Estimating the Frequency of a Clustered Signal

Xue Chen*
xue.chen1@northwestern.edu
Northwestern University

Eric Price†
ecprice@cs.utexas.edu
The University of Texas at Austin

July 7, 2019

Abstract

We consider the problem of locating a signal whose frequencies are "off grid" and clustered in a narrow band. Given noisy sample access to a function $g(t)$ with Fourier spectrum in a narrow range $[f_0 - \Delta, f_0 + \Delta]$, how accurately is it possible to identify f_0? We present generic conditions on g that allow for efficient, accurate estimates of the frequency. We then show bounds on these conditions for k-Fourier-sparse signals that imply recovery of f_0 to within $\Delta + \widetilde{O}(k^3)$ from samples on $[-1, 1]$. This improves upon the best previous bound of $O\big(\Delta + \widetilde{O}(k^5)\big)^{1.5}$. We also show that no algorithm can do better than $\Delta + \widetilde{O}(k^2)$.

In the process we provide a new $\widetilde{O}(k^3)$ bound on the ratio between the maximum and average value of continuous k-Fourier-sparse signals, which has independent application.

1 Introduction

A natural question, dating at least to the work of Prony in 1795, is to estimate a signal from samples, assuming the signal has a k-sparse Fourier representation, i.e., that the signal is a sum of k complex exponentials: $g(t) = \sum_{j=1}^k v_j e^{2\pi \mathbf{i} f_j t}$ for some set of frequencies f_j and coefficients v_j.

If the frequencies are located on a discrete grid (giving a sparse discrete Fourier transform), then a long line of work has studied efficient algorithms for recovering the signal (e.g., [Man92, GGI+02, AGS03, GMS05, HIKP12, IK14]). If the frequencies are not on a grid, then Prony's method from 1795 [Pro95] or matrix pencil [BM86] can still identify them in the absence of noise. With noise, however, one cannot robustly recover frequencies that are too close together: if one listens to a signal for the interval $[-T, T]$ then any two frequencies θ and $\theta + \varepsilon/T$ will be $O(\varepsilon)$-close to each other, and so cannot be distinguished with noise. As shown in [Moi15], this nonrobustness grows exponentially in k. On the other hand, [Moi15] also showed that recovery with polynomially small noise is possible if all the frequencies have separation $1/2T$, and [PS15] showed that a constant fraction of noise is tolerable with separation $\log^{O(1)}(FT)/T$, where F is the bandlimit of the signal.

So what *is* possible for arbitrary Fourier-sparse signals, without any assumption of frequency separation? One cannot hope to identify the frequencies exactly, but one can still estimate the *signal itself*. If two frequencies are similar enough to be indistinguishable over the sampled interval, we

---

do not need to distinguish them. In [CKPS16], this led to an algorithm for an arbitrary k-Fourier-sparse signal that used $\text{poly}(k, \log(FT))$ samples to estimate it with only a constant factor increase in the noise. However, this polynomial is fairly poor.

Since prior work could handle the case of well-separated frequencies, a key challenge in [CKPS16] is the setting with all the frequencies in a narrow cluster. Formally, consider the following subproblem: if all the frequencies f_i of the signal lie in a narrow band $[f_0 - \Delta, f_0 + \Delta]$, how accurately can we estimate f_0? Note that while we would like an efficient algorithm that takes a small number of samples, the key question is *information theoretic*. And we can ask this question more generally: if the signal is not k-sparse, but still has all its frequencies in a narrow band, can we locate that band?

Question 1.1. *Let $g(t)$ be a signal with Fourier transform supported on $[f_0 - \Delta, f_0 + \Delta]$, for some $f_0 \in [-F, F]$. Suppose that we can sample from $y(t) = g(t) + \eta(t)$ at points in $[-T, T]$, where $\eta(t)$ could be any ℓ_2 bounded noise on $[-T, T]$ with*

$$\mathop{\mathbb{E}}_{t \in [-T,T]} \left[|\eta(t)|^2 \right] \leq \varepsilon \mathop{\mathbb{E}}_{t \in [-T,T]} \left[|g(t)|^2 \right]$$

for a small constant ε. Under what conditions on g can we estimate f_0, and how accurately?

One might expect to be able to estimate f_0 to $\pm(\Delta + O(\frac{1}{T}))$ for all functions g; after all, g is just a combination of individual frequencies, each of which points to some frequency in the right range, and each individual frequency in isolation can be estimated to within $\pm O(\frac{1}{T})$ in the presence of noise. Unfortunately, this intuition is false.

To see this, consider the family of k-sparse Fourier functions with $f_j = \varepsilon j$, i.e.,

$$\text{span}(e^{2\pi \mathbf{i}(j\varepsilon)t} \mid j \in [k]).$$

By sending $\varepsilon \to 0$ and taking a Taylor expansion, this family can get arbitrarily close to any degree $k - 1$ polynomial, on any interval $[-T', T']$. Thus, to solve the question, one would also need to solve it when $g(t)$ is a polynomial even for arbitrarily small Δ.

There are two ways in which $g(t)$ being a degree d polynomial can lead to trouble. The first is that $g(t)$ could itself be a Taylor expansion of $e^{\pi \mathbf{i} ft}$. If $d \gtrsim fT$, this Taylor approximation will be quite accurate on $[-T, T]$; with the noise η, the observed signal can *equal* $e^{\pi \mathbf{i} ft}$. Thus the algorithm has to output f, which can be $\Theta(d/T)$ far from the "true" answer $f_0 = 0$.

The second way in which $g(t)$ can lead to trouble is by removing most of the signal energy. If $g(t)$ is the (slightly shifted) Chebyshev polynomial $g(t) = T_d\big(t/T + O(\frac{\log^2 d}{d^2})\big)$, then $|g(t)| \leq 1$ for $t \leq \big(1 - O(\frac{\log^2 d}{d^2})\big)T$, while $g(t) \geq d$ for $t \geq \big(1 - O(\frac{\log^2 d}{d^2})\big)T$. That is to say, the majority of the ℓ_2 energy of g can lie in the final $O(\frac{\log^2 d}{d^2})$ fraction of the interval. In such a case, a small constant noise level η can make samples outside that $T \cdot \widetilde{O}(1/d^2)$ size region equal to zero, and hence completely uninformative; and samples in that region still have to tolerate noise. This leads to an "effective" interval size of $T' = T \cdot \widetilde{O}(\frac{1}{d^2})$, leading to accuracy $O(1/T') = \widetilde{O}(d^2)/T$.

Our main result is that, in a sense, these two types of difficulties are the only ones that arise. We can measure the second type of difficulty by looking at how much larger the maximum value of g is than its average:

$$R := \frac{\sup_{t \in [-T,T]} |g(t)|^2}{\mathbb{E}_{t \in [-T,T]} |g(t)|^2}.$$

We can measure the former by observing that while a polynomial may approximate a complex exponential on a bounded region, as $t \to \infty$ the polynomial will blow up. In particular, we take

2

the S such that

$$|g(t)|^2 \leq \mathsf{poly}(R) \cdot \mathop{\mathbb{E}}_{t \in [-T,T]} \left[|g(t)|^2 \right] \cdot |\frac{t}{T}|^S$$

for all $|t| \geq T$. We show that if R and S are bounded, one can estimate f_0 to within $\Delta + \widetilde{O}(R+S)/T$, which is almost tight from the above discussion of polynomials. Moreover, the time and number of samples required are fairly efficient:

Theorem 1.2. *Given any $T > 0, F > 0, \Delta > 0, R,$ and $S > 0$, let $g(t)$ be a signal with the following properties:*

1. $\mathsf{supp}(\widehat{g}) \subseteq [f_0 - \Delta, f_0 + \Delta]$ *where* $f_0 \in [-F, F]$.

2. $\displaystyle\sup_{t \in [-T,T]} \left[|g(t)|^2 \right] \leq R \cdot \mathop{\mathbb{E}}_{t \in [-T,T]} \left[|g(t)|^2 \right].$

3. $|g(t)|^2$ *grows as at most* $\mathsf{poly}(R) \cdot \mathop{\mathbb{E}}_{t \in [-T,T]} \left[|g(t)|^2 \right] \cdot |\frac{t}{T}|^S$ *for* $t \notin [-T, T]$.

Let $y(t) = g(t) + \eta(t)$ be the observable signal on $[-T, T]$, where $\displaystyle\mathop{\mathbb{E}}_{t \in [-T,T]} \left[|\eta(t)|^2 \right] \leq \epsilon \cdot \mathop{\mathbb{E}}_{t \in [-T,T]} \left[|g(t)|^2 \right]$ for a sufficiently small constant ϵ. For $\Delta' = \Delta + \frac{\widetilde{O}(R+S)}{T}$ and any $\delta > 0$, there exists an efficient algorithm that takes $O(R \log \frac{F}{\Delta' \cdot \delta})$ samples from $y(t)$ and outputs $\widetilde{f}$ satisfying $|f_0 - \widetilde{f}| \leq O(\Delta')$ with probability at least $1 - \delta$.

Application to sparse Fourier transforms Specializing to k-Fourier-sparse signals, we give bounds on R and S for this family. Since (as described above) this family can approximate degree-$(k-1)$ polynomials, we know that $R \gtrsim k^2$ and $S \gtrsim k$; we show that $R \lesssim k^3 \log^2 k$ and $S \lesssim k^2 \log k$. Thus, whenever R is between k^2 and $\widetilde{O}(k^3)$, we can identify k-Fourier-sparse signals to within $\Delta + \widetilde{O}(R)/T$. This is an improvement over the results in [CKPS16] in several ways.

Formally, for a given sparsity level k, we consider signals in

$$\mathcal{F} := \left\{ g(t) = \sum_{j=1}^{k} v_j e^{2\pi i f_j t} \middle| f_j \in [-F, F] \right\}.$$

Theorem 1.3. *For any k and T,*

$$R := \sup_{g \in \mathcal{F}} \frac{\displaystyle\sup_{x \in [-T,T]} |g(x)|^2}{\displaystyle\mathop{\mathbb{E}}_{x \in [-T,T]} [|g(x)|^2]} = O(k^3 \log^2 k). \tag{1}$$

It was previously known that $R \lesssim k^4 \log^3 k$ [CKPS16], and this fact was used in [AKM+19]. (Thus, our improved bound on R immediately implies an improvement in Theorem 8 of [AKM+19], from $s_{\mu,\varepsilon}^5 \log^3 s_{\mu,\varepsilon}$ to $s_{\mu,\varepsilon}^4 \log^2 s_{\mu,\varepsilon}$.)

Next we bound the growth $S = \widetilde{O}(k^2)$ for any $|t| \geq T$.

Theorem 1.4. *There exists $S = O(k^2 \log k)$ such that for any $|t| > T$ and $g(t) = \sum_{j=1}^{k} v_j \cdot e^{2\pi i f_j t}$,*
$|g(t)|^2 \leq \mathsf{poly}(k) \cdot \displaystyle\mathop{\mathbb{E}}_{x \in [-T,T]} [|g(x)|^2] \cdot |\frac{t}{T}|^S.$

This is analogous to Theorem 5.5 of [CKPS16], which proves a bound of $(kt)^k$ rather than $t^{\widetilde{O}(k^2)}$. These bounds are incomparable, but the $t^{\widetilde{O}(k^2)}$ bound is actually more useful for this problem: what really matters is showing that $g(t)$ is not too large just outside the interval. Theorem 1.4 gives the "correct" polynomial dependence at $t = (1 + 1/k^2)T$.

We can now apply Theorem 1.2 to get an efficient algorithm to recover the center of a cluster of k frequencies within accuracy $\tilde{O}(R)$.

Theorem 1.5. *Given F, T, and k, let R be the ratio between the maximum and average value of continuous k-Fourier-sparse signals defined in (1). Given Δ, let $g(t)$ be a k-Fourier-sparse signal centered around f_0: $g(t) = \sum_{i \in [k]} v_i \cdot e^{2\pi \mathbf{i} f_i t}$ where each $f_i \in [f_0 - \Delta, f_0 + \Delta]$ and $y(t) = g(t) + \eta(t)$ be the observable signal on $[-T, T]$, where $\underset{t \in [-T,T]}{\mathbb{E}}\big[|\eta(t)|^2\big] \leq \epsilon \cdot \underset{t \in [-T,T]}{\mathbb{E}}\big[|g(t)|^2\big]$ for a sufficiently small constant ϵ.*

For any $\delta > 0$, there exist $\Delta' = \Delta + \frac{\tilde{O}(R)}{T}$ and an efficient algorithm that takes $O(k \log^2 k \log \frac{F}{\Delta' \cdot \delta})$ samples from $y(t)$ and outputs $\widetilde{f}$ satisfying $|f_0 - \widetilde{f}| \leq O(\Delta')$ with probability at least $1 - \delta$.

Note that the sample complexity here is $\widetilde{O}(k)$ not $\widetilde{O}(R)$. This is because, based on the structure of the problem, we can use a nonuniform sampling procedure that performs better. Otherwise this theorem is just Theorem 1.2 applied to the R and S from Theorems 1.3 and 1.4.

Theorem 1.5 is a direct improvement on Theorem 7.5 of [CKPS16], which for $T = 1$ could estimate to within $O\left(\Delta + \widetilde{O}(k^5)\right)^{1.5}$ accuracy and used $\text{poly}(k)$ samples. In particular, in addition to improving the additive $\text{poly}(k)$ term, our result avoids a multiplicative increase in the bandwidth Δ of g.

The main technical lemma in proving Theorems 1.2 and 1.5 is a filter function H with a compact supported Fourier transform $\widehat{H}$ that simulates a box function on $[-T, T]$ for any g satisfying the conditions in Theorem 1.2.

Lemma 1.6. *Given any T, S, and R, there exists a filter function H with $\left|\text{supp}(\widehat{H})\right| \leq \frac{\tilde{O}(R+S)}{T}$ such that for any $g(t)$ satisfying the second and third conditions in Theorem 1.2,*

1. *H is close to a box function on $[-T, T]$: $\int_{-T}^{T} |g(t) \cdot H(t)|^2 \mathrm{d}t \geq 0.9 \int_{-T}^{T} |g(t)|^2 \mathrm{d}t$.*

2. *The tail of $H(t) \cdot g(t)$ is small: $\int_{-T}^{T} |g(t) \cdot H(t)|^2 \mathrm{d}t \geq 0.95 \int_{-\infty}^{\infty} |g(t) \cdot H(t)|^2 \mathrm{d}t$.*

Organization We introduce some notation and tools in Section 2. Then we provide a technical overview in Section 3. We show our filter function and prove Lemma 1.6 in Section 4. Next we present the algorithm about frequency estimation of Theorem 1.2 in Section 5. Finally we prove the results about sparse Fourier transform — Theorem 1.3 and Theorem 1.4 in Section 6.

2 Preliminaries

In the rest of this work, we fix the observation interval to be $[-1, 1]$ and define

$$\|g\|_2 = \Big(\underset{x \sim [-1,1]}{\mathbb{E}} |g(x)|^2 \Big)^{1/2}, \tag{2}$$

because we could rescale $[-T, T]$ to $[-1, 1]$ and $[-F, F]$ to $[-FT, FT]$.

We first review several facts about the Fourier transform. The Fourier transform $\widehat{g}(f)$ of an integrable function $g : \mathbb{R} \to \mathbb{C}$ is

$$\widehat{g}(f) = \int_{-\infty}^{+\infty} g(t)e^{-2\pi i f t}\mathrm{d}t \text{ for any real } f.$$

We use $g \cdot h$ to denote the pointwise dot product $g(t) \cdot h(t)$ and g^k to denote $\underbrace{g(t) \cdots g(t)}_{k}$. Similarly, we use $g * h$ to denote the convolution of g and h: $\int_{-\infty}^{+\infty} g(x) \cdot h(t-x)\mathrm{d}x$. In this work, we always set g^{*k} as the convolution $\underbrace{g(t) * \cdots * g(t)}_{k}$. Notice that $\mathsf{supp}(g \cdot h) = \mathsf{supp}(g) \cap \mathsf{supp}(h)$ and $\mathsf{supp}(g * h) = \mathsf{supp}(g) + \mathsf{supp}(h)$.

We define the box function and its Fourier transform sinc function as follows. Given a width $s > 0$, the box function $\mathrm{rect}_s(t) = 1/s$ iff $|t| \leq s/2$; and its Fourier transform is $\mathrm{sinc}(sf) = \frac{\sin(\pi f s)}{\pi f s}$ for any f.

We state the Chernoff bound for random sampling [Che52].

Lemma 2.1. *Let* $X_1, X_2, \cdots, X_n$ *be independent random variables in* $[0, R]$ *with expectation* 1. *For any* $\varepsilon < 1/2$ *and* $n \gtrsim \frac{R}{\epsilon^2}$, $X = \frac{\sum_{i=1}^{n} X_i}{n}$ *with expectation* 1 *satisfies*

$$\Pr[|X - 1| \geq \varepsilon] \leq 2\exp(-\frac{\varepsilon^2}{3} \cdot \frac{n}{R}).$$

3 Proof Overview

We first outline the proofs of Lemma 1.6 and Theorem 1.2. Then we show the proof sketch of $R = \tilde{O}(k^3)$ and $S = \tilde{O}(k^2)$ of k-Fourier-sparse signals.

The filter functions $(H, \widehat{H})$ **in Lemma 1.6.** Ideally, to satisfy the two claims in Lemma 1.6, we could set $H(t)$ to be the box function $2\,\mathrm{rect}_2(t)$ on $[-1, 1]$. However, by the uncertainty principle, it is impossible to make its Fourier transform $\widehat{H}$ compact using such an $H(t)$. Hence our construction of $(H, \widehat{H})$ is in the inverse direction: we build $\widehat{H}(f)$ by box functions and $H(t)$ by the Fourier transform of box functions — the sinc function. In the rest of this discussion, we focus on using the sinc function to prove Lemma 1.6 given the properties of g in Theorem 1.2.

We first notice that any H with the following two properties is effective in Lemma 1.6 for g satisfying $|g(t)|^2 \leq R \cdot \|g\|_2^2$ for any $|t| \leq 1$ and $|g(t)|^2 \leq \mathsf{poly}(R)\|g\|_2^2 \cdot |t|^S$ for $|t| > 1$:

1. $H(t) = 1 \pm 0.01$ for any $t \in [-1 + \frac{1}{C \cdot R}, 1 - \frac{1}{C \cdot R}]$ of a large constant C. This shows

$$\int_{-1}^{1} |H(t) \cdot g(t)|^2\mathrm{d}t \geq 0.99^2 \int_{-1+\frac{1}{C \cdot R}}^{1-\frac{1}{C \cdot R}} |g(t)|^2\mathrm{d}t.$$

   Because $|g(t)|^2 \leq R \cdot \|g\|_2^2$ for any $t \in [-1, 1] \setminus [-1 + \frac{1}{C \cdot R}, 1 - \frac{1}{C \cdot R}]$, the constant on the R.H.S. is at least $0.99^2 \cdot (1 - \frac{1}{C}) \geq 0.9$, which implies the first claim of Lemma 1.6.

2. $H(t)$ declines to $\frac{1}{\mathsf{poly}(R) \cdot t^{2S}}$ for any $|t| > 1$. This shows

$$\int_{1}^{\infty} |H(t) \cdot g(t)|^2\mathrm{d}t \leq 0.01 \int_{-1}^{1} |g(t)|^2\mathrm{d}t,$$

   which implies the second claim.

For ease of exposition, we start with $S = 0$. We plan to design a filter $H_0(t)$ with compact $\widehat{H_0}$ dropping from 0.99 at $t = 1 - \frac{1}{C \cdot R}$ to $\frac{1}{\mathsf{poly}(R)}$ at $t = 1$ in a small range $\frac{1}{CR}$ using the sinc function. To apply the sinc function, we notice that

$$\mathrm{sinc}(CR \cdot t)^{O(\log R)} = \left(\frac{\sin(\pi CR \cdot t)}{\pi CR \cdot t} \right)^{O(\log R)}$$

decays from 1 at $t = 0$ to $1/\mathsf{poly}(R)$ at $t = \frac{1}{C \cdot R}$, which matches the dropping of $H_0(t)$ from $t = 1 - \frac{1}{C \cdot R}$ to $t = 1$.

Then, to make $H(t) \approx 1$ for any $|t| \leq 1 - \frac{1}{C \cdot R}$, let us consider a convolution of $\mathrm{rect}_1(t)$ and $\mathrm{sinc}(CR \cdot t)^{O(\log R)}$. Because most of the mass of the latter is in $[-\frac{1}{CR}, \frac{1}{CR}]$, this convolution keeps almost the same value in $[-\frac{1}{2} + \frac{1}{CR}, \frac{1}{2} - \frac{1}{CR}]$ and drops down to $1/\mathsf{poly}(R)$ at $t = \frac{1}{2} + \frac{1}{CR}$. At the same time, it will keep the compactness of $\widehat{H_0}$ since it corresponds to the dot product on the Fourier domain. By normalizing and scaling, this gives the desired $(H_0, \widehat{H_0})$ for $S = 0$.

Next we describe the construction of $S > 0$. The high level idea is to consider the decays of $H(t)$ in $\log_2 S + O(1)$ segments rather than one segment of $S = 0$:

$$(1 - \frac{1}{CR}, 1], (1, 1 + \frac{1}{S}], (1 + \frac{1}{S}, 1 + \frac{2}{S}], \ldots, (1 + \frac{2^j}{S}, 1 + \frac{2^{j+1}}{S}], \ldots, (1 + \frac{S/2}{S}, 2], (2, +\infty).$$

For each segment, we provide a power of sinc functions matching its decay in $H(t)$ like the construction of H_0 on $(1 - \frac{1}{CR}, 1]$. The final construction is the convolution of the dot product of all sinc powers and a box function, which appears in Section 4.

The Algorithm of Theorem 1.2. Now we show how to estimate f_0 given the observable signal $y = g + \eta$ where $\mathsf{supp}(\widehat{g}) \subseteq [f_0 - \Delta, f_0 + \Delta]$ and $\|\eta\|_2^2 \leq \varepsilon \|g\|_2^2$ (with ℓ_2 norm taken over $[-T, T]$ defined in (2)). We instead consider $y_H(t) = y(t) \cdot H(t)$ with the filter function $(H, \widehat{H})$ from Lemma 1.6 and the corresponding dot products $g_H = g \cdot H$ and $\eta_H = \eta \cdot H$. The starting point is that for a sufficiently small β, we expect

$$y_H(t + \beta) \approx e^{2\pi i f_0 \beta} \cdot y_H(t)$$

because y_H has Fourier spectrum concentrated around f_0. This does not hold for *all* t, but it does hold on average:

$$\int_{-1}^{1} |y_H(t + \beta) - e^{2\pi i f_0 \beta} \cdot y_H(t)|^2 \mathrm{d}t \lesssim \varepsilon \cdot \int_{-1}^{1} |y_H(t)|^2 \mathrm{d}t. \tag{3}$$

This is because we can use Parseval's identity to replace these integrals by an integral over Fourier domain—Parseval's identity would apply if the integrals were from $-\infty$ to ∞, but because of the filter function H, relatively little mass in y_H lies outside $[-1, 1]$. Then, the Fourier transform of the term inside the left square is $e^{2\pi i f \beta} \cdot \widehat{y_H}(f) - e^{2\pi i f_0 \beta} \cdot \widehat{y_H}(f)$. Note that $\widehat{y_H} = \widehat{g_H} + \widehat{\eta_H}$ has most of its ℓ_2 mass in $\mathsf{supp}(g_H) \subseteq [f_0 - \Delta', f_0 + \Delta']$ for $\Delta' = \Delta + |\mathsf{supp}(\widehat{H})|$, and every such frequency shrinks in the left by a factor $|e^{2\pi i f \beta} - e^{2\pi i f_0 \beta}| = O(\beta \Delta')$. Thus, for $\beta \ll 1/\Delta'$, (3) holds.

To learn f_0 through $e^{2\pi i f_0 \beta}$, we design a sampling procedure to output α satisfying

$$|y_H(\alpha + \beta) - e^{2\pi i f_0 \beta} y_H(\alpha)| \leq 0.3 \cdot y_H(\alpha) \text{ with probability } \textit{more than half}.$$

Even though the above discussion shows the left hand side is smaller than the R.H.S. on average, a uniformly random $\alpha \sim [-1, 1]$ may not satisfy it with good probability: $|y_H(\alpha)| \geq \|y_H\|_2$ may

be only true for $1/R$ fraction of $\alpha \in [-1, 1]$, while the corruption by adversarial noise η has $\|\eta\|_2^2 \gtrsim \varepsilon \|y_H\|_2^2$ for a constant $\varepsilon \gg 1/R$. At the same time, even for many points $\alpha_1, \ldots, \alpha_m$ where some of them satisfy the above inequality, it is infeasible to verify such an α_i given f_0 is unknown. We provide a solution by adopting the importance sampling: for $m = O(R)$ random samples $\alpha_1, \ldots, \alpha_m \in [-1, 1]$, we output α with probability proportional to the weight $|y_H(\alpha_i)|^2$.

We prove the correctness of this sampling procedure in Lemma 5.2 in Section 5.

Finally, learning $e^{2\pi \mathbf{i} f_0 \beta}$ is not enough to learn f_0: because of the noise, we only learn $e^{2\pi \mathbf{i} f_0 \beta}$ to within a constant ε, which gives f_0 to within $\pm O(\varepsilon/\beta)$; and because of the different branches of the complex logarithm, this is only up to integer multiples of $1/\beta$. Therefore to fully learn f_0, we repeat the sampling procedure at logarithmically many different scales of β, from $\beta = 1/2F$ to $\beta = \frac{\Theta(1)}{\Delta'}$.

k-Fourier-sparse signals. Finally, we show $R = \widetilde{O}(k^3)$ and $S = \widetilde{O}(k^2)$ such that for any $g(t) = \sum_{j=1}^k v_j \cdot e^{2\pi \mathbf{i} f_j t}$ — not necessarily one with the f_j clustered together—

$$\frac{\sup_{t \in [-1,1]} |g(t)|^2}{\|g\|_2^2} \le R \text{ and } |g(t)|^2 \le \mathsf{poly}(R) \cdot \|g\|_2^2 \cdot |t|^S.$$

We first review the previous argument of $R = \widetilde{O}(k^4)$ [CKPS16]. The key point is to show for some $d = \widetilde{O}(k^2)$ that $g(1)$ is a linear combination of $g(1-\theta), \ldots, g(1-d \cdot \theta)$ using bounded integer coefficients $c_1, \ldots, c_d = O(1)$ for any $\theta \le \frac{2}{d}$. Then

$$g(1) = \sum_{j \in [d]} c_j \cdot g(1 - j \cdot \theta) \text{ implies } |g(1)|^2 \le (\sum_{j \in [d]} |c_j|^2) \cdot (\sum_{j \in [d]} |g(1 - j \cdot \theta)|^2). \tag{4}$$

If we think of $g(1)$ as the supremum and $g(1 - j \cdot \theta)$ as the average $\|g\|_2$—which we can formally do up to logarithmic factors by averaging over θ—this shows $|g(1)|^2 \le \widetilde{O}(d^2)\|g\|_2^2$. One natural idea to improve it is to use a smaller value d and a shorter linear combination [CP19]. However, $d = \tilde{\Omega}(k^2)$ for such a combination when g is approximately the degree $k-1$ Chebyshev polynomial. In this work, we use a geometric sequence to control c_j such that $\sum_j |c_j|^2 = O(d/k)$ instead of $O(d)$, which provides an improvement of a factor $\widetilde{O}(k)$ on R.

Then we bound $S = \widetilde{O}(k^2)$ for $g(t)$ at $|t| > 1$. The intuition is that given (4) holds for any $g(t)$ in terms of $g(t-\theta), \ldots, g(t-d \cdot \theta)$ with $\theta = \frac{2}{d}$, it implies $|g(t)|^2 \le \mathsf{poly}(k) \cdot \|g\|_2^2 \cdot e^{(t-1) \cdot O(d)}$ for $t > 1$. Combining this with an alternate bound $|g(t)|^2 \le \mathsf{poly}(k) \cdot \|g\|_2^2 \cdot (k \cdot t)^{O(k)}$ for $t > 1 + 1/k$, it completes the proof of Theorem 1.4 about S.

Finally we notice that we could improve the sample complexity in Theorem 1.5 to $\widetilde{O}(k) \log \frac{F}{\Delta'}$ using a biased distribution [CP19] to generate α. These results about k-Fourier-sparse signals appear in Section 6.

4 Our Filter Function

The main result is an explicit filter function H with compact support $\widehat{H}$ that is close to the box function on $[-1, 1]$ for any g satisfying the conditions in Theorem 1.2.

We show our filter function as follows.

Definition 4.1. *Given R, the growth rate S and an even constant C, we define the filter function*

$$H(t) = s_0 \cdot \left(\mathrm{sinc}(CR \cdot t)^{C \log R} \cdot \mathrm{sinc}\left(C \cdot S \cdot t \right)^C \cdot \mathrm{sinc}\left(\frac{C \cdot S}{2} \cdot t \right)^{2C} \cdots \mathrm{sinc}\left(C \cdot t \right)^{C \cdot S} \right) * \mathrm{rect}_2(t)$$

where $s_0 \in \mathbb{R}^+$ is a parameter to normalize $H(0) = 1$. On the other hand, its Fourier transform is

$$\widehat{H}(f) = s_0 \cdot \left(\operatorname{rect}_{CR}(f)^{*C \log R} * \operatorname{rect}_{C \cdot S}(f)^{*C} * \operatorname{rect}_{\frac{C \cdot S}{2}}(f)^{*2C} * \cdots * \operatorname{rect}_C(f)^{*CS} \right) \cdot \operatorname{sinc}(2t),$$

whose support size is $O(CR \cdot C \log R + CS \cdot C + \cdots + C \cdot C \cdot S) = O(R \log R + S \log S)$.

We prove Lemma 1.6 using $H(\alpha x)$ with a large constant C and a scale parameter $\alpha = \frac{1}{2} + \frac{1.2}{\pi CR}$. For convenience, we state the full version of Lemma 1.6 for $T = 1$ as follows.

Theorem 4.2. *Let $R, S > 0$, let C be a large even constant, and define $\alpha = (\frac{1}{2} + \frac{1.2}{\pi CR})$. Consider any function g satisfying the following two conditions:*

1. $\sup\limits_{t \in [-1,1]} |g(t)|^2 \leq R \cdot \|g\|_2^2$

2. And $|g(t)|^2 \leq \operatorname{poly}(R) \cdot \|g\|_2^2 \cdot |t|^S$ for $t \notin [-1, 1]$,

Then the filter function $H(\alpha x)$ is such that $H(\alpha x) \cdot g(x)$ satisfies

1. $\int_{-1}^{1} |g(x) \cdot H(\alpha x)|^2 \mathrm{d}x \geq 0.9 \int_{-1}^{1} |g(x)|^2 \mathrm{d}x$.

2. $\int_{-1}^{1} |g(x) \cdot H(\alpha x)|^2 \mathrm{d}x \geq 0.95 \int_{-\infty}^{\infty} |g(x) \cdot H(\alpha x)|^2 \mathrm{d}x$.

3. $|H(x)| \leq 1.01$ for any x.

For completeness, we show a few properties of H and finish the proof of Theorem 4.2 in Appendix 7.

5 Frequency Estimation

We show the algorithm for frequency estimation and prove Theorem 1.2 in this section. We fix $T = 1$ and use the definition $\|h\|_2^2 = \mathop{\mathbb{E}}\limits_{x \sim [-1,1]}[|h(x)|^2]$ to restate the theorem.

Theorem 5.1. *Given any $F > 0, \Delta > 0, R$, and $S > 0$, let $g(t)$ be a signal with the following properties:*

1. $\operatorname{supp}(\widehat{g}) \subseteq [f_0 - \Delta, f_0 + \Delta]$ where $f_0 \in [-F, F]$.

2. $\sup\limits_{t \in [-1,1]} [|g(t)|^2] \leq R \cdot \|g\|_2^2$.

3. $|g(t)|^2$ grows as at most $\operatorname{poly}(R) \cdot \|g\|_2^2 \cdot |t|^S$ for $t \notin [-1, 1]$.

Let $y(t) = g(t) + \eta(t)$ be the observable signal on $[-1, 1]$, where $\|\eta\|_2^2 \leq \epsilon \cdot \|g\|_2^2$ for a sufficiently small constant ϵ. For $\Delta' = \Delta + \widetilde{O}(R + S)$ and any δ, there exists an efficient algorithm that takes $O(R \log \frac{F}{\Delta' \cdot \delta})$ samples from $y(t)$ and outputs $\widetilde{f}$ satisfying $|f_0 - \widetilde{f}| \leq O(\Delta')$ with probability at least $1 - \delta$.

For convenience, we set $h_H(t) = h(t) \cdot H(\alpha t)$ for any signal $h(t)$ with the filter function H defined in Theorem 4.2 such that $y_H(t) = y(t) \cdot H(\alpha t)$.

Given the observation $y(t)$ with most Fourier mass concentrated around f_0, the main technical result in this section is an estimation of $e^{2\pi \mathbf{i} \beta f_0}$ through $y_H(\alpha) e^{2\pi i f_0 \beta} \approx y_H(\alpha + \beta)$.

Lemma 5.2. *Given parameters F, R, S, and Δ, let g be a signal satisfying the three conditions in Theorem 1.2 for some $f_0 \in [-F, F]$ and $\Delta' = \Delta + O(R \log R + S \log S)$.*

Let $y(t) = g(t) + \eta(t)$ be the observable signal on $[-1, 1]$ where the noise $\|\eta\|_2^2 \leq \epsilon \|g\|_2^2$ for a sufficiently small constant ϵ. There exist a constant γ and an algorithm such that for any $\beta \leq \frac{\gamma}{\Delta'}$, it takes $O(R)$ samples to output α satisfying $|y_H(\alpha)e^{2\pi i f_0 \beta} - y_H(\alpha + \beta)| \leq 0.3|y_H(\alpha)|$ with probability at least 0.6.

We show our algorithm in Algorithm 1. We finish the proof of Theorem 1.5 here and defer the proof of Lemma 5.2 to Section 5.1.

Algorithm 1 Obtain one good α

1: **procedure** OBTAINONEGOODSAMPLE$(R, y(t))$
2:    Let $m = C \cdot R$ for a large constant C.
3:    Take m random samples $x_1, \cdots, x_m$ uniform in $[-1, 1]$.
4:    Query $y(x_i)$ and compute $y_H(x_i) = y(x_i) \cdot H(x_i)$ for each i.
5:    Set a distribution D_m proportional to $|y_H(x_i)|^2$, i.e., $D_m(x_i) = \frac{|y_H(x_i)|^2}{\sum_{j=1}^{m} |y_H(x_j)|^2}$.
6:    Output $\alpha \sim D_m$.
7: **end procedure**

Proof of Theorem 5.1. From Lemma 5.2, $\frac{y_H(\alpha+\beta)}{y_H(\alpha)}$ gives a good estimation of $e^{2\pi i f_0 \beta}$ with probability 0.6 for any $\beta \leq \frac{\gamma}{\Delta'}$. We use the frequency search algorithm of Lemma 7.3 in [CKPS16] with the sampling procedure in Lemma 5.2. Because the algorithm in [CKPS16] uses the sampling procedure $O(\log \frac{F}{\Delta' \cdot \delta})$ times to return a frequency $\widetilde{f}$ satisfying $|\widetilde{f} - f_0| \leq \Delta'$ with prob. at least $1 - \delta$, the sample complexity is $O(R \cdot \log \frac{F}{\Delta' \cdot \delta})$. $\qquad\square$

5.1 Proof of Lemma 5.2

For $y_H(x) = g_H(x) + \eta_H(x)$, we have the following concentration lemma for estimation $g_H(x)$.

Claim 5.3. *Given any g satisfying the three conditions in Theorem 1.2 and any ε and δ, there exists $m = O(R \log \frac{1}{\delta}/\varepsilon^2)$ such that for m random samples $x_1, \ldots, x_m \sim [-1, 1]$, with probability $1 - \delta$,*

$$\frac{\sum_{i=1}^{m} |g_H(x_i)|^2}{m} \in [1 - \varepsilon, 1 + \varepsilon] \cdot \mathop{\mathbb{E}}_{x \sim [-1,1]}[|g_H(x)|^2].$$

Proof. Notice that $\frac{\sup_{x \sim [-1,1]}[|g_H(x)|^2]}{\mathop{\mathbb{E}}_{x \sim [-1,1]}[|g_H(x)|^2]} \leq 2R$. From the Chernoff bound in Lemma 2.1, $m = O(R \log \frac{1}{\delta}/\varepsilon^2)$ suffices to estimate $\|g_H\|_2^2$. $\qquad\square$

Next we consider the effect of noise $\eta_H(x_i)$ and $y_H(x_i)$.

Claim 5.4. *With probability 0.9 over m random samples in $[-1, 1]$, $\sum_{i=1}^{m} |y_H(x_i)|^2/m \geq 0.8\|g\|_2^2$.*

Proof. From Theorem 4.2, $\|g_H\|_2^2 \geq 0.95\|g\|_2^2$. Thus Claim 5.3 implies $\sum_{i=1}^{m} |g_H(x_i)|^2/m \geq 0.98 \cdot 0.95\|g\|_2^2$ for $m = O(R)$ with probability 0.99.

At the same time, because $\mathbb{E}[\sum_{i=1}^{m} |\eta_H(x_i)|^2/m] = \|\eta_H\|_2^2$, $\sum_{i=1}^{m} |\eta_H(x_i)|^2/m \leq 14\|\eta_H\|_2^2$ with probability at least $1 - \frac{1}{14}$ from the Markov inequality. This is also less than $14 \cdot 1.02^2 \|\eta\|_2^2 \leq 15\epsilon\|g\|_2^2$ from the upper bound on $H(t)$.

9

We have

$$\frac{1}{m}\sum_{i=1}^{m}|y_H(x_i)|^2 \geq \frac{1}{m}\sum_{i=1}^{m}\left(|g_H(x_i)|^2 - 2|g_H(x_i)|\cdot|\eta_H(x_i)| + |\eta_H(x_i)|^2\right).$$

By the Cauchy-Schwartz inequality, the cross term $\sum_{i=1}^{m}|g_H(x_i)|\cdot|\eta_H(x_i)| \leq (\sum_{i=1}^{m}|g_H(x_i)|^2)^{1/2}\cdot$ $(\sum_{i=1}^{m}|\eta_H(x_i)|^2)^{1/2}$. From all discussion above,

$$\frac{1}{m}\sum_{i=1}^{m}|y_H(x_i)|^2 \geq \left(0.93 - 2\sqrt{0.93\cdot15\epsilon}\right)\|g\|_2^2.$$

When ε is a small constant, it is at least $0.8\cdot\|g\|_2^2$. $\qquad\square$

We set $z(t) = y_H(t)\cdot e^{2\pi i f_0\beta} - y_H(t+\beta)$ for convenience and bound it as follows.

Claim 5.5. *Given any small constant γ, $\Delta' = \Delta + \mathsf{supp}(H)$, and $z(t) = y_H(t)\cdot e^{2\pi i f_0\beta} - y_H(t+\beta)$ for $\beta \leq \frac{\gamma}{\Delta'}$, $\|z\|_2^2 \lesssim (\gamma^2 + \epsilon)\|g\|_2^2$.*

Proof. Notice that $y_H = g_H + \eta_H$ where $\mathsf{supp}(\widehat{g_H}) \in [f_0 - \Delta, f_0 + \Delta]$ such that

$$\int_{f\notin[f_0-\Delta',f_0+\Delta']}|\widehat{y}(f)|^2\mathrm{d}f \leq \int_{-\infty}^{\infty}|\widehat{\eta_H}(f)|^2\mathrm{d}f = \int_{-\infty}^{\infty}|\eta_H(t)|^2\mathrm{d}t \leq 1.02^2\epsilon\int_{-1}^{1}|g(t)|^2\mathrm{d}t.$$

We bound $\|z\|_2^2$ through

$$\int_{-1}^{1}|z(t)|^2\mathrm{d}t \leq \int_{-\infty}^{\infty}|z(t)|^2\mathrm{d}t = \int_{-\infty}^{\infty}|\widehat{z}(f)|^2\mathrm{d}f = \int_{f_0-\Delta'}^{f_0+\Delta'}|\widehat{z}(f)|^2\mathrm{d}f + \int_{f\notin[f_0-\Delta',f_0+\Delta']}|\widehat{z}(f)|^2\mathrm{d}f.$$

Therefore we write

$$\int_{f_0-\Delta'}^{f_0+\Delta'}|\widehat{z}(f)|^2\mathrm{d}f = \int_{f_0-\Delta'}^{f_0+\Delta'}|\widehat{y_H}(f)\cdot e^{2\pi i f_0\beta} - \widehat{y_H}(f)\cdot e^{2\pi i f\beta}|^2\mathrm{d}f \leq \int_{f_0-\Delta'}^{f_0+\Delta'}|\widehat{y_H}(f)|^2\cdot|e^{2\pi i f_0\beta} - e^{2\pi i f\beta}|^2\mathrm{d}f.$$

Because $f \in [f_0 - \Delta', f_0 + \Delta']$ and $\beta \leq \frac{\gamma}{\Delta'}$, $|e^{2\pi i f_0\beta} - e^{2\pi i f\beta}| \leq 4\pi\gamma$. So

$$\int_{f_0-\Delta'}^{f_0+\Delta'}|\widehat{z}(f)|^2\mathrm{d}f \lesssim \gamma^2\int_{-\infty}^{+\infty}|\widehat{y_H}(f)|^2\mathrm{d}f = \gamma^2\int_{-\infty}^{+\infty}|y_H(t)|^2\mathrm{d}t \lesssim \gamma^2(1+2\epsilon)\int_{-1}^{1}|g(t)|^2\mathrm{d}t.$$

On the other hand,

$$\int_{f\notin[f_0-\Delta',f_0+\Delta']}|\widehat{z}(f)|^2\mathrm{d}f = \int_{f\notin[f_0-\Delta',f_0+\Delta']}|\widehat{y_H}(f)\cdot e^{2\pi i f_0\beta} - \widehat{y_H}(f)\cdot e^{2\pi i f\beta}|^2\mathrm{d}f$$

$$\leq 4\int_{f\notin[f_0-\Delta',f_0+\Delta']}|\widehat{y_H}(f)|^2\mathrm{d}f$$

$$\leq 4\int_{-\infty}^{+\infty}|\widehat{\eta_H}(f)|^2\mathrm{d}f = 4\int_{-\infty}^{+\infty}|\widehat{\eta_H}(t)|^2\mathrm{d}t$$

which is less than $5\epsilon\int_{-1}^{1}|g(t)|^2\mathrm{d}t$.

From all discussion above, $\int_{-1}^{1}|z(t)|^2\mathrm{d}t \lesssim (\gamma^2 + \epsilon)\int_{-1}^{1}|g(t)|^2\mathrm{d}t$. $\qquad\square$

For sufficiently small γ and ε, by Markov inequality, we have the following corollary.

10

Corollary 5.6. *For sufficiently small constants γ and ϵ, with probability 0.9 over m random samples in $[-1, 1]$, $\sum_{i=1}^{m} |z(x_i)|^2 \leq 0.01\|g\|_2^2$.*

Finally we finish the proof of Lemma 5.2.

Proof of Lemma 5.2. We assume Claim 5.4 and Corollary 5.6 hold in this proof, i.e.,

$$\sum_{i=1}^{m} |y_H(x_i)|^2/m \geq 0.8\|g\|_2^2 \text{ and } \sum_{i=1}^{m} |z(x_i)|^2/m \leq 0.01\|g\|_2^2.$$

For a random sample $\alpha \sim D_m$, we bound

$$\underset{\alpha \sim D_m}{\mathbb{E}} \left[\frac{|y_H(\alpha)e^{2\pi i f_0 \beta} - y_H(\alpha + \beta)|^2}{|y_H(\alpha)|^2} \right] = \underset{\alpha \sim D_m}{\mathbb{E}} \left[\frac{|z(\alpha)|^2}{|y_H(\alpha)|^2} \right] = \sum_{i=1}^{m} \frac{|z(x_i)|^2}{|y_H(x_i)|^2} \cdot \frac{|y_H(x_i)|^2}{\sum_{j=1}^{m} |y_H(x_j)|^2}.$$

This is $\frac{\sum_{i=1}^{m} |z(x_i)|^2}{\sum_{j=1}^{m} |y_H(x_j)|^2} \leq \frac{0.01}{0.8}$. Thus with probability 0.8, $\frac{|y_H(\alpha)e^{2\pi i f_0 \beta} - y_H(\alpha+\beta)|^2}{|y_H(\alpha)|^2}$ is less than $0.05/0.8 \leq$ 0.09. From all discussion above, $\frac{|y_H(\alpha)e^{2\pi i f_0 \beta} - y_H(\alpha+\beta)|}{|y_H(\alpha)|} \leq 0.3$ with probability 0.6. $\qquad\square$

6   Bounds on Fourier-sparse Signals

We consider $g(t) = \sum_{j=1}^{k} v_j e^{2\pi i f_j t}$ where each $f_j \in [f_0 - \Delta, f_0 + \Delta]$ in this section. The main result is to prove $R = \tilde{O}(k^3)$ and $S = \tilde{O}(k^2)$ for k arbitrary real frequencies. We restate Theorem 1.5 after fixing $T = 1$.

Theorem 6.1. *Given F, Δ, and k, let $g(t)$ be a k-Fourier-sparse signal centered around $f_0 \in [-F, F]$: $g(t) = \sum_{i \in [k]} v_i \cdot e^{2\pi i f_i t}$ where $f_i \in [f_0 - \Delta, f_0 + \Delta]$ and $y(t) = g(t) + \eta(t)$ be the observable signal on $[-1, 1]$, where $\|\eta\|_2^2 \leq \epsilon \cdot \|g\|_2^2$ for a sufficiently small constant ϵ.*

For any $\delta > 0$, there exist $\Delta' = \Delta + \tilde{O}(R)$ and an efficient algorithm that takes $O(k \log^2 k \log \frac{F}{\Delta' \cdot \delta})$ samples from $y(t)$ and outputs $\tilde{f}$ satisfying $|f_0 - \tilde{f}| \leq O(\Delta')$ with probability at least $1 - \delta$.

The main improvement is a biased distribution that saves the sample complexity from $O(R) \cdot \log \frac{F}{\Delta' \cdot \delta}$ to $\tilde{O}(k) \cdot \log \frac{F}{\Delta' \cdot \delta}$.

We provide the main technical lemma here and defer the proofs of Theorem 1.3, 1.4, and 6.1 to Appendix 8.

Theorem 6.2. *Given $z_1, \ldots, z_k$ with $|z_1| = |z_2| = \cdots = |z_k| = 1$, there exists a degree $d = O(k^2 \log k)$ polynomial $P(z) = \sum_{j=0}^{d} c(j) \cdot z^j$ satisfying*

1. *$P(z_i) = 0$ for each $i \in [k]$.*

2. *Coefficients $c(0) = \Omega(1)$, $c(j) = O(1)$ and $\sum_{j=1}^{d} |c(j)|^2 = O(k) \cdot |c(0)|^2$.*

Corollary 6.3. *Given any $g(t) = \sum_{j=1}^{k} v_j e^{2\pi i f_j t}$ and $\theta > 0$, there exist $d = O(k^2 \log k)$ and a sequence of coefficients $(\alpha_1, \ldots, \alpha_d)$ such that*

1. *$\alpha_j = O(1)$ for any $j = 1, \ldots, d$.*

2. *For any x (not necessarily in $[-1, 1]$), $g(x) = \sum_{j=1}^{d} \alpha_j \cdot g(x - j\theta)$.*

Proof. Given θ, we set $z_i = e^{-2\pi i f_j \theta}$ and apply Theorem 6.2 to obtain coefficients $c(0), \ldots, c(d)$. Then we set $\alpha_j = -c(j)/c(0)$. It is straightforward to verify the second property because of

$$e^{2\pi i f_j x} - \sum_j \alpha_j \cdot e^{2\pi i f_j (x - j\theta)} = 0.$$

□

The proof of Theorem 6.2 requires the following bound on the coefficients of residual polynomials, which is stated as Lemma 5.3 in [CKPS16].

Lemma 6.4. *Given $z_1, \ldots, z_k$, for any integer n, let $r_{n,k}(z) = \sum_{i=0}^{k-1} r_{n,k}^{(i)} \cdot z^i$ denote the residual polynomial of $r_{n,k} \equiv z^n \mod \prod_{j=1}^{k} (z - z_j)$. Then each coefficient in $r_{n,k}$ is bounded: $|r_{n,k}^{(i)}| \leq \binom{k-1}{i} \cdot \binom{n}{k-1}$ for $n \geq k$ and $|r_{n,k}^{(i)}| \leq \binom{k-1}{i} \cdot \binom{|n|+k-1}{k-1}$ for $n < 0$.*

We finish the proof of Theorem 6.2 here.

Proof. Let C_0 be a large constant and $d = 5 \cdot k^2 \log k$. We use $\mathcal{P}$ to denote the following subset of polynomials with bounded coefficients:

$$\left\{ \sum_{j=0}^{d} \alpha_j \cdot 2^{-j/k} \cdot z^j \,\middle|\, \alpha_0, \ldots, \alpha_d \in [-C_0, C_0] \cap \mathbb{Z} \right\}.$$

For each polynomial $P(z) \in \mathcal{P}$, we rewrite $P(z) \mod \prod_{j=1}^{k}(z - z_j)$ as

$$\sum_{j=0}^{d} \alpha_j \cdot 2^{-j/k} \cdot \left(z^j \mod \prod_{j=1}^{k}(z - z_j) \right) = \sum_{i=0}^{k-1} \left(\sum_{j=0}^{d} \alpha_j \cdot 2^{-j/k} \cdot r_{n,k}^{(i)} \right) z^i.$$

The coefficient $\sum_{j=0}^{d} \alpha_j \cdot 2^{-j/k} \cdot r_{n,k}^{(i)}$ is bounded by

$$\sum_{j=0}^{d} C_0 \cdot 2^{-j/k} \cdot 2^k j^{k-1} \leq d \cdot C_0 \cdot 2^k \cdot d^k \leq d^{2k}.$$

Then we apply the pigeonhole principle on the $(2C_0 + 1)^d$ polynomials in $\mathcal{P}$ after module $\prod_{j=1}^{d}(z - z_j)$: there exist $m > (2C_0 + 1)^{0.9d}$ polynomials $P_1, \ldots, P_m$ such that each coefficient of $(P_i - P_j) \mod \prod_{j=1}^{k}(z - z_j)$ is d^{-2k} small from the counting

$$\frac{(2C_0 + 1)^d}{(d^{2k}/4d^{-2k})^k} > (2C_0 + 1)^{0.9d}.$$

Because $m > (2C_0 + 1)^{0.9d}$, there exists $j_1 \in [m]$ and $j_2 \in [m] \setminus \{j_1\}$ such that the lowest monomial z^l with different coefficients in P_{j_1} and P_{j_2} satisfies $l \leq 0.1d$. Eventually we set

$$P(z) = z^{-l} \cdot \left(P_{j_1}(z) - P_{j_2}(z) \right) - \left(z^{-l} \mod \prod_{j=1}^{k}(z - z_j) \right) \cdot \left(P_{j_1}(z) - P_{j_2}(z) \mod \prod_{j=1}^{k}(z - z_j) \right)$$

to satisfy the first property $P(z_1) = P(z_2) = \cdots = P(z_k) = 0$. We prove the second property in the rest of this proof.

We bound every coefficient in $\left(z^{-l} \mod \prod_{j=1}^{k}(z - z_j)\right) \cdot \left(P_{j_1}(z) - P_{j_2}(z) \mod \prod_{j=1}^{k}(z - z_j)\right)$ by

$$k \cdot \text{max-coefficient}\left(z^{-l} \mod \prod_{j=1}^{k}(z - z_j)\right) \cdot \text{max-coefficient}\left(P_{j_1}(z) - P_{j_2}(z) \mod \prod_{j=1}^{k}(z - z_j)\right),$$

which is less than $k \cdot 2^k (l+k)^{k-1} \cdot d^{-2k} \leq d \cdot 2^k d^{k-1} \cdot d^{-2k} \leq d^{-0.5k}$ from Lemma 6.4 and the above discussion.

On the other hand, the constant coefficient in $z^{-l} \cdot \left(P_{j_1}(z) - P_{j_2}(z)\right)$ is at least $2^{-l/k} \geq 2^{-0.1d/k} = k^{-0.5k}$ because z^l is the smallest monomial with different coefficients in P_{j_1} and P_{j_2} from $\mathcal{P}$. Thus the constant coefficient $|C(0)|^2$ of $P(z)$ is at least $0.5 \cdot 2^{-2l/k}$.

Next we upper bound the sum of the rest of the coefficients $\sum_{j=1}^{d} |C(j)|^2$ by

$$\sum_{j=1}^{d} (2C_0 \cdot 2^{-(l+j)/k} + d^{-0.5k})^2 \leq 2 \cdot 4C_0^2 \sum_{j=1}^{d} 2^{-2(l+j)/k} + 2 \cdot \sum_{j=1}^{d} d^{-0.5k \cdot 2} \lesssim k \cdot 2^{-2l/k},$$

which demonstrates the second property after normalizing $C(0)$ to 1. $\qquad\square$

Acknowledgement

References

[AGS03] A. Akavia, S. Goldwasser, and S. Safra. Proving hard-core predicates using list decoding. *FOCS*, 44:146–159, 2003.

[AKM+19] Haim Avron, Michael Kapralov, Cameron Musco, Christopher Musco, Ameya Velingker, and Amir Zandieh. A universal sampling method for reconstructing signals with simple fourier transforms. In *Proceedings of the 51st annual ACM symposium on Theory of computing (STOC 2019)*, 2019.

[BM86] Y. Bresler and A. Macovski. Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(5):1081–1089, Oct 1986.

[Che52] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23:493–507, 1952.

[CKPS16] Xue Chen, Daniel M. Kane, Eric Price, and Zhao Song. Fourier-sparse interpolation without a frequency gap. In *Foundations of Computer Science(FOCS), 2016 IEEE 57th Annual Symposium on*, 2016.

[CP19] Xue Chen and Eric Price. Active regression via linear-sample sparsification. In *the 32nd Annual Conference on Learning Theory (COLT 2019)*, 2019.

[GGI+02] Anna C Gilbert, Sudipto Guha, Piotr Indyk, S Muthukrishnan, and Martin Strauss. Near-optimal sparse Fourier representations via sampling. In *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*, pages 152–161. ACM, 2002.

[GMS05] Anna C Gilbert, S Muthukrishnan, and Martin Strauss. Improved time bounds for near-optimal sparse Fourier representations. In *Optics & Photonics 2005*, pages 59141A–59141A. International Society for Optics and Photonics, 2005.

[HIKP12] Haitham Hassanieh, Piotr Indyk, Dina Katabi, and Eric Price. Simple and practical algorithm for sparse Fourier transform. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 1183–1194. SIAM, 2012.

[IK14] Piotr Indyk and Michael Kapralov. Sample-optimal Fourier sampling in any constant dimension. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 514–523. IEEE, 2014.

[Man92] Y. Mansour. Randomized interpolation and approximation of sparse polynomials. *ICALP*, 1992.

[Moi15] Ankur Moitra. The threshold for super-resolution via extremal functions. In *STOC*, 2015.

[Pro95] R Prony. Essai experimental et analytique. *J. de l'Ecole Polytechnique*, 1795.

[PS15] Eric Price and Zhao Song. A robust sparse Fourier transform in the continuous setting. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 583–600. IEEE, 2015.

7 Properties of the Filter function

We show basic properties of our filter function in Appendix 7.1 and prove Theorem 4.2 in Appendix 7.2.

7.1 Properties of H

We use two bounds on the sinc function:

1. For any $|x| \geq \frac{1.2}{\pi}$, $\text{sinc}(x) \leq \frac{1}{\pi|x|}$.

2. For any $|x| \leq \frac{1.2}{\pi}$, $\text{sinc}(x) \in [1 - \frac{\pi^2 |x|^2}{6}, 1 - \frac{\pi^2 |x|^2}{10}]$.

Without loss of generality, we assume both R and C are powers of 2 and $R \geq S$ (otherwise set $R = S$). Recall that C is even in this section.

We use $g(t)$ to denote the product of sinc functions in $H(t)$ for convenience:

$$g(t) = \left(\text{sinc}(CR \cdot t)^{C \log R} \cdot \text{sinc} \left(C \cdot S \cdot t \right)^C \cdot \text{sinc} \left(\frac{C \cdot S}{2} \cdot t \right)^{2C} \cdots \text{sinc} \left(C \cdot t \right)^{C \cdot S} \right)$$

We fix $l = \log_2(S)$ in this section and rewrite $g(t)$ as

$$\text{sinc}(CR \cdot t)^{C \log R} \cdot \prod_{j=0}^{l} \text{sinc} \left(2^{-j} \cdot C \cdot S \cdot t \right)^{2^j \cdot C}.$$

Before we show the properties of H, we consider the tail of $g(t)$.

Claim 7.1. *1. $g(t) = \Theta(1)$ for $|t| \le \frac{1.2}{\pi CR \cdot \sqrt{C \log R}}$.*

2. $g(t) = e^{-\Theta(|CR \cdot t|^2 \log R)}$ for $|t| \in [\frac{1.2}{\pi CR \cdot \sqrt{C \log R}}, \frac{1.2}{\pi CR}]$.

3. $g(t) \le (\frac{1}{\pi \cdot CR \cdot |t|})^{C \log R}$ for $|t| \in [\frac{1.2}{\pi CR}, \frac{1.2}{\pi C \cdot S}]$.

4. For any $i \in [l]$, $g(t) \le (\frac{1}{\pi \cdot CR \cdot |t|})^{C \log R} \cdot 1.2^{-(2^{i+1}-1)C}$ for any $|t| \in [\frac{1.2 \cdot 2^{i-1}}{\pi C \cdot S}, \frac{1.2 \cdot 2^i}{\pi C \cdot S}]$.

5. $g(t) \le (\frac{1}{\pi CR \cdot t})^{C \log R} \cdot \prod_{j=0}^{l} (\frac{1}{\pi 2^{-j} \cdot C \cdot S \cdot t})^{2^j \cdot C}$ for $|t| \ge \frac{1.2 \cdot 2^l}{\pi C \cdot S} = \frac{1.2}{C\pi}$.

Proof. We first bound $\text{sinc}(CR \cdot t)^{C \log R}$ then bound $\prod_{j=0}^{l} \text{sinc}\left(2^{-j} \cdot C \cdot S \cdot t\right)^{2^j \cdot C}$.

1. For $|t| \le \frac{1.2}{\pi CR}$, from the second property of sinc functions,

$$\text{sinc}(CR \cdot t) \in \left[1 - \frac{\pi^2 |CRt|^2}{6}, 1 - \frac{\pi^2 |CRt|^2}{10}\right] \Rightarrow \text{sinc}(CR \cdot t)^{C \log R} = \Theta(1) \text{ for } |t| \le \frac{1.2}{\pi CR \cdot \sqrt{C \log R}}$$

and

$$\text{sinc}(CR \cdot t)^{C \log R} = e^{-\Theta(|CR \cdot t|^2 \log R)} \text{ for } t \in [\frac{1.2}{\pi CR \cdot \sqrt{C \log R}}, \frac{1.2}{\pi CR}].$$

2. For $|t| \ge \frac{1.2}{\pi CR}$, from the first property of sinc functions,

$$\text{sinc}(CR \cdot t)^{C \log R} \le \left(\frac{1}{\pi \cdot CR \cdot |t|}\right)^{C \log R}.$$

Then we bound the tail of the product of sinc functions.

1. For $|t| \le \frac{1.2}{\pi C \cdot S}$,

$$\text{sinc}\left(2^{-j} \cdot C \cdot S \cdot t\right)^{2^j \cdot C} \in \left[\left(1 - \frac{\pi^2 \cdot |2^{-j} \cdot C \cdot S \cdot t|^2}{6}\right)^{2^j \cdot C}, \left(1 - \frac{\pi^2 \cdot |2^{-j} \cdot C \cdot S \cdot t|^2}{10}\right)^{2^j \cdot C}\right].$$

Notice that $\pi^2 \cdot |2^{-j} \cdot C \cdot S \cdot t|^2$ is less than $1.2^2 \cdot 2^{-2j}$. Thus $\text{sinc}\left(2^{-j} \cdot C \cdot S \cdot t\right)^{2^j \cdot C} = \left(1 - \Theta(2^{-j})\right)^C$ and their products over j is

$$\prod_{j=0}^{l} \text{sinc}\left(2^{-j} \cdot C \cdot S \cdot t\right)^{2^j \cdot C} = \left(1 - \Theta\left(\sum_{j=0}^{l} 2^{-j}\right)\right)^C = \Theta(1)^C = \Theta(1).$$

2. Let us fix $i \le l$ and consider $\text{sinc}\left(2^{-j} \cdot C \cdot S \cdot t\right)^{2^j \cdot C}$ for $|t| \in [\frac{1.2 \cdot 2^{i-1}}{\pi C \cdot S}, \frac{1.2 \cdot 2^i}{\pi C \cdot S}]$. By the first property of sinc function, for $j \le i$,

$$\text{sinc}\left(2^{-j} \cdot C \cdot S \cdot t\right)^{2^j \cdot C} \le \left(\frac{1}{\pi \cdot 2^{-j} \cdot C \cdot S \cdot |t|}\right)^{2^j \cdot C} \le \left(\frac{1}{1.2 \cdot 2^{-j+i}}\right)^{2^j \cdot C} \le 1.2^{-2^j \cdot C}.$$

For $j > i$, we use the same analysis with the second property of the sinc function:

$$\text{sinc}\left(2^{-j} \cdot C \cdot S \cdot t\right)^{2^j \cdot C} \in \left[\left(1 - \frac{\pi^2 \cdot |2^{-j} \cdot C \cdot S \cdot t|^2}{6}\right)^{2^j \cdot C}, \left(1 - \frac{\pi^2 \cdot |2^{-j} \cdot C \cdot S \cdot t|^2}{10}\right)^{2^j \cdot C}\right]$$

where $\pi^2 \cdot |2^{-j} \cdot C \cdot S \cdot t|^2$ is at least $1.2^2 \cdot 2^{-2(j-i)}$. Hence the product is

$$\prod_{j=0}^{l} \text{sinc}\left(2^{-j} \cdot C \cdot S \cdot t\right)^{2^j \cdot C} \le 1.2^{-\sum_{j=0}^{i} 2^j \cdot C} \cdot \prod_{j=i+1}^{l} \left(1 - \frac{1.2^2 \cdot 2^{-2(j-i)}}{6}\right)^{2^j \cdot C} \le 1.2^{-(2^{i+1}-1)C}.$$

15

We get the tail bounds by combining the above discussion of $\text{sinc}(CR \cdot t)^{C \log R}$ and $\prod_{j=0}^{l} \text{sinc}\left(2^{-j} \cdot C \cdot S \cdot t\right)^{2^j \cdot C}$ together. □

Since $H(t) = s_0 \cdot g(t) * \text{rect}_2(t) = s_0 \cdot \int_{t-1/2}^{t+1/2} g(x) \mathrm{d}x$, we have the following bounds on $H(t)$ based on Claim 7.1.

Lemma 7.2. *For any constant $C \geq 4$,*

1. $s_0 = \Theta(\pi C R \cdot \sqrt{C \log R})$.

2. $H(t) = 1 \pm 0.01$ *for* $|t| \leq \frac{1}{2} - \frac{1.2}{\pi C R}$.

3. $H(t) \lesssim \frac{s_0}{S} \cdot R^{-C/4}$ *for* $|t| \in [\frac{1}{2} + \frac{1.2}{\pi C R}, \frac{1}{2} + \frac{1.2}{\pi C \cdot S}]$.

4. $H(t) \lesssim s_0 \cdot R^{-C/4} \cdot 1.2^{-2^i C}$ *for* $|t| \in [\frac{1}{2} + \frac{1.2 \cdot 2^{i-1}}{\pi C \cdot S}, \frac{1}{2} + \frac{1.2 \cdot 2^i}{\pi C \cdot S}]$ *of any* $i \leq [l]$.

5. $H(t) \leq s_0 \cdot \left(\frac{1}{1.2 \pi C R \cdot (|t| - \frac{1}{2})}\right)^{C \log R} \cdot \left(\frac{1}{C \pi \cdot (|t| - \frac{1}{2})}\right)^{CS}$ *for* $t \geq \frac{1}{2} + \frac{1.2}{C\pi}$.

Proof. We bound the integration of different intervals of $g(t)$ as follows:

1. $\int_{\frac{-1.2}{\pi C R}}^{\frac{1.2}{\pi C R}} g(x) \mathrm{d}x = \int_{\frac{-1.2}{\pi C R \cdot \sqrt{C \log R}}}^{\frac{1.2}{\pi C R \cdot \sqrt{C \log R}}} g(x) \mathrm{d}x + 2 \int_{\frac{1.2}{\pi C R \cdot \sqrt{C \log R}}}^{\frac{1.2}{\pi C R}} e^{-\Theta(|CR \cdot x|^2 \log R)} \mathrm{d}x = \Theta\left(\frac{1}{\pi C R \cdot \sqrt{C \log R}}\right)$.

2. $\int_{\frac{1.2}{\pi C R}}^{\frac{1.2}{\pi C \cdot S}} g(x) \mathrm{d}x \leq \int_{\frac{1.2}{\pi C R}}^{\frac{1.2}{\pi C \cdot S}} \left(\frac{1}{\pi \cdot C R \cdot x}\right)^{C \log R} \mathrm{d}x \leq \frac{1.2}{\pi C \cdot S} \cdot 1.2^{-C \log R}$.

3. For any $i \in [l]$ of $l = \log_2 S$,

$$
\int_{\frac{1.2 \cdot 2^{i-1}}{\pi C \cdot S}}^{\frac{1.2 \cdot 2^i}{\pi C \cdot S}} g(x) \mathrm{d}x \leq \int_{\frac{1.2 \cdot 2^{i-1}}{\pi C \cdot S}}^{\frac{1.2 \cdot 2^i}{\pi C \cdot S}} \left(\frac{1}{\pi \cdot CR \cdot x}\right)^{C \log R} \cdot 1.2^{-(2^{i+1}-1)C} \mathrm{d}x
$$

$$
\leq \frac{1.2 \cdot 2^i}{\pi C \cdot S} \cdot \left(\frac{S}{1.2 \cdot 2^{i-1} R}\right)^{C \log R} \cdot 1.2^{-(2^{i+1}-1)C}
$$

$$
\leq \frac{1.2 \cdot 2^i}{\pi C \cdot S} \cdot R^{-C/4} \cdot 1.2^{-2^i C}.
$$

4. For $|t| \geq \frac{1.2}{C\pi}$,

$$
\int_t^{t+1} g(x) \mathrm{d}x \leq \int_t^{t+1} \left(\frac{1}{\pi CR \cdot x}\right)^{C \log R} \cdot \prod_{j=0}^{l} \left(\frac{1}{\pi 2^{-j} \cdot C \cdot S \cdot x}\right)^{2^j \cdot C} \mathrm{d}x
$$

$$
\leq \left(\frac{1}{\pi CR \cdot t}\right)^{C \log R} \cdot \left(\frac{1}{\pi C \cdot t}\right)^{2^l \cdot C}
$$

$$
\leq \left(\frac{1}{\pi CR \cdot t}\right)^{C \log R} \cdot \left(\frac{1}{\pi C \cdot t}\right)^{CS}.
$$

Next we prove all claims in this lemma.

1. For s_0, notice that

$$
\int_{-1/2}^{1/2} g(x) \mathrm{d}x \leq \int_{\frac{-1.2}{\pi C R}}^{\frac{1.2}{\pi C R}} g(x) \mathrm{d}x + \int_{|x| \in (\frac{1.2}{\pi C R}, 1/2]} g(x) \mathrm{d}x = \Theta\left(\frac{1}{\pi CR \cdot \sqrt{C \log R}}\right) + O\left(\frac{1.2}{\pi C \cdot S} \cdot 1.2^{-C \log R}\right),
$$

which also indicates $s_0 \in [1, 1 + 10^{-3}] \cdot 1/\left(\int_{\frac{-1.2}{\pi C R}}^{\frac{1.2}{\pi C R}} g(x) \mathrm{d}x\right)$.

16

2. When $|t| < \frac{1}{2} - \frac{1.2}{\pi CR}$, $H(t) = s_0 \cdot \left(\int_{\frac{-1.2}{\pi CR}}^{\frac{1.2}{\pi CR}} g(x)\mathrm{d}x + \int_{[t-1/2,t+1/2]\setminus[\frac{-1.2}{\pi CR},\frac{1.2}{\pi CR}]} g(x)\mathrm{d}x \right)$, which is in

$s_0 \cdot [1, 1 + 10^{-3}] \cdot \int_{\frac{-1.2}{\pi CR}}^{\frac{1.2}{\pi CR}} g(x)\mathrm{d}x \subseteq [1 - 0.01, 1 + 0.01]$.

3. When $|t| \in [\frac{1}{2} - \frac{1.2}{\pi CR}, \frac{1}{2} + \frac{1.2}{\pi CR}]$, $H(t) \in [0, 1]$.

4. When $|t| \in [\frac{1}{2} + \frac{1.2}{\pi CR}, \frac{1}{2} + \frac{1.2}{\pi C \cdot S}]$,

$$H(t) \leq s_0 \cdot \left(\int_{\frac{1.2}{\pi CR}}^{\frac{1.2}{\pi C \cdot S}} g(x)\mathrm{d}x + \sum_{j=1}^{l} \int_{\frac{1.2 \cdot 2^{j-1}}{\pi C \cdot S}}^{\frac{1.2 \cdot 2^{j}}{\pi C \cdot S}} g(x)\mathrm{d}x + \int_{\frac{1.2}{\pi}}^{\frac{1.2}{\pi}+1} g(x)\mathrm{d}x \right) \leq 2s_0 \cdot \int_{\frac{1.2}{\pi CR}}^{\frac{1.2}{\pi C \cdot S}} g(x)\mathrm{d}x.$$

5. When $|t| \in [\frac{1}{2} + \frac{1.2 \cdot 2^{i-1}}{\pi C \cdot S}, \frac{1}{2} + \frac{1.2 \cdot 2^{i}}{\pi C \cdot S}]$ of a positive integer $i < l$,

$$H(t) \leq s_0 \cdot \left(\sum_{j=i}^{l} \int_{\frac{1.2 \cdot 2^{j-1}}{\pi C \cdot S}}^{\frac{1.2 \cdot 2^{j}}{\pi C \cdot S}} g(x)\mathrm{d}x + \int_{\frac{1.2}{\pi}}^{\frac{1.2}{\pi}+1} g(x)\mathrm{d}x \right) \leq 2s_0 \cdot \frac{1.2}{C\pi} \cdot R^{-C/4} \cdot 1.2^{-2^{i}C}.$$

6. When $t > \frac{1}{2} + \frac{1.2}{C\pi}$, we use the bound in the last item of the above discussion.

$\square$

7.2 Proof of Theorem 4.2

We finish the proof of Theorem 4.2 using Lemma 7.2 for $\alpha = \frac{1}{2} + \frac{1.2}{\pi CR}$. Without loss of generality, we assume $R \geq S$ in this proof (otherwise set $R = S$).

We first show

$$\int_{-1}^{1} |g(x) \cdot H(\alpha x)|^2 \mathrm{d}x \geq 0.9 \int_{-1}^{1} |g(x)|^2 \mathrm{d}x.$$

From the second property of H in Lemma 7.2, $H(\alpha x) \geq 1 - 0.01$ for any $|x| \leq \frac{\frac{1}{2} - \frac{1.2}{\pi CR}}{\alpha} = 1 - \frac{2.4}{\pi CR + 2.4}$ such that

$$\int_{-1 + \frac{2.4}{\pi CR/2 + 1.2}}^{1 - \frac{2.4}{\pi CR/2 + 1.2}} |g(x) \cdot H(\alpha x)|^2 \mathrm{d}x \geq 0.99^2 \int_{-1 + \frac{2.4}{\pi CR/2 + 1.2}}^{1 - \frac{2.4}{\pi CR/2 + 1.2}} |g(x)|^2 \mathrm{d}x.$$

At the same time, $|g(t)|^2 \leq R \cdot \mathbb{E}_{x \sim [-1,1]}[|g(x)|^2] = R/2 \cdot \int_{-1}^{1} |g(x)|^2 \mathrm{d}x$ for any $t \in [-1, 1]$. This indicates

$$\int_{-1 + \frac{2.4}{\pi CR + 2.4}}^{1 - \frac{2.4}{\pi CR + 2.4}} |g(x)|^2 \mathrm{d}x \geq (1 - \frac{R/2 \cdot 2.4}{\pi CR + 2.4}) \int_{-1}^{1} |g(x)|^2 \mathrm{d}x.$$

The first property follows from these two inequalities.

In the rest of this proof, we apply Lemma 7.2 to prove:

$$\int_{-\infty}^{-1} |g(x) \cdot H(\alpha x)|^2 \mathrm{d}x + \int_{1}^{\infty} |g(x) \cdot H(\alpha x)|^2 \mathrm{d}x \leq 0.04 \int_{-1}^{1} |g(x)|^2 \mathrm{d}x.$$

We split $\int_{1}^{\infty} |g(x) \cdot H(\alpha x)|^2 \mathrm{d}x$ into several intervals:

$$\int_{1}^{(\frac{1}{2} + \frac{1.2}{\pi C \cdot S})/\alpha} |g(x) \cdot H(\alpha x)|^2 \mathrm{d}x + \sum_{i=1}^{\log_2 S} \int_{(\frac{1}{2} + \frac{1.2 \cdot 2^{i-1}}{\pi C \cdot S})/\alpha}^{(\frac{1}{2} + \frac{1.2 \cdot 2^{i}}{\pi C \cdot S})/\alpha} |g(x) \cdot H(\alpha x)|^2 \mathrm{d}x + \int_{(\frac{1}{2} + \frac{1.2}{\pi C})/\alpha}^{\infty} |g(x) \cdot H(\alpha x)|^2 \mathrm{d}x.$$

17

In the first two terms, we rewrite $|g(t)| \leq \mathsf{poly}(R) \cdot \|g\|_2 \cdot t^S$ as $\mathsf{poly}(R) \cdot \|g\|_2 \cdot e^{(t-1)S}$. By the third and fourth properties of $H(t)$ in Lemma 7.2, their summations is less than $0.01\|g\|_2^2$. For the last term, given the last property of $H(t)$ in Lemma 7.2 and a large constant C, we have

$$H(\alpha t) \leq s_0 \cdot \left(\frac{1}{1.2R}\right)^{C \log R} \cdot \left(\frac{1}{2t}\right)^S \text{ when } t \geq \left(\frac{1}{2} + \frac{1.2}{\pi C}\right)/\alpha.$$

It is straightforward to verify that $\int_1^\infty |g(x) \cdot H(\alpha x)|^2 \mathrm{d}x \leq 0.02 \cdot \|g\|_2^2$.

The last property follows from the upper bounds in Lemma 7.2.

8 Omitted Proofs in Section 6

We first prove Theorem 1.5 then finish the proof of Theorem 1.3 and 1.4 in Appendix 8.2 and 8.3 separately.

8.1 Proof of Theorem 1.5

We finish the proof of Theorem 1.5 in this section. The only difference compared to Theorem 1.2 is to use a biased distribution D such that we could improve the sample complexity to $\widetilde{O}(k \log \frac{F}{\Delta' \epsilon})$.

How to Generate Samples. We will use a distribution D not uniform on $[-1, 1]$ to generate the random samples. For m samples $x_1, \cdots, x_m \sim D$, we assign a weight $w_i = \frac{1}{2m \cdot D(x_i)}$ for each sample x_i such that for any function h,

$$\mathop{\mathbb{E}}_{x_1, \cdots, x_m \sim D}\left[\sum_{i=1}^m w_i |h(x_i)|^2\right] = \sum_{i=1}^m \mathop{\mathbb{E}}_{x_i \sim D}\left[\frac{1}{2m \cdot D(x_i)}|h(x_i)|^2\right] = \sum_{i=1}^m \mathop{\mathbb{E}}_{x \sim [-1,1]}\left[\frac{1}{m}|h(x_i)|^2\right] = \|h\|_2^2.$$

[CP19] presented an explicit distribution D such that $\tilde{O}(k)$ samples could guarantee $\sum_{i=1}^m w_i|g(x_i)|^2$ is close to $\|g\|_2^2$ with high probability. For completeness, we show it with our improved bound R.

Lemma 8.1. *Given the sparsity k, there exists a constant c such that the distribution*

$$D_\mathcal{F}(x) = \begin{cases} \frac{c}{(1-|x|)\log k}, & \text{for } |x| \leq 1 - \frac{1}{k^2 \log^2 k} \\ c \cdot k^2 \log k, & \text{for } |x| > 1 - \frac{1}{k^2 \log^2 k} \end{cases}$$

guarantees for any k-Fourier-sparse signal g, $\sup\limits_{x \in [-1,1]} \frac{1}{2D(x)} \cdot \frac{|g(x)|^2}{\|g\|_2^2} = O(k \log^2 k)$.

Moreover, $m = O(\frac{k \log^2 k \log \frac{1}{\delta}}{\epsilon^2})$ samples $x_1, \cdots, x_m$ from D with weights $w_i = \frac{1}{2m \cdot D(x_i)}$ for $i \in [m]$ guarantee that, with probability at least $1 - \delta$,

$$\sum_{i=1}^m w_i \cdot |g(x_i)|^2 \in [1 \pm \epsilon] \cdot \|g\|_2^2.$$

Proof. Given D and the k-Fourier-sparse signal g, let $z(x)$ denote $\frac{|g(x)|^2}{2D(x)}$ for $x \in [-1, 1]$. We have $\mathbb{E}_{x \sim D}[z(x)] = \mathbb{E}_{x \sim [-1,1]}[|g(x)|^2] = \|g\|_2^2$ and $\sup\limits_{x \in \mathsf{supp}(D)} \frac{z(x)}{\mathbb{E}_{x' \sim D}[z(x')]} = O(k \log^2 k)$. We apply the Chernoff bound in Lemma 2.1 on the random variables $z(x_1), \cdots, z(x_m)$ to obtain the statement. $\square$

18

Similar to Lemma 5.2, we state the following version for Fourier-sparse signals.

Lemma 8.2. *Given the sparsity k, f_0 and Δ, let g be a k-Fourier-sparse signal $g(t) = \sum_{i \in [k]} v_i \cdot e^{2\pi \mathbf{i} f_i t}$ with $f_i \subseteq [f_0 - \Delta, f_0 + \Delta]$ and $\Delta' = \Delta + O(\frac{R \log k + k^2 \log^2 k}{T})$.*

Let $y(t) = g(t) + \eta(t)$ be the observable signal on $[-1, 1]$ where the noise $\|\eta\|_2^2 \leq \epsilon \|g\|_2^2$ for a sufficiently small constant ϵ. There exist a constant γ and an algorithm such that for any $\beta \leq \frac{\gamma}{\Delta'}$, it takes $O(k \log^2 k)$ samples to output α satisfying $|y_H(\alpha)e^{2\pi \mathbf{i} f_0 \beta} - y_H(\alpha + \beta)| \leq 0.3|y_H(\alpha)|$ with probability at least 0.6.

We show our algorithm in Algorithm 2. We finish the proof of Theorem 1.5.

Algorithm 2 Obtain one good α

1: **procedure** OBTAINONEGOODSAMPLE($k, y(t)$)
2:     Let $m = C \cdot k \log^2 k$ for a large constant C.
3:     Take m samples $x_1, \cdots, x_m$ from the distribution D in Lemma 8.1.
4:     Assign a weight $w_i = \frac{1}{2m \cdot D(x_i)}$ for each sample x_i.
5:     Query $y(x_i)$ and compute $y_H(x_i) = y(x_i) \cdot H(x_i)$ for each i.
6:     Set a distribution D_m proportional to $w_i \cdot |y_H(x_i)|^2$, i.e., $D_m(x_i) = \frac{w_i \cdot |y_H(x_i)|^2}{\sum_{j=1}^m w_j \cdot |y_H(x_j)|^2}$.
7:     Output $\alpha \sim D_m$.
8: **end procedure**

Proof of Theorem 6.1.   From Lemma 8.2, $\frac{y_H(\alpha + \beta)}{y_H(\alpha)}$ gives a good estimation of $e^{2\pi \mathbf{i} f_0 \beta}$ with probability 0.6 for any $\beta \leq \frac{\gamma}{\Delta'}$. We use the frequency search algorithm of Lemma 7.3 in [CKPS16] with the sampling procedure in Lemma 8.2. Because the algorithm in [CKPS16] uses the sampling procedure $O(\log \frac{F}{\Delta' \cdot \delta})$ times to return a frequency $\widetilde{f}$ satisfying $|\widetilde{f} - f_0| \leq \Delta'$ with prob. at least $1 - \delta$, the sample complexity is $O(k \log^2 k \cdot \log \frac{F}{\Delta' \cdot \delta})$. $\qquad\square$

8.2 Proof of Theorem 1.3

We bound R of k-sparse-Fourier signals in this section. We first state the technical result to prove the upper bound R.

Theorem 8.3. *Given any $k > 0$, there exists $d = O(k^2 \log k)$ such that for any $g(x) = \sum_{j=1}^k v_j \cdot e^{2\pi \mathbf{i} f_j \cdot x}$, any $t \in \mathbb{R}$, and any $\theta > 0$,*

$$|g(t)|^2 \leq O(k) \cdot \left(\sum_{j=1}^d |g(t + j \cdot \theta)|^2 \right).$$

Proof of Theorem 8.3.   Given k frequencies $f_1, \cdots, f_k$ and θ, we set $z_1 = e^{2\pi \mathbf{i} f_1 \cdot \theta}, \cdots, z_k = e^{2\pi \mathbf{i} f_k \cdot \theta}$. Let $C(0), \cdots, C(d)$ be the coefficients of the degree d polynomial $P(z)$ in Theorem 6.2. We have

$$\sum_{j=0}^d C(j) \cdot g(t + j \cdot \theta) = \sum_{j=0}^d C(j) \sum_{j' \in [k]} v_{j'} \cdot e^{2\pi \mathbf{i} f_{j'} (t + j\theta)}$$

$$= \sum_{j=0}^d C(j) \sum_{j' \in [k]} v_{j'} \cdot e^{2\pi \mathbf{i} f_{j'} t} \cdot z_{j'}^j = \sum_{j' \in [k]} v_{j'} \cdot e^{2\pi \mathbf{i} f_{j'} t} \sum_{j=0}^d C(j) \cdot z_{j'}^j = 0.$$

19

Hence for every $i \in [k]$,

$$- C(0) \cdot g(t) = \sum_{j=1}^{d} C(j) \cdot g(t + j \cdot \theta). \tag{5}$$

By Cauchy-Schwartz inequality, we have

$$|C(0)|^2 \cdot |g(t)|^2 \leq \left(\sum_{j=1}^{d} |C(j)|^2 \right) \cdot \left(\sum_{j=1}^{d} |g(t + j \cdot \theta)|^2 \right). \tag{6}$$

From the second property of $C(0), \cdots, C(d)$ in Theorem 6.2, $|g(t)|^2 \leq O(k) \cdot \left(\sum_{j=1}^{d} |g(t + j \cdot \theta|^2 \right).$ $\qquad \square$

We finish the proof of Theorem 1.3 bounding R by the above relation. For convenience, we restate it for $T = 1$.

Theorem 8.4. *For any* $g(t) = \sum_{j=1}^{k} v_j e^{2\pi i f_j t}$,

$$\frac{\sup_{x \in [-1,1]} |g(x)|^2}{\mathbb{E}_{x \in [-1,1]}[|g(x)|^2]} = O(k^3 \log^2 k).$$

Proof. We prove

$$|g(t)|^2 = O(k^3 \log^2 k) \int_{t}^{1} |g(x)|^2 \mathrm{d}x \text{ for any } t \leq 0,$$

which indicates $|g(t)|^2 = O(k^3 \log^2 k) \cdot \mathbb{E}_{x \sim [-1,1]}[|g(x)|^2]$. By symmetry, it also implies that $|g(t)|^2 = O(k^3 \log^2 k) \cdot \mathbb{E}_{x \sim [-1,1]}[|g(x)|^2]$ for any $t \geq 0$.

We use Theorem 8.3 on $g(t)$:

$$\frac{1-t}{d} \cdot |g(t)|^2 \leq O(k) \cdot \int_{\theta=0}^{\frac{1-t}{d}} \sum_{j \in [d]} |g(t + j\theta)|^2 \mathrm{d}\theta$$

$$\lesssim k \sum_{j \in [d]} \int_{\theta=0}^{\frac{1-t}{d}} |g(t + j\theta)|^2 \mathrm{d}\theta$$

$$\lesssim k \sum_{j \in [d]} \frac{1}{j} \cdot \int_{\theta'=0}^{\frac{(1-t)j}{d}} |g(t + \theta')|^2 \mathrm{d}\theta'$$

$$\lesssim k \sum_{j \in [d]} \frac{1}{j} \cdot \int_{x=-1}^{1} |g(x)|^2 \mathrm{d}x$$

$$\lesssim k \log k \cdot \int_{x=-1}^{1} |g(x)|^2 \mathrm{d}x.$$

From all discussion above, we have $|g(t)|^2 \lesssim dk \log k \cdot \mathbb{E}_{x \in [-1,1]}[|g(x)|^2]$. $\qquad \square$

8.3 Growth outside of the observation

We prove Theorem 1.4 which bounds $S = \tilde{O}(k^2)$ in this section. We divide the proof into two parts for $|x| \leq 1 + 1/k$ and $|x| > 1 + 1/k$ separately after fixing $T = 1$.

Lemma 8.5. *For any $g(t) = \sum_{j=1}^{k} v_j \cdot e^{2\pi i f_j t}$, there exists a constant C_1 such that for any $x \geq 1$, $|g(x)| \lesssim \mathsf{poly}(k) \cdot \|g\|_2 \cdot C_1^{(x-1) \cdot k^2 \log k}$.*

Remark 8.6. *The growth of Chebyshev polynomial is $e^{\Theta(k\sqrt{x-1})}$ for $x = 1 + O(1/k^2)$.*

Proof. Let $d = O(k^2 \log k)$ denote the length of the linear combination in Corollary 6.3 and $\theta = \frac{2}{d}$. Given $g(t)$ and θ, we use $\alpha_1, \cdots, \alpha_d$ to denote the coefficients of the linear combination of $g(t), g(t - \theta), \ldots, g(t - d\theta)$ in Corollary 6.3. For convenience, we use C_0 to denote the upper bound on the coefficients α_j.

We use induction to prove that for some $C = O(1)$, for any l,

$$\text{for any } x \in (1, 1 + \frac{2l}{d}], |g(x)| \leq C \cdot dk^{1.5} \log k \cdot \|g\|_2 \cdot (2C_0)^l. \tag{7}$$

For base case $l = 1$, from Corollary 6.3, $g(x) = \sum_{j=1}^{d} \alpha_j \cdot g(x - j\theta)$ where $x - j\theta \in [-1, 1]$. Because each $|g(x - j\theta)| \leq C \cdot k^{1.5} \log k \cdot \|g\|_2$ from Theorem 1.3, we have

$$|g(x)| \leq \sum_{j=1}^{d} |\alpha_j| \cdot |g(x - j\theta)| \leq C \cdot C_0 \cdot d \cdot k^{1.5} \log k \cdot \|g\|_2.$$

Suppose (7) is true for any $x \in (1, 1 + \frac{2l}{d}]$. Let us consider $x \in (1 + \frac{2l}{d}, 1 + \frac{2(l+1)}{d}]$. We still have $g(x) = \sum_{j=1}^{d} \alpha_j \cdot g(x - j\theta)$ where each $x - j\theta \in (1 + \frac{2(l-j)}{d}, 1 + \frac{2(l+1-j)}{d}]$. This indicates

$$|g(x)| \leq \sum_{j=1}^{d} |\alpha_j| \cdot |g(x - j\theta)|$$

$$\leq C_0 \sum_{j=1}^{d} |g(x - j\theta)|$$

$$\leq C_0 \sum_{j=1}^{l} |g(x - j\theta)| + C_0 \sum_{j=l+1}^{d} |g(x - j\theta)|$$

$$\leq C_0 \sum_{j=1}^{l} C \cdot dk^{1.5} \log k \cdot \|g\|_2 \cdot (2C_0)^{l+1-j} + C_0(d - j) \cdot C \cdot k^{1.5} \log k \cdot \|g\|_2.$$

$$\leq C_0^{l+1} \cdot C \cdot dk^{1.5} \log k \cdot \|g\|_2 \cdot \sum_{j=1}^{l} 2^{l+1-j} + C_0 d \cdot C \cdot k^{1.5} \log k \cdot \|g\|_2.$$

$$\leq C_0^{l+1} \cdot C \cdot dk^{1.5} \log k \cdot \|g\|_2 (2^{l+1} - 2) + C_0 d \cdot C \cdot k^{1.5} \log k \cdot \|g\|_2$$

$$\leq C_0^{l+1} \cdot C \cdot dk^{1.5} \log k \cdot \|g\|_2 \cdot 2^{l+1}.$$

$\square$

For completeness, we bound the growth rate of $|t| > 1 + 1/k$ here, which is a reformulation of Lemma 5.5 in [CKPS16].

Lemma 8.7. *For any $g(t) = \sum_{j=1}^{k} v_j e^{2\pi i f_j t}$ and any $|t| > 1$,*

$$|g(t)|^2 \lesssim k^3 \cdot (3k \cdot t)^k \cdot \|g\|_2^2.$$

Proof. We fix $t > 1$ in this proof. Let $\theta = 1/k$ and $n = \left[(t + 1/2)/\theta\right]$ such that $t - n\theta \in [-1/2, -1/2 + \theta]$ and $t - (n-k)\theta \in [1/2, 1/2 + \theta]$. We first show the coefficients $C_0, \cdots, C_{k-1}$ in

$$\sum_{j=0}^{k-1} C_j \cdot z^j = z^n \mod \prod_{j=1}^{k} (z - e^{2\pi i f_j \theta})$$

satisfy $g(t) = \sum_{l=0}^{k-1} C_j \cdot g\big(t - (n - l)\theta\big)$. Let $z_j = e^{2\pi i f_j \theta}$ such that $z_j^n = \sum_{j=0}^{k-1} C_j \cdot z^j$. For $g(t) = \sum_{j=1}^{k} v_j e^{2\pi i f_j t}$, we rewrite it as

$$\sum_{j=1}^{k} v_j e^{2\pi i f_j (t-n\theta)} \cdot e^{2\pi i f_j n\theta} = \sum_{j=1}^{k} v_j e^{2\pi i f_j (t-n\theta)} \cdot z_j^n$$

$$= \sum_{j=1}^{k} v_j e^{2\pi i f_j (t-n\theta)} \cdot \sum_{l=0}^{k-1} C_l \cdot z_j^l$$

$$= \sum_{l=0}^{k-1} C_l \cdot \sum_{j=1}^{k} v_j e^{2\pi i f_j (t-n\theta)} z_j^l$$

$$= \sum_{l=0}^{k-1} C_l \cdot g(t - n\theta + l\theta).$$

Thus $|g(t)|^2 \leq (\sum_{j=0}^{k-1} |C_j|^2) \cdot (\sum_{l=0}^{k-1} |g(t - n\theta + l\theta)|^2)$.

Since $g(t - n\theta + l\theta) \in [-2/3, 2/3]$, $|g(t - n\theta + l\theta)|^2 \lesssim k \operatorname*{\mathbb{E}}_{x \in [-1,1]}[|g(x)|^2]$ from [CP19]. On the other hand, $|C_j| \leq \binom{k-1}{j}\binom{n}{k-1} \leq (2n)^{k-1}$ from Lemma 6.4.

From all discussion above,

$$|g(t)|^2 \lesssim k \cdot (2n)^{k-1} \cdot k^2 \operatorname*{\mathbb{E}}_{x \in [-1,1]}[|g(x)|^2] \lesssim k^3 (3kt)^k \cdot \operatorname*{\mathbb{E}}_{x \in [-1,1]}[|g(x)|^2].$$

$\square$

Proof of Theorem 1.4. We combine Lemma 8.5 and 8.7: For $x \leq 1 + 1/k$, $C_1^{(x-1)k^2 \log k} = e^{(x-1)k^2 \log k \log C_1} = x^{O(k^2 \log k)}$. For $x > 1 + 1/k$, $(3kx)^k$ is still less than $x^{O(k^2 \log k)}$. $\square$