Algorithmica
DOI 10.1007/s00453-014-9918-0

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

Petros Boufounos · Volkan Cevher · Anna C. Gilbert · Yi Li · Martin J. Strauss

Received: 7 December 2012 / Accepted: 4 July 2014

Abstract We design a sublinear Fourier sampling algorithm for a case of sparse *off-grid* frequency recovery. These are signals with the form $f(t) = \sum_{j=1}^{k} a_j e^{i\omega_j t} + \int \nu(\omega) e^{i\omega t} d\mu(\omega)$; 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, for any $\epsilon \in (0, \eta/k]$, which takes $O(k \log k \log(1/\epsilon)(\log k + \log(\|a\|_1/\|\nu\|_1))$ samples of $f(t)$ and runs in time

A preliminary version of this paper appeared in the *Proceedings of RANDOM/APPROX 2012*, *LNCS* 7408, pp. 61–72.

P. Boufounos
Mitsubishi Electric Research Labs, 201 Broadway, Cambridge, MA 02139, USA
e-mail: petrosb@merl.com

V. Cevher
Laboratory for Information and Inference Systems, EPFL, Lausanne, Switzerland
e-mail: volkan.cevher@epfl.ch

A. C. Gilbert
Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA
e-mail: annacg@umich.edu

Y. Li (✉)
Department 1, Max-Planck-Institut für Informatik, Campus E1 4, 66123 Saarbrücken, Germany
e-mail: yli@mpi-inf.mpg.de

M. J. Strauss
Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA

M. J. Strauss
Department of Electrical Engineering and Computer Science, University of Michigan,
Ann Arbor, MI 48109, USA
e-mail: martinjs@umich.edu

Published online: 22 July 2014

proportional to number of samples, recovering $\omega'_j \approx \omega_j$ and $a'_j \approx a_j$ such that, with probability $\Omega(1)$, the approximation error satisfies $|\omega'_j - \omega_j| \leq \epsilon$ and $|a_j - a'_j| \leq \|\nu\|_1/k$ for all j with $|a_j| \geq \|\nu\|_1/k$. We apply our model and algorithm to bearing estimation or source localization and discuss their implications for receiver array processing.

Keywords Sparse signal recovery · Fourier sampling · Sublinear algorithms

Mathematics Subject Classification 94A20 · 68W20

1 Introduction

Many natural and man-made signals can be described as having a few degrees of freedom relative to their size as a result of their natural parameterization or constraints. For example, AM, FM, and other communication signals and per-flow traffic measurements of the Internet are signals with few degrees of freedom. 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).

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 (signals that are sparse in the discrete Fourier basis), beginning with Kushilevitz and Mansour [14], including [6,7,13], and continuing through in the recent work of Hassanieh, et al. [8,9]. 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 [1,4].

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 [6–9,13,14].

A natural discretization approach, which takes a large number of samples at equidistant time points and reduces the signal to this finite discrete signal of samples, is complicated because there are inconveniences such as interpreting the result of the discrete version in the continuous setting and locating a real-valued frequency from a cluster of components in the discretized signal. Therefore, instead of developing a discretization reduction in this paper, we take a more direct approach of extending the existing techniques for the discrete setting, such as isolation by hashing and estimation, to the

continuous setting, despite the fact that this approach introduces an additional $\log k$ factor into sampling duration.

A concrete application of our approach 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 [17].

Paper organization In the rest of present section, we shall define our problem and discuss the difference from the traditional models. At the end of the current section we shall detail our results. In Sect. 2, we define the notations and give a high-level overview of our approach. In Sect. 3, we present our algorithm and its analysis. In Sect. 4 we give an application of this result to bearing estimation problems. In Sect. 5, we conclude with a discussion of a simple discretization approach and compare its consequences with our result.

1.1 Our Model

We define a spectrally sparse function f with *off-grid* frequencies as a function $f : \mathbb{R} \to \mathbb{C}$ with k frequencies $\omega_1, \ldots, \omega_k \in \mathbb{S}^1$ (hereinafter $\mathbb{S}^1$ is identified with $(-\pi, \pi]$ and the arithmetic is to be modular), and we allow for noise ν in the spectrum that is supported on a measurable set $I_\nu \subset \mathbb{S}^1$. We fix a minimum frequency resolution $\eta > 0$ and assume that $\{[\omega_j - \eta/2, \omega_j + \eta/2)\}_{j=1}^k$ and $I_\nu + [-\eta/2, \eta/2)$ are all mutually disjoint on $\mathbb{S}^1$. 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 $|\omega_j| \geq \eta$ without loss of generality.

Formally, we assume that f is of the form

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

where I_ν is a measurable set on $\mathbb{S}^1$ and $\nu \in L^1(I_\nu)$. Without loss of generality, we can assume that $a_j \neq 0$ for all j.

Fix $\epsilon_1, \epsilon_2 \in (0, 1]$. Our goal is to find all (a_j, ω_j) with

$$|a_j| \geq \frac{\epsilon_1}{k} \|a\|_1 + \frac{\epsilon_2}{k} \int_{I_\nu} |\nu(\omega)| d\omega \tag{2}$$

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 (2) *significant*.[1]

[1] To connect this formulation with the notion of 'spectrum', one can view f as a tempered distribution, whose Fourier transform $\hat{f}$ is also a tempered distribution. The spectrum of $\hat{f}$ is the support of $\hat{f}$, which captures $\{\omega_j\}$. An alternative way to define the spectrum of a bounded function can be found in [11].

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.

1.2 Prony's Method and Associated Model Problem

There is a substantial body of work (several hundred years' worth) on a closely related problem, that of fitting a linear combination of exponentially decaying but oscillatory functions to data. The problem and several well-known solution methods are conflated and generally referred to as Prony's method. The signal model is defined as follows: let $k \geq 1$ be an integer, $\omega_j \in (-\infty, 0] + \mathrm{i}[-\pi, \pi)$ for $j = 1, \ldots, k$ be distinct complex numbers, and we assume that the coefficients $a_j \in \mathbb{C} \setminus \{0\}$ have magnitude greater than ϵ, $|a_j| > \epsilon$ for some $0 < \epsilon \ll 1$. The signal is a linear combination of such decaying exponential functions

$$f(t) = \sum_{j=1}^{k} a_j \mathrm{e}^{\omega_j t} \quad \text{for } t \geq 0.$$

In almost all formulations of the problem, the signal is sampled at positions $t_1, \ldots, t_N$, uniformly spaced, with $N \geq 2k$. There are a number of different algorithms for estimating the parameters a_j, k, and ω_j.

We summarize the main approaches:

Traditional Prony's Method:

1. Given the N signal samples, write the signal $f(t_m)$ as a linear combination of the previous k time steps $f(t_{m-1})$, $f(t_{m-2})$, $\ldots$, $f(t_{m-k})$,

$$f(t_m) = b_1 f(t_{m-1}) + b_2 f(t_{m-2}) + \ldots b_k f(t_{m-k}),$$

   and solve a linear prediction problem for the coefficients b_k. Note that the linear prediction system is an over-determined Toeplitz system (assuming one orders the columns accordingly).
2. Solve for the complex exponentials $z_j = \mathrm{e}^{\omega_j}$ as the roots of the characteristic polynomial formed by the linear prediction coefficients.
3. Solve the Vandermonde system generated by powers of the exponentials z_j for the coefficients a_j.

While the accuracy of this algorithm is acceptable in exact arithmetic and with noiseless samples, the algorithm fails to remain accurate with noisy samples or finite precision arithmetic as the last two steps are ill-conditioned.

SVD:

1. Compute the SVD of the Hankel matrix formed as in the traditional method (except with the column order reversed):

$$\begin{bmatrix} f(t_1) & f(t_2) & f(t_3) & \dots & f(t_J) \\ f(t_2) & f(t_3) & f(t_4) & \dots & f(t_{J+1}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f(t_{N-J}) & f(t_{K-J+1}) & f(t_{K-J+2}) & \dots & f(t_N) \end{bmatrix}, \quad J = k+1.$$

   Let that decomposition be of the form $U\Sigma V$. Alternatively, factor the matrix into its QR decomposition and use the R matrix in the next step.
2. Form the truncated version $\widetilde{V}$ of the unitary matrix V by selecting the first k rows only and form the matrix $F = (W_0^*)^\dagger (W_1^*)$, where W_0 is the first k columns of $\widetilde{V}$, W_1 the last k columns of $\widetilde{V}$, and $\dagger$ denotes the pseudo-inverse of the matrix. Solve for the complex exponentials $z_j^{t_2-t_1}$ as the eigenvalues of F.
3. Solve the Vandermonde system generated by powers of the exponentials z_j for the coefficients a_j.

There is a substantial body of literature on the analysis of the stability of perturbed Prony methods, in which the signal samples are perturbed by an unknown but bounded amount [16]. There has also been some recent work [10] on choosing sample points according to a pseudo-random distribution (rather than deterministically and uniformly). The sample points are on random arithmetic progressions, thus effectively hashing large frequency components randomly into different signals, which can then be more easily analyzed. Also, the notion of a matrix pencil (a generalized version of Prony's method) [2,12] has been re-introduced by Demanet and Chiu for a sparse FFT algorithm [3].

Clearly, our signal model is different from that of the traditional Prony's model; we do not assume that the oscillatory pieces of the signal decay exponentially which places fewer restrictions on our sampling scheme. We do not, for instance, assume that samples are drawn for $t \geq 0$. Furthermore, we assume that there is noise present in the signal from the outset, although there are restrictions on this noise (its ℓ_1 norm is bounded and its support is bounded away from the significant spectral components). One major difference between our algorithm and all of those referred to as Prony's method is that we determine the frequencies present one frequency at a time. Furthermore, we do not set up the estimation and identification of the frequencies as a linear system (solved one equation at a time or simultaneously).

1.3 Main Result

Theorem 1 *There exist a probability 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 $f(t)$ as in (1), with probability $\geq 1-\delta$, the algorithm returns a list $\Lambda = \{(a'_j, \omega'_j)\}_{j=1}^k$ of coefficients and frequencies such that*

1. *For each a_j that satisfies (2) there exists $\omega'_j \in \Lambda$ such that*

$$|\omega_j - \omega'_j| \leq \frac{\epsilon_2}{k}\eta.$$

2. *Let $\Lambda_0 = \{\omega'_j \in \Lambda : \exists \omega_{j_0}$ such that $\left|\omega_{j_0} - \omega'_j\right| \leq \frac{\epsilon_2 \eta}{k}$ and $|a_{j_0}|$ satisfies (2)$\}$. Then for each $\omega'_j \in \Lambda_0$ it holds that*

$$|a'_j - a_j| \leq \frac{\epsilon_1}{k}\|a\|_1 + \frac{\epsilon_2}{k}\|\nu\|_1.$$

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

$$|a'_j| \leq \frac{\epsilon_1}{k}\|a\|_1 + \frac{\epsilon_2}{k}\|\nu\|_1.$$

The algorithm takes

$$O\left(\frac{k}{\epsilon_2}\log\frac{k}{\delta}\log\frac{1}{\epsilon_2\eta}\log\frac{k}{\min\{\epsilon_1,\epsilon_2\}}\right)$$

samples and runs in time

$$O\left(\frac{k}{\epsilon_2}\log\frac{k}{\delta}\left(\log\frac{1}{\epsilon_2\eta}\log\frac{k}{\min\{\epsilon_1,\epsilon_2\}} + \log\frac{1}{\delta}\right)\right).$$

Furthermore, the size of the support of $\mathcal{D}$, i.e., the total duration of sampling, is

$$O\left(\frac{k}{\epsilon_2\eta}\log\frac{k}{\min\{\epsilon_1,\epsilon_2\}}\right).$$

This result gives ℓ_∞ bounds on both frequency estimates and coefficient estimates. In comparison with the discrete setting, the form of the frequency estimates is new, while the coefficient estimates are of the same type as in [9]. We also note a distinction in the accuracy or resolution of the output frequencies in any algorithm on a discrete grid and our continuous setting. The signal model on a discrete grid of size N assumes that frequencies are separated by a resolution 1 and the algorithms return frequencies with integer values in $1, \ldots, N$. In other words, there is no distinction between the frequency resolution of the input signal model and the output. Our algorithm, by comparison, assumes that the input signal frequencies are separated by a resolution of η and returns frequency values that are accurate up to a smaller resolution of η/k. We argue in 3.2.1 that this "super-resolution" is a *necessary* consequence of the algorithmic approach.

2 Preliminaries

In this section, we define the notation we use throughout the paper and we review typical sublinear-time approaches for sparse Fourier sampling algorithms.

2.1 Kernels

Let Ω be a domain (which can be either continuous or discrete). Roughly speaking, we call a function $K : \Omega \to \mathbb{R}$ 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 signal resulting from applying filter K to signal f (viewed as a function on Ω) is the pointwise product $K \cdot f$.

Let $K_{m,\epsilon,\alpha}$ (often abbreviated as $K_{m,\epsilon}$ or K_m when there is no ambiguity on the parameters) be a kernel defined on $\mathbb{S}^1$ that satisfies the following properties:

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

A Dolph-Chebyshev filter convolved with the characteristic function of an interval meets these criteria. See Fig. 1 for a plot of K_m. 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 [9] and [8] on a discrete domain with the only difference that their kernel was constructed by a Gaussian kernel convolved with the characteristic function of an interval. In this paper, we use a kernel with a finite Fourier expansion for an efficient algorithm.

2.2 Sampling with Respect to a Kernel

In our algorithm, we shall repeatedly query the input signal to obtain values of the form

$$g_K(x,\theta) = \sum_{j=1}^{k} a_j K(\omega_j x - \theta) + \int_{I_v} v(\omega) K(\omega x - \theta) d\omega \tag{3}$$

for some kernel function K (in particular, for two different kernels in one computation). While we cannot obtain $g(x,\theta)$ directly, we can compute it from appropriate samples of $f(t)$. Assume that K has a finite Fourier expansion

$$K(x) = \sum_{n=-N}^{N} \kappa_n e^{inx}.$$

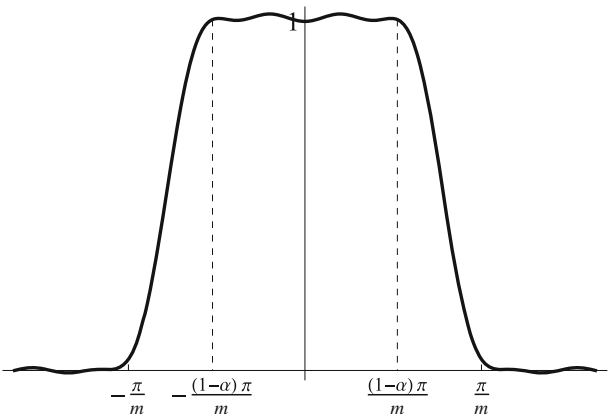

Fig. 1 Plot of K_m with $m = 15$ and $\alpha = 0.476$

Then

$$\begin{aligned} g_K(x) &= \sum_{j=1}^{k} \sum_{n=-N}^{N} \kappa_n a_j e^{in\omega_j x - in\theta} + \int_{I_\nu} \sum_{n=-N}^{N} \kappa_n \nu(\omega) e^{in\omega_j x - in\theta} d\mu \\ &= \sum_{n=-N}^{N} \kappa_n e^{-in\theta} \left(\sum_{j=1}^{k} a_j e^{in\omega_j x} + \int_{I_\nu} \nu(\omega) e^{in\omega_j x} d\mu \right) \\ &= \sum_{n=-N}^{N} e^{-in\theta} \kappa_n f(nx), \end{aligned}$$

which means that $g(x)$ is a weighted sum of samples $\{f(nx)\}_{n=-N}^{N}$, where the weights are the Fourier coefficients of the kernel $K(x)$. Remark 1 explains how, with appropriate signal sample positions, we can compute this quantity efficiently via a short FFT.

2.3 Sublinear Time Algorithms: Overview

The sparse recovery problem under the Fourier basis asks to recover the discrete Fourier transform $\hat{f}$ of some input vector f under the assumption that $\hat{f}$ is sparse, i.e., $\hat{f}$ has very few coordinates of large magnitude. Almost all sublinear-time algorithms perform the the following: (1) randomly hash frequencies into buckets via filters and identify the frequencies whose Fourier coefficients are large in magnitude; (2) estimate the Fourier coefficients of the frequencies identified in the previous step. Since the representation is sparse, it is likely that each bucket contains exactly one coefficient and a small amount of noise so that its position can be found and its value estimated. If not all frequencies with large Fourier coefficients have been recovered, the algorithm will then subtract the partial Fourier representation that has been identified and repeat

the two steps above (called one iteration). Generally, each iteration is guaranteed to recover a substantial part of $\hat{f}$, usually by recovering either a constant fraction of the large coefficients in $\hat{f}$ or a constant fraction of the energy of $\hat{f}$. In either case, the sparsity of $\hat{f}$ keeps improving from one iteration to the next, and a small number of repetitions will suffice for an accurate approximation to $\hat{f}$.

Our algorithm also follows this recipe of hash-and-estimate. However, it 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. We do not need to iterate in the estimation procedure, because we use good kernel for hashing, as in [9].

In fact, for the continuous case, it is difficult to devise an iterative algorithm. A main difference between the discrete and continuous case is that, in the continuous case, it is impossible to recover a frequency exactly (from finite samples) so that one can subtract off recovered signals at exact positions. In the case where the loop invariant is on the number of large coefficients (as in [5,8]), 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, the more inaccurate the frequency estimate is. Unless the estimation procedure not only estimates the coefficient at given frequency but also improves the frequency estimate, we would have to increase the distance d between samples from $O(k/\eta)$ to $O(k^2/\eta)$ in order to the achieve the same accuracy for the final frequency estimate, (i.e., we must increase the duration over which samples are collected.)

In the other case (as in [7]) where 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 [7]. Besides, using a "non-leaking" kernel like the one used in [8,9] or the one used in this paper unfortunately introduces a factor $\log \|a\|$ into the number of samples in order to decrease the noise in a bucket.

3 Algorithm and Analysis

3.1 Recovery Algorithm

We give detailed pseudocode for our algorithm in Algorithm 1. Before we analyze the algorithm, we provide some intuition about the overall architecture. First, this algorithm, unlike many of the other sublinear time signal recovery algorithms, is not an iterative one. It is similar to other sublinear time algorithms in that there are two separate sets of samples, one for identifying the frequencies ω and a second for estimating the coefficients a_ω. In the pseudocode in Algorithm 1, we sample the function as needed in each function for ease of exposition (to avoid a large number of

Algorithm 1 The overall recovery algorithm

function MAIN
  $y \leftarrow$ signal samples from f
  $L \leftarrow$ IDENTIFY(y)
  $\Lambda \leftarrow$ ESTIMATE(L)
  return $\sum_{\omega \in \Lambda} a_\omega e^{i\omega t}$

function IDENTIFY(y)
  $L \leftarrow \emptyset$
  for $t \leftarrow 1$ **to** $\Theta(\log \frac{k}{\delta})$ **do**
    Choose a random d uniformly from $[D, 2D]$
    $b_i \leftarrow 0$ for all $i = 0, \ldots, m-1$
    for $r \leftarrow 1$ **to** $\lceil \log_2(1/\eta) \rceil$ **do**
      Collect samples $z_{p,q} = f(qd + \frac{pd}{2^r})$ for $p = -N, \ldots, N$ and $|q| = -M, \ldots, M$,
        where $M = \max \operatorname{supp} \widehat{K_m}$ and $N = \max \operatorname{supp} \widehat{K_n}$.
      Compute $\{u_\ell\}_{\ell=0}^{m-1}$ and $\{v_\ell\}_{\ell=0}^{m-1}$ using a short FFT (according to Remark 1)
        where $u_\ell = \sum_j a_j K_m\left(\omega_j d - \frac{2\pi\ell}{m}\right) K_n\left(\frac{\omega_j d}{2^r} - \frac{2\pi}{2^r m}\ell - \frac{2b_\ell \pi}{2^r}\right)$
        and $v_\ell = \sum_j a_j K_m\left(\omega_j d - \frac{2\pi\ell}{m}\right) K_n\left(\frac{\omega_j d}{2^r} - \frac{2\pi}{2^r m}\ell - \frac{2b_\ell \pi}{2^r} - \pi\right)$
      for $\ell \leftarrow 0$ **to** $m-1$ **do**
        if $|v_\ell| > |u_\ell|$ **then**
          $b_\ell \leftarrow b_\ell + 2^{r-1}$
    for $\ell \leftarrow 0$ **to** $m-1$ **do**
      $L \leftarrow L \cup \{\frac{2\pi\ell}{md} + \frac{2b_\ell \pi}{d}\}$
  return L

function ESTIMATE(L)
  Choose hash families $\mathcal{H}_1$ and $\mathcal{H}_2$ as described in Equation (6).
  for $r \leftarrow 1$ **to** $\Theta(\log \frac{m}{\delta})$ **do**
    for each $\omega \in L$ **do**
      $a_\omega^{(r)} \leftarrow$ measurement w.r.t. $\mathcal{H}_1$
      $b_\omega^{(r)} \leftarrow$ measurement w.r.t. $\mathcal{H}_2$
  for each $\omega \in L$ **do**
    $a_\omega \leftarrow \text{median}_t\, a_\omega^{(r)}$
    $b_\omega \leftarrow \text{median}_t\, b_\omega^{(r)}$
  $L' \leftarrow \{x \in L : |b_\omega| \geq |a_\omega|/2\}$.
  $\Lambda \leftarrow \{(\omega, a_\omega) : \omega \in L'\}$.
  Cluster $\Lambda = \{(\omega, a_\omega)\}$ by ω and retain only one element in the cluster.
  Retain top k ones (w.r.t. a_ω) in Λ
  return Λ

parameters in the pseudocode) but we emphasize that the algorithm is non-adaptive; we can simply pass in these samples to the IDENTIFY and ESTIMATE functions.

In IDENTIFY, we generate a list L of candidate frequencies by repeatedly sampling from a filtered version of the input signal. The filtering includes both a random hashing into $O(k)$ buckets and a bit-testing procedure in parallel (hence, the two kernels K_m and K_n in Line 8 of the pseudo-code). The bit-testing procedure tests in Line 10 whether the ℓth bit (from least significant to most significant) is 0 or 1. Once we have determined the identify of the frequency, it is added to the list.

Remark 1 describes how to compute an efficient sampling of the input signal with respect to the two kernels K_m and K_n in Line 8 via a short Fourier transform using

equally spaced points instead of a direct multiplication. The method described is similar to that in [7].

In ESTIMATE, we estimate the coefficient of each frequency in the list L by sampling (again) from two different filtered or hashed versions of the input signal (Lines 5–6). We use the two different coefficient estimates from the two filters to detect which estimates are accurate and we retain those and their associated frequencies (Lines 10–11). We then cluster the list of frequencies by the output resolution (Line 12) and prune the list, keeping only the frequencies and coefficients corresponding to the k largest coefficients.

3.2 Analysis of Algorithm

Our analysis of the algorithm follows a modular characterization, starting with showing that the algorithm is correct and satisfies the announced failure probability—it isolates significant frequencies and then identifies those that are isolated, after which the algorithm estimates the coefficients with the desired accuracy and failure probability. We show that the number of samples total is as announced. We end with analyzing the run-time of the algorithm, the duration of the sampling procedure (not only how many samples but what their extent is), and the error metric we use to assess the output.

3.2.1 Isolation

This portion of the analysis is similar to that of [8] in that we use similar kernels for filtering or hashing the signal into buckets but in our analysis, we emphasize the continuous frequency setting.

Let K_m be the kernel as described in Sect. 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 [8] and [7], we shall consider m-translations of K_m, denoted by $\{K_m^{(j)}\}_{j=0}^{m-1}$, where $K_m^{(j)}(x) = K_m\left(x - \frac{2\pi j}{m}\right)$ $(x \in \mathbb{S}^1)$, 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 $b(\omega, d)$.

In the discrete setting, we hash N elements into m buckets so that each bucket contains N/m elements. Here we hash $(-\pi, \pi]$ into m buckets so that each bucket has measure $2\pi/m$. Since there are k frequencies to isolate, we would need to hash them into $\Omega(k)$ buckets so that each frequency is isolated with probability $O(1/k)$ such that we can afford to take a union bound over all k frequencies in the end to conclude that all of them are isolated with constant probability. We shall see later that we pick $m = \Theta(k)$. Furthermore, to separate two frequencies which are η apart, we

have to dilate $\mathbb{S}^1$ by a factor of $\Omega(1/\eta)$, so the collision probability remains $O(1/k)$, small enough for a union bound. This effectively splits a bucket of length $O(1/k)$ into $\Omega(1/\eta)$ pieces, each of length $O(\eta/k)$. Thus we obtain a precision of $O(\eta/k)$ for frequency estimation, at the cost of a sampling duration of $O(k/\eta)$. The sampling duration cannot be shortened following our approach, nor can we reduce the resolution of the frequency estimate to $\Theta(\eta)$.

The idea above will be elaborated though the rest of this section. Similar to [8] and [7], the next lemmata show that this hashing is effective, immitating Claim 3.1, Claim 3.2 of [8] and Lemma 3.1 of [7].

The first lemma tells us that the probability of collision of two well-separated frequencies under a random hash function $h_d \in \mathcal{H}$ is small.

Lemma 1 *Suppose that* $|\omega' - \omega| \geq \eta$. *Then*

$$\Pr\{b(\omega, d) = b(\omega', d)\} \leq \frac{1}{m}\left(2 + \frac{1}{m}\right),$$

where $c > 0$ *is a universal constant.*

Proof Without loss of generality, assume that $\Delta\omega = \omega' - \omega > 0$. Write $\omega d = x + \xi$ with $|\xi| < \pi/m$. Then the probability is equal to

$$\Pr\left\{(\Delta\omega)d \in \left[2s\pi - \xi - \frac{\pi}{m}, 2s\pi - \xi + \frac{\pi}{m}\right]\right\} \leq \frac{1}{D} \cdot \frac{2\pi}{m\Delta\omega} \cdot |I|,$$

where I is the set of possible s's,

$$I = \left\{\left\lfloor \frac{(\Delta\omega)D + \xi - \frac{\pi}{m}}{2\pi} \right\rfloor + 1, \ldots, \left\lfloor \frac{2(\Delta\omega)D + \xi + \frac{\pi}{m}}{2\pi} \right\rfloor\right\}.$$

The desired upper bound of the probability follows immediately from that

$$|I| \leq \frac{(\Delta\omega)D}{2\pi} + 1 + \frac{1}{m}.$$

□

While Lemma 1 guarantees that well-separated frequencies do not collide under our hash function, because we are in the continuous setting, there is some probability that a frequency is hashed into the transition region of the kernel K_m. The following lemma shows that with high probability, a frequency bounded away from zero is mapped to the region of the kernel that is very close to 1.

Lemma 2 *Assume that* $\omega \geq \eta$ *and let* $0 < \alpha < 1/2$ *be as given in the definition of the kernel* K_m. *Then*

$$\Pr\left\{\omega d \in \left[\frac{2s\pi}{m} - (1-\alpha)\frac{\pi}{m}, \frac{2s\pi}{m} + (1-\alpha)\frac{\pi}{m}\right] \text{ for some } s \in \mathbb{Z}\right\}$$
$$\geq (1-\alpha)\left(1 - \frac{1}{m}\right).$$

Proof It is clear that the probability is at least

$$\frac{1}{D} \cdot (1-\alpha)\frac{2\pi}{\omega m} \cdot |I|,$$

where I is the set of possible s's,

$$I = \left\{ \left\lfloor \omega D \frac{m}{2\pi} - \frac{1-\alpha}{2} \right\rfloor + 1, \ldots, \left\lfloor 2\omega D \frac{m}{2\pi} + \frac{1-\alpha}{2} \right\rfloor \right\}.$$

Then

$$|I| \geq \frac{\omega D m}{2\pi} - 1.$$

and the result follows immediately. □

The next lemma will allow us to estimate the coefficient of an isolated frequency and to bound the inaccuracy of its estimate in terms of the noise $\|\nu\|_1$.

Lemma 3 *Suppose that ξ is a random variable such that $|\xi| \leq \pi/m$ almost surely and the parameter ϵ of $K_{m,\epsilon}$ satisfies $\epsilon \leq c/m$ for some constant c. Let $\omega \geq \eta$. Then*

$$\mathbb{E}_d[|K_m(\omega d + \xi)|] \leq \frac{2(c+2)}{m}.$$

Proof Define

$$\tilde{K}_m(x) = \sup_{|y-x| \leq \pi/m} |K_m(y)|,$$

it is not difficult to see that

$$\|\tilde{K}_m\|_1 \leq 2\pi\epsilon + \frac{4\pi}{m} \leq \frac{2\pi C}{m}$$

where $C = c + 2$. Let d be uniformly chosen from some interval I. Then

$$\mathbb{E}_d[|K_m(\omega d + \xi)|] \leq \mathbb{E}_d[|\tilde{K}_m(\omega d)|] = \frac{1}{|I|}\int_I |\tilde{K}_m(\omega t)|dt = \frac{1}{\omega|I|}\int_{\omega I} |\tilde{K}_m(x)|dx.$$

From

$$\int_{\omega I} |\tilde{K}_m(x)|dx \leq \frac{2C\pi}{m}\left\lceil \frac{\omega|I|}{2\pi} \right\rceil$$

it follows that

$$\mathbb{E}_d[\tilde{K}_m(\omega d + \xi)] \leq \frac{C}{m} \cdot \frac{\lceil \frac{\omega|I|}{2\pi} \rceil}{\frac{\omega|I|}{2\pi}}.$$

For $I = [D, 2D]$,

$$\mathbb{E}_d[|\tilde{K}_m(\omega d)|] \le \frac{C}{m}\left(1 + \frac{1}{[\frac{\omega D}{2\pi}]}\right).$$

Let $D = 2\pi/\eta$, since $\omega \ge \eta$, it holds that $[\omega D/(2\pi)] \ge 1$ and the desired result follows. □

Now we are ready to show 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 3 that the expected contribution of ν to the bucket is at most $c\|\nu\|_1/m$ for some constant $c > 0$. Therefore we conclude by Markov's inequality that

Lemma 4 *Conditioned on ω_{j_0} being well-isolated under $h_d \in \mathcal{H}$, w.p. $\Omega(1)$,*

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

for some constants $C_1, C_2 > 0$ that depend on the failure probability.

Proof Note that $h_d(\omega) = K_m((\omega - \omega_{j_0})d + \xi_d)$, where ξ_d is piecewise continuous on $[D, 2D]$. By Lemma 3 and Fubini's Theorem,

$$\begin{aligned}\mathbb{E}\left[\left|\int_{I_\nu} \nu(\omega) h_d(\omega) d\mu\right|\right] &\le \mathbb{E}\left[\int_{I_\nu} |\nu(\omega)|\,|h_d(\omega)| d\mu\right] \\ &= \int_{I_\nu} \nu(\omega)\mathbb{E}\,|h_d(\omega)|\, d\omega \\ &\le \frac{c}{m}\int_{I_\nu} |\nu(\omega)| d\mu,\end{aligned}$$

Since ω_j $(j \ne j_0)$ land in different bucket from ω_{j_0},

$$|h_d(\omega_j)| \le \epsilon, \quad j \ne j_0$$

thus

$$\mathbb{E}\left[\left|\sum_{j \ne j_0} a_j h_d(\omega_j) + \int_{I_\nu} \nu(\omega) h_d(\omega) d\mu\right|\right] \le \epsilon\|a\|_1 + \frac{c}{m}\|\nu\|_1$$

The result follows from Markov's inequality. □

3.2.2 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 $\bigcup_{n=0}^{N-1} [2n\pi - \zeta, 2n\pi + \zeta]$ for some small $\zeta = \pi/m$ and (large) N. Suppose that $\omega_{j_0} d \in [2s\pi - \zeta, 2s\pi + \zeta]$. Note that s is the integer closest to $(\omega_{j_0} d)/(2\pi)$. We shall find this integer s (to be referred to as chunk index), and thus recover ω_{j_0}. Assume that ω_{j_0} is significant; i.e., a_{j_0} satisfies (2).

We recover s from the least significant bit to the most significant bit, as in [7]. 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' (a small constant). The following lemma shows that Line 6–14 of IDENTIFY gives the correct s.

Lemma 5 *Suppose that the lowest r bits of s are* 0*, let*

$$G_1(x) = K_m(x) K_n\left(\frac{x}{2^r}\right), \quad G_2(x) = K_m(x) K_n\left(\frac{x}{2^r} - \pi\right)$$

and u be the sample taken using G_1 and v using G_2. Then $|u| > |v|$ if $s \equiv 0 \pmod{2^{r+1}}$ and $|u| < |v|$ if $s \equiv 2^r \pmod{2^{r+1}}$, provided that

$$m \geq \frac{C}{\epsilon_2} k \quad \text{and} \quad \epsilon \leq C \cdot \min\left\{\frac{\epsilon_1}{k}, \frac{\epsilon_2}{k}\right\}$$

for some $C > 0$.

Proof It is straightforward from the isolation discussion. When $s \equiv 0 \pmod{2^r}$,

$$|u| \geq (1-\epsilon)(1-\epsilon')|a_{j_0}| - (1+\epsilon')\left(C_1 \epsilon \|a\|_1 - \frac{C_2}{m}\|v\|_1\right). \tag{4}$$

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

$$|u| \leq (1+\epsilon)\epsilon'|a_{j_0}| + (1+\epsilon')\left(C_1 \epsilon \|a\|_1 + \frac{C_2}{m}\|v\|_1\right). \tag{5}$$

Similar bounds hold for $|v|$. Thus it suffices to choose

$$m \geq \frac{2(1+\epsilon')C_2}{1-\epsilon-2\epsilon'} \cdot \frac{k}{\epsilon_2}.$$

□

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 $O\left(|\operatorname{supp} \widehat{G_1}| + |\operatorname{supp} \widehat{G_2}|\right) =$

$O\left(|\operatorname{supp}\widehat{K_m}| \cdot |\operatorname{supp}\widehat{K_n}|\right) = O\left(\frac{k}{\epsilon_2}\log\frac{1}{\epsilon}\right)$, so the total number of samples used in a single execution of Line 7 of IDENTIFY is $O(\frac{k}{\epsilon_2}\log\frac{1}{\epsilon}\log\frac{1}{\eta})$.

The precision of $\omega_{j_0}d$ will be $\zeta = \pi/m$, and thus the precision of ω_{j_0} will be $\pi/(md) \le \pi/(mD) = \eta/m$.

In summary, the hashing process guarantees that

Lemma 6 *With probability at least* $1 - O(\delta)$, IDENTIFY *returns a list* L *such that for each* ω_j *with* a_j *satisfying* (2)*, there exists* $\omega' \in L$ *such that* $|\omega' - \omega_j| \le \eta/m$.

Remark 1 Notice that the number of non-zero Fourier coefficients of the filter K_m is $2M + 1$ for some integer $M > 0$. We shall show that, similar to [7], despite Line 8 of IDENTIFY (for m translations altogether) requires mr numbers, each of which is a sum of $2M + 1$ terms, this process can be done in $O((M + m\log m)r)$ time instead of $O(Mmr)$ time.

Suppose that at some step, in the j-th translation bucket ($0 \le j \le m - 1$) lies a hashed frequency with chunk index s_j. Let b_j be the integer such that the lowest bits of $s_j - b_j$ are 0. By Sect. 2.2, we shall take $\Theta(Mn)$ samples, corresponding to the the spectrum of $K_m K_n$, so we index the samples by (p, q) with $|p| \le \Theta(n)$ and $|q| \le M$. We need the numbers

$$\sum_{p=-\Theta(n)}^{\Theta(n)} e^{-2\pi i(b_j+\frac{j}{m})\frac{p}{2^r}} \sum_{q=-M}^{M} e^{-2\pi i\frac{jq}{m}} w_{p,q} z_{p,q}, \quad j = 0, \ldots, m-1,$$

where $z_{p,q}$ is the sample with index (p, q) and $w_{p,q}$ are the Fourier coefficients of $K_m K_n$. Notice that the inner sum can be rewritten as

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

which can be done in $O(M + m\log m)$ time using FFT. The outer sum has only constantly many terms. Hence to compute m numbers, each is a double sum as above, takes only $O(M + m\log m)$ times. There are r steps, so the total time complexity is $O((M + m\log m)r)$.

3.2.3 Coefficient Estimation

The isolation procedure generates a list L of candidate frequencies. Like [8], we estimate the coefficient at each position in L by hashing the frequencies 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 7 *Let $D = \Theta(1/\eta)$ and $\delta > 0$. Choose d uniformly at random from $[\theta_1 D, \theta_2 D]$.*

1. *if $|\omega - \omega'| \le \beta_1 \delta / D \le \eta$ then*

$$\Pr\left\{b(\omega', d) = b(\omega, d)\right\} \ge 1 - \beta_1 \theta_2.$$

2. *if $|\omega - \omega'| \ge \beta_2 \delta / D$ then*

$$\Pr\left\{b(\omega', d) = b(\omega, d)\right\} \le \frac{1}{\beta_2(\theta_2 - \theta_1)} + \frac{c\delta}{D}$$

for some universal constant $c > 0$.

Proof Without loss of generality assume $\omega' > \omega$. Then the probability in case (1) can be rewritten as

$$\sum_{s \in \mathbb{Z}} \Pr\left\{\left[\frac{\omega' d}{2\delta} + \frac{1}{2}\right] = \left[\frac{\omega d}{2\delta} + \frac{1}{2}\right] = s\right\} = \sum_{\text{possible} s} \frac{1}{(\theta_2 - \theta_1) D} \cdot m\left(\left[\frac{2s\delta - \delta}{\omega}, \frac{2s\delta + \delta}{\omega'}\right]\right),$$

where $m(E)$ denotes the Lebesgue measure of set E. Note that

$$\begin{aligned} m\left(\left[\frac{2s\delta - \delta}{\omega}, \frac{2s\delta + \delta}{\omega'}\right]\right) &= \delta \cdot \frac{\omega + \omega' - 2s(\omega' - \omega)}{\omega' \omega} \\ &\ge \delta \cdot \frac{\omega + \omega' - \omega\theta_2\beta_1 - \frac{\beta_1 \delta}{D}}{\omega' \omega} \\ &\ge \delta \cdot \frac{1 - \theta_2 \beta_1}{\omega'} \end{aligned}$$

Thus the sum of probabilities can be bounded from below by

$$\frac{1}{(\theta_2 - \theta_1) D} \cdot \delta \cdot \frac{1 - \theta_2 \beta_1}{\omega'} \left(\left[\frac{\omega' \theta_2 D}{2\delta} + \frac{1}{2}\right] - \left[\frac{\omega' \theta_1 D}{2\delta} + \frac{1}{2}\right] + 1\right) \ge 1 - \theta_2 \beta_1.$$

The other case can be proved similarly as Lemma 1. □

Now consider a fixed significant frequency, say, ω_1. Assume that ω_1 is isolated from other frequencies under hashing $\omega \mapsto \omega d$. If ω_1 lands in the plateau of the kernel K_m, that is, $K_m(\omega_1 d) \approx 1$, then the bucket value

$$\int K_m(xd) f(x) dx \approx a_1 K_m(\omega_1 d) \approx a_1,$$

is a good estimate to the coefficient. Similarly if it lands outside the pass region, the bucket value would be close 0. The only challenging situation is when it lands in the

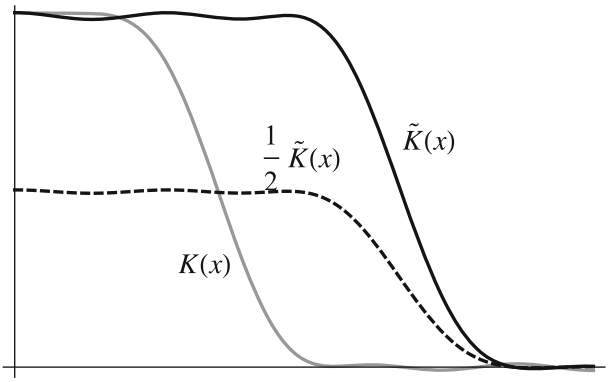

Fig. 2 K and $\tilde{K}$

transition of K_m, in which case we may have a significant bucket value but much smaller than the desired a_1. Our plan is to detect the estimates from the transition region and remove them. To this end, consider two kernels K and $\tilde{K}$ such that the pass region of K falls within the plateau of $\tilde{K}$ (see Fig. 2). Observe that the set $\{x : K(x) \geq \tilde{K}(x)/2\}$ consists of two parts: one part falls completely within the plateau of $\tilde{K}$, the other outside the pass region of K. With respect to the two kernels, we obtain two bucket values

$$a = \int K(xd) f(x) dx \approx a_1 K(\omega_1 d)$$

and

$$a' = \int \tilde{K}(xd) f(x) dx \approx a_1 \tilde{K}(\omega_1 d)$$

If $|a| > |a'|/2$, we know that either $\omega_1 d$ falls in the plateau of $\tilde{K}$, which indicates that a' is a reliable estimate, or $\omega_1 d$ falls outside the pass region of K, which indicates that a is small (and thus is a'). Hence we can drop the estimates with $|a| < |a'|/2$ and retain reliable estimates (either in the plateau of $\tilde{K}$ or outside the pass region of K). We always take a' as our final estimate, which would be either significant or small. Furthermore, the plateau of K is contained in the set $\{x : K(x) \geq \tilde{K}(x)/2\}$, so we will always obtain a good estimate to a_1 if we can guarantee that $\omega_1 d$ will land in the plateau of K at least once.

This idea, together with the frequency estimate guarantee, is formalized in the next few lemmata.

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

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

Let $\omega \in \mathbb{S}^1$ and $h \in \mathcal{H}$. Suppose that $\omega \in [\frac{(2\ell-1)\pi}{m}, \frac{(2\ell+1)\pi}{m}]$ and $h = h_d(\omega)$. We call $a_\omega = g_{K_m}(d, \frac{2\ell\pi}{m})$ (see (3)) a measurement with respect to h. When h is randomly chosen (meaning that d is a random number), we simply say a_ω is a measurement with respect to $\mathcal{H}$.

It then holds that

Lemma 8 *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 $|a_{\omega'} - a_{j_0}| \leq \epsilon\|a\|_1 + c'\|v\|_1/m$ 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 $|a_{\omega'}| \leq \epsilon\|a\|_1 + c'\|v\|_1/m$ for some $c' > 0$ dependent on the failure probability.*

Proof As in hashing, ω' is separated from all other ω_j ($j \neq j_0$) with probability $\Omega(1)$, so ω_j does not land in the bucket for all $j \neq j_0$. It follows from Lemma 7 that

1. If $|\omega' - \omega_{j_0}| \leq \beta_1 C\eta/m$ then ω_{j_0} falls in the plateau of the same bucket as ω' except with probability $\beta_1\theta_2 + \alpha$.
2. If $|\omega' - \omega_{j_0}| \geq \beta_2 C\eta/m$ then ω_{j_0} falls in a different bucket from ω' except with probability $1/(\beta_2(\theta_2 - \theta_1)) + cC/m$.

Upon the success of either case, the noise in the bucket is at most $\epsilon\|a\|_1 + C\|v\|_1/m$ (by the argument in isolation the section) and the conclusion follows. □

Let $\Delta = \epsilon\|a\|_1 + c'\|v\|_1/m$, where c' is a constant dependent on the failure probability guaranteed in the preceding lemma and ϵ satisfies the condition in Lemma 5. Take $C_1 > C_2$ (and thus $D_1 > D_2$) such that $\beta_1 C_2 \geq 1$ and $C_2\beta_2 \leq C_1\beta_1$. Define hash families $\mathcal{H}_i$ ($i = 1, 2$) as

$$\mathcal{H}_i = \{K_m(\omega d) = h_d(\omega) | d \in [\theta_1 D_i, \theta_2 D_i]\}, \quad i = 1, 2. \tag{6}$$

It then follows that

Lemma 9 *Upon termination of execution of line 10 in* ESTIMATE*, with probability $\geq 1 - O(\delta)$, 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 $|a_{\omega'}| \leq 2\Delta$.*

Proof Case (1) and (2) follow from the previous lemma. For ω' in case (3) it holds that $a_{\omega'} \leq 2b_{\omega'}$ and $b_{\omega'} \leq \Delta$ since $|\omega' - \omega_{j_0}| \geq \beta_2 C_2\eta/m$. □

Loosely speaking, Lemma 9 guarantees a multiplicative gap between the coefficient estimates for the "good" estimates of significant frequencies and the coefficient estimates for all other frequency estimates. Next, we merge estimates of the same true source. In increasing order, for each $\omega' \in L'$ with coefficient 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 any element from I as representative of all elements in I and add it to Λ. Continue this process from the next $\omega' \in L$ that is larger than I. Retain the top k items of Λ.

Lemma 10 *Suppose that* ESTIMATE *is called with argument L. With probability $\geq 1 - O(\delta)$, 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*

$$|\omega'' - \omega_j| \leq \frac{C_1\beta_1\eta}{m}, \quad |a_{\omega''} - a_j| \leq \Delta.$$

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 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. □

Now we are ready to prove our main result.

Proof (Proof of Theorem 1) It suffices to show that MAIN returns a desirable result with probability $\geq 1 - \delta$. Choose ϵ in the estimation procedure to be $\epsilon = \min\{\epsilon_1, \epsilon_2\}/(\gamma k)$ and $m \geq \gamma c' k$, then $\Delta \leq \|\nu\|_1/(\gamma k)$ and thus whenever $|a_j|$ satisfies (2) it holds that $|a_j| \geq \gamma\Delta$. Combining Lemmas 6 and 10 shows the correctness of the recovery guarantees.

Next we bound number of samples, runtime and duration of sampling.

Number of Samples. There are $O(\log \frac{k}{\delta})$ repetitions in isolation and each takes $O(\frac{k}{\epsilon_2} \log \frac{1}{\epsilon} \log \frac{1}{\eta})$ samples, hence the isolation procedure takes $O(\frac{k}{\epsilon_2} \log \frac{k}{\delta} \log \frac{1}{\epsilon} \log \frac{1}{\eta})$ samples in total.

The input of ESTIMATE is a list L of size $|L| = O(m \log \frac{k}{\delta})$. Use the same trick as in isolation, it takes $O(M) = O(\frac{k}{\epsilon_2} \log \frac{1}{\epsilon})$ samples for each of $O(\log \frac{|L|}{\delta}) = O(\log \frac{m}{\delta}) = O(\log \frac{k}{\delta\epsilon_2})$ repetitions. Hence the estimation takes $O(\frac{k}{\epsilon_2} \log \frac{k}{\delta\epsilon_2} \log \frac{1}{\epsilon})$ samples.

The total number of samples is therefore

$$O\left(\frac{k}{\epsilon_2} \log \frac{1}{\epsilon} \left(\log \frac{k}{\delta} \log \frac{1}{\eta} + \log \frac{k}{\delta\epsilon_2}\right)\right) = O\left(\frac{k}{\epsilon_2} \log \frac{k}{\delta} \log \frac{1}{\epsilon_2\eta} \log \frac{k}{\min\{\epsilon_1, \epsilon_2\}}\right).$$

Runtime. It follows from Remark 1 that each isolation repetition takes $O((M + m \log m)r) = O(\frac{k}{\epsilon_2} \log \frac{k}{\epsilon\epsilon_2} \log \frac{1}{\eta}) = O(\frac{k}{\epsilon_2} \log \frac{1}{\epsilon} \log \frac{1}{\eta})$ time. There are $O(\log \frac{k}{\delta})$ repetitions so the total time for isolation is $O(\frac{k}{\epsilon_2} \log \frac{k}{\delta} \log \frac{1}{\epsilon} \log \frac{1}{\eta})$.

The input of ESTIMATE is a list L of size $|L| = O(m \log \frac{k}{\delta})$. Use the same trick in Remark 1, it stakes $O(M + m \log m + |L|)$ time 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 3–6 of ESTIMATE takes time $O((M+m \log m+|L|) \log \frac{m}{\delta}) = O(\frac{k}{\epsilon_2} \log \frac{1}{\epsilon} \log \frac{k}{\delta\epsilon_2})$ time. Thus estimation takes time $O(\frac{k}{\epsilon_2} \log \frac{1}{\epsilon} \log \frac{k}{\delta\epsilon_2} + |L| \log \frac{m}{\delta} + |L| \log |L|) = O(\frac{k}{\epsilon_2} \log \frac{k}{\delta} (\log \frac{1}{\epsilon} \log \frac{1}{\epsilon_2} + \log \frac{1}{\delta}))$.

The total running time is therefore

$$O\left(\frac{k}{\epsilon_2} \log \frac{k}{\delta} \left(\log \frac{1}{\epsilon} \log \frac{1}{\eta} + \log \frac{1}{\epsilon} \log \frac{1}{\epsilon_2} + \log \frac{1}{\delta}\right)\right)$$
$$= O\left(\frac{k}{\epsilon_2} \log \frac{k}{\delta} \left(\log \frac{1}{\epsilon_2 \eta} \log \frac{k}{\min\{\epsilon_1, \epsilon_2\}} + \log \frac{1}{\delta}\right)\right).$$

Sample Duration It is clear that the sample duration is

$$O(Md) = O\left(\frac{M}{\eta}\right) = O\left(\frac{k}{\epsilon_2 \eta} \log \frac{k}{\min\{\epsilon_1, \epsilon_2\}}\right).$$

The proof of the main theorem is now complete. □

4 Application to Bearing Estimation

A source on the plane emits a sine wave at a single frequency ω and this wave travels at speed c isotropically in this medium. If we ignore the decay of the amplitude of the wave as it travels, the source is localized by a single bearing parameter $\theta \in \mathbb{S}^1$, if we were to express its position in polar coordinates. Formally, a source at angle θ produces a wave field

$$F_\theta(x, t) = a_\theta \exp\left(i\omega\left(t + \frac{\langle x, n_\theta \rangle}{c}\right)\right), \quad x \in \mathbb{R}^2, t \in \mathbb{R},$$

where $n_\theta = (\cos\theta, \sin\theta)$ is the unit vector in the direction θ. Restricting the wave field to $x \in \mathbb{R}$, the horizontal axis, we have

$$F_\theta(x, t) = a_\theta \exp\left(i\omega\left(t + \frac{x \cos\theta}{c}\right)\right), \quad x \in \mathbb{R}, t \in \mathbb{R},$$

or, writing $\omega_\theta = \omega \cos\theta$ and assuming without loss of generality $c = 1$,

$$F_\theta(x, t) = a_\theta e^{i\omega t + i\omega_\theta x} = a_\theta e^{i\omega t} e^{i\omega_\theta x}.$$

On the horizontal axis, the wavefield oscillates in both time t and in position $x \in \mathbb{R}$, separately.

Suppose that there are k sources, each transmitting sine waves at the same frequency ω and at angles $\theta_1, \ldots, \theta_k \in \mathbb{S}^1$, and there is background noise at frequency ω supported on $I_\nu \subset \mathbb{S}^1$. We assume that $\{[\theta_j - \eta/2, \theta_j + \eta/2)\}_{j=1}^k$ and $[I_\nu - \eta/2, I_\nu + \eta/2)$

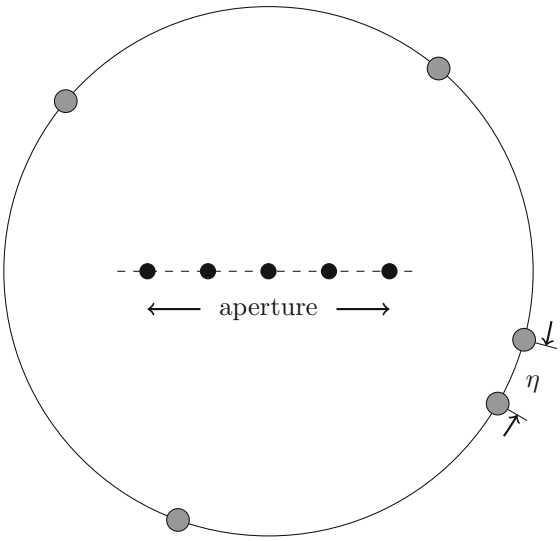

Fig. 3 Receiver array and source configuration. The *black* nodes are receivers and the *grey* ones sources

are all mutually disjoint. That is, the sources are not on a fixed, discrete grid but they are separated from each other and from the noise by a minimum resolution angle η.

To simplify notation, we denote ω_{θ_j} by ω_j and observe that $|\omega_j| \leq \omega$. A single receiver at position x on the horizontal axis observes the wavefield as

$$f(x,t) = \sum_{j=1}^{k} a_j \mathrm{e}^{i\omega t} \mathrm{e}^{i\omega_j x} + \mathrm{e}^{i\omega t} \left\{ \int_{I_v} a(\theta) \mathrm{e}^{i\omega \cos\theta \cdot x} d\mu \right\}, \quad x, t \in \mathbb{R},$$

where μ is a measure on $\mathbb{S}^1$ that satisfies the assumptions prescribed in Sect. 1.1.

The goal of the bearing estimation problem is to construct a (distribution over) placements x_m of M receivers and, from observations $y(x_m, t_0)$ at a *fixed time* t_0, find the amplitudes $\{a_j\}$ and positions or angles $\{\omega_j\}$ of the sources. This problem has the form of an off-grid Fourier sampling problem where we seek the identity of the unknown k "frequencies" ω_j from M samples of a sparse signal plus background noise. Figure 3 shows the configuration of sources and receivers.

Choose $0 < \beta < \pi/2$ and consider $\theta \in J = [-(\pi - \beta), -\beta] \cup [\beta, \pi - \beta]$, whence $|\sin\theta| \geq \sin\beta$. Furthermore, we have

$$|\omega \cos\theta_1 - \omega \cos\theta_2| \geq \omega (\sin\beta) |\theta_1 - \theta_2|$$

for $\theta_1, \theta_2 \in J$. Thus, it follows for $\theta_j \in J$ that ω_j is separated from the other frequencies and the noise by at least $\omega\eta \sin\beta$. Because there is nothing special about placing the receivers on the horizontal axis, we can also consider receivers distributed on a line that is rotated a constant number of times with respect to the horizontal axes so that the translations of J cover $\mathbb{S}^1$ altogether, it suffices to find ω_j (and the associated a_j) with $\theta_j \in J$ for a fixed β. We also re-define the angular resolution η to

be $\eta \sin \beta$. We might hope that, in this way, we can reduce the problem to recovery of $\{\omega_j\}$, or, $\{\cos\theta_j\}$; ambiguity, however, arises in the position of sources since $\cos\theta = \cos(-\theta)$, that is, sources symmetric around the horizontal line cannot be distinguished. Also a source that is close to the symmetric image of the other source may ruin the minimum separation in the reduced problem. Therefore, we make the following additional assumption: there exists an integer $q > 4$ such that for each pair of

$$E_p = \left\{\theta_j : \theta_j \in \left[\frac{\pi}{4} + (2p-1)\frac{\pi}{q} - 2\eta, \frac{\pi}{4} + (2p+1)\frac{\pi}{q} + 2\eta\right]\right\}$$

and

$$E'_p = \left\{\theta_j : \theta_j \in \left[-\frac{\pi}{4} + (2p-1)\frac{\pi}{q} - 2\eta, -\frac{\pi}{4} + (2p+1)\frac{\pi}{q} + 2\eta\right]\right\},$$

it holds that $d\left(-E_p + \{\frac{2p\pi}{q}\}, E'_p - \{\frac{2p\pi}{q}\}\right) \geq \eta$ and that $d\left(-E_p + \{\frac{2p\pi}{q}\}, I_v\right) \geq \eta$, where $d(\cdot,\cdot)$ is the metric on $\mathbb{S}^1$ and $p = 0, \ldots, q-1$.

Consider a filter K on $\mathbb{S}^1$ with a finite Fourier transform $\hat{K}$, supported on $I \subset \mathbb{Z}$. Then, placing receivers down on a line at positions $\{nx\}_{n\in I}$ associated with weights $\{\hat{K}(n)\}_{n\in I}$, we find that the wavefield these receivers observe is

$$\begin{aligned}\sum_{n\in I} \hat{K}(n) y(nx) &= \sum_{j=1}^{k} a_j \sum_{n\in I} \hat{K}(n) \exp(i\omega_j n x) \\ &\quad + \int_{I_v} a(\theta) \sum_{n\in I} \hat{K}(n) \exp\left(i\omega(\cos\theta) n x\right) d\mu \\ &= \sum_{j=1}^{k} a_j K(\omega_j x) + \int_{I_v} a(\theta) K(\omega(\cos\theta)x) d\mu.\end{aligned}$$

It is clear that any translation $K(u+\cdot)$ can be achieved by the same receiver array with associated weights $\{\hat{K}(n)e^{iun}\}_{n\in I}\}$ and that scaling $K(\alpha x)$ can be achieved by scaling the receiver array by the same factor α. Thus, we can perform all of the required measurement techniques for sampling in a receiver array.

The following result is an immediate application of our main result.

Theorem 2 *There is a distribution $\mathcal{D}$ on uniform receiver arrays and an algorithm $\mathcal{A}$ such that for each wavefield emanating from k sources $f(s) = \sum_{j=1}^{k} a_j e^{i\omega_j s} + \int_{I_v} a(\theta) e^{i\omega s} d\mu$, with constant probability, given observations from the receiver arrays, the algorithm returns a list Λ of amplitudes and bearings $\Lambda = \{(a'_j, \omega'_j)\}_{j=1}^{k}$ such that*

1. *For each a_j that satisfies (2) there exists $\omega'_j \in \Lambda$ such that*

$$|\omega_j - \omega'_j| \leq \frac{\epsilon_2}{k}\eta.$$

2. *Let* $\Lambda_0 = \{\omega'_j \in \Lambda : \exists \omega_{j_0}$ *such that* $\left|\omega_{j_0} - \omega'_j\right| \leq \frac{\epsilon_2 \eta}{k}$ *and* $|a_{j_0}|$ *satisfies* (2)$\}$, *then for each* $\omega'_j \in \Lambda_0$ *it holds that*

$$|a'_j - a_j| \leq \epsilon_1 \|a\|_1 + \frac{\epsilon_2}{k} \|\nu\|_1.$$

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

$$|a'_j| \leq \epsilon_1 \|a\|_1 + \frac{\epsilon_2}{k} \|\nu\|_1.$$

The algorithm places

$$O\left(\frac{k}{\epsilon_2} \log k \log \frac{1}{\epsilon_2 \eta} \log \frac{1}{\min\{\epsilon_1, \epsilon_2\}}\right)$$

receivers and runs in time proportional to number of samples. Furthermore, the receiver aperture size is

$$O\left(\frac{k}{\epsilon_2 \eta} \max\left\{\log \frac{1}{\epsilon_1}, \log \frac{k}{\epsilon_2}\right\}\right).$$

Proof By our assumption on E_p and E'_p, we can recover the sources in

$$\left[\frac{\pi}{4} + (2p-1)\frac{\pi}{q}, \frac{\pi}{4} + (2p+1)\frac{\pi}{q}\right] \cup \left[-\frac{\pi}{4} + (2p-1)\frac{\pi}{q}, -\frac{\pi}{4} + (2p+1)\frac{\pi}{q}\right]$$

with symmetry ambiguity. With rotations of the receiver arrays, the only ambiguity left will be to distinguish a source at θ from $\theta + \pi$. Suppose that $\theta \in E_{p_0}$. This ambiguity can be resolved by rotating the receiver array for E_{p_0} by η, the correctness of which is guaranteed by our assumption on E_p and E'_p again, noting the 2η brim on each side of each interval. □

We remark that it is possible to handle more configurations of sources, such as sources at the vertices of a regular polygon, by starting with a random direction instead of x-axis.

5 Conclusion and Open Problems

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 and as-yet unanswered questions with directly applying the discrete sublinear Fourier sampling techniques, both algorithmic and mathematical which we summarize first and then discuss open problems and conjectures that stem from our continuous approach.

5.1 Discretization

As mentioned in the introduction, one may be tempted to reduce the continuous setting to the discrete setting, assuming

$$f(t) = \sum_{i=j}^{k} a_j e^{i\omega_j t}, \quad a_j \in \mathbb{C},$$

by simply taking samples of $f(t)w(t)$ for some window function $W(t)$ at N equidistant points $t = 0, \Delta t, 2\Delta t, \ldots, (N-1)\Delta t$. The discrete Fourier transform (DFT) of the samples are approximately

$$\frac{1}{\Delta t} \sum_{j=1}^{k} a_j \widehat{W}\left(\omega_j - \frac{\ell}{N\Delta t}\right), \quad \ell = 0, 1, \ldots, N-1$$

by observing that

$$\sum_{k=0}^{N-1} e^{i\omega k\Delta t} W(k\Delta t) e^{-i\frac{k\ell}{N}} \approx \frac{1}{\Delta t} \int_0^T e^{i\omega x} W(x) e^{-i\ell\frac{x}{N\Delta t}} dt \approx \frac{1}{\Delta t} \widehat{W}\left(\omega - \frac{\ell}{N\Delta t}\right)$$

provided that $\Delta t \lesssim 1/\pi$ (so the Riemann sum is a good approximation to the integral) and $W(t)$ is supported on, or negligible outside, $[0, N\Delta t]$. A typical choice of $\widehat{W}$ is also a window function. Suppose that the pass region of $\widehat{W}$ has width $\sigma \lesssim \eta$ to avoid the interference of two different frequencies. The pass region of W is typically $1/\sigma \lesssim N\Delta t$, hence $1/\eta \lesssim 1/\sigma \lesssim N\Delta t$. Take $\widehat{W}(\xi) = \sin^2(\xi/\sigma)/(\xi/\sigma)^2$ (See Fig. 4). Consider ω_1 and let ℓ be the nearest integer to $N\Delta t \cdot \omega_1$. Then the ℓ-th coefficient in the DFT is

$$\frac{1}{\Delta t}\left(a_1 \widehat{W}\left(\omega_1 - \frac{\ell}{N\Delta t}\right) + \sum_{j\neq 1} a_j \widehat{W}\left(\omega_j - \frac{\ell}{N\Delta t}\right)\right).$$

The first term in the bracket is close to a_1 because $|\omega_1 - \ell/(N\Delta t)| \le 1/(2N\Delta t) \lesssim \sigma$ and thus $\widehat{W}$ is close to 1. To bound the second term, notice that for $j \neq 1$,

$$\left|\omega_j - \frac{\ell}{N\Delta t}\right| \ge |\omega_j - \omega_1| - \frac{1}{2N\Delta t} \ge |\omega_j - \omega_1| - \frac{\eta}{C}$$

for some absolute constant $C_1 > 0$. Thus the second term is bounded by (after rearranging the indices)

$$\frac{|a_2| + |a_3|}{\left(1 - \frac{1}{C}\right)^2 \left(\frac{\eta}{\sigma}\right)^2} + \frac{|a_4| + |a_5|}{\left(2 - \frac{1}{C}\right)^2 \left(\frac{\eta}{\sigma}\right)^2} + \frac{|a_6| + |a_7|}{\left(3 - \frac{1}{C}\right)^2 \left(\frac{\eta}{\sigma}\right)^2} + \cdots \lesssim \left(\frac{\sigma}{\eta}\right)^2 \cdot \max_{i\ge 2} |a_i|$$

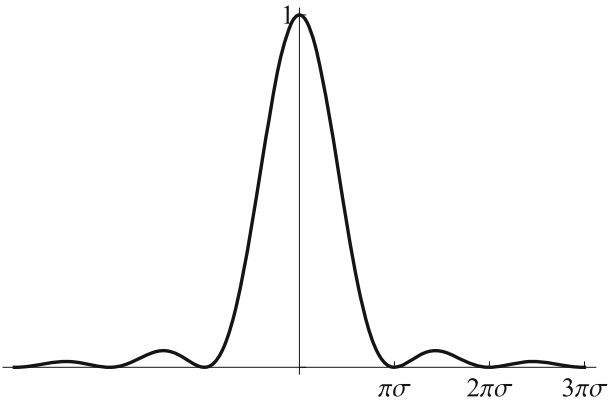

Fig. 4 $\widehat{W}(\xi) = \sin^2(\xi/\sigma)/(\xi/\sigma)^2$

This means that the ℓ-th coefficient in the DFT is proportional to the coefficient associated with ω_1 corrupted by contributions from other coefficients associated with other frequencies. It is therefore conceivable that there exists a constant $c > 0$ (depending on W) such that the top ck coefficients of the N DFT coefficients include constant approximations to those $\{a_i\}$'s which are at least a constant fraction of $\|a\|_1$ with the choice of $\sigma \lesssim \eta$. The sample duration will thus be $\Theta(1/\eta)$. Using existing sparse recovery results for discrete setting, it seems probable to recover the frequencies with coefficients at least a constant fraction of $\|a\|_1$ with sample duration of $N\Delta t \simeq 1/\eta$. Improving the guarantee to recovering frequencies with coefficients at least $1/k$ fraction of $\|a\|_1$ will increase the sample duration to $\Theta(k/\eta)$ with σ shrunk from $\Theta(\eta)$ to $\Theta(\eta/k)$.

This approach looks promising yet there are some inconveniences compared with the preceding direct approach. For instance, typical discrete case results give an ℓ_2/ℓ_2 error bound, that is, $\|x - x'\|_2 \leq (1+\epsilon)\|x - x_k\|$, where x' is the approximation of x and x_k the best k-term approximation. It is not obvious how to interpret such result in the continuous setting. On the other hand, under the discretization scenario, a real-valued frequency spreads around so one may encounter a cluster of significant components in the discretized signal and thus an additional step of locating the frequency from a cluster of them is needed. This is not obvious either provided only the ℓ_2 error guarantee. An ℓ_∞ error guarantee is more desirable, however, it increases the complexity of the algorithms for the discrete case.

5.2 Open Problems

First we ask a question about output evaluation metric. Since we do not expect to recover the frequencies exactly, the typical approximation error of the form

$$\left\| \sum_j \left(a_j \mathrm{e}^{i\omega_j x} - a'_j \mathrm{e}^{i\omega'_j x} \right) + \nu(x) \right\|_p \quad \text{for } p = 1, 2$$

contains both the coefficient 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 coefficients $|a_j - a'_j|$ 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., [15]), with which our result is comparable.

An alternative method to express the output error is to treat the exponential sum $\sum a_i e^{i\omega x}$ as a distribution and consider

$$\sup_{g \in \mathcal{F}} |\langle f_1, g\rangle - \langle f_2, g\rangle|,$$

where f_1 and f_2 are two exponential sums and $\mathcal{F}$ some class of test functions. For instance, when f_1 and f_2 are probability measures, it gives total variation distance when $\mathcal{F}$ is the set of functions bounded by 1 and Wasserstein distance (or earth-mover distance) when $\mathcal{F}$ consists of all 1-Lipschitz functions. But our algorithm has no guarantee that $\|a\| = \|a'\|$ so it is generally not a metric. Obviously this bound can be expressed in terms of $|a_i - a'_i|$ and $|\omega_i - \omega'_i|$ (for example, with $\mathcal{F}$ being the set of 1-Lipschitz function we obtain the L^1 norm) and can thus be bounded if the estimates of individual source are bounded. It remains unclear what class of test functions would give a less common but interesting bound for this problem.

Besides, we leave the following problems open and make conjectures.

- 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.
- 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. We think $\Theta(k/\eta)$ is the tight bound with $k \log^{O(1)}(1/\eta)$ samples while our algorithm takes $O((k/\eta)\log k)$ samples for $\epsilon_1 = \Theta(1/k)$ and $\epsilon_2 = \Theta(1)$. It remains future work to devise a sublinear time algorithm with smaller sample duration, for which a good reduction to discrete case looks a promising approach.
- If we do not reduce to discrete case, a better algorithm would be an iterative one so the runtime can be lowered by an $O(\log k)$ factor. Notice that the number of samples is always bounded by the runtime from above.

Acknowledgments The authors would like to thank an anonymous reviewer of *APPROX/RANDOM* 2012 for the suggestion of improving the running time. V. Cevher was partially supported by Faculty Fellowship at Rice University, MIRG-268398, ERC Future Proof, and DARPA KeCoM program #11-DARPA-1055. A. C. Gilbert was partially supported by NSF DMS 0743372 and DARPA ONR N66001-06-1-2011. Y. Li was partially supported by NSF DMS 0743372 when the author was at University of Michigan, Ann Arbor. M. J. Strauss was partially supported by NSF DMS 0743372 and DARPA ONR N66001-06-1-2011.

References

1. Baraniuk, R., Cevher, V., Wakin, M.: Low-dimensional models for dimensionality reduction and signal recovery: a geometric perspective. Proc. IEEE **98**(6), 959–971 (2010)
2. Chahine, K., Baltazart, V., Wang, Y.: Interpolation-based matrix pencil method for parameter estimation of dispersive media in civil engineering. Signal Process. **90**(8), 2567–2580 (2010)
3. Chui, D.: Personal communication (2013).
4. Gilbert, A., Indyk, P.: Sparse recovery using sparse matrices. Proc. IEEE **98**(6), 937–947 (2010)

5. Gilbert, A., Li, Y., Porat, E., Strauss, M.: Approximate sparse recovery: optimizing time and measurements. SIAM J. Comput. **41**(2), 436–453 (2012)
6. Gilbert, A.C., Guha, S., Indyk, P., Muthukrishnan, S., Strauss, M.: Near-optimal sparse Fourier representations via sampling. In: Proceedings of the thiry-fourth annual ACM Symposium on Theory of Computing. STOC '02, pp. 152–161. ACM, New York (2002)
7. Gilbert, A.C., Muthukrishnan, S., Strauss, M.: Improved time bounds for near-optimal sparse Fourier representations. In: Proceedings of Wavelets XI conference, pp. 398–412 (2005).
8. Hassanieh, H., Indyk, P., Katabi, D., Price, E.: Nearly optimal sparse Fourier transform. In: Proceedings of the 44th Symposium on Theory of Computing, STOC '12, pp. 563–578. ACM, New York (2012)
9. Hassanieh, H., Indyk, P., Katabi, D., Price, E.: Simple and practical algorithm for sparse Fourier transform. In: Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '12, pp. 1183–1194. SIAM (2012).
10. Heider, S., Kunis, S., Potts, D., Veit, M.: A sparse Prony FFT. In: Proceedings of Tenth International Conference on Sampling Theory and Applications, SampTA 2013, pp. 572–575 (2013)
11. Helson, H.: Harmonic Analysis (2nd edn.). Hindustan Book Agency, New Delhi (1995).
12. Hua, Y., Sarkar, T.: On SVD for estimating generalized eigenvalues of singular matrix pencil in noise. In: Proceedings of the IEEE Transactions on Signal Processing, vol. 39(4), pp. 892–900 (1991)
13. Iwen, M.: Combinatorial sublinear-time Fourier algorithms. Found. Comput. Math. **10**(3), 303–338 (2009)
14. Kushilevitz, E., Mansour, Y.: Learning decision trees using the Fourier spectrum. SIAM J. Comput. **22**(6), 1331–1348 (1993)
15. Peter, T., Potts, D., Tasche, M.: Nonlinear approximation by sums of exponentials and translates. SIAM J. Sci. Comput. **33**(4), 1920–1947 (2011)
16. Potts, D., Tasche, M.: Parameter estimation for exponential sums by approximate Prony method. Signal Process. **90**(5), 1631–1642 (2010)
17. Vetterli, M., Marziliano, P., Blu, T.: Sampling signals with finite rate of innovation. In: Proceedings of the IEEE Transactions on Signal Processing, vol. 50(6), pp. 1417–1428 (2002)