pass region of K_m is $[-\frac{\pi}{m}, \frac{\pi}{m}]$ and we define the transition region to be the complement of plateau in the pass region. A similar kernel was used in [5] and [6] with the only difference that their kernel was constructed by a Gaussian kernel convolved with the characteristic function of an interval.

2.3 Main result

Theorem 1. *There is a distribution \mathcal{D} on a set of sampling points $t \in \mathbb{R}$ and an algorithm \mathcal{A} such that for each perturbed exponential polynomial $f(t) = \sum_{j=1}^k a_j e^{i\omega_j t} + \hat{\nu}(t)$, with constant probability, the algorithm returns a list $\Lambda = \{(a'_j, \omega'_j)\}_{j=1}^k$ of coefficients and frequencies such that*

1. For each $|a_j| \geq \|\nu\|_1/k$ there exists $\omega'_j \in \Lambda$ such that

$$|\omega_j - \omega'_j| \leq \frac{\eta}{k}.$$

2. Let $\Lambda_0 = \left\{ \omega'_j \in \Lambda : \exists \omega_{j_0} \text{ such that } |\omega_{j_0} - \omega'_j| \leq \frac{\eta}{k} \text{ and } |a_{j_0}| \geq \frac{\|\nu\|_1}{k} \right\}$, then for each $\omega'_j \in \Lambda_0$ it holds that

$$|a'_j - a_j| \leq \frac{\|\nu\|_1}{k}.$$

3. For each $\omega'_j \in \Lambda \setminus \Lambda_0$, it holds that

$$|a'_j| \leq \frac{\|\nu\|_1}{k}.$$

The algorithm takes $O(k \log k \log(1/\eta)(\log k + \log(\|a\|_1/\|\nu\|_1)))$ samples and runs in time proportional to number of samples. Furthermore, the size of the support of \mathcal{D} , i.e., the total duration of sampling, is $O(k/\eta(\log k + \log(\|a\|_1/\|\nu\|_1)))$.

3 Analysis

Almost all sublinear sparse recovery algorithms (including both the Fourier and canonical basis) randomly hash frequencies or vector elements into buckets. Since the representation of the vector is sparse (in either the Fourier or the canonical basis), it is likely that each bucket contains exactly one coefficient and small noise so that the position of the “heavy hitter” can be found and then its value estimated. At a high level, our algorithm also follows this recipe. Some of these sublinear algorithms are iterative (i.e., non-adaptive hashing and estimation of the difference between the original vector and significant frequencies found in previous iterations) to use fewer samples or measurements or to refine inaccurate estimates. In contrast, our algorithm is not iterative. We hash the range of the frequencies into buckets and repeat sufficiently many times so that all frequencies are isolated, then we locate the frequency and estimate its amplitude.

A main difference between the discrete and continuous Fourier sampling problems is that, in the continuous frequency setting, it is impossible to recover a frequency exactly (from finite samples) so that one can subtract off recovered signals at exact positions. Typically in the discrete setting, an iterative algorithm uses a loop invariant either as in [8, 6] or in [3]. In the former case [8, 6], the number of buckets decreases per round as the number of remaining heavy hitters decreases. In the continuous case, however, the accuracy of the frequency estimates produced by location procedure are dependent on the width the pass region of the filter: the wider the pass region is, the more inaccurate the frequency estimate is. Unless the algorithm not only estimates the coefficient at a given frequency but also improves the frequency estimate, we must increase the distance d between samples from $O(k/\eta)$ to $O(k^2/\eta)$ to achieve the same accuracy for the final frequency estimate, i.e., we must increase the duration over which samples are collected.

In the latter case [3], the number of buckets is kept the same at each round while the energy of the residual signal drops, and there are typically $\log \|a\|$ rounds. In hashing, we need to bound the inaccuracy $|K(h(\omega)) - K(h(\omega'))|$, where ω' is the recovered estimate of some real frequency ω , h the hash function and K the kernel. We can achieve this with a kernel that does not have a significant portion of its total energy outside of its pass region (i.e., a “non-leaking” kernel), but it is not obvious how to achieve such an accurate estimate using a Dirichlet or Fejér kernel which was used in [3]. Unfortunately, using a “non-leaking” kernel like the one used in [5, 6] or the one used in this paper introduces a factor $\log \|a\|$ into the number of samples in order to decrease the noise in a bucket.

3.1 Recovery algorithm

See Algorithm 1 for detailed pseudo-code.

3.2 Analysis of algorithm

In this subsection, we provide a modular characterization of the algorithm.

Isolation. This portion of the analysis is similar to that of [6] but we emphasize the continuous frequency setting.

Let K_m be the kernel as described in Sec. 2 and set $D = 2\pi/\eta$. Define

$$\mathcal{H} = \{K_m(\omega d) = h_d(\omega) | d \in [D, 2D]\}$$

to be a family of hash functions. We choose h_d randomly from \mathcal{H} by drawing d from the interval $[D, 2D]$ uniformly at random. Observe that the map $\omega \mapsto \omega d$ is a random dilation of \mathbb{S}^1 . Similar to [6] and [3], we shall consider m -translations of K_m , denoted by $\{K_m^{(j)}\}_{j=0}^{m-1}$, where $K_m^{(j)}(x) = K_m(x + \frac{2\pi j}{m})$ ($x \in \mathbb{S}^1$), so that their pass regions cover \mathbb{S}^1 . The pass regions will be referred to as *buckets* and the pass region of $K_m^{(j)}$ as j -th bucket. For convenience we shall also call the plateau of $K_m^{(j)}$ the plateau of the j -th bucket. It is clear that each frequency ω , under the random dilation $\omega \mapsto \omega d$, will land in some bucket with index $b(\omega, d)$. Similar to the hashing in [6], our hashing scheme guarantees that

- (small collision) Suppose that $|\omega - \omega'| \geq \eta$ then $\Pr\{b(\omega, d) = b(\omega', d)\} \leq c/m$ for some absolute constant $c > 0$.
- (good plateau landing) Suppose that $\omega \geq \eta$ and let $0 < \alpha < 1/2$ be as given in the definition of K_m , then ω lands in the plateau of the bucket with index $b(\omega, d)$ with probability $\geq (1 - \alpha)(1 - 1/m)$.

If a bucket contains exactly one frequency ω_{j_0} , we say that ω_{j_0} is isolated. Furthermore, if ω_{j_0} lands in the plateau of the bucket, we say that ω_{j_0} is well-isolated. Notice that when ω_{j_0} is isolated, it holds that $|h_d(\omega_j)| \leq \epsilon$ for all $j \neq j_0$.

The next lemma, an imitation of Lemma 3.1 in [3], allows us to bound the inaccuracy of its estimate in terms of the noise $\|\nu\|_1$.

Lemma 1. *Suppose that ξ is a random variable on $[D, 2D]$ such that $|\xi| \leq \pi/m$. Let $\omega \geq \eta$. Then $\mathbb{E}_d[|K_m(\omega d + \xi)|] \leq c/m$ for some absolute constant $c > 0$.*

Now we are ready to show that our algorithm isolates frequencies.

Fix j_0 and choose $m = \Omega(k)$. The hashing guarantees that ω_{j_0} is well-isolated with probability $\Omega(1)$ by taking a union bound. Also, it follows immediately from Lemma 1 that the expected contribution of ν to the bucket is at most $c\|\nu\|_1/m$. Therefore we conclude by Markov's inequality that

Lemma 2. *Conditioned on ω_{j_0} being well-isolated under $h_d \in \mathcal{H}$, w.p. $\Omega(1)$,*

$$\left| \sum_{j \neq j_0} a_j h_d(\omega_j) + \int_{I_\nu} \nu(\omega) h_d(\omega) d\omega \right| \leq C_1 \epsilon \|a\|_1 + \frac{C_2}{m} \|\nu\|_1$$

for some constants C_1, C_2 that depend on the failure probability.

Algorithm 1 The overall recovery algorithm

```

1: function MAIN
2:    $y \leftarrow$  signal samples
3:    $L \leftarrow$  IDENTIFY( $y$ )
4:    $\Lambda \leftarrow$  ESTIMATE( $L$ )
5:   return  $\sum_{\omega \in \Lambda} a_\omega e^{i\omega t}$ 
6: end function

1: function IDENTIFY( $y$ )
2:    $L \leftarrow \emptyset$ 
3:   for  $t \leftarrow 1$  to  $\Theta(\log m)$  do
4:     Choose a random  $d$  as described
5:     Collect  $z_{st}$ , the sample taking at time point with index  $(s, t)$ 
6:      $b_i \leftarrow 0$  for all  $i = 0, \dots, m-1$ 
7:     for  $r \leftarrow 1$  to  $\lceil \log_2(1/\eta) \rceil$  do
8:       Compute  $\{u_\ell\}_{\ell=0}^{m-1}$  and  $\{v_\ell\}_{\ell=0}^{m-1}$  according to Remark 1
          where  $u_\ell = \sum_j a_j K_m(\omega_j d - \frac{2\pi\ell}{m}) K_n\left(\frac{\omega_j d}{2^r} - \frac{2\pi}{2^r m} \ell - \frac{2b_\ell \pi}{2^r}\right)$ 
          and  $v_\ell = \sum_j a_j K_m(\omega_j d - \frac{2\pi\ell}{m}) K_n\left(\frac{\omega_j d}{2^r} - \frac{2\pi}{2^r m} \ell - \frac{2b_\ell \pi}{2^r} - \pi\right)$ 
9:       for  $\ell \leftarrow 0$  to  $m-1$  do
10:        if  $|v_\ell| > |u_\ell|$  then
11:           $b_r \leftarrow b_r + 2^{r-1}$ 
12:        end if
13:      end for
14:    end for
15:    for  $\ell \leftarrow 0$  to  $m-1$  do
16:       $L \leftarrow L \cup \{\frac{2\pi\ell}{md} + \frac{2b_\ell \pi}{d}\}$ 
17:    end for
18:  end for
19:  return  $L$ 
20: end function

1: function ESTIMATE( $L$ )
2:   Choose hash families  $\mathcal{H}_1$  and  $\mathcal{H}_2$  as described.
3:   for  $r \leftarrow 1$  to  $\Theta(\log k)$  do
4:     for each  $\omega \in L$  do
5:        $a_\omega^{(r)} \leftarrow$  measurement w.r.t.  $\mathcal{H}_1$ 
6:        $b_\omega^{(r)} \leftarrow$  measurement w.r.t.  $\mathcal{H}_2$ 
7:     end for
8:   end for
9:   for each  $\omega \in L$  do
10:     $a_\omega \leftarrow$  median $_t a_\omega^{(r)}$ 
11:     $b_\omega \leftarrow$  median $_t b_\omega^{(r)}$ 
12:   end for
13:    $L' \leftarrow \{x \in L : |b_x| \geq |a_x|/2\}$ .
14:    $\Lambda \leftarrow \{(\omega, a_\omega) : \omega \in L'\}$ .
15:   Cluster  $\Lambda = \{(\omega, a_\omega)\}$  by  $x$  and retain only one element in the cluster.
16:   Retain top  $k$  ones (w.r.t.  $a_\omega$ ) in  $\Lambda$ 
17:   return  $\Lambda$ 
18: end function

```

Bit Testing. The isolation procedure above reduces the problem to the following: The parameter d is known, and exactly one of $\{\omega_j d\}_{j=1}^k$, say $\omega_{j_0} d$, belongs to $\bigcup_{n=0}^{N-1} [2n\pi - \delta, 2n\pi + \delta]$ for some small δ and (large) N . Suppose that $\omega_{j_0} d \in [2s\pi - \delta, 2s\pi + \delta]$. We shall find s and thus recover ω_{j_0} . Assume that ω_{j_0} is significant, i.e., a_{j_0} satisfies (1).

We recover s from the least significant bit to the most significant bit, as in [3]. Assume we have already recovered the lowest r bits of s , and by translation, the lowest r bits of s are 0s. We shall now find the $(r+1)$ -st lowest bit.

Let K_n (n is a constant, possibly $n=3$) be another kernel with parameter ϵ' . The following lemma shows that Line 6–14 of IDENTIFY gives the correct s .

Lemma 3. *Suppose that the lowest r bits of s are 0, let $G_1 = K_m(x)K_n\left(\frac{x}{2^r}\right)$, $G_2 = K_m(x)K_n\left(\frac{x}{2^r} - \pi\right)$ and u be the sample taken using G_1 and v using G_2 , then $|u| > |v|$ if $s \equiv 0 \pmod{2^r}$ and $|u| < |v|$ if $s \equiv 2^r \pmod{2^{r+1}}$, provided that $m = \Omega(k)$ and $\epsilon \leq \|\nu\|_1 / (m\|a\|_1)$.*

Proof. We leverage the isolation discussion. By Lemma 2, when $s \equiv 0 \pmod{2^r}$,

$$|u| \geq (1 - \epsilon)(1 - \epsilon')|a_{j_0}| - (1 + \epsilon') \left(C_1 \epsilon \|a\|_1 - \frac{C_2}{m} \|\nu\|_1 \right). \quad (2)$$

and when $s \equiv 2^{r-1} \pmod{2^r}$,

$$|u| \leq (1 + \epsilon)\epsilon'|a_{j_0}| + (1 + \epsilon') \left(C_1 \epsilon \|a\|_1 + \frac{C_2}{m} \|\nu\|_1 \right). \quad (3)$$

Similar bounds hold for $|v|$. Thus it suffices to choose $m \geq \frac{2(1+\epsilon')(C_1+C_2)}{1-\epsilon-2\epsilon'} k$. \square

Repeat this process until $r = \log_2(\pi D) = O(\log(\pi/\eta))$ to recover all bits of s . At each iteration step the number of samples needed is $O(|\text{supp } \widehat{G}_1| + |\text{supp } \widehat{G}_2|) = O(|\text{supp } \widehat{K}_m| \cdot |\text{supp } \widehat{K}_n|) = O(k \log \frac{1}{\epsilon})$, so the total number of samples used in a single execution of Line 8 of IDENTIFY is $O(k \log \frac{1}{\epsilon} \log \frac{1}{\eta})$.

The precision of $\omega_{j_0} d$ will be $\delta = \pi/m$ and thus the precision of ω_{j_0} will be $\delta/d \leq \pi/(mD) = \eta/m$. In summary, the hashing process guarantees that

Lemma 4. *With probability $\Omega(1)$, IDENTIFY returns a list L such that for each ω_j with a_j satisfying (1), there exists $\omega' \in L$ such that $|\omega' - \omega_j| \leq \eta/m$.*

Remark 1. Notice that $\sigma(K_m) \subseteq [-M, M] \cap \mathbb{Z}$ for integer $M = O(\frac{k}{\alpha} \log \frac{1}{\epsilon})$. We shall show that, similar to [3], despite Line 6–14 of IDENTIFY (for m translations altogether) requires mr numbers, each of which is a sum of $O(M)$ terms, this process can be done in $O((M + m \log m)r)$ time instead of $O(Mmr)$ time.

Suppose that at step r , the translation that shifts the lowest bits of s_j to 0 is b_j ($0 \leq j \leq m-1$). In Line 8 of IDENTIFY, each u_j or v_j has the form

$$\sum_{s=-\Theta(n)}^{\Theta(n)} e^{-2\pi i(b_j + \frac{j}{m}) \frac{s}{2^r}} \sum_{t=-M}^M e^{-2\pi i \frac{jt}{m}} w_{st} z_{st}, \quad j = 0, \dots, m-1,$$

where z_{st} is the sample at time with index (s, t) and the associated weight is w_{st} . Notice that the inner sum can be rewritten as

$$\sum_{\ell=0}^m e^{-2\pi i \frac{j\ell}{m}} \sum_{t \in (m\mathbb{Z} + \{\ell\}) \cap [-M, M]} w_{st} z_{st},$$

which can be done in $O(M + m \log m)$ time using FFT. The outer sum has only constantly many terms. Hence Line 8 of IDENTIFY takes $O(M + m \log m)$ times. There are r steps, so the total time complexity is $O((M + m \log m)r)$.

Amplitude Estimation. The isolation procedure generates a list L of candidate frequencies. Like [6], we estimate the amplitude at each position in L by hasing it into buckets using the same kernel but with possibly different parameters. We shall show how to extract good estimates and eliminate unreliable estimates among $|L|$ estimates.

The following lemma states that if a frequency candidate is near a true frequency then they fall in the same bucket with a good probability and if a frequency candidate is adequately away from a true frequency then they fall in different buckets with a good probability.

Lemma 5. *Let $D = \Theta(1/\eta)$ and $\delta > 0$. Choose d uniformly at random from $[\theta_1 D, \theta_2 D]$.*

1. *if $|\omega - \omega'| \leq \beta_1 \delta / D \leq \eta$ then $\Pr \{b(\omega', d) = b(\omega, d)\} \geq 1 - \beta_1 \theta_2$. Thus except with probability $\leq \beta_1 \theta_2 + \alpha$ it holds that ω falls in the same bucket as ω' ;*
2. *if $|\omega - \omega'| \geq \beta_2 \delta / D$ then $\Pr \{b(\omega', d) = b(\omega, d)\} \leq 1/(\beta_2(\theta_2 - \theta_1)) + c\delta D$ for some universal constant $c > 0$.*

Choose parameters $0 < \beta_1 < \beta_2$, $0 < \theta_1 < \theta_2$ such that $\beta_1 \theta_2 + \alpha < 1/3$ and $1/(\beta_2(\theta_2 - \theta_1)) < 1/3$. Let $D = C\pi/\eta$. Define a hash family

$$\mathcal{H} = \{K_m(\omega d) = h_d(\omega) | d \in [\theta_1 D, \theta_2 D]\}.$$

As a direct corollary of Lemma 5 we have

Lemma 6. *Let $\omega' \geq \eta$ and $j_0 = \arg \min_j |\omega' - \omega_j|$. Obtain a measurement $a_{\omega'}$ w.r.t. $h_d \in \mathcal{H}$.*

1. *If $|\omega' - \omega_{j_0}| \leq \beta_1 C\eta/m$, with probability $\Omega(1)$, it holds that $|a_{\omega'} - a_{j_0}| \leq \epsilon \|a\|_1 + c' \|\nu\|_1 / m$ for some $c' > 0$ dependent on the failure probability;*
2. *If $|\omega' - \omega_{j_0}| \geq \beta_2 C\eta/m$, with probability $\Omega(1)$, it holds that $|a_{\omega'}| \leq \epsilon \|a\|_1 + c' \|\nu\|_1 / m$ for some $c' > 0$ dependent on the failure probability.*

Let $\Delta = \epsilon \|a\|_1 + c' \|\nu\|_1 / m$, where c' is a constant dependent on the failure probability guaranteed in the lemma.

Take different $C_1 > C_2$ (and thus different D_1 and D_2) such that $\beta_1 C_2 \geq 1$ and that $C_2 \beta_2 \leq C_1 \beta_1$. Define hash families \mathcal{H}_i ($i = 1, 2$) as

$$\mathcal{H}_i = \{K_m(\omega d) = h_d(\omega) | d \in [\theta_1 D_i, \theta_2 D_i]\}, \quad i = 1, 2.$$

It then follows that

Lemma 7. *Upon termination of execution of line 13 in ESTIMATE, with probability $\Omega(1)$, for each $\omega' \in L'$ let $j_0 = \arg \min_j |\omega' - \omega_j|$ it holds that*

1. *If $|\omega' - \omega_{j_0}| \leq \beta_1 C_1 \eta / m$, then $|a_{\omega'} - a_{j_0}| \leq \Delta$;*
2. *If $|\omega' - \omega_{j_0}| \geq \beta_2 C_1 \eta / m$, then $|a_{\omega'}| \leq \Delta$*
3. *If $\beta_1 C_1 \eta / m \leq |\omega' - \omega_{j_0}| \leq \beta_2 C_1 \eta / m$, then $|a_{\omega'}| \leq 2\Delta$.*

Loosely speaking, Lemma 7 guarantees a multiplicative gap between the amplitude estimates for the “good” estimates of significant frequencies and the amplitudes estimates for all other frequency estimates. Next, we merge estimates of the same true source utilizing the gap as follows. In increasing order, for each $\omega' \in L'$ with amplitude estimate $a_{\omega'}$, find

$$I(\omega') = \left\{ \omega \in L' : \omega' \leq \omega \leq \omega' + \frac{C_1 \beta_1 \eta}{m} \text{ and } \frac{2}{\gamma - 1} |a_{\omega'}| < |a_\omega| < \frac{\gamma - 1}{2} |a_{\omega'}| \right\},$$

where $\gamma > 3$ is a constant to be determined later.

Choose an arbitrary element from I as the representative of all elements in I and add it to Λ . Continue this process from the next $\omega' \in L'$ that is larger than all elements in I . Retain the top k items of Λ .

Lemma 8. *Suppose that ESTIMATE is called with argument L . With probability $\Omega(1)$, it produces a list Λ such that*

1. *For each j with $|a_j| \geq \gamma \Delta$ for some $\gamma > 2 + \sqrt{5}$, if there exists $\omega' \in L$ such that $|\omega' - \omega_j| \leq \pi / m$, then there exists $(\omega'', a_{\omega''}) \in \Lambda$ (we say that $\omega'' \in \Lambda$ is paired) such that $|\omega'' - \omega_j| \leq C_1 \beta_1 \eta / m$ and $|a_{\omega''} - a_j| \leq \Delta$.*
2. *For each unpaired $\omega \in \Lambda$ it holds that $|a_\omega| \leq 2\Delta$.*

Proof. In case (1), for all $\omega \in L'$ such that $|\omega - \omega_j| \leq C_1 \beta_1 \eta / m$ it holds that $|a_\omega| \geq (\gamma - 1)\Delta$ while for other ω it holds that $|a_\omega| \leq 2\Delta$. There is a multiplicative gap so the merging process does not mix frequencies that are close to and far away from a true source. It is easy to verify that $\omega \in L'$ upon termination of line 13 since $C_2 \beta_1 \geq 1$. The rest is obvious. \square

Our main result is now ready.

Proof (of Theorem 1). We show that MAIN returns the claimed result with probability $\Omega(1)$. Choose ϵ in the estimation procedure to be $\epsilon = \|\nu\|_1 / (2\gamma k \|a\|_1)$ and $m \geq \gamma c' k$, then $\Delta \leq \|\nu\|_1 / (\gamma k)$ and thus whenever $|a_j|$ satisfies (1) it holds that $|a_j| \geq \gamma \Delta$. Combining Lemma 4 and Lemma 8 completes the proof. \square

Number of Samples. There are $O(\log k)$ repetitions in isolation and each takes $O(k \log \frac{1}{\epsilon} \log \frac{1}{\eta})$ samples, hence the isolation procedure takes $O(k \log k \log \frac{1}{\epsilon} \log \frac{1}{\eta})$ samples in total.

The input of ESTIMATE is a list L of size $|L| = O(m \log m) = O(k \log k)$. Use the same trick as in isolation, it takes $O(M) = O(k \log(1/\epsilon))$ samples for each of $O(\log k)$ repetitions. Hence the estimation takes $O(k \log k \log \frac{1}{\epsilon} \log \frac{1}{\eta})$ samples.

The total number of samples is therefore

$$O\left(k \log k \log \frac{1}{\epsilon} \log \frac{1}{\eta}\right) = O\left(k \log k \left(\log \frac{\|a\|_1}{\|\nu\|_1} + \log k\right) \log \frac{1}{\eta}\right).$$

Run time. It follows from Remark 1 that each isolation repetition takes $O((M + m \log m)r) = O(k \log \frac{k}{\epsilon} \log \frac{1}{\eta})$ time. There are $O(\log m) = O(\log k)$ repetitions so the total time for isolation is $O(k \log k \log \frac{k}{\epsilon} \log \frac{1}{\eta})$.

The input of ESTIMATE is a list L of size $|L| = O(k \log k)$. Use the same trick as in isolation, it takes $O(M + m \log m + |L|)$ to obtain values for all buckets and compute $a_\omega^{(s)}$ and $b_\omega^{(s)}$ for all $\omega \in L$ and each s . Hence line 3–8 of ESTIMATE takes time $O((M + m \log m + |L|) \log k) = O(k \log k \log(k/\epsilon))$ time. Thus estimation takes time $O(k \log k \log(k/\epsilon)) + |L| \log k + |L| \log |L| = O(k \log k \log(k/\epsilon))$.

The total running time is dominated by that of isolation, which is proportional to the number of samples taken.

Output Evaluation Metric. Since we do not expect to recover the frequencies exactly, the typical approximation error of the form

$$\left\| \sum_j a_j e^{i\omega_j t} - a'_j e^{i\omega'_j t} + \nu(t) \right\|_p$$

contains both the amplitude approximation error $\|a - a'\|$ and a term of the form $\sum |a_j| |\omega_j - \omega'_j|$, rather than the more usual bound in terms of the noise alone $\|\nu\|_p$ in the discrete case. Given bounds on both the amplitudes $|a_j - a'_j|$ and the frequencies $|\omega_j - \omega'_j|$, it is possible to compute the two terms in the error. This is standard in the literature of polynomial-time algorithms to recover real frequencies (e.g., [9], with which our result is comparable).

4 Conclusion

In this paper, we define a mathematically rigorous and practical signal model for sampling sparse Fourier signals with continuously placed frequencies and devise a sublinear time algorithm for recovering such signals. There are a number of technical difficulties in this model with directly applying the discrete sublinear Fourier sampling techniques, both algorithmic and mathematical. In particular, several direct techniques incur the penalty of extra measurements. We do not know if these additional measurements are necessary, if they are inherent in the model. Furthermore, unlike the discrete case, the “duration” of the sampling or the extent of the samples is a resource for which we have no lower bounds.

Acknowledgements. The authors would like to thank an anonymous reviewer for a suggestion that improves the running time.

References

1. Kushilevitz, E., Mansour, Y.: Learning decision trees using the Fourier spectrum. In: STOC. (1991) 455–464

2. Gilbert, A., Guha, S., Indyk, P., Muthukrishnan, M., Strauss, M.: Near-optimal sparse fourier representations via sampling. In: STOC. (2002) 152–161
3. Gilbert, A.C., Muthukrishnan, S., Strauss, M.: Improved time bounds for near-optimal sparse Fourier representations. In: Proceedings of Wavelets XI conference. (2005)
4. Iwen, M.: Combinatorial sublinear-time Fourier algorithms. Foundations of Computational Mathematics **10**(3) (2009) 303–338
5. Hassanieh, H., Indyk, P., Katabi, D., Price, E.: Simple and practical algorithm for sparse Fourier transform. In: SODA. (2012) 1183–1194
6. Hassanieh, H., Indyk, P., Katabi, D., Price, E.: Nearly optimal sparse Fourier transform. In: STOC. (2012) 563–578
7. Vetterli, M., Marziliano, P., Blu, T.: Sampling signals with finite rate of innovation. Signal Processing, IEEE Transactions on **50**(6) (2002) 1417–1428
8. Gilbert, A., Li, Y., Porat, E., Strauss, M.: Approximate sparse recovery: Optimizing time and measurements. SIAM J. Comput. **41**(2) (2012) 436–453
9. Peter, T., Potts, D., Tasche, M.: Nonlinear approximation by sums of exponentials and translates. SIAM J. Sci. Comput. **33**(4) (2011) 1920–1947