Fourier-sparse interpolation without a frequency gap

Xue Chen[*]
xchen@cs.utexas.edu
The University of Texas at Austin

Daniel M. Kane
dakane@cs.ucsd.edu
University of California, San Diego

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

Zhao Song
zhaos@utexas.edu
The University of Texas at Austin

January 17, 2017

Abstract

We consider the problem of estimating a Fourier-sparse signal from noisy samples, where the sampling is done over some interval $[0, T]$ and the frequencies can be "off-grid". Previous methods for this problem required the gap between frequencies to be above $1/T$, the threshold required to robustly identify individual frequencies. We show the frequency gap is not necessary to estimate the signal as a whole: for arbitrary k-Fourier-sparse signals under ℓ_2 bounded noise, we show how to estimate the signal with a constant factor growth of the noise and sample complexity polynomial in k and logarithmic in the bandwidth and signal-to-noise ratio.

As a special case, we get an algorithm to interpolate degree d polynomials from noisy measurements, using $O(d)$ samples and increasing the noise by a constant factor in ℓ_2.

Contents

1   Introduction

In an interpolation problem, one can observe $x(t) = x^*(t) + g(t)$, where $x^*(t)$ is a structured signal and $g(t)$ denotes noise, at points t_i of one's choice in some interval $[0, T]$. The goal is to recover an estimate $\widetilde{x}$ of x^* (or of x). Because we can sample over a particular interval, we would like our approximation to be good on that interval, so for any function $y(t)$ we define

$$\|y\|_T^2 = \frac{1}{T} \int_0^T |y(t)|^2 \mathrm{d}t.$$

to be the ℓ_2 error on the sample interval. For some parameters C and δ, we would then like to get

$$\|\widetilde{x} - x^*\|_T \leq C \|g\|_T + \delta \|x^*\|_T \tag{1}$$

while minimizing the number of samples and running time. Typically, we would like C to be $O(1)$ and to have δ be very small (either zero, or exponentially small). Note that, if we do not care about changing C by $O(1)$, then by the triangle inequality it doesn't matter whether we want to estimate x^* or x (i.e. we could replace the LHS of (1) by $\|\widetilde{x} - x\|_T$).

Of course, to solve an interpolation problem one also needs x^* to have structure. One common form of structure is that x^* have a sparse Fourier representation. We say that a function x^* is k-Fourier-sparse if it can be expressed as a sum of k complex exponentials:

$$x^*(t) = \sum_{j=1}^k v_j e^{2\pi \mathbf{i} f_j t}.$$

for some $v_j \in \mathbb{C}$ and $f_j \in [-F, F]$, where F is the "bandlimit". Given F, T, and k, how many samples must we take for the interpolation (1)?

If we ignore sparsity and just use the bandlimit, then Nyquist sampling and Shannon-Whittaker interpolation uses $FT + 1/\delta$ samples to achieve (1). Alternatively, in the absence of noise, x^* can be found from $O(k)$ samples by a variety of methods, including Prony's method from 1795 or Reed-Solomon syndrome decoding [Mas69], but these methods are not robust to noise.

If the signal is periodic with period T—i.e., the frequencies are multiples of $1/T$—then we can use sparse discrete Fourier transform methods, which take $O(k \log^c(FT/\delta))$ time and samples (e.g. [GGI+02, HIKP12a, IKP14]). If the frequencies are not multiples of $1/T$ (are "off the grid"), then the discrete approximation is only k/δ sparse, making the interpolation less efficient; and even this requires that the frequencies be well separated.

A variety of algorithms have been designed to recover off-grid frequencies directly, but they require the minimum gap among the frequencies to be above some threshold. With frequency gap at least $1/T$, we can achieve a k^c approximation factor using $O(FT)$ samples [Moi15], and with gap above $O(\log^2 k)/T$ we can get a constant approximation using $O(k \log^c(FT/\delta))$ samples and time [PS15].

Having a dependence on the frequency gap is natural. If two frequencies are very close together—significantly below $1/T$—then the corresponding complex exponentials will be close on $[0, T]$, and hard to distinguish in the presence of noise. In fact, from a lower bound in [Moi15], below $1/T$ frequency gap one cannot recover the frequencies in the presence of noise as small as $2^{-\Omega(k)}$. The lower bound proceeds by constructing two signals using significantly different frequencies that are exponentially close over $[0, T]$.

But if two signals are so close, do we need to distinguish them? Such a lower bound doesn't apply to the interpolation problem, it just says that you can't solve it by finding the frequencies.

Our question becomes: can we benefit from Fourier sparsity in a regime where we can't recover the individual frequencies?

We answer in the affirmative, giving an algorithm for the interpolation using $O(\text{poly}(k \log(FT/\delta)))$ samples. Our main theorem is the following:

Theorem 1.1. *Let* $x(t) = x^*(t) + g(t)$, *where* x^* *is* k-*Fourier-sparse signal with frequencies in* $[-F, F]$. *Given samples of* x *over* $[0, T]$ *we can output* $\widetilde{x}(t)$ *such that with probability at least* $1 - 2^{-\Omega(k)}$,

$$\|\widetilde{x} - x^*\|_T \lesssim \|g\|_T + \delta \|x^*\|_T.$$

Our algorithm uses $\text{poly}(k, \log(1/\delta)) \cdot \log(FT)$ *samples and* $\text{poly}(k, \log(1/\delta)) \cdot \log^2(FT)$ *time. The output* $\widetilde{x}$ *is* $\text{poly}(k, \log(1/\delta))$-*Fourier-sparse signal.*

Relative to previous work, this result avoids the need for a frequency gap, but loses a polynomial factor in the sample complexity and time. We lose polynomial factors in a number of places; some of these are for ease of exposition, but others are challenging to avoid.

Degree d polynomials are the special case of d-Fourier-sparse functions in the limit of $f_j \to 0$, by a Taylor expansion. This is a regime with no frequency gap, so previous sparse Fourier results would not apply but Theorem 1.1 shows that $\text{poly}(d \log(1/\delta))$ samples suffices. In fact, in this special case we can get a better polynomial bound:

Theorem 1.2. *For any degree* d *polynomial* $P(t)$ *and an arbitrary function* $g(t)$, *Procedure* RO-BUSTPOLYNOMIALLEARNING *in Algorithm 5 takes* $O(d)$ *samples from* $x(t) = P(t) + g(t)$ *over* $[0, T]$ *and reports a degree* d *polynomial* $Q(t)$ *in time* $O(d^\omega)$ *such that, with probability at least* $99/100$,

$$\|P(t) - Q(t)\|_T^2 \lesssim \|g(t)\|_T^2.$$

where $\omega < 2.373$ *is matrix multiplication exponent [Str69],[CW87],[Wil12].*

We also show how to reduce the failure probability to an arbitrary $p > 0$ with $O(\log(1/p))$ independent repetitions, in Theorem 4.5.

Although we have not seen such a result stated in the literature, our method is quite similar to one used in [CDL13]. Since d samples are necessary to interpolate a polynomial without noise, the result is within constant factors of optimal.

One could apply Theorem 1.2 to approximate other functions that are well approximated by polynomials or piecewise polynomials. For example, a Gaussian of standard deviation at least σ can be approximated by a polynomial of degree $O((\frac{T}{\sigma})^2 + \log(1/\delta))$; hence the same bound applies as the sample complexity of improper interpolation of a positive mixture of Gaussians.

1.1 Related work

Sparse discrete Fourier transforms. There is a large literature on sparse discrete Fourier transforms. Results generally are divided into two categories: one category of results that carefully choose measurements that allow for sublinear recovery time, including [GGI+02, GMS05, HIKP12b, Iwe13, HIKP12a, IK14, IKP14, Kap16]. The other category of results expect randomly chosen measurements and show that a generic recovery algorithm such as ℓ_1 minimization will work with high probability; these results often focus on proving the Restricted Isometry Property [CRT06, RV08, Bou14, HR15]. At the moment, the first category of results have better theoretical sample complexity and running time, while results in the second category have better failure probabilities and empirical performance. Our result falls in the first category. The best results here can achieve $O(k \log n)$ samples [IK14], $O(k \log^2 n)$ time [HIKP12b], or within $\log \log n$ factors of both [IKP14].

For signals that are not periodic, the discrete Fourier transform will not be sparse: it takes k/δ frequencies to capture a $1 - \delta$ fraction of the energy. To get a better dependence on δ, one has to consider frequencies "off the grid", i.e. that are not multiples of $1/T$.

Off the grid. Finding the frequencies of a signal with sparse Fourier transform off the grid has been a question of extensive study. The first algorithm was by Prony in 1795, which worked in the noiseless setting. This was refined by classical algorithms like MUSIC [Sch81] and ESPRIT [RPK86], which empirically work better with noise. Matrix pencil [BM86] is a method for computing the maximum likelihood signal under Gaussian noise and evenly spaced samples. The question remained how accurate the maximum likelihood estimate is; [Moi15] showed that it has an $O(k^c)$ approximation factor if the frequency gap is at least $1/T$.

Now, the above results all use FT samples, which is analogous to n in the discrete setting. This can be decreased down till $O(k)$ by only looking at a subset of time, i.e. decreasing T; but doing so increases the frequency gap needed for decent robustness results.

A variety of works have studied how to adapt sparse Fourier techniques from the discrete setting to get sublinear sample complexity; they all rely on the minimum separation among the frequencies to be at least c/T for $c \geq 1$. [TBSR13] showed that a convex program can recover the frequencies exactly in the noiseless setting, for $c \geq 4$. This was improved in [CF14] to $c \geq 2$ for complex signals and $c \geq 1.87$ for real signals. [CF14] also gave a result for $c \geq 2$ that was stable to noise, but this required the signal frequencies to be placed on a finely spaced grid. [YX15] gave a different convex relaxation that empirically requires smaller c in the noiseless setting. [DB13] used model-based compressed sensing when $c = \Omega(1)$, again without theoretical noise stability. Note that, in the noiseless setting, exact recovery can be achieved without any frequency separation using Prony's method or Berlekamp-Massey syndrome decoding [Mas69]; the benefit of the above results is that a convex program might be robust to noise, even if it has not been proven to be so.

In the noisy setting, [FL12] gave an extension of Orthogonal Matching Pursuit (OMP) that can recover signals when $c = \Omega(k)$, with an approximation factor $O(k)$, and a few other assumptions. Similarly, [BCG$^+$14] gave a method that required $c = \Omega(k)$ and was robust to certain kinds of noise. [HK15] got the threshold down to $c = O(1)$, in multiple dimensions, but with approximation factor $O(FTk^{O(1)})$.

[TBR15] shows that, under Gaussian noise and with separation $c \geq 4$, a semidefinite program can optimally estimate $x^*(t_i)$ at evenly spaced sample points t_i from observations $x^*(t_i) + g(t_i)$. This is somewhat analogous to our setting, the differences being that (a) we want to estimate the signal over the entire interval, not just the sampled points, (b) our noise g is adversarial, so we cannot hope to reduce it—if g is also k-Fourier-sparse, we cannot distinguish x^* and g, and of course (c) we want to avoid requiring frequency separation.

In [PS15], we gave the first algorithm with $O(1)$ approximation factor, finding the frequencies when $c \gtrsim \log(1/\delta)$, and the signal when $c \gtrsim \log(1/\delta) + \log^2 k$.

Now, all of the above results algorithms are designed to recover the frequencies; some of the ones in the noisy setting then show that this yields a good approximation to the overall signal (in the noiseless setting this is trivial). Such an approach necessitates $c \geq 1$: [Moi15] gave a lower bound, showing that any algorithm finding the frequencies with approximation factor $2^{o(k)}$ must require $c \geq 1$.

Thus, in the current literature, we go from not knowing how to get any approximation for $c < 1$, to getting a polynomial approximation at $c = 1$ and a constant approximation at $c \gtrsim \log^2 k$. In this work, we show how to get a constant factor approximation to the signal regardless of c.

Polynomial interpolation. Our result is a generalization of robust polynomial interpolation, and in Theorem 1.2 we construct an optimal method for polynomial interpolation as a first step toward interpolating Fourier-sparse signals.

Our result here can be seen as essentially an extension of a technique shown in [CDL13]. The focus of [CDL13] is on the setting where sample points x_i are chosen independently, so $\Theta(d \log d)$ samples are necessary. One of their examples, however, shows essentially the same thing as our Corollary 4.2. From this, getting our theorem is not difficult.

The recent work [GZ16] looks at robust polynomial interpolation in a different noise model, featuring ℓ_∞ bounded noise with some outliers. In this setting they can get a stronger ℓ_∞ guarantee on the output than is possible in our setting.

Nyquist sampling. The classical method for learning bandlimited signals uses Nyquist sampling—i.e., samples at rate $1/F$, for FT points—and interpolates them using Shannon-Nyquist interpolation. This doesn't require any frequency gap, but also doesn't benefit from sparsity like sparse Fourier transform-based techniques. As discussed in [PS15], on the signal $x(t) = 1$ it takes $FT + O(1/\delta)$ samples to get δ error on average. Our dependence is logarithmic on both those terms.

1.2 Our techniques

Previous results on sparse Fourier transforms with robust recovery all required a frequency gap. So consider the opposite situation, where all the frequencies converge to zero and the coefficients are adjusted to keep the overall energy fixed. If we take a Taylor expansion of each complex exponential, then the signal will converge to a degree k polynomial. So robust polynomial interpolation is a necessary subproblem for our algorithm.

Polynomial interpolation. Let $P(x)$ be a degree d polynomial, and suppose that we can query $f(x) = P(x) + g(x)$ over the interval $[-1, 1]$, where g represents adversarial noise. We would like to query f at $O(d)$ points and output a degree d polynomial $Q(x)$ such that $\|P - Q\| \lesssim \|g\|$, where we define $\|h\|^2 := \int_{-1}^{1} |h(x)|^2 \mathrm{d}x$.

One way to do this would be to sample points $S \subset [-1, 1]$ uniformly, then output the degree d polynomial Q with the smallest empirical error

$$\|P + g - Q\|_S^2 := \frac{1}{|S|} \sum_{x \in S} |(P + g - Q)(x)|^2$$

on the observed points. If $\|R\|_S \approx \|R\|$ for all degree d polynomials R, in particular for $P - Q$, then since usually $\|g\|_S \lesssim \|g\|$ by Markov's inequality, the result follows.

This has two problems: first, uniform sampling is poor because polynomials like Chebyshev polynomials can have most of their energy within $O(1/d^2)$ of the edges of the interval. This necessitates $\Omega(d^2)$ uniform samples before $\|R\|_S \approx \|R\|$ with good probability on a single polynomial. Second, the easiest method to extend from approximating one polynomial to approximating all polynomials uses a union bound over a net exponential in d, which would give an $O(d^3)$ bound.

To fix this, we need to bias our sampling toward the edges of the interval and we need our sampling to not be iid. We partition $[-1, 1]$ into $O(d)$ intervals $I_1, \ldots, I_n$ so that the interval containing each x has width at most $O(\sqrt{1 - x^2})$, except for the $O(1/d^2)$ size regions at the edges. For any degree d polynomial R and any choice of n points $x_i \in I_i$, the appropriately weighted empirical energy is close to $\|R\|$. This takes care of both issues with uniform sampling. If the points

are chosen uniformly at random from within their intervals, then $\|g\|$ is probably bounded as well, and the empirically closest degree d polynomial Q will satisfy our requirements.

This result is shown in Section 4.

Clusters. Many previous sparse Fourier transform algorithms start with a one-sparse recovery algorithm, then show how to separate frequencies to get a k-sparse algorithm by reducing to the one-sparse case. Without a frequency gap, we cannot hope to reduce to the one-sparse case; instead, we reduce to individual clusters of nearby frequencies.

Essentially the problem is that one *cannot* determine all of the high-energy frequencies of a function x only by sampling it on a bounded interval, as some of the frequencies might cancel each other out on this interval. We also cannot afford to work merely with the frequencies of the truncation of x to the interval $[0, T]$, as the truncation operation will spread the frequencies of x over too wide a range. To fix this problem, we must do something in between the two. In particular, we instead study $x \cdot H$ for a judiciously chosen function H. We want H to approximate the indicator function of the interval $[0, T]$ and have small Fourier-support, $\mathrm{supp}(\widehat{H}) \subset [-k^c/T, k^c/T]$. By using some non-trivial lemmas about the growth rate of x^*, we can show that the difference between $x \cdot H$ on $\mathbb{R}$ and the truncation of x to $[0, T]$ has small L^2 mass, so that we can use the former as a substitute for the latter.

On the other hand, the Fourier transform of $x \cdot H$ is the convolution $\widehat{x} * \widehat{H}$, which has most of its mass within $\mathrm{poly}(k)/T$ of the frequencies of x^*. Although it is impossible to determine the individual frequencies of x^*, we can hope to identify $O(k)$ intervals each of length $\mathrm{poly}(k)/T$ so that all but a small fraction of the energy of $\widehat{x}$ is contained within these intervals.

Note that many of these intervals will represent not individual frequencies of x^*, but small clusters of such frequencies. Furthermore, some frequencies of x^* might not show up in these intervals either because they are too small, or because they cancel out other frequencies when convolved with $\widehat{H}$.

One-cluster recovery. Given our notion of clusters, we start looking at Fourier-sparse interpolation in the special case of *one-cluster recovery*. This is a generalization of one-sparse recovery where we can have multiple frequencies, but they all lie in $[f - \Delta, f + \Delta]$ for some base frequency f and bandwidth $\Delta = k^c/T$. Because all the frequencies are close to each other, values $x(a)$ and $x(a + \beta)$ will tend to have ratio close to $e^{2\pi \mathbf{i} f \beta}$ when β is small enough. We find that $\beta < \frac{1}{\Delta \sqrt{T\Delta}}$ is sufficient, which lets us figure out a frequency $\widetilde{f}$ with $|\widetilde{f} - f| \leq \Delta\sqrt{T\Delta} = k^{O(1)}/T$.

Once we have the frequency $\widetilde{f}$, we can consider $x'(t) = x(t)e^{-2\pi \mathbf{i} \widetilde{f}}$. This signal is k-Fourier-sparse with frequencies bounded by $k^{O(1)}/T$. By taking a Taylor approximation to each complex exponential[1], can show x^* is δ-close to $P(t)e^{2\pi \mathbf{i} \widetilde{f}}$ for a degree $d = O(k^c + k\log(1/\delta))$ polynomial P. Thus we could apply our polynomial interpolation algorithm to recover the signal.

k-cluster frequency estimation. Reminiscent of algorithms such as [HIKP12a, PS15], we choose random variables $\sigma \approx T/k^c$, $a \in [0, 1]$, and $b \in [0, 1/\sigma]$ and look at $v \in \mathbb{C}^{k^c}$ given by

$$v_i = (x \cdot H)(\sigma(i - a))e^{-2\pi \mathbf{i} \sigma b i}G(i)$$

---

[1]There is a catch here, that the coefficients of the exponentials are potentially unbounded, if the frequencies are arbitrarily close together. We first use Gram determinants to show that the signal is δ-close to one with frequency gap $\delta 2^{-k}$, and coefficients at most $2^k/\delta$.

where G is a filter function. That is, G has compact support ($\mathrm{supp}(G) \subset [-k^c, k^c]$), and $\widehat{G}$ approximates an interval of length $\Theta(\frac{2\pi}{k})$. In other words, G is the same as $\widehat{H}$ with different parameters: an interval convolved with itself k^c times, multiplied by a sinc function.

We alias v down to $O(k)$ dimensions and take the discrete Fourier transform, getting $\widehat{u}$. It has been implicit in previous work—and we make it explicit—that $\widehat{u}_j$ is equal to $z_{\sigma a}$ for a vector z defined by

$$\widehat{z} = (\widehat{x} * \widehat{H}) \cdot \widehat{G}^{(j)}_{\sigma,b}$$

where $\widehat{G}^{(j)}_{\sigma,b}$ is a particular permutation of $\widehat{G}$. In particular, $\widehat{G}^{(j)}_{\sigma,b}$ has period $1/\sigma$, and approximates an interval of size $\frac{1}{\sigma B}$ within each period.

In previous work, when σ and b were chosen randomly, each individual frequency would have a good chance of being the only frequency preserved in $\widehat{z}$, and we could apply one-sparse recovery by choosing a variety of a. Without a frequency gap we can't quite say that: we pick $1/\sigma \gg \Delta$ so that the entire cluster usually lands in the same bin, but then nearby clusters can also often land in the same bin. Fortunately, it is still usually true that only nearby clusters will collide. Since our 1-cluster algorithm works when the signal frequencies are nearby, we apply it to find a frequency approximation within $\frac{\sqrt{T/\sigma}}{\sigma} = k^{O(1)}/T$ of the cluster.

The above algorithm recovers each individual frequency with constant probability. By repeating it $O(\log k)$ times, with high probability we find a list L of $O(k)$ frequencies within $k^{O(1)}/T$ of each significant cluster.

k-sparse recovery. Because different clusters aren't anywhere close to orthogonal, we can't simply approximate each cluster separately and add them up. Instead, given the list L of candidate frequencies, we consider the $O(kd)$-dimensional space of functions

$$\widetilde{x}(t) := \sum_{\widetilde{f} \in L} \sum_{i=0}^{d} \alpha_{\widetilde{f},i} t^i e^{2\pi \mathbf{i} \widetilde{f} t}$$

where $d = O(k^{O(1)} + \log(1/\delta))$. We then take a bunch of random samples of x, and choose the $\widetilde{x}(t)$ minimizing the empirical error using linear regression. This regression can be made slightly faster using oblivious subspace embeddings [CW13], [NN13], [Woo14],[CNW15].

Our argument to show this works is analogous to the naive method we considered for polynomial recovery. Similarly to the one-cluster setting, using Taylor approximations and Gram determinants, we can show that this space includes a sufficiently close approximation to x. Since polynomials are the limit of sparse Fourier as frequencies tend to zero, these functions are arbitrarily close to $O(kd)$-Fourier-sparse functions. Hence we know that the maximum of $|\widetilde{x}(t)|$ is at most a $\mathrm{poly}(kd)$ factor larger than its average over $[0, T]$. Using a net argument, this shows $\mathrm{poly}(kd)$ samples are sufficient to find a good approximation to the nearest function in our space.

Growth rate of Fourier-sparse signals. We need that $\frac{1}{\sqrt{T}} \|x^* \cdot H\|_2 \approx \|x^*\|_T$, where H approximates the interval $1_{[0,T]}$. Because H has support size k^c/T, it has a transition region of size $T/k^{c'}$ at the edges, and it decays as $(t/T)^{-k^{c''}}$ for $t \gg T$. The difference between $\frac{1}{\sqrt{T}} \|x^* \cdot H\|_2$ and $\|x^*\|_T$ involves two main components: mass in the transition region that is lost, and mass outside the sampling interval that is gained. To show the approximation, we need that $|x^*(t)| \lesssim \widetilde{O}(k^2) \|x^*\|_T$ within the interval and $|x^*(t)| \lesssim (kt/T)^{O(k)} \|x^*\|_T$ outside.

8

We outline the bound of $\max_{t \in [0,T]} |x^*(t)|$ in terms of its average $\|x^*\|_T$ to bound $|x^*(t)|$ within the interval. Notice that we can assume $|x^*(0)| = \max_{t \in [0,T]} |x^*(t)|$: if $t^* = \arg\max_{t \in [0,T]} |x^*(t)|^2$ is not 0 or T, we can rescale the two intervals $[0, t^*]$ and $[t^*, T]$ to $[0, T]$ separately. Then we show that for any t', there exist $m = \widetilde{O}(k^2)$ and constants $C_1, \cdots, C_m$ such that $x^*(0) = \sum_{j \in [m]} C_j \cdot x^*(j \cdot t')$. Then we take the integration of t' over $[0, T/m]$ to bound $|x^*(0)|^2$ by its average. For any outside $t > T$, we follow this approach to show $x^*(t) = \sum_{j \in [k]} C_j \cdot x^*(t_j)$ where $t_j \in [0, T]$ and $|C_j| \le \mathrm{poly}(k) \cdot (kt/T)^{O(k)}$ for each $j \in [k]$. These results are shown in Section 5.

1.3 Organization

This paper is organized as follows. We provide a brief overview about signal recovery in Section 2. We introduce some notations and tools in Section 3. Then we show our main Theorem 1.2 about polynomial interpolation in Section 4. For signals with k-sparse Fourier transform, we show two bounds on their growth rate in Section 5 and describe the hash functions and filter functions in Section 6. We provide the algorithm for frequency estimation and its proof in Section 7. In Section 8, we describe the algorithm for one-cluster recovery. In Section 9, we show the proof of Theorem 1.1. We defer several technical proofs in Appendix A. Appendix B gives a summary of several well-known facts are existing in literature. We provide the analysis of hash functions and filter functions in Appendix C.

2 Proof Sketch

We first consider one-cluster recovery centered at zero, i.e., $x^*(t) = \sum_{j=1}^{k} v_j \cdot e^{2\pi \mathbf{i} f_j t}$ where every f_j is in $[-\Delta, \Delta]$ for some small $\Delta > 0$. The road map is to replace x^* by a low degree polynomial P such that $\|x^*(t) - P(t)\|_T^2 \lesssim \delta \|x^*\|_T^2$ then recover a polynomial Q to approximate P through the observation $x(t) = P(t) + g'(t)$ where $g'(t) = g(t) + (x^*(t) - P(t))$.

A natural way to replace $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$ by a low degree polynomial $P(t)$ is the Taylor expansion. To bound the error after taking the low degree terms in the expansion by $\delta \|x^*\|_T$, we show the existence of $x'(t) = \sum_{j=1}^{k} v'_j e^{2\pi \mathbf{i} f'_j t}$ approximating x^* on $[0, T]$ with an extra property—any coefficient v'_j in $x'(t)$ has an upper bound in terms of $\|x'\|_T^2 = \frac{1}{T} \int_0^T |x'(t)|^2 \mathrm{d}t$. We prove the existence of $x'(t)$ via two more steps, both of which rely on the estimation of some Gram matrix constituted by these k signals.

The first step is to show the existence of a k-Fourier-sparse signal $x'(t)$ with frequency gap $\eta \ge \frac{\exp(-\mathrm{poly}(k)) \cdot \delta}{T}$ that is sufficiently close to $x^*(t)$.

Lemma 2.1. *There is a universal constant $C_1 > 0$ such that, for any $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$ and any $\delta > 0$, there always exist $\eta \ge \frac{\delta}{T} \cdot k^{-C_1 k^2}$ and $x'(t) = \sum_{j=1}^{k} v'_j e^{2\pi \mathbf{i} f'_j t}$ satisfying*

$$\|x'(t) - x^*(t)\|_T \le \delta \|x^*(t)\|_T$$

with $\min_{i \ne j} |f'_i - f'_j| \ge \eta$ and $\max_{j \in [k]} \{|f'_j - f_j|\} \le k\eta$.

9

We outline our approach and defer the proof to Section 8. We focus on the replacement of one frequency f_k in $x^* = \sum_{j \in [k]} v_j e^{2\pi \mathbf{i} f_j t}$ by a new frequency $f_{k+1} \neq f_k$ and its error. The idea is to consider every signal $e^{2\pi \mathbf{i} f_j t}$ as a vector and prove that for any vector x^* in the linear subspace $\text{span}\{e^{2\pi \mathbf{i} f_j t} | j \in [k]\}$, there exists a vector in the linear subspace $\text{span}\{e^{2\pi \mathbf{i} f_{k+1} t}, e^{2\pi \mathbf{i} f_j t} | j \in [k-1]\}$ with distance at most $\exp(k^2) \cdot (|f_k - f_{k+1}|T) \cdot \|x^*\|_T$ to x^*.

The second step is to lower bound $\|x'\|_T^2$ by its coefficients through the frequency gap η in x'.

Lemma 2.2. *There exists a universal constant $c > 0$ such that for any $x(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$ with frequency gap $\eta = \min_{i \neq j} |f_i - f_j|$,*

$$\|x(t)\|_T^2 \geq k^{-ck^2} \min\left((\eta T)^{2k}, 1\right) \sum_{j=1}^{k} |v_j|^2.$$

Combining Lemma 2.1 and Lemma 2.2, we bound $|v_j'|$ by $\exp(\text{poly}(k)) \cdot \delta^{-O(k)} \cdot \|x'\|_T$ for any coefficient v_j' in x'. Now we apply the Taylor expansion on $x'(t)$ and keep the first $d = O(\Delta T + \text{poly}(k) + k \log \frac{1}{\delta})$ terms of every signal $v_j' \cdot e^{2\pi \mathbf{i} f_j' t}$ in the expansion to obtain a polynomial $P(t)$ of degree at most d. To bound the distance between $P(t)$ and $x'(t)$, we observe that the error of every point $t \in [0, T]$ is at most $(\frac{2\pi \Delta \cdot T}{d})^d \sum_j |v_j'|$, which can be upper bounded by $\delta \|x'(t)\|_T$ via the above connection. We summarize all discussion above as follows.

Lemma 2.3. *For any $\Delta > 0$ and any $\delta > 0$, let $x^*(t) = \sum_{j \in [k]} v_j e^{2\pi \mathbf{i} f_j t}$ where $|f_j| \leq \Delta$ for each $j \in [k]$. There exists a polynomial $P(t)$ of degree at most*

$$d = O(T\Delta + k^3 \log k + k \log 1/\delta)$$

such that

$$\|P(t) - x^*(t)\|_T^2 \leq \delta \|x^*\|_T^2.$$

To recover $x^*(t)$, we observe $x(t)$ as a degree d polynomial $P(t)$ with noise. We use properties of the Legendre polynomials to design a method of random sampling such that we only need $O(d)$ random samples to find a polynomial $Q(t)$ approximating $P(t)$.

Theorem 1.2. *For any degree d polynomial $P(t)$ and an arbitrary function $g(t)$, Procedure RO-BUSTPOLYNOMIALLEARNING in Algorithm 5 takes $O(d)$ samples from $x(t) = P(t) + g(t)$ over $[0, T]$ and reports a degree d polynomial $Q(t)$ in time $O(d^\omega)$ such that, with probability at least $99/100$,*

$$\|P(t) - Q(t)\|_T^2 \lesssim \|g(t)\|_T^2.$$

where $\omega < 2.373$ is matrix multiplication exponent [Str69],[CW87],[Wil12].

We can either report the polynomial $Q(t)$ or transfer $Q(t)$ to a signal with d-sparse Fourier transform. We defer the technical proofs and the formal statements to Section 8 and discuss the recovery of k clusters from now on.

As mentioned before, we apply the filter function $(H(t), \widehat{H}(f))$ on x^* such that $\widehat{x^* \cdot H}$ has at most k clusters given $\widehat{x^*}$ with k-sparse Fourier transform. First, we show that all frequencies in the "heavy" clusters of $\widehat{x^* \cdot H}$ constitute a good approximation of x^* in Section 9.

Definition 2.4. *Given* $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$, *any* $\mathcal{N} > 0$, *and a filter function* $(H, \widehat{H})$ *with bounded support in frequency domain. Let* L_j *denote the interval of* $\mathrm{supp}(\widehat{e^{2\pi \mathbf{i} f_j t} \cdot H})$ *for each* $j \in [k]$.

Define an equivalence relation $\sim$ *on the frequencies* f_i *by the transitive closure of the relation* $f_i \sim f_j$ *if* $L_i \cap L_j \neq \emptyset$. *Let* $S_1, \ldots, S_n$ *be the equivalence classes under this relation.*

Define $C_i = \underset{f \in S_i}{\cup} L_i$ *for each* $i \in [n]$. *We say* C_i *is a "heavy" cluster iff* $\int_{C_i} |\widehat{H \cdot x^*}(f)|^2 \mathrm{d}f \geq T \cdot \mathcal{N}^2 / k$.

Claim 2.5. *Given* $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$ *and any* $\mathcal{N} > 0$, *let* H *be the filter function defined in Appendix C.1 and* $C_1, \cdots, C_l$ *be the heavy clusters from Definition 2.4. For*

$$S = \left\{ j \in [k] \Big| f_j \in C_1 \cup \cdots C_l \right\},$$

we have $x^{(S)}(t) = \sum_{j \in S} v_j e^{2\pi \mathbf{i} f_j t}$ *approximating* x^* *within distance* $\|x^{(S)}(t) - x^*(t)\|_T^2 \lesssim \mathcal{N}^2$.

Hence it is enough to recover $x^{(S)}$ for the recovery of x^*. Let Δ_h denote the bandwidth of $\widehat{H}$. In Section 7, we choose $\Delta > k \cdot \Delta_h$ such that for any $j \in S$, $\int_{f_j - \Delta}^{f_j + \Delta} |\widehat{H \cdot x^*}(f)|^2 \mathrm{d}f \geq T \cdot \mathcal{N}^2 / k$ from the fact $|C_i| \leq k \cdot \Delta_h$. Then we prove Theorem 2.6 in Section 7, which finds $O(k)$ frequencies to cover all heavy clusters of $\widehat{x^* \cdot H}$.

Theorem 2.6. *Let* $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$ *and* $x(t) = x^*(t) + g(t)$ *be our observable signal where* $\|g(t)\|_T^2 \leq c\|x^*(t)\|_T^2$ *for a sufficiently small constant* c. *Then Procedure* FREQUENCYRECOVERYKCLUSTER *returns a set* L *of* $O(k)$ *frequencies that covers all heavy clusters of* x^*, *which uses* $\mathrm{poly}(k, \log(1/\delta)) \log(FT)$ *samples and* $\mathrm{poly}(k, \log(1/\delta)) \log^2(FT)$ *time. In particular, for* $\Delta = \mathrm{poly}(k, \log(1/\delta))/T$ *and* $\mathcal{N}^2 := \|g(t)\|_T^2 + \delta\|x^*(t)\|_T^2$, *with probability* $1 - 2^{-\Omega(k)}$, *for any* f^* *with*

$$\int_{f^* - \Delta}^{f^* + \Delta} |\widehat{x \cdot H}(f)|^2 \mathrm{d}f \geq T \mathcal{N}^2 / k, \tag{2}$$

there exists an $\widetilde{f} \in L$ *satisfying*

$$|f^* - \widetilde{f}| \lesssim \Delta \sqrt{\Delta T}.$$

Let $L = \{\widetilde{f}_1, \cdots, \widetilde{f}_l\}$ be the list of frequencies from the output of Procedure FREQUENCYRECOVERYKCLUSTER in Theorem 2.6. The guarantee is that, for any f_j in $x^{(S)}$, there exists some $p_j \in [l]$ such that $|\widetilde{f}_{p_j} - f_j| \lesssim \Delta \sqrt{\Delta T}$ for $\Delta = \mathrm{poly}(k, \log(1/\delta))/T$. Hence we rewrite $x^{(S)}(t) = \sum_{i \in [l]} e^{2\pi \mathbf{i} \widetilde{f}_i t} (\sum_{j \in S: p_j = i} e^{2\pi \mathbf{i}(f_j - \widetilde{f}_i)t})$. For each $i \in [l]$, we apply Lemma 2.3 of one-cluster recovery on $\sum_{j \in S: p_j = i} e^{2\pi \mathbf{i}(f_j - \widetilde{f}_i)t}$ to approximate it by a degree d polynomial $P_i(t)$.

Now we consider $x(t) = \sum_{i \in [l]} e^{2\pi \mathbf{i} \widetilde{f}_i t} \cdot P_i(t) + g''(t)$ where $\|g''(t)\|_T \lesssim \|g(t)\|_T + \delta\|x^*(t)\|_T$. To recover $\sum_{i \in [l]} e^{2\pi \mathbf{i} \widetilde{f}_i t} \cdot P_i(t)$, we treat it as a vector in the linear subspace

$$V = \mathrm{span}\left\{ e^{2\pi \mathbf{i} \widetilde{f}_i t} \cdot t^j \Big| j \in \{0, \cdots, d\}, i \in [l] \right\}$$

with dimension at most $l(d+1)$ and find a vector in this linear subspace approximating it.

11

We show that for any $v \in V$, the average of $\text{poly}(kd)$ random samples on v is enough to estimate $\|v\|_T^2$. In particular, any vector in this linear subspace satisfies that the maximum of it in $[0,T]$ has an upper bound in terms of its average in $[0,T]$. Then we apply the Chernoff bound to prove that $\text{poly}(kd)$ random samples are enough for the estimation of one vector $v \in V$.

Claim 2.7. *For any $\vec{u} \in \text{span} \left\{ e^{2\pi \mathbf{i} \widetilde{f}_i t} \cdot t^j \,\middle|\, j \in \{0, \cdots, d\}, i \in [l] \right\}$, there exists some universal constants $C_1 \leq 4$ and $C_2 \leq 3$ such that*

$$\max_{t \in [0,T]} \{|\vec{u}(t)|^2\} \lesssim (ld)^{C_1} \log^{C_2}(ld) \cdot \|\vec{u}\|_T^2$$

At last we use an ϵ-net to argue that $\text{poly}(kd)$ random samples from $[0,T]$ are enough to interpolate $x(t)$ by a vector $v \in V$. Because the dimension of this linear subspace is at most $l(d+1) = O(kd)$, there exists an ϵ-net in this linear subspace for unit vectors with size at most $\exp(kd)$. Combining the Chernoff bound on all vectors in the ϵ-net and Claim 2.7, we know that $\text{poly}(kd)$ samples are sufficient to estimate $\|v\|_T^2$ for any vector $v \in V$. In Section 9, we show that a vector $v \in V$ minimizing the distance on $\text{poly}(kd)$ random samples is a good approximation for $\sum_{i \in [l]} e^{2\pi \mathbf{i} \widetilde{f}_i t} \cdot P_i(t)$, which is a good approximation for $x^*(t)$ from all discussion above.

Theorem 1.1. *Let $x(t) = x^*(t) + g(t)$, where x^* is k-Fourier-sparse signal with frequencies in $[-F, F]$. Given samples of x over $[0,T]$ we can output $\widetilde{x}(t)$ such that with probability at least $1 - 2^{-\Omega(k)}$,*

$$\|\widetilde{x} - x^*\|_T \lesssim \|g\|_T + \delta \|x^*\|_T .$$

Our algorithm uses $\text{poly}(k, \log(1/\delta)) \cdot \log(FT)$ samples and $\text{poly}(k, \log(1/\delta)) \cdot \log^2(FT)$ time. The output $\widetilde{x}$ is $\text{poly}(k, \log(1/\delta))$-Fourier-sparse signal.

3 Preliminaries

We first provide some notations in Section 3.1 and basic Fourier facts in Section 3.2. Then we review some probability inequalities in Section 3.3. At last, we introduce Legendre polynomials in Section 3.4 and review some basic properties of Gram matrix and its determinant in Section 3.5.

3.1 Notation

For any function f, we define $\widetilde{O}(f)$ to be $f \cdot \log^{O(1)}(f)$. We use $[n]$ to denote $\{1, 2, \cdots, n\}$. Let $\mathbf{i}$ denote $\sqrt{-1}$. For any Complex number $z = a + \mathbf{i}b \in C$, where $a, b \in \mathbb{R}$. We define $\overline{z}$ to be $a - \mathbf{i}b$ and $|z| = \sqrt{a^2 + b^2}$ such that $|z|^2 = z\overline{z}$. For any function $f(t): \mathbb{R} \to \mathbb{C}$, we use $\text{supp}(f)$ to denote the support of f.

For convenience, we define the sinc function and the Gaussian distribution $\text{Gaussian}_{\mu,\sigma}$ on $\mathbb{R}$ with expectation μ and variance σ^2 as follows:

$$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}, \quad \text{Gaussian}_{\mu,\sigma}(t) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}.$$

For a fixed $T > 0$, we define the inner product of two functions $x, y : [0,T] \to \mathbb{C}$ as

$$\langle x, y \rangle_T = \frac{1}{T} \int_0^T x(t) \overline{y}(t) \mathrm{d}t.$$

We define the $\| \cdot \|_T$ norm as

$$\|x(t)\|_T = \sqrt{\langle x(t), x(t) \rangle_T} = \sqrt{\frac{1}{T} \int_0^T |x(t)|^2 \mathrm{d}t}.$$
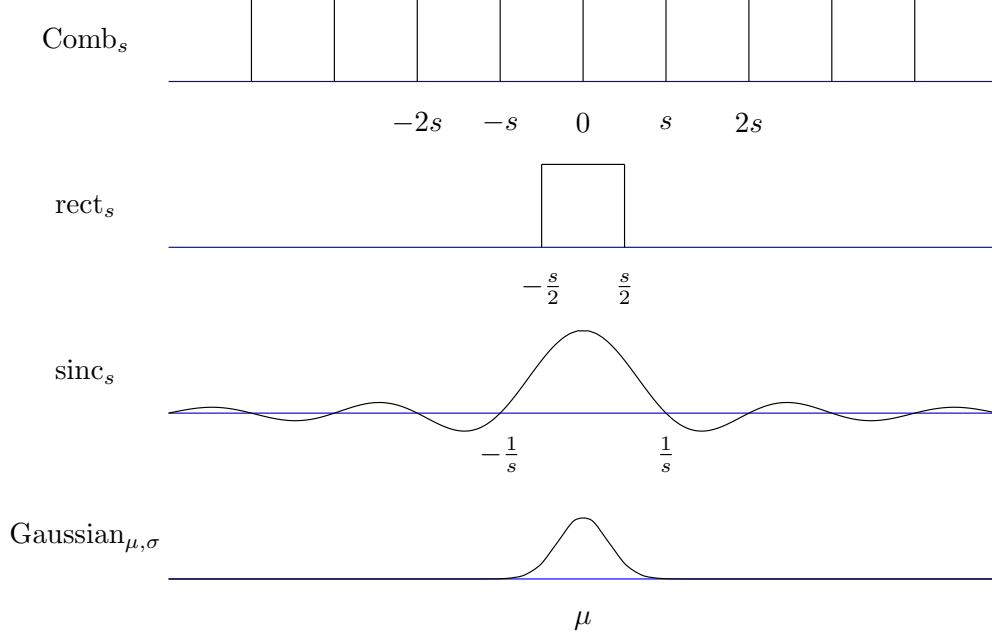
Figure 1: A picture of a Comb_s, rect_s, sinc_s, $\mathrm{Gaussian}_{\mu,\sigma}$.

3.2   Facts about the Fourier transform

In this work, we always use $x(t)$ to denote a signal from $\mathbb{R} \to \mathbb{C}$. The Fourier transform $\widehat{x}(f)$ of an integrable function $x : \mathbb{R} \to \mathbb{C}$ is defined as

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

Similarly, $x(t)$ is determined from $\widehat{x}(f)$ by the inverse transform:

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

Let CFT denote the continuous Fourier transform, DTFT denote the discrete-time Fourier transform, DFT denote the discrete Fourier transform, and FFT denote the fast Fourier transform. For any signal $x(t)$ and $n \in \mathbb{N}_+$, we define $x^{*n}(t) = \underbrace{x(t) * \cdots * x(t)}_{n}$ and $\widehat{x}^n(f) = \underbrace{\widehat{x}(f) \cdot \cdots \cdot \widehat{x}(f)}_{n}$.

Fact 3.1. *Let $\delta_\Delta(f)$ denote the Dirac delta at Δ. Then*

$$\widehat{\delta}_\Delta(t) = \int_{-\infty}^{+\infty} \delta_\Delta(f)e^{2\pi \mathbf{i} ft}\mathrm{d}f = e^{2\pi \mathbf{i} t\Delta}.$$

Fact 3.2. *For any $s > 0$, let $\mathrm{Comb}_s(t) = \sum_{j \in \mathbb{Z}} \delta_{js}(t)$. Then the Fourier transform of $\mathrm{Comb}_s(t)$ is*

$$\widehat{\mathrm{Comb}}_s(f) = \frac{1}{s}\mathrm{Comb}_{1/s}(f).$$

The following fact says the the Fourier transform of a rectangle function is a sinc function.

Fact 3.3. *We use*

$$\mathrm{rect}_s(t) = \begin{cases} 1 & \text{if } |t| \leq \frac{s}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

Then the Fourier transform of $\mathrm{rect}_s(t)$ *is* $\widehat{\mathrm{rect}_s}(f) = \frac{\sin(\pi fs)}{\pi fs} = \mathrm{sinc}(fs).$

The Fourier transform of a Gaussian function is another Gaussian function.

Fact 3.4. *For* $\mathrm{Gaussian}_{\mu,\sigma}(t) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}.$ *Then the Fourier transform is*

$$\widehat{\mathrm{Gaussian}}_{\mu,\sigma}(f) = e^{-2\pi i fu}\frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{Gaussian}_{0,\sigma'}(f) \ \text{for } \sigma' = 1/(2\pi\sigma).$$

Proof. From the definition of the Fourier transform,

$$
\begin{aligned}
\widehat{\mathrm{Gaussian}}_{\mu,\sigma}(f) &= \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}e^{-2\pi i ft}\mathrm{d}t \\
&= e^{-2\pi i fu}\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{t^2}{2\sigma^2}}e^{-2\pi i ft}\mathrm{d}t \\
&= e^{-2\pi i fu}\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t+2\pi i\sigma^2 f)^2}{2\sigma^2}-2\pi^2 f^2\sigma^2}\mathrm{d}t \\
&= e^{-2\pi i fu}e^{-\frac{f^2}{2\sigma'^2}}
\end{aligned}
$$

where $\sigma' = 1/(2\sigma\pi)$, which is $e^{-2\pi i fu}\cdot\sigma'\sqrt{2\pi}\cdot\mathrm{Gaussian}_{0,\sigma'}(f).$ □

3.3 Tools and inequalities

From the Chernoff Bound (Lemma B.2), we show that if the maximum of a signal is bounded by d times its energy over some fixed interval, then taking more than d samples (each sample is drawn i.i.d. over that interval) suffices to approximate the energy of the signal on the interval with high probability.

Lemma 3.5. *Given any function* $x(t): \mathbb{R} \to \mathbb{C}$ *with* $\max_{t\in[0,T]}|x(t)|^2 \leq d\|x(t)\|_T^2.$ *Let S denote a set of points from 0 to T. If each point of S is chosen uniformly at random from $[0,T]$, we have*

$$\Pr\left[\left|\frac{1}{|S|}\sum_{i\in S}|x(t_i)|^2 - \|x(t)\|_T^2\right| \geq \epsilon\|x(t)\|_T^2\right] \leq e^{-\Omega(\epsilon^2|S|/d)}$$

We provide a proof in Appendix A.5.

Because $d\cdot\frac{1}{2d} + \frac{1}{2}\cdot(1-\frac{1}{2d}) \leq 1$, we have the following inequality when the maximum of $|x(t)|^2$ is at most d times its average.

Lemma 3.6. *Given any function* $x(t): \mathbb{R} \to \mathbb{C}$ *with* $\max_{t\in[0,T]}|x(t)|^2 \leq d\|x(t)\|_T^2.$ *Let S denote a set of points from 0 to T. For any point a is sampled uniformly at random from $[0,T]$, we have,*

$$\Pr_{a\sim[0,T]}\left[|x(a)|^2 \geq \frac{1}{2}\|x(t)\|_T^2\right] \geq \frac{1}{2d}.$$

14

3.4 Legendre polynomials

We provide an brief introduction to Legendre polynomials (please see [Dun10] for a complete introduction). For convenience, we fix $\|f(t)\|_T^2 = \frac{1}{2}\int_{-1}^1 |f(t)|^2 dt$ in this section.

Definition 3.7. *Let $L_n(x)$ denote the Legendre polynomials of degree n, the solution to Legendre's differential equation:*

$$\frac{d}{dx}\left[(1-x^2)\frac{d}{dx}L_n(x)\right] + n(n+1)L_n(x) = 0 \tag{3}$$

We will the following two facts about the Legendre polynomials in this work.

Fact 3.8. *$L_n(1) = 1$ for any $n \geq 0$ in the Legendre polynomials.*

Fact 3.9. *The Legendre polynomials constitute an orthogonal basis with respect to the inner product on interval $[-1, 1]$:*

$$\int_{-1}^1 L_m(x)L_n(x)dx = \frac{2}{2n+1}\delta_{mn}$$

where δ_{mn} denotes the Kronecker delta, i.e., it equals to 1 if $m = n$ and to 0 otherwise.

For any polynomial $P(x)$ of degree at most d with complex coefficients, there exists a set of coefficients from the above properties such that

$$P(x) = \sum_{i=0}^d \alpha_i \cdot L_i(x), \text{ where } \alpha_i \in \mathbb{C}, \forall i \in \{0, 1, 2, \cdots, d\}.$$

Lemma 3.10. *For any polynomial $P(t)$ of degree at most d from R to $\mathbb{C}$, for any interval $[S, T]$,*

$$\max_{t \in [S,T]} |P(t)|^2 \leq (d+1)^2 \cdot \frac{1}{T-S}\int_S^T |P(t)|^2 dx.$$

We provide a proof in Appendix A.6.

3.5 Gram matrix and its determinant

We provide an brief introduction to Gramian matrices (please see [Haz01] for a complete introduction). We use $\langle x, y \rangle$ to denote the inner product between vector x and vector y.

Let $\vec{v}_1, \cdots, \vec{v}_n$ be n vectors in an inner product space and span$\{\vec{v}_1, \cdots, \vec{v}_n\}$ be the linear subspace spanned by these n vectors with coefficients in $\mathbb{C}$, i.e., $\left\{\sum_{i \in [n]} \alpha_i \vec{v}_i | \forall i \in [n], \alpha_i \in \mathbb{C}\right\}$. The Gram matrix Gram$_n$ of $\vec{v}_1, \cdots, \vec{v}_n$ is an $n \times n$ matrix defined as Gram$_n(i, j) = \langle \vec{v}_i, \vec{v}_j \rangle$ for any $i \in [n]$ and $j \in [n]$.

Fact 3.11. *det(Gram$_n$) is the square of the volume of the parallelotope formed by $\vec{v}_1, \cdots, \vec{v}_n$.*

Let Gram$_{n-1}$ be the Gram matrix of $\vec{v}_1, \cdots, \vec{v}_{n-1}$. Let $\vec{v}_n^{\parallel}$ be the projection of v_n onto the linear subspace span$\{\vec{v}_1, \cdots, \vec{v}_{n-1}\}$ and $\vec{v}_n^{\perp} = \vec{v}_n - \vec{v}_n^{\parallel}$. We use $\|\vec{v}\|$ to denote the length of $\vec{v}$ in the inner product space, which is $\sqrt{\langle \vec{v}, \vec{v} \rangle}$.

Claim 3.12.

$$\|\vec{v}_n^{\perp}\|^2 = \frac{\det(\text{Gram}_{n-1})}{\det(\text{Gram}_n)}.$$

Proof.

$$\det(\text{Gram}_n) = \text{volume}^2(\vec{v}_1, \cdots, \vec{v}_n) = \text{volume}^2(\vec{v}_1, \cdots, \vec{v}_{n-1}) \cdot \|\vec{v}_n^{\perp}\|^2 = \det(\text{Gram}_n) \cdot \|\vec{v}_n^{\perp}\|^2.$$

$\square$

4 Robust Polynomial Interpolation Algorithm

In Section 4.1, we show how to learn a low degree polynomial by using linear number of samples, running polynomial time, and achieving constant success probability. In Section 4.2, we show to how boost the success probability by rerunning previous algorithm several times.

4.1 Constant success probability

We show how to learn a degree-d polynomial P with $n = O(d)$ samples and prove Theorem 1.2 in this section. For convenience, we first fix the interval to be $[-1, 1]$ and use $\|f\|^2_{[-1,1]} = \frac{1}{2} \int_{-1}^{1} |f(t)|^2 \mathrm{d}t$.

Lemma 4.1. *Let $d \in \mathbb{N}$ and $\epsilon \in \mathbb{R}^+$, there exists an efficient algorithm to compute a partition of $[-1, 1]$ to $n = O(d/\epsilon)$ intervals $I_1, \cdots, I_n$ such that for any degree d polynomial $P(t) : \mathbb{R} \to \mathbb{C}$ and any n points $x_1, \cdots, x_n$ in the intervals $I_1, \cdots, I_n$ respectively, the function $Q(t)$ defined by*

$$Q(t) = P(x_j) \quad if \quad t \in I_j$$

approximates P by

$$\|Q - P\|_{[-1,1]} \leq \epsilon \|P\|_{[-1,1]}. \tag{4}$$

One direct corollary from the above lemma is that observing $n = O(d/\epsilon)$ points each from $I_1, \cdots, I_n$ provides a good approximation for all degree d polynomials. For any set $S = \{t_1, \cdots, t_m\}$ where each $t_i \in [-1, 1]$ and a distribution with support $\{w_1, \cdots, w_m\}$ on S where $\sum\limits_{i=1}^{m} w_i = 1$ and $w_i \geq 0$ for each $i \in [m]$, we define $\|x\|_{S,w} = (\sum_{i=1}^{m} w_i \cdot |x(t_i)|^2)^{1/2}$.

Corollary 4.2. *Let $I_1, \cdots, I_n$ be the intervals in the above lemma and $w_j = |I_j|/2$ for each $j \in [n]$. For any $x_1, \cdots, x_n$ in the intervals $I_1, \cdots, I_n$ respectively, we consider $S = \{x_1, \cdots, x_n\}$ with the distribution $w_1, \cdots, w_n$. Then for any degree d polynomial P, we have*

$$\|P\|_{S,w} \in \left[(1 - \epsilon)\|P\|_{[-1,1]}, (1 + \epsilon)\|P\|_{[-1,1]} \right].$$

We first state the main technical lemma and finish the proof of the above lemma (we defer the proof of Lemma 4.3 to Appendix A.3).

Lemma 4.3. *For any degree d polynomial $P(t) : \mathbb{R} \to \mathbb{C}$ with derivative $P'(t)$, we have,*

$$\int_{-1}^{1} (1 - t^2)|P'(t)|^2 \mathrm{d}t \leq 2d^2 \int_{-1}^{1} |P(t)|^2 \mathrm{d}t. \tag{5}$$

Proof of Lemma 4.1. We set $m = 10d/\epsilon$ and show a partition of $[-1, 1]$ into $n \leq 20m$ intervals. We define $g(t) = \frac{\sqrt{1-t^2}}{m}$ and $y_0 = 0$. Then we choose $y_i = y_{i-1} + g(y_{i-1})$ for $i \in \mathbb{N}^+$. Let l be the first index of y such that $y_l \geq 1 - \frac{9}{m^2}$. We show $l \lesssim m$.

Let j_k be the first index in the sequence such that $y_{j_k} \geq 1 - 2^{-k}$. Notice that

$$j_2 \leq \frac{3/4}{\frac{\sqrt{1-(3/4)^2}}{m}} \leq 1.5m$$

and

$$y_i - y_{i-1} = g(y_{i-1}) = \frac{\sqrt{1 - y_{i-1}^2}}{m} \geq \frac{\sqrt{1 - y_{i-1}}}{m}.$$

16

Then for all $k > 2$, we have

$$j_k - j_{k-1} \leq \frac{2^{-k}}{\frac{\sqrt{1-y_{(j_k-1)}}}{m}} \leq 2^{-k/2}m.$$

Therefore $j_k \leq \left(1.5 + (2^{-3/2} + \cdots 2^{-k/2})\right)m$ and $l \leq 10m$.

Because $y_{l-1} \leq 1 - \frac{9}{m^2}$, for any $j \in [l]$ and any $x \in [y_{i-1}, y_i]$, we have the following property:

$$\frac{1-x^2}{m^2} \geq \frac{1}{2} \cdot \frac{(1-y_{i-1}^2)}{m^2} = (y_i - y_{i-1})^2/2. \tag{6}$$

Now we set n and partition $[-1,1]$ into $I_1, \cdots, I_n$ as follows:

1. $n = 2(l+1)$.

2. For $j \in [l]$, $I_{2j-1} = [y_{j-1}, y_j]$ and $I_{2j} = [-y_j, -y_{j-1}]$.

3. $I_{2l+1} = [y_l, 1]$ and $I_{2l+2} = [-1, -y_l]$.

For any $x_1, \cdots, x_n$ where $x_j \in I_j$ for each $j \in [n]$, we rewrite the LHS of (4) as follows:

$$\underbrace{\sum_{j=1}^{n-2} \int_{I_j} |P(x_j) - P(t)|^2 \, dt}_{A} + \underbrace{\int_{I_{n-1}} |P(x_{n-1}) - P(t)|^2 \, dt + \int_{I_n} |P(x_n) - P(t)|^2 \, dt}_{B}. \tag{7}$$

For A in Equation (7), from the Cauchy-Schwarz inequality, we have

$$\sum_{j=1}^{n-2} \int_{I_j} |P(x_j) - P(t)|^2 \, dt = \sum_{j=1}^{n-2} \int_{I_j} \left| \int_{x_j}^{t} P'(y) dy \right|^2 dt \leq \sum_{j=1}^{n-2} \int_{I_j} |t - x_j| \int_{x_j}^{t} |P'(y)|^2 dy dt.$$

Then we swap dt with dy and use Equation (6):

$$\sum_{j=1}^{n-2} \int_{I_j} |P'(y)|^2 \int_{t \notin (x_j, y)} |t - x_j| dt dy \leq \sum_{j=1}^{n-2} \int_{I_j} |P'(t)|^2 \cdot |I_j|^2 dt \leq \sum_{j=1}^{n-2} \int_{I_j} |P'(t)|^2 \frac{2(1-t^2)}{m^2} dt.$$

We use Lemma 4.3 to simplify it by

$$\sum_{j=1}^{n-2} \int_{I_j} |P(x_j) - P(t)|^2 \, dt \leq \int_{-1}^{1} |P'(t)|^2 \frac{2(1-t^2)}{m^2} dt \leq \frac{2d^2}{m^2} \int_{-1}^{1} |P(t)|^2 dt.$$

For B in Equation (7), notice that $|I_{n-1}| = |I_n| = 1 - y_l \leq 9m^{-2}$ and for $j \in \{n-1, n\}$

$$|P(t) - P(x_j)|^2 \leq 4 \max_{t \in [-1,1]} |P(t)|^2 \leq 4(d+1)^2 \|P\|_{[-1,1]}^2$$

from the properties of degree-d polynomials, i.e., Lemma 3.10. Therefore B in Equation (7) is upper bounded by $2 \cdot 4(d+1)^2(9m^{-2})\|P(t)\|_{[-1,1]}^2$.

From all discussion above, $\|Q(t) - P(t)\|_{[-1,1]}^2 \leq \frac{99d^2}{m^2} \leq \epsilon^2.$ $\qquad\square$

Now we use the above lemma to provide a faster learning algorithm for polynomials on interval $[-1,1]$ with noise instead of using the ϵ-nets argument. Algorithm ROBUSTPOLYNOMIALLEARN-INGFIXEDINTERVAL works as follows:

1. Let $\epsilon = 1/20$ and $I_1, \cdots, I_n$ be the intervals for d and ϵ in Lemma 4.1.

2. Random choose $x_j \in I_j$ for every $j \in [n]$ and define $S = \{x_1, \cdots, x_n\}$ with weight $w_1 = \frac{|I_1|}{2}, \cdots, w_n = \frac{|I_n|}{2}$.

3. Find the degree d polynomial $Q(t)$ that minimizes $\|P(t) - Q(t)\|_{S,w}$ using Fact B.3.

Lemma 4.4. *For any degree d polynomial $P(t)$ and an arbitrary function $g(t)$, Algorithm* ROBUST-POLYNOMIALLEARNINGFIXEDINTERVAL *takes $O(d)$ samples from $x(t) = P(t) + g(t)$ over $[-1,1]$ and reports a degree d polynomial $Q(t)$ in time $O(d^\omega)$ such that, with probability at least $99/100$,*

$$\|P(t) - Q(t)\|^2_{[-1,1]} \lesssim \|g(t)\|^2_{[-1,1]}.$$

Proof. Notice that $n = O(d/\epsilon) = O(d)$ and the running time depends on solving a linear regression problem(Fact B.3), which takes $O(d^\omega)$ time. It is enough to bound the distance between P and Q:

$$
\begin{aligned}
&\|P - Q\|_{[-1,1]} \\
\leq\ & 1.09\|P - Q\|_{S,w} && \text{by Corollary 4.2} \\
=\ & 1.09\|x - g - Q\|_{S,w} && \text{by } x = P + g \\
\leq\ & 1.09\|g\|_{S,w} + 1.09\|x - Q\|_{S,w} && \text{by triangle inequality} \\
\leq\ & 1.09\|g\|_{S,w} + 1.09\|x - P\|_{S,w} && Q = \underset{\text{degree-}d\ R}{\arg\min} \|R - x\|_{S,w} \\
\leq\ & 2.2\|g\|_{S,w}
\end{aligned}
$$

Because $\mathbb{E}_S[\|g\|^2_{S,w}] = \|g\|^2_{[-1,1]}$, we know that $\|P - Q\|_{[-1,1]} \leq 2200\|g\|_{[-1,1]}$ with probability $\geq .999$ by using Markov's inequality. $\qquad\square$

For any function $f : [0,T] \to \mathbb{C}$, let $\widetilde{f}(t) = f(\frac{2t-T}{T})$. Then $\|\widetilde{f}\|_{[-1,1]} = \|f\|_T$ from the definition. Hence we can switch any interval $[0,T]$ to $[-1,1]$ and use Lemma 4.4.

Theorem 1.2. *For any degree d polynomial $P(t)$ and an arbitrary function $g(t)$, Procedure* ROBUSTPOLYNOMIALLEARNING *in Algorithm 5 takes $O(d)$ samples from $x(t) = P(t) + g(t)$ over $[0,T]$ and reports a degree d polynomial $Q(t)$ in time $O(d^\omega)$ such that, with probability at least $99/100$,*

$$\|P(t) - Q(t)\|^2_T \lesssim \|g(t)\|^2_T.$$

where $\omega < 2.373$ is matrix multiplication exponent [Str69],[CW87],[Wil12].

4.2 Boosting success probability

Notice that the success probability of Theorem 1.2 is only constant, and the proof technique of obtaining that result cannot be modified to $1 - 1/\text{poly}(d)$ or $1 - 2^{-\Omega(d)}$ success probability due to using Markov's inequality. However, we can use that algorithm as a black box, and rerun it $O(\log(1/p))$ (for any $p > 0$) times on fresh samples. Using the careful median analysis from [MP14] gives

Theorem 4.5. *For any degree d polynomial $P(t)$, an arbitrary function $g(t)$, and any $p > 0$, Procedure* ROBUSTPOLYNOMIALLEARNING$^+$ *in Algorithm 5 takes $O(d\log(1/p))$ samples from $x(t) = P(t) + g(t)$ over $[0,T]$ and reports a degree d polynomial $Q(t)$ in time $O(d^\omega \log(1/p))$ such that, with probability at least $1 - p$,*

$$\|P(t) - Q(t)\|^2_T \lesssim \|g(t)\|^2_T.$$

where $\omega < 2.373$ is matrix multiplication exponent.

Proof. We run algorithm ROBUSTPOLYNOMIALLEARNING R rounds with $O(d)$ independent and fresh samples per round. We will obtain R degree-d polynomials $Q_1(t), Q_2(t), \cdots, Q_R(t)$. We say a polynomial $Q_i(t)$ is good if $\|Q_i(t) - P(t)\|_T^2 \lesssim \|g(t)\|_T^2$. Using the Chernoff bound, with probability at least $1 - 2^{-\Omega(R)}$, at least a $3/4$ fraction of the polynomials are "good". We output polynomial $Q(t) = Q_{j^*}(t)$ such that

$$j^* = \underset{j \in [R]}{\arg\min}(\text{median}\{\|Q_j(t) - Q_1(t)\|_T^2, \|Q_j(t) - Q_2(t)\|_T^2, \cdots, \|Q_j(t) - Q_R(t)\|_T^2\}) \quad (8)$$

The Equation (8) can be solved in following straightforward way. For $i \neq j$, it takes $O(d)$ time to compute $\|Q_j(t) - Q_i(t)\|_T^2$. Because of the number of pairs is $O(R^2)$, thus it takes $O(R^2 d)$ time write down a $R \times R$ matrix. For each column, we run linear time 1-median algorithm. This step takes $O(R^2)$ time. At the end, j^* is index of the column that has the smallest median value. Thus, polynomial $Q(t) = Q_{j^*}(t)$ 1 the 0 with probability at least $1 - p$ by choosing $R = O(\log(1/p))$. The running time is not optimized yet.

To improve the dependence on R for running time, we replace the step of solving Equation (8) by an approach that is similar to [MP14]. We choose a new set of samples S, say $S = \{t_1, t_2, \cdots, t_n\}$ and $n = O(d)$. Using Fact B.4, we can compute $Q_i(t_j)$ for all $i, j \in [R] \times [n]$ in $O(Rd \operatorname{poly}(\log(d)))$ time. Define

$$\widetilde{Q}_j = \underset{i \in [R]}{\text{median }} Q_i(t_j), \forall j \in [n]. \quad (9)$$

Our algorithm will output a degree-d polynomial Q which is the optimal solution of this problem, $\underset{\text{degree-}d \ Q'}{\min} \|Q' - \widetilde{Q}\|_{S,w}$.[2] In the rest of the proof, we will show that $\|Q - P\|_T \lesssim \|g\|_T$ with probability at least $1 - 2^{-\Omega(R)}$.

Notice that Equation (9) implies that $\widetilde{Q}_j - P(t_j) = \underset{i \in [R]}{\text{median}}(Q_i(t_j) - P(t_j))$. Fix a coordinate j and applying the proof argument of Lemma 6.1 in [MP14], we have

$$(\widetilde{Q}_j - P(t_j))^2 \lesssim \underset{\text{good } i}{\text{mean}}(Q_i(t_j) - P(t_j))^2$$

Taking the weighted summation over all the coordinates j, we have

$$\|\widetilde{Q} - P\|_{S,w}^2 \lesssim \underset{\text{good } i}{\text{mean}}\|Q_i - P\|_{S,w}^2$$

Using Corollary 4.2, for each good i,

$$\|Q_i - P\|_{S,w}^2 \lesssim \|Q_i - P\|_T^2$$

Combining the above two inequalities gives

$$\|\widetilde{Q} - P\|_{S,w}^2 \lesssim \underset{\text{good } i}{\text{mean}}\|Q_i - P\|_T^2 \lesssim \|g\|_T^2 \quad (10)$$

Because Q is the optimal solution for $\widetilde{Q}$, then

$$\|\widetilde{Q} - Q\|_{S,w}^2 \leq \|\widetilde{Q} - P\|_{S,w}^2 \lesssim \|g\|_T^2 \quad (11)$$

Using Corollary 4.2 and for any good i, i', $\|Q_i - Q_{i'}\|_T \lesssim \|g\|_T$, we can replace P by $Q_{i'}$ in the Equation (10). Thus, for any $Q_{i'}$ where i' is good,

$$\|\widetilde{Q} - Q_{i'}\|_{S,w}^2 \lesssim \|g\|_T^2 \quad (12)$$

---

[2] Outputting $Q = \underset{\text{degree-}d \ Q'}{\arg\min} \|Q' - x\|_{S,w}$ is not good enough, because it only gives constant success probability.

19

For any good i',

$$\|Q_{i'} - Q\|_T$$
$$\lesssim \|Q_{i'} - Q\|_{S,w} \qquad\qquad\qquad \text{by Corollary 4.2}$$
$$\leq \|Q_{i'} - \widetilde{Q}\|_{S,w} + \|\widetilde{Q} - Q\|_{S,w} \qquad\qquad \text{by triangle inequality}$$
$$\lesssim \|g\|_T \qquad\qquad\qquad\qquad \text{by Equation (11) and (12)}$$

Thus, our algorithm takes $O(dR)$ samples from $x(t) = P(t) + g(t)$ over $[0, T]$ and reports a polynomial $Q(t)$ in time $O(Rd^\omega)$ such that, with probability at least $1 - 2^{-\Omega(R)}$, $\|P(t) - Q(t)\|_T^2 \lesssim \|g(t)\|_T^2$. Choosing $R = O(\log(1/p))$ completes the proof. $\qquad\qquad\square$

5 Bounding the Magnitude of a Fourier-sparse Signal in Terms of Its Average Norm

The main results in this section are two upper bounds, Lemma 5.1 on $\max_{t \in [0,T]} |x(t)|^2$ and Lemma 5.5 on $|x(t)|^2$ for $t > T$, in terms of the typical signal value $\|x\|_T^2 = \frac{1}{T} \int_0^T |x(t)|^2 dt$. We prove Lemma 5.1 in Section 5.1 and Lemma 5.5 in Section 5.2

5.1 Bounding the maximum inside the interval

The goal of this section is to prove Lemma 5.1.

Lemma 5.1. *For any k-Fourier-sparse signal $x(t) : \mathbb{R} \to \mathbb{C}$ and any duration T, we have*

$$\max_{t \in [0,T]} |x(t)|^2 \lesssim k^4 \log^3 k \cdot \|x\|_T^2$$

Proof. Without loss of generality, we fix $T = 1$. Then $\|x\|_T^2 = \int_0^1 |x(t)|^2 dt$. Because $\|x\|_T^2$ is the average over the interval $[0, T]$, if $t^* = \arg\max_{t \in [0,T]} |x(t)|^2$ is not 0 or $T = 1$, we can rescale the two intervals $[0, t^*]$ and $[t^*, T]$ to $[0, 1]$ and prove the desired property separately. Hence we assume $|x(0)|^2 = \max_{t \in [0,T]} |x(t)|^2$ in this proof.

Claim 5.2. *For any k, there exists $m = O(k^2 \log k)$ such that for any k-Fourier-sparse signal $x(t)$, any $t_0 \geq 0$ and $\tau > 0$, there always exist $C_1, \cdots, C_m \in \mathbb{C}$ such that the following properties hold,*

$$\text{Property I} \qquad |C_j| \leq 11 \quad for \quad all \ j \in [m],$$
$$\text{Property II} \qquad x(t_0) = \sum_{j \in [m]} C_j \cdot x(t_0 + j \cdot \tau).$$

We first use this claim to finish the proof of Lemma 5.1. We choose $t_0 = 0$ such that $\forall \tau > 0$, there always exist $C_1, \cdots, C_m \in \mathbb{C}$, and

$$x(0) = \sum_{j \in [m]} C_j \cdot x(j \cdot \tau).$$

By the Cauchy-Schwarz inequality, it implies that for any τ,

$$|x(0)|^2 \ \leq \ m \sum_{j \in [m]} |C_j|^2 |x(j \cdot \tau)|^2$$
$$\lesssim \ m \sum_{j \in [m]} |x(j \cdot \tau)|^2. \qquad\qquad (13)$$

At last, we obtain

$$
\begin{aligned}
|x(0)|^2 &= m \int_0^{1/m} |x(0)|^2 \mathrm{d}\tau \\
&\lesssim m \cdot \int_0^{1/m} \left(m \sum_{j=1}^m |x(j \cdot \tau)|^2\right) \mathrm{d}\tau \\
&= m^2 \cdot \sum_{j=1}^m \int_0^{1/m} |x(j \cdot \tau)|^2 \mathrm{d}\tau \\
&= m^2 \cdot \sum_{j=1}^m \frac{1}{j} \int_0^{j/m} |x(\tau)|^2 \mathrm{d}\tau \\
&\le m^2 \cdot \sum_{j=1}^m \frac{1}{j} \cdot \int_0^1 |x(\tau)|^2 \mathrm{d}\tau \\
&\lesssim m^2 \log m \cdot \|x\|_T^2
\end{aligned}
$$

where the first inequality follows by Equation (13), the second inequality follows by $j/m \le 1$ and the last step follows by $\sum_{i=1}^m \frac{1}{i} = O(\log m)$. From $m = O(k^2 \log k)$, we obtain $|x(0)|^2 = O(k^4 \log^3 k \|x\|_T^2)$. $\qquad \square$

To prove Claim 5.2, we use the following lemmas about polynomials. We defer their proofs to Appendix A.2.

Lemma 5.3. *Let $Q(z)$ be a degree k polynomial, all of whose roots are complex numbers with absolute value 1. For any integer n, let $r_{n,k}(z) = \sum_{l=0}^{k-1} r_{n,k}^{(l)} \cdot z^l$ denote the residual polynomial of*

$$
r_{n,k}(z) \equiv z^n \pmod{Q(z)}.
$$

Then, each coefficient of $r_{n,k}$ is bounded: $|r_{n,k}^{(l)}| \le 2^k n^{k-1}$ for any l.

Lemma 5.4. *For any $k \in \mathbb{Z}$ and any $z_1, \cdots, z_k$ on the unit circle of $\mathbb{C}$, there always exists a degree $m = O(k^2 \log k)$ polynomial $P(z) = \sum_{j=0}^m c_j z^j$ with the following properties:*

$$
\begin{aligned}
\text{Property I} &\quad P(z_i) = 0, \forall i \in \{1, \cdots, k\}, \\
\text{Property II} &\quad c_0 = 1, \\
\text{Property III} &\quad |c_j| \le 11, \forall j \in \{1, \cdots, m\}.
\end{aligned}
$$

Proof of Claim 5.2. For $x(t) = \sum_{i=1}^k v_i e^{2\pi \mathbf{i} f_i t}$, we fix t_0 and τ then rewrite $x(t_0 + j \cdot \tau)$ as a polynomial of $b_i = v_i \cdot e^{2\pi \mathbf{i} f_i t_0}$ and $z_i = e^{2\pi \mathbf{i} f_i \tau}$ for each $i \in [k]$.

$$
\begin{aligned}
x(t_0 + j \cdot \tau) &= \sum_{i=1}^k v_i e^{2\pi \mathbf{i} f_i \cdot (t_0 + j \cdot \tau)} \\
&= \sum_{i=1}^k v_i e^{2\pi \mathbf{i} f_i t_0} \cdot e^{2\pi \mathbf{i} f_i \cdot j\tau} \\
&= \sum_{i=1}^k b_i \cdot z_i^j.
\end{aligned}
$$

Given k and $z_1, \cdots, z_k$, let $P(z) = \sum_{j=0}^{m} c_j z^j$ be the degree m polynomial in Lemma 5.4.

$$
\begin{aligned}
\sum_{j=0}^{m} c_j x(t_0 + j\tau) &= \sum_{j=0}^{m} c_j \sum_{i=1}^{k} b_i \cdot z_i^j \\
&= \sum_{i=1}^{k} b_i \sum_{j=0}^{m} c_j \cdot z_i^j \\
&= \sum_{i=1}^{k} b_i P(z_i) \\
&= 0, \hspace{4cm} (14)
\end{aligned}
$$

where the last step follows by Property I of $P(z)$ in Lemma 5.4. From the Property II and III of $P(z)$, we obtain $x(t_0) = -\sum_{j=1}^{m} c_j x(t_0 + j\tau)$. $\qquad \square$

5.2 Bounding growth outside the interval

Here we show signals with sparse Fourier transform cannot grow too quickly outside the interval.

Lemma 5.5. *Let $x(t)$ be a k-Fourier-sparse signal. For any $T > 0$ and any $t > T$,*

$$
|x(t)|^2 \leq k^7 \cdot (2kt/T)^{2.5k} \cdot \|x\|_T^2.
$$

Proof. For any $t > T$, let $t = t_0 + n \cdot \tau$ such that $t_0 \in [0, T/k], \tau \in [0, T/k]$ and $n \leq \frac{2kt}{T}$. We define $b_i = v_i e^{2\pi \mathbf{i} f_i t_0}$, and $z_i = e^{2\pi \mathbf{i} f_i \tau}$ such that $x(t_0 + n \cdot \tau) = \sum_{j \in [k]} b_j z_j^n$.

By Lemma 5.3, we have for any $z_1, z_2, \cdots, z_k$ and any n,

$$
z^n \equiv \sum_{i=0}^{k-1} a_i z^i \pmod{\prod_{i=1}^{k}(z - z_i)},
$$

where $|a_i| \leq 2^k \cdot n^k, \forall i \in \{0, 1, \cdots, k-1\}$. Thus, we obtain

$$
x(t_0 + n\tau) = \sum_{j=1}^{k} b_j z_j^n = \sum_{j=1}^{k} b_j \left(\sum_{i=0}^{k-1} a_i z_j^i \right).
$$

From the fact that $x(t_0 + i \cdot \tau) = \sum_{j \in [k]} b_j z_j^i$, we simplify it to be

$$
x(t_0 + n\tau) = \sum_{i=0}^{k-1} a_i \sum_{j=1}^{k} b_j z_j^i = \sum_{i=0}^{k-1} a_i x(t_0 + i \cdot \tau).
$$

Because $(t_0 + i \cdot \tau) \in [0, T]$ for any $i = 0, \cdots, k-1$, we have $|x(t_0 + i\tau)|^2 \leq \max_{t \in [0,T]} |x(t)|^2 \lesssim k^4 \log^3 k \|x\|_T^2$ from Lemma 5.1. Hence

$$
\begin{aligned}
|x(t_0 + n \cdot \tau)|^2 &\leq k \sum_{i=0}^{k-1} |a_i|^2 \cdot |x(t_0 + i \cdot \tau)|^2 \\
&\leq k \sum_{i=0}^{k-1} n^{2.2k} \cdot \max_{t \in [0,T]} |x(t)|^2 \\
&\leq k^7 \cdot (2kt/T)^{2.2k} \|x\|_T^2.
\end{aligned}
$$

Thus, we complete the proof. $\qquad \square$

6 Hash Functions and Filter Functions

6.1 Permutation function and hash function

We first review the permutation function $P_{\sigma,a,b}$ and the hash function $h_{\sigma,b}$ in [PS15], which translates discrete settings to the continuous setting.

Definition 6.1. *For any signal $x(t) : \mathbb{R} \to \mathbb{C}$ and $a, b, \sigma \in \mathbb{R}$, let $(P_{\sigma,a,b}x)(t) = x\big(\sigma(t-a)\big)e^{-2\pi \mathbf{i}\sigma bt}$.*

Lemma 6.2. $\widehat{P_{\sigma,a,b}x}(\sigma(f-b)) = \frac{1}{\sigma}e^{-2\pi \mathbf{i}\sigma a f}\widehat{x}(f)$ *and* $\widehat{P_{\sigma,a,b}x}(f) = \frac{1}{\sigma}e^{-2\pi \mathbf{i}\sigma a(f/\sigma+b)}\widehat{x}(f/\sigma+b)$

For completeness, we provide a proof of Lemma 6.2 in Appendix A.4.

Definition 6.3. *[PS15] Let $\pi_{\sigma,b}(f) = 2\pi\sigma(f-b) \pmod{2\pi}$ and $h_{\sigma,b}(f) = \text{round}(\pi_{\sigma,b}(f) \cdot \frac{B}{2\pi})$ be the hash function that maps frequency $f \in [-F, F]$ into bins $\{0, \cdots, B-1\}$.*

Claim 6.4. *[PS15] For any $\Delta > 0$, let σ be a sample uniformly at random from $[\frac{1}{B\Delta}, \frac{2}{B\Delta}]$.*
  (I) If $\Delta \le |f^+ - f^-| < \frac{(B-1)\Delta}{2}$, then $\Pr[h_{\sigma,b}(f^+) = h_{\sigma,b}(f^-)] = 0$
  (II) If $\frac{(B-1)\Delta}{2} \le |f^+ - f^-|$, then $\Pr[h_{\sigma,b}(f^+) = h_{\sigma,b}(f^-)] \lesssim \frac{1}{B}$

From previous work [HIKP12b, HIKP12a, PS15], uniformly sampling from $[A, 2A]$ for some large $A \ge \widetilde{T}$ provides an almost uniform sample on $[0, \widetilde{T}]$ when taken modulo over $\widetilde{T}$.

Lemma 6.5. *For any $\widetilde{T}$, and $0 \le \widetilde{\epsilon}, \widetilde{\delta} \le \widetilde{T}$, if we sample $\widetilde{\sigma}$ uniformly at random from $[A, 2A]$, then*

$$\frac{2\widetilde{\epsilon}}{\widetilde{T}} - \frac{2\widetilde{\epsilon}}{A} \le \Pr\left[\widetilde{\sigma} \pmod{\widetilde{T}} \in [\widetilde{\delta} - \widetilde{\epsilon}, \widetilde{\delta} + \widetilde{\epsilon}\,]\right] \le \frac{2\widetilde{\epsilon}}{\widetilde{T}} + \frac{4\widetilde{\epsilon}}{A}. \tag{15}$$

6.2 Filter function

We state the properties of filter function $(H(t), \widehat{H}(f))$ and $(G(t), \widehat{G}(f))$, the details of proofs are presented in Appendix C.1 and C.2.

Lemma 6.6. *Given $s_0, s_1, 0 < s_3 < 1, \ell > 1, 0 < \delta < 1$, where $\ell = \Theta(k \log(k/\delta))$. The filter function $(H(t), \widehat{H}(f))$ has the following properties,*

$$\text{Property I}: \quad H(t) \in [1 - \delta, 1], \text{ when } |t| \le (\frac{1}{2} - \frac{2}{s_1})s_3.$$

$$\text{Property II}: \quad H(t) \in [0, 1], \text{ when } (\frac{1}{2} - \frac{2}{s_1})s_3 \le |t| \le \frac{1}{2}s_3.$$

$$\text{Property III}: \quad H(t) \le s_0 \cdot (s_1(\frac{|t|}{s_3} - \frac{1}{2}) + 2)^{-\ell}, \forall |t| > \frac{1}{2}s_3.$$

$$\text{Property IV}: \quad \text{supp}(\widehat{H}(f)) \subseteq [-\frac{s_1\ell}{2s_3}, \frac{s_1\ell}{2s_3}].$$

For any exact k-Fourier-sparse signal $x^*(t)$, we shift the interval from $[0, T]$ to $[-1/2, 1/2]$ and consider $x^*(t)$ for $t \in [-1/2, 1/2]$ to be our observation, which is also $x^*(t) \cdot \text{rect}_1(t)$.

$$\text{Property V}: \quad \int_{-\infty}^{+\infty}|x^*(t) \cdot H(t) \cdot (1 - \text{rect}_1(t))|^2 \mathrm{d}t < \delta \int_{-\infty}^{+\infty}|x^*(t) \cdot \text{rect}_1(t)|^2 \mathrm{d}t.$$

$$\text{Property VI}: \quad \int_{-\infty}^{+\infty}|x^*(t) \cdot H(t) \cdot \text{rect}_1(t)|^2 \mathrm{d}t \in [1 - \epsilon, 1] \cdot \int_{-\infty}^{+\infty}|x^*(t) \cdot \text{rect}_1(t)|^2 \mathrm{d}t.$$

for arbitrarily small constant ϵ.
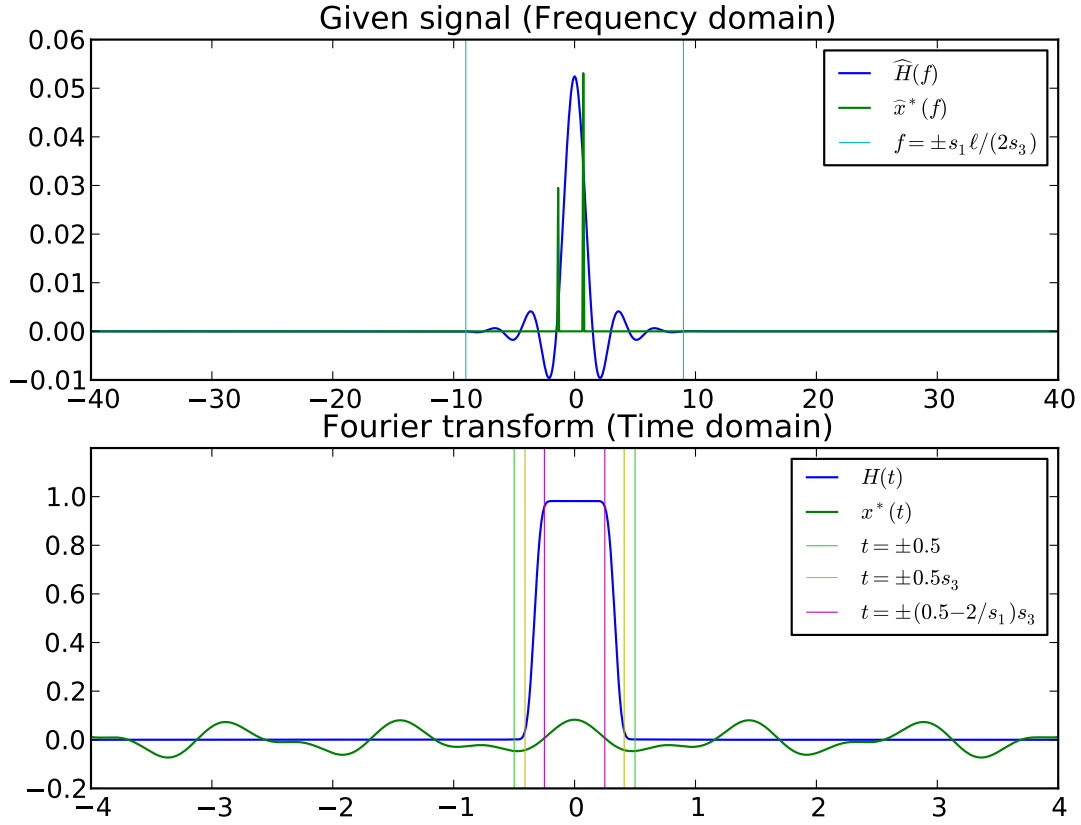
Figure 2: The filter function $(H(t), \widehat{H}(f))$ with a k-Fourier-sparse signal. The property I, II and III are presented in the bottom one, the property IV is presented in the top one.

Lemma 6.7. *Given $B > 1$, $\delta > 0$, $\alpha > 0$, we set $l = \Omega(\log(\delta/k))$. The filter function $(G(t), \widehat{G}(f))[B, \delta, \alpha, l]$ satisfies the following properties,*

$$\text{Property I:} \quad \widehat{G}(f) \in [1 - \delta/k, 1], \ \ if \ |f| \leq (1 - \alpha)\frac{2\pi}{2B}.$$

$$\text{Property II:} \quad \widehat{G}(f) \in [0, 1], \ \ if \ (1 - \alpha)\frac{2\pi}{2B} \leq |f| \leq \frac{2\pi}{2B}.$$

$$\text{Property III:} \quad \widehat{G}(f) \in [-\delta/k, \delta/k], \ \ if \ |f| > \frac{2\pi}{2B}.$$

$$\text{Property IV:} \quad \text{supp}(G(t)) \subset [\frac{l}{2} \cdot \frac{-B}{\pi\alpha}, \frac{l}{2} \cdot \frac{B}{\pi\alpha}].$$

$$\text{Property V:} \quad \max_t |G(t)| \lesssim \text{poly}(B, l).$$

6.3 HASHTOBINS

We first define two functions $G_{\sigma,b}^{(j)}(t)$ and $\widehat{G}_{\sigma,b}^{(j)}(f)$, then show the result returned by Procedure HASHTOBINS in Algorithm 6 satisfying some nice properties. The details of proofs are presented in Appendix C.4.
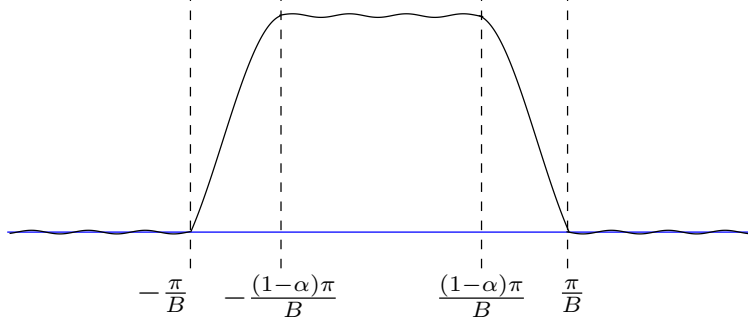
Figure 3: G and $\widehat{G}$. [PS15]

Definition 6.8. $\forall \sigma > 0, b$ and $j \in [B]$. Define,

$$G_{\sigma,b}^{(j)}(t) = \frac{1}{\sigma}G(t/\sigma)e^{2\pi \mathbf{i}t(j/B-\sigma b)/\sigma}$$

$$\widehat{G}_{\sigma,b}^{(j)}(f) = \widehat{G}^{\text{dis}}(\frac{j}{B} - \sigma f - \sigma b) = \sum_{i\in\mathbb{Z}}\widehat{G}(i + \frac{j}{B} - \sigma f - \sigma b)$$

Lemma 6.9. *Let* $u \in \mathbb{C}^B$ *be the result of* HashToBins *under permutation* $P_{\sigma,a,b}$, *and let* $j \in [B]$. *Define*

$$\widehat{z} = \widehat{x \cdot H} \cdot \widehat{G}_{\sigma,b}^{(j)},$$

so

$$z = (x \cdot H) * G_{\sigma,b}^{(j)}.$$

Let vector $\widehat{u} \in \mathbb{C}^B$ *denote the B-dimensional DFT of* u, *then* $\forall j \in [B]$,

$$\widehat{u}[j] = z_{\sigma a}.$$

7   Frequency Recovery

The goal of this section is to prove Theorem 2.6, which is able to recover the frequencies of a signal x^* has k-sparse Fourier transform under noise.

Theorem 2.6. *Let* $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i}f_j t}$ *and* $x(t) = x^*(t) + g(t)$ *be our observable signal where* $\|g(t)\|_T^2 \leq c\|x^*(t)\|_T^2$ *for a sufficiently small constant* c. *Then Procedure* FrequencyRecov-eryKCluster *returns a set* L *of* $O(k)$ *frequencies that covers all heavy clusters of* x^*, *which uses* $\text{poly}(k, \log(1/\delta))\log(FT)$ *samples and* $\text{poly}(k, \log(1/\delta))\log^2(FT)$ *time. In particular, for* $\Delta = \text{poly}(k, \log(1/\delta))/T$ *and* $\mathcal{N}^2 := \|g(t)\|_T^2 + \delta\|x^*(t)\|_T^2$, *with probability* $1 - 2^{-\Omega(k)}$, *for any* f^* *with*

$$\int_{f^*-\Delta}^{f^*+\Delta} |\widehat{x \cdot H}(f)|^2 \mathrm{d}f \geq T\mathcal{N}^2/k, \tag{16}$$

there exists an $\widetilde{f} \in L$ *satisfying*

$$|f^* - \widetilde{f}| \lesssim \Delta\sqrt{\Delta T}.$$

25

7.1 Overview

We give an overview of proving Theorem 2.6. Instead of starting with k-cluster recovery, we first show how to achieve one-cluster recovery.

One-cluster recovery. we start with $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$ where there exists f_0 and Δ such that f_j is in $[f_0 - \Delta, f_0 + \Delta]$ for each $j \in [k]$ and consider its properties for frequency recovery.

Definition 7.1 ((ϵ, Δ)-one-cluster signal). *We say that a signal $z(t)$ is an (ϵ, Δ)-one-cluster signal around f_0 iff $z(t)$ and $\widehat{z}(f)$ satisfy the following two properties:*

$$
\text{Property I} \quad : \quad \int_{f_0 - \Delta}^{f_0 + \Delta} |\widehat{z}(f)|^2 \mathrm{d}f \geq (1 - \epsilon) \int_{-\infty}^{+\infty} |\widehat{z}(f)|^2 \mathrm{d}f
$$

$$
\text{Property II} \quad : \quad \int_{0}^{T} |z(t)|^2 \mathrm{d}t \geq (1 - \epsilon) \int_{-\infty}^{+\infty} |z(t)|^2 \mathrm{d}t.
$$

The main result of one-cluster recovery is to prove that the two properties in Definition 7.1 with a sufficiently small constant ϵ are sufficient to return $\widetilde{f}_0$ close to f_0 with high probability, which provides a black-box for k-cluster recovery algorithm.

We first prove that the pair of conditions, Property I and Property II in Definition 7.1, are sufficient to obtain an estimation of $e^{2\pi \mathbf{i} f_0}$ in Section 7.2. We also provide the proof of the correctness of Procedures GETLEGAL1SAMPLE and GETEMPIRICAL1ENGERGY in Section 7.2.

Lemma 7.2. *For a sufficiently small constant $\epsilon > 0$, any $f_0 \in [-F, F]$, and $\Delta > 0$, given $\widehat{\beta} \approx \frac{1}{\Delta \sqrt{\Delta T}}$ and an (ϵ, Δ)-one-cluster signal $z(t)$ around f_0, Procedure GETLEGAL1SAMPLE in Algorithm 3 with any $\beta \leq 2\widehat{\beta}$ takes $O((T\Delta)^3)$ samples to output $\alpha \in \mathbb{R}$ satisfying*

$$
|z(\alpha + \beta) - z(\alpha) e^{2\pi \mathbf{i} f_0 \beta}| \leq 0.08(|z(\alpha)| + |z(\alpha + \beta)|),
$$

with probability at least 0.6.

The following lemma shows that for any (ϵ, Δ)-one-cluster signal $z(t)$ around f_0, we could use the above procedure to find a frequency $\widetilde{f}_0$ approximating f_0 with high probability.

Lemma 7.3. *For a sufficiently small constant $\epsilon > 0$, any $f_0 \in [-F, F]$, and $\Delta > 0$, given an (ϵ, Δ)-one-cluster signal $z(t)$ around f_0 , Procedure FREQUENCYRECOVERY1CLUSTER in Algorithm 4 returns $\widetilde{f}_0$ with $|\widetilde{f}_0 - f_0| \lesssim \Delta \cdot \sqrt{\Delta T}$ with probability at least $1 - 2^{-\Omega(k)}$.*

We provide a proof of Lemma 7.3 in Section 7.4. We show $z(t) = (x^*(t) + g(t)) \cdot H(t)$ satisfy Properties I and II (Definition 7.1) when all frequencies in $\widehat{x}^*$ are in a small range in Section 7.3.

Lemma 7.4. *For any $f_0 \in [-F, F]$, $\Delta' > 0$, and $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_t}$ with $|f_j - f_0| \leq \Delta'$ for all $j \in [k]$, let $x(t) = x^*(t) + g(t)$ be our observable signal whose noise $\|g\|_T^2 \leq c\|x^*\|_T^2$ for a sufficiently small constant c and $H(t)$ be the filter function defined in Section 6 with $|\mathrm{supp}(\widehat{H})| = \Delta_h$. Then $z = H \cdot x$ is an $(O(\sqrt{c}), \Delta_h + \Delta')$-one-cluster signal around f_0.*

From all discussion above, we summarize the result of frequency recovery when $\widehat{x}^*$ is in one cluster.

Theorem 7.5. *For any $f_0 \in [-F, F]$, $\Delta' > 0$, and $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$ with $|f_j - f_0| \leq \Delta'$ for all $j \in [k]$, let $x(t) = x^*(t) + g(t)$ be our observable signal whose noise $\|g\|_T^2 \leq c\|x^*\|_T^2$ for a sufficiently small constant c and $H(t)$ be the filter function defined in Section 6 with $|\mathrm{supp}(\widehat{H})| = \Delta_h$. Then Procedure FREQUENCYRECOVERY1CLUSTER in Algorithm 4 with $\Delta = \Delta' + \Delta_h$ takes $\mathrm{poly}(k, \log(1/\delta)) \cdot \log(FT)$ samples, runs in $\mathrm{poly}(k, \log(1/\delta)) \cdot \log^2(FT)$ time, returns a frequency $\widetilde{f}_0$ satisfying $|\widetilde{f}_0 - f_0| \lesssim \Delta \sqrt{\Delta T}$ with probability at least $1 - 2^{-\Omega(k)}$.*

k-cluster recovery. Given any $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi i f_j t}$, we plan to convolve the filter function $G(t)$ on $x(t) \cdot H(t)$ and use Lemma 7.3 as a black box to find a list of frequencies that covers $\{f_1, \cdots, f_k\}$.

We fix $\Delta = \text{poly}(k, \log(1/\delta))/T$, $B = \Theta(k)$ and sample σ uniformly at random from $[\frac{1}{B\Delta}, \frac{2}{B\Delta}]$ for k-cluster recovery. We will cover all $f^* \in [-F, F]$ with the following property :

$$\int_{f^*-\Delta}^{f^*+\Delta} |\widehat{x \cdot H}(f)|^2 \mathrm{d}f \geq T\mathcal{N}^2/k, \tag{17}$$

We consider one frequency $f^* \in [-F, F]$ satisfying (17) and use $j = h_{\sigma,b}(f^*)$ to denote its index in $[B]$ after hashing (σ, b). Recall that for $j \in [B]$, any $\sigma > 0$ and any b,

$$G_{\sigma,b}^{(j)}(t) = \frac{1}{\sigma}G(t/\sigma)e^{2\pi it(j/B-\sigma b)/\sigma} \text{ such that } \widehat{G}_{\sigma,b}^{(j)}(f) = \sum_{i \in \mathbb{Z}} \widehat{G}(i + \frac{j}{B} - \sigma f - \sigma b).$$

We set $\widehat{z} = \widehat{x \cdot H} \cdot \widehat{G}_{\sigma,b}^{(j)}$ and $z = (x \cdot H) * G_{\sigma,b}^{(j)}$ for f^* and $j = h_{\sigma,b}(f^*)$. In Section 7.5, we show that with high probability over the hashing (σ, b), $(z, \widehat{z})$ satisfies Property I with $[f^* - \Delta, f^* + \Delta]$ and Property II in Definition 7.1 such that we could use Lemma 7.3 on z to recover f^*.

Lemma 7.6. *Let $f^* \in [-F, F]$ satisfy (17). For a random hashing (σ, b), let $j = h_{\sigma,b}(f^*)$ be the bucket that f^* maps to under the hash such that $z = (x \cdot H) * G_{\sigma,b}^{(j)}$ and $\widehat{z} = \widehat{x \cdot H} \cdot \widehat{G}_{\sigma,b}^{(j)}$. With probability at least 0.9, $z(t)$ is an (ϵ, Δ)-one-cluster signal around f^* .*

Combining Lemma 7.6 and Lemma 7.3, we could recover any heavy frequency f^* satisfying (17) with probability at least 0.8. Then we repeat this procedure to guarantee that we cover all heavy frequencies and finish the proof of the main frequency recovery Theorem 2.6 in Section 7.6.

7.2 Analysis of GetLegal1Sample and GetEmpirical1Energy

Let $I = [f_0 - \Delta, f_0 + \Delta]$ and $\overline{I} = (-\infty, +\infty) \setminus I$ in this proof. We define $(z^I(t), \widehat{z}^I(f))$ and $(z^{\overline{I}}(t), \widehat{z}^{\overline{I}}(f))$ as follows:

$$\widehat{z}^I(f) = \begin{cases} \widehat{z}(f) & \text{if } f \in I \\ 0 & \text{if } f \in \overline{I} \end{cases}, \quad \widehat{z}^{\overline{I}}(f) = \begin{cases} 0 & \text{if } f \in I \\ \widehat{z}(f) & \text{if } f \in \overline{I} \end{cases}$$

We consider $z^I(t)$ as the "signal" to recover f_0 and treat $z^{\overline{I}}(t)$ as the "noise". We first show some basic properties of $z^I(t)$.

Claim 7.7. *For $z^{\overline{I}}(t)$, we have $\int_0^T |z^{\overline{I}}(t)|^2 \mathrm{d}t \leq \epsilon \int_{-\infty}^{+\infty} |z(t)|^2 \mathrm{d}t$. For $z^I(t)$, we have*

$$\int_0^T |z^I(t)|^2 \mathrm{d}t \geq (1 - 5\sqrt{\epsilon}) \int_{-\infty}^{+\infty} |z(t)|^2 \mathrm{d}t \text{ and } \int_0^T |z^I(t)|^2 \mathrm{d}t \geq (1 - 6\sqrt{\epsilon}) \int_{-\infty}^{+\infty} |z^I(t)|^2 \mathrm{d}t.$$

Proof. From the definition and Property I in Definition 7.1, we know

$$z(t) = z^I(t) + z^{\overline{I}}(t) \quad \text{and} \quad \int_{-\infty}^{+\infty} |\widehat{z}^{\overline{I}}(f)|^2 \mathrm{d}f \leq \epsilon \int_{-\infty}^{+\infty} |\widehat{z}(f)|^2 \mathrm{d}f.$$

Notice that Property I(in Definition 7.1) indicates that

$$\int_0^T |z^{\overline{I}}(t)|^2 \mathrm{d}t \leq \int_{-\infty}^{+\infty} |z^{\overline{I}}(t)|^2 \mathrm{d}t = \int_{-\infty}^{+\infty} |\widehat{z}^{\overline{I}}(f)|^2 \mathrm{d}f \leq \epsilon \int_{-\infty}^{+\infty} |\widehat{z}(f)|^2 \mathrm{d}f.$$

On the other hand, from Property II(in Definition 7.1), we know

$$(1-\epsilon)\int_{-\infty}^{+\infty}|z(t)|^2\mathrm{d}t \le \int_0^T|z^I(t)+z^{\overline{I}}(t)|^2\mathrm{d}t \le \int_0^T|z^I(t)|^2\mathrm{d}t+2\int_0^T|z^I(t)|\cdot|z^{\overline{I}}(t)|\mathrm{d}t+\int_0^T|z^{\overline{I}}(t)|^2\mathrm{d}t.$$

We have $\int_0^T|z^I(t)|^2\mathrm{d}t \le 2\int_{-\infty}^{+\infty}|z(t)|^2\mathrm{d}t$ from the above inequality. From $\int_0^T|z^{\overline{I}}(t)|^2\mathrm{d}t \le \epsilon\int_{-\infty}^{+\infty}|\widehat{z}(f)|^2\mathrm{d}f$, we bound

$$\int_0^T|z^I(t)|\cdot|z^{\overline{I}}(t)|\mathrm{d}t \le \sqrt{2\epsilon}\int_{-\infty}^{+\infty}|z(t)|^2\mathrm{d}t$$

by the Cauchy-Schwartz inequality and have

$$\int_0^T|z^I(t)|^2\mathrm{d}t \ge (1-5\sqrt{\epsilon})\int_{-\infty}^{+\infty}|z(t)|^2\mathrm{d}t. \tag{18}$$

Because $\int_{-\infty}^{+\infty}|z^{\overline{I}}(t)|^2\mathrm{d}t \le \epsilon\int_{-\infty}^{+\infty}|z(t)|^2\mathrm{d}t$, inequality (18) also indicates that

$$\int_0^T|z^I(t)|^2\mathrm{d}t \ge (1-6\sqrt{\epsilon})\int_{-\infty}^{+\infty}|z^I(t)|^2\mathrm{d}t.$$

$\square$

One useful property of $z^I(t)$ is that its maximum can be bounded by its average on $[0,T]$.

Claim 7.8. $\forall t \in [0,T], |z^I(t)| \le 2\sqrt{\Delta T}\cdot\|z^I\|_T.$

Proof. From the definition $|z^I(t)|$, it is upper bounded by $\int_{f_0-\Delta}^{f_0+\Delta}|\widehat{z^I}(f)|\mathrm{d}f$ for any $t \in [0,T]$. On the other hand,

$$\int_{f_0-\Delta}^{f_0+\Delta}|\widehat{z^I}(f)|\mathrm{d}f \le \sqrt{2\Delta}(\int_{f_0-\Delta}^{f_0+\Delta}|\widehat{z^I}(f)|^2\mathrm{d}f)^{1/2}$$
$$= \sqrt{2\Delta}(\int_{-\infty}^{+\infty}|z^I(t)|^2\mathrm{d}t)^{1/2}$$
$$\le 2\sqrt{\Delta}(\int_0^T|z^I(t)|^2\mathrm{d}t)^{1/2}$$
$$= 2\sqrt{\Delta T}\|z^I\|_T.$$

$\square$

Claim 7.9. *Given* $\widehat{\beta} = \frac{C_\beta}{\Delta\cdot\sqrt{\Delta T}}$ *with a sufficiently small constant* C_β*, for any two* $\widehat{\beta}$*-close samples in* $z^I(t)$*, we have that*

$$\forall\alpha \in [0,T], \ \forall\beta \in [\widehat{\beta},2\widehat{\beta}], \quad |z^I(\alpha)e^{2\pi\mathbf{i}f_0\beta} - z^I(\alpha+\beta)| \le 0.01\cdot\|z^I\|_T.$$

Proof. From the definition of the Fourier transform, we have

$$|z^I(a+\beta) - z^I(a)e^{2\pi \mathbf{i} f_0 \beta}| = \left| \int_{f_0-\Delta}^{f_0+\Delta} \widehat{z^I}(f)e^{2\pi \mathbf{i}(fa+f_0\beta)}(e^{2\pi \mathbf{i}(f-f_0)\beta} - 1)\mathrm{d}f \right|$$

$$\leq 2 \cdot (2\pi \Delta \beta) \cdot \int_{f_0-\Delta}^{f_0+\Delta} |\widehat{z^I}(f)|\mathrm{d}f \qquad \text{by Taylor expansion}$$

$$\leq 4\pi \beta \Delta \cdot \sqrt{2\Delta} \left(\int_{f_0-\Delta}^{f_0+\Delta} |\widehat{z^I}(f)|^2 \mathrm{d}f \right)^{\frac{1}{2}} \qquad \text{by Hölder inequality}$$

$$\leq 10\pi \widehat{\beta} \Delta \cdot \sqrt{2\Delta} \left(\int_0^T |z^I(t)|^2 \mathrm{d}t \right)^{\frac{1}{2}} \qquad \text{by inequality (18)}$$

$$\leq 10^{-2}\|z^I\|_T.$$

$\square$

We consider how to output an α such that $e^{2\pi \mathbf{i} f_0 \beta} \approx z(\alpha+\beta)/z(\alpha)$ with high probability in the rest of this section.

If we can sample from $z^I(t)$, we already know $|z^I(\alpha)e^{2\pi \mathbf{i} f_0 \beta} - z^I(\alpha+\beta)| \leq 0.01\|z^I\|_T$ from Claim 7.9. Then it is enough to find any α such that $|z^I(\alpha)| \geq 0.5\|z^I\|_T$. From Claim 7.8, we can take $O(\sqrt{\Delta T})$ samples $(z^I(\alpha), z^I(\alpha+\beta))$ where each α is uniformly sampled from $[0,T]$ such that with high probability, the sample $z^I(\alpha)$ with the largest norm $|z^I(\alpha)|$ satisfies $|z^I(\alpha)| \geq 0.5\|z^I\|_T$. Then we have $e^{2\pi \mathbf{i} f_0 \beta} \approx z^I(\alpha+\beta)/z^I(\alpha)$.

Next, we move to $z(t) = z^I(t) + z^{\overline{I}}(t)$ and plan to output $\alpha \in [0,T]$ with probability at least 0.5 such that $|z^{\overline{I}}(\alpha)| \leq 0.1|z^I(\alpha)|$ and $|z^{\overline{I}}(\alpha+\beta)| \leq 0.1|z^I(\alpha+\beta)|$. Because the "noise" $z^{\overline{I}}(t)$ has $\|z^{\overline{I}}(t)\|_T^2 \geq \epsilon\|z^I(t)\|_T^2$ for a constant ϵ and the bound $\sqrt{\Delta T}$ in Claim 7.8 is a polynomial in k, the approach for $z^I(t)$ cannot guarantee that $z(\alpha+\beta)/z(\alpha) \approx e^{2\pi \mathbf{i} f_0 \beta}$ with probability more than $1/2$.

The key observation is as follows:

Observation 7.10. *For a sufficiently small ϵ and $\|z^{\overline{I}}\|_T^2 \leq \epsilon\|z\|_T^2$, let D_T be the weighted distribution on $[0,T]$ according to $|z(t)|^2$, i.e., $D_T(t) = \frac{|z(t)|^2}{T\|z\|_T^2}$. If we sample $\alpha \in [0,T]$ from the distribution D_T instead of the uniform distribution on $[0,T]$, $|z^{\overline{I}}(\alpha)| \leq 0.01|z^I(\alpha)|$ with probability 0.9.*

It follows from the fact that

$$\mathbb{E}_{\alpha \sim D_T} \frac{|z^{\overline{I}}(\alpha)|^2}{|z(\alpha)|^2} = \int_0^T \frac{|z^{\overline{I}}(\alpha)|^2}{|z(\alpha)|^2} \cdot \frac{|z(\alpha)|^2}{T\|z\|_T^2}\mathrm{d}\alpha = \frac{\int_0^T |z^{\overline{I}}(\alpha)|^2\mathrm{d}\alpha}{T\|z\|_T^2} \leq \epsilon.$$

In Procedure GETLEGAL1SAMPLE, we collect $(\Delta T)^2$ samples (in expectation) $(z(\alpha), z(\alpha+\beta))$ in S_{heavy} with $|z(\alpha)| \geq 0.49\|z\|_T$ and resample one α from these samples according to their norm $|z(\alpha)|^2 + |z(\alpha+\beta)|^2$. We show its correctness as follows.

Because we do not know $0.5\|z\|_T$, we use z_{emp} to approximate it.

Claim 7.11. *Procedure GETEMPIRICAL1ENERGY in Algorithm 3 takes $O((T\Delta)^2)$ samples to output z_{emp} such that $z_{\text{emp}} \in [0.8\|z\|_T, 1.2\|z\|_T]$ with prob. 0.9.*

Proof. We know $z_{\text{emp}}^2 = \mathbb{E}_{i \in [R_{\text{est}}]}[|z(\alpha_i)|^2] = \mathbb{E}_{i \in [R_{\text{est}}]}[|z^I(\alpha_i) + z^{\overline{I}}(\alpha_i)|^2]$.

Notice that $\mathbb{E}_{i \in [R_{\text{est}}]}[|z^I(\alpha_i)|^2]$ is in $[0.99\|z^I\|_T, 1.01\|z^I\|_T]$ with prob. 0.99 from the Chernoff bound and Claim 7.8.

At the same time, $\mathbb{E}_{\alpha_i}[|z^{\overline{I}}(\alpha_i)|^2] = \|z^{\overline{I}}\|_T^2$. With prob. 0.92, $\mathbb{E}_{i\in[R_{\text{est}}]}[|z^{\overline{I}}(\alpha_i)|^2] \leq 13\|z^{\overline{I}}\|_T^2$. For a sufficiently small ϵ and $\|z^{\overline{I}}\|_T^2 \leq \epsilon\|z^I\|_T^2$, $\mathbb{E}_{i\in[R_{\text{est}}]}[|z^{\overline{I}}(\alpha_i)|^2] \leq 13\epsilon\|z^I\|_T^2$.

At last, we bound the cross terms of $|z^I(\alpha_i) + z^{\overline{I}}(\alpha_i)|^2$ by the Cauchy-Schwartz inequality,

$$\mathbb{E}_{i\in R_{\text{est}}}[|\overline{z}^I(\alpha_i)z^{\overline{I}}(\alpha_i)| + |z^I(\alpha_i)\overline{z}^{\overline{I}}(\alpha_i)|]$$

$$\leq 2\,\mathbb{E}_{i\in R_{\text{est}}}[|z^I(\alpha_i)| \cdot |z^{\overline{I}}(\alpha_i)|]$$

$$\leq 2\left(\mathbb{E}_{i\in[R_{\text{est}}]}[|z^I(\alpha_i)|^2] \cdot \mathbb{E}_{i\in[R_{\text{est}}]}[|z^{\overline{I}}(\alpha_i)|^2]\right)^{1/2}$$

$$\leq 10\sqrt{\epsilon}\|z^I\|_T^2.$$

For a sufficiently small ϵ, we have $\mathbb{E}_{i\in[R_{\text{est}}]}[|z(\alpha_i)|^2]^{1/2}$ is in $[0.9\|z^I\|_T, 1.1\|z^I\|_T]$, which is also in $[0.8\|z\|_T, 1.2\|z\|_T]$ because of Property II. $\square$

We assume $z_{\text{emp}} \in [0.8\|z\|_T, 1.2\|z\|_T]$ and focus on $U = \{t \in [0, T]\,\big|\,|z(t)| \geq 0.5z_{\text{emp}}\}$. Notice that

$$\int_U |z(t)|^2\mathrm{d}t = \int_0^T |z(t)|^2\mathrm{d}t - \int_{[0,T]\setminus U} |z(t)|^2\mathrm{d}t \geq (1 - 0.6^2)\int_0^T |z(t)|^2\mathrm{d}t.$$

Let $R_{\text{heavy}} = |S_{\text{heavy}}|$. From Claim 7.8 and ϵ, $\mathbb{E}[R_{\text{heavy}}] \geq R_{\text{repeat}}/(T\Delta)$. So we assume $R_{\text{heavy}} \geq 0.01R_{\text{repeat}}/(T\Delta) = 0.01(T\Delta)^2$ in the rest of this section and think each $\alpha_i \in S_{\text{heavy}}$ is a uniform sample from U over the randomness on S_{heavy}.

Claim 7.12. *With probability 0.95, $\sum_{i\in S_{\text{heavy}}}(|z^{\overline{I}}(\alpha_i)|^2 + |z^{\overline{I}}(\alpha_i+\beta)|^2) \leq 10^{-4}\sum_{i\in S_{\text{heavy}}}(|z(\alpha_i)|^2 + |z(\alpha_i+\beta)|^2)$ for a sufficiently small ϵ and $\|z^{\overline{I}}\|_T^2 \leq \epsilon\|z\|_T^2$.*

Proof. At first,

$$\mathbb{E}_{S_{\text{heavy}}}\left[\sum_{i\in S_{\text{heavy}}}(|z(\alpha_i)|^2 + |z(\alpha_i+\beta)|^2)\right] \geq R_{\text{heavy}} \cdot \mathbb{E}_{t\sim U}[|z(t)|^2] = R_{\text{heavy}} \cdot \frac{\int_U |z(t)|^2\mathrm{d}t}{|U|}.$$

At the same time,

$$\mathbb{E}_{S_{\text{heavy}}}\left[\sum_{i\in S_{\text{heavy}}}[|z^{\overline{I}}(\alpha_i)|^2 + |z^{\overline{I}}(\alpha_i+\beta)|^2]\right] = R_{\text{heavy}} \cdot \mathbb{E}_{t\sim U}[|z^{\overline{I}}(t)|^2 + |z^{\overline{I}}(t+\beta)|^2] \leq \frac{2\int_T^0 |z^{\overline{I}}(t)|^2\mathrm{d}t}{|U|}.$$

From $\int_U |z(t)|^2\mathrm{d}t \geq 0.64\int_0^T |z(t)|^2\mathrm{d}t$ and $\int_T^0 |z^{\overline{I}}(t)|^2\mathrm{d}t \leq \epsilon\int_0^T |z(t)|^2\mathrm{d}t$, we get the conclusion. $\square$

We assume all results in the above claims hold and prove that the sample from S_{heavy} is a good sample such that $z^{\overline{I}}(\alpha)$ is small.

Claim 7.13. *If we sample $i \in S_{\text{heavy}}$ according to the weight $|z(\alpha_i)|^2 + |z(\alpha_i+\beta)|^2$, with prob. at least 0.9, $|z^{\overline{I}}(\alpha_i)| + |z^{\overline{I}}(\alpha_i+\beta)| \leq 0.05(|z(\alpha_i)| + |z(\alpha_i+\beta)|)$.*

Proof. Similar to the proof of the key observation, we compute the expectation of $\frac{|z^{\overline{I}}(\alpha_i)|^2+|z^{\overline{I}}(\alpha_i+\beta)|^2}{|z(\alpha_i)|^2+|z(\alpha_i+\beta)|^2}$ over the sampling in S_{heavy}:

$$\sum_{i\in S_{\text{heavy}}}\frac{|z(\alpha_i)|^2+|z(\alpha_i+\beta)|^2}{\sum_{j\in S_{\text{heavy}}}|z(\alpha_j)|^2+|z(\alpha_j+\beta)|^2}\cdot\frac{|z^{\overline{I}}(\alpha_i)|^2+|z^{\overline{I}}(\alpha_i+\beta)|^2}{|z(\alpha_i)|^2+|z(\alpha_i+\beta)|^2}$$

$$=\frac{\sum_{i\in S_{\text{heavy}}}|z^{\overline{I}}(\alpha_i)|^2+|z^{\overline{I}}(\alpha_i+\beta)|^2}{\sum_{i\in S_{\text{heavy}}}|z(\alpha_i)|^2+|z(\alpha_i+\beta)|^2}$$

$$\leq\ 10^{-4}.$$

By Markov's inequality, when we sample $i\in S_{\text{heavy}}$ according to the weight $|z(\alpha_i)|^2+|z(\alpha_i+\beta)|^2$, $\frac{|z^{\overline{I}}(\alpha_i)|^2+|z^{\overline{I}}(\alpha_i+\beta)|^2}{|z(\alpha_i)|^2+|z(\alpha_i+\beta)|^2}\leq 10^{-3}$ with probability 0.9. We have that with prob. at least 0.9, $|z^{\overline{I}}(\alpha_i)|+|z^{\overline{I}}(\alpha_i+\beta)|\leq 0.05(|z(\alpha_i)|+|z(\alpha_i+\beta)|)$. $\qquad\square$

We assume all above claims hold and finish the proof by setting $\alpha=\alpha_i$. From Claim 7.9, we know that
$$|z^I(\alpha)e^{2\pi\mathbf{i}f_0\beta}-z^I(\alpha+\beta)|\leq 0.01\cdot\mathop{\mathbb{E}}_{t\in[0,T]}[|z^I(\alpha)|^2]^{1/2}\leq 0.03|z^I(\alpha)|.$$

Now we add back the noise $z^{\overline{I}}(\alpha)$ and $z^{\overline{I}}(\alpha+\beta)$ to get

$$|z(\alpha)e^{2\pi\mathbf{i}f_0\beta}-z(\alpha+\beta)|\leq|z^I(\alpha)e^{2\pi\mathbf{i}f_0\beta}-z^I(\alpha+\beta)|+|z^{\overline{I}}(\alpha)|+|z^{\overline{I}}(\alpha+\beta)|\leq 0.08(|z(\alpha)|+|z(\alpha+\beta)|).$$

7.3 A cluster of frequencies, times H, is a one-cluster signal per Definition 7.1

The goal of this section is to prove Lemma 7.4. Without loss of generality, we assume $g(t)=0$ for any $t\notin[0,T]$ and notice that $\text{supp}(\widehat{H}*\widehat{x}^*)\subseteq f_0+[-\Delta,\Delta]$ for $\Delta=\Delta'+\Delta_h$ from the definition of $\widehat{H}$. From the Property VI (presented in Lemma 6.6) of $(H,\widehat{H})$,

$$\int_0^T|x^*(t)|^2\mathrm{d}t=(1\pm c)\int_{-\infty}^{+\infty}|H(t)\cdot x^*(t)|^2\mathrm{d}t.$$

From the first two properties of $(H,\widehat{H})$, we bound the energy of $g\cdot H$:

$$\int_{-\infty}^{+\infty}|H(t)\cdot g(t)|^2\mathrm{d}t\leq(1+c)\int_0^T|g(t)|^2\mathrm{d}t.$$

Let $z(t)=(x^*(t)+g(t))H(t)$. We use the triangle inequality on the above two inequalities:

$$\int_0^T|z(t)|^2\mathrm{d}t$$

$$\geq\int_0^T|H(t)\cdot x^*(t)|^2\mathrm{d}t-\int_0^T|H(t)\cdot g(t)|^2\mathrm{d}t-2\int_0^T|H(t)\cdot x^*(t)|\cdot|H(t)\cdot g(t)|\mathrm{d}t$$

$$\geq(1-c)\int_0^T|x^*(t)|^2\mathrm{d}t-(1+c)\int_0^T|g(t)|^2\mathrm{d}t-2\sqrt{(1+c)^2\int_0^T|g(t)|^2\mathrm{d}t\int_0^T|x^*(t)|^2\mathrm{d}t}\cdot$$

$$\geq(1-5\sqrt{c})\int_0^T|x^*(t)|^2\mathrm{d}t,$$

31

where we use the Cauchy-Schwarz inequality and $\int_0^T |g(t)|^2 dt \leq c \int_0^T |x^*(t)|^2 dt$ in the last step. Similarly,

$$\int_{-\infty}^{+\infty} |z(t)|^2 dt$$

$$\leq (1+c) \int_0^T |x^*(t)|^2 dt + (1+c) \int_0^T |g(t)|^2 dt + 2 \sqrt{(1+c)^2 \int_0^T |x^*(t)|^2 dt \int_0^T |g(t)|^2 dt}$$

$$\leq (1+5\sqrt{c}) \int_0^T |x^*(t)|^2 dt.$$

Hence we obtain Property II(in Definition 7.1) when c is sufficiently small.

Then we observe that

$$\int_{f_0-\Delta_h}^{f_0+\Delta_h} |\widehat{z}(f)|^2 df$$

$$\geq \int_{f_0-\Delta_h}^{f_0+\Delta_h} |H \cdot \widehat{(x^* + g)}|^2 df$$

$$\geq \int_{f_0-\Delta_h}^{f_0+\Delta_h} |\widehat{H \cdot x^*}|^2 - |\widehat{H \cdot g}|^2 - 2|\widehat{H \cdot x^*}| \cdot |\widehat{H \cdot g}| df$$

$$\geq \int_{f_0-\Delta_h}^{f_0+\Delta_h} |\widehat{H \cdot x^*}|^2 df - \int_{-\infty}^{+\infty} |\widehat{H \cdot g}|^2 df - 2 \sqrt{\int_{f_0-\Delta_h}^{f_0+\Delta_h} |\widehat{H \cdot x^*}|^2 df \int_{-\infty}^{+\infty} |\widehat{H \cdot g}|^2 df}$$

$$= \int_{-\infty}^{+\infty} |H \cdot x^*|^2 dt - \int_{-\infty}^{+\infty} |H \cdot g|^2 dt - 2 \sqrt{\int_{f_0-\Delta_h}^{f_0+\Delta_h} |H \cdot x^*|^2 dt \int_{-\infty}^{+\infty} |H \cdot g|^2 dt}$$

$$\geq \frac{(1-c) - c(1+c) - 3\sqrt{c}}{1 + 5\sqrt{c}} \int_{-\infty}^{+\infty} |z(t)|^2 dt.$$

Thus we have Property I(in Definition 7.1) for z.

7.4 Frequency recovery of one-cluster signals

The goal of this section is prove Theorem 7.5. We first show the correctness of Procedure LO-CATE1INNER. Second, we analyze the Procedure LOCATE1SIGNAL. At end, we rerun Procedure LOCATE1SIGNAL and use median analysis to boost the constant success probability.[3]

Lemma 7.14. *Let $f_0 \in \text{region}(q')$. Let β is sampled from $[\frac{st}{4\Delta}, \frac{st}{2\Delta l}]$ and let γ denote the output of Procedure GETLEGAL1SAMPLE in Algorithm 4. Then using the pair of samples $z(\gamma + \beta)$ and $z(\gamma)$, we have*

I. for the q' with probability at least $1 - s$, $v_{q'}$ will increase by one.

II. for any q such that $|q - q'| > 3$, with probability at least $1 - 15s$, v_q will not increase.

Proof. We replace f_0 by θ in the rest of the proof. By Lemma 7.2, we have that for any $\widehat{\beta} \leq \beta \leq 2\widehat{\beta}$, Procedure GETLEGAL1SAMPLE outputs a $\gamma \in [0, T]$ satisfying

$$|z(\gamma + \beta) - z(\gamma)e^{2\pi i f_0 \beta}| \leq 0.1(|z(\gamma)| + |z(\gamma + \beta)|)$$

with probability at least 0.6.

---

[3]The proofs in this section are identical to [HIKP12b] and [PS15].

Furthermore, there exists such some constant $g \in (0, 1)$ such that with probability $1 - g$,

$$\|\phi(z(\gamma + \beta)) - (\phi(z(\gamma)) - 2\pi\beta\theta)\|_{\bigcirc} \lesssim \sin^{-1}(\frac{1}{g}),$$

where $\|x - y\|_{\bigcirc} = \min_{z \in \mathbb{Z}} |x - y + 2\pi z|$ denote the "circular distance" between x and y. We can set $s = \Theta(g^{-1})$. There exists some constant $p = \Theta(s)$, with probability at least $1 - p$,

$$\|o - 2\pi\beta\theta\|_{\bigcirc} < s\pi/2$$

where $o := \phi(z(\gamma + \beta)/z(\gamma))$. The above equation shows that o is a good estimate for $2\pi\beta\theta$ with good probability. We will now show that this means the true region $Q_{q'}$ gets a vote with large probability.

For each q' with $\theta \in [l - \frac{\Delta l}{2} + \frac{q'-1}{t}\Delta l, l - \frac{\Delta l}{2} + \frac{q'}{t}\Delta l] \subset [-F, F]$, we have that $\theta_{q'} = l - \frac{\Delta l}{2} + \frac{q'-0.5}{t}\Delta l$ satisfies that

$$\theta - \theta_{q'} \le \frac{\Delta l}{2t}.$$

Note that we sample β uniformly at random from $[\widehat{\beta}, 2\widehat{\beta}]$, then $2\widehat{\beta} = \frac{st}{2\Delta l} \le \frac{cT}{10A^{\frac{3}{2}}}$ (Note that A is some constant > 1), which implies that $2\pi\beta\frac{\Delta l}{2t} \le \frac{s\pi}{2}$. Thus, we can show the observation o is close to the true region in the following sense,

$$
\begin{aligned}
&\|o - 2\pi\beta\theta_{q'}\|_{\bigcirc} \\
\le \quad &\|o - 2\pi\beta\theta\|_{\bigcirc} + \|2\pi\beta\theta - 2\pi\beta\theta_{q'}\|_{\bigcirc} \text{ by triangle inequality} \\
\le \quad &\frac{s\pi}{2} + 2\pi\|\beta\theta - \beta\theta_{q'}\|_{\bigcirc} \\
\le \quad &s\pi.
\end{aligned}
$$

Thus, $v_{q'}$ will increase in each round with probability at least $1 - s$.

On the other side, consider q with $|q - q'| > 3$. Then $|\theta - \theta_q| \ge \frac{7\Delta l}{2t}$, and (assuming $\beta \ge \frac{st}{4\Delta l}$) we have

$$2\pi\beta|\theta - \theta_q| \ge 2\pi\frac{st}{4\Delta l}|\theta - \theta_q| = \frac{s\pi t}{2\Delta l}|\theta - \theta_q| \ge \frac{7s\pi}{4} > \frac{3s\pi}{2}.$$

There are two cases: $|\theta - \theta_q| \le \frac{\Delta l}{st}$ and $|\theta - \theta_q| > \frac{\Delta l}{st}$.

First, if $|\theta - \theta_q| \le \frac{\Delta l}{st}$. In this case, from the definition of β it follows that

$$2\pi\beta|\theta - \theta_q| \le \frac{s\pi t}{\Delta l}|\theta - \theta_q| \le \pi$$

Combining the above equations implies that

$$\Pr\left[2\pi\beta(\theta - \theta_q) \pmod{2\pi} \in [-\frac{3s}{4}2\pi, \frac{3s}{4}2\pi]\right] = 0$$

Second, if $|\theta - \theta_q| > \frac{\Delta l}{st}$. We show this claim is true : $\Pr[2\pi\beta(\theta - \theta_q) \pmod{2\pi} \in [-\frac{3s}{4}2\pi, \frac{3s}{4}2\pi]] \lesssim s$. To prove it, we apply Lemma 6.5 by setting $\widetilde{T} = 2\pi$, $\widetilde{\sigma} = 2\pi\beta$, $\widetilde{\delta} = 0$, $\epsilon = \frac{3s}{4}2\pi$, $A = 2\pi\widehat{\beta}$, $\Delta f = |\theta - \theta_q|$. By upper bound of Lemma 6.5, the probability is at most

$$\frac{2\widetilde{\epsilon}}{\widetilde{T}} + \frac{4\widetilde{\epsilon}}{A\Delta f} = \frac{3s}{2} + \frac{3s}{\widehat{\beta}\Delta f} \le \frac{3s}{2} + \frac{3s}{\frac{st}{4\Delta l}\frac{\Delta l}{st}} < 15s$$

Then in either case, with probability at least $1 - 15s$, we have

$$\|2\pi\beta\theta_q - 2\pi\beta\theta\|_{\bigcirc} > \frac{3s}{4}2\pi$$

which implies that v_q will not increase. $\qquad\square$

Lemma 7.15. *Procedure* LOCATE1INNER *in Algorithm 4 uses* R_{loc} *"legal" samples, and then after Procedure* LOCATE1SIGNAL *in Algorithm 4 running Procedure* LOCATE1INNER $D_{\max}$ *times, it outputs a frequency* $\widetilde{f}_0$ *such that*

$$|\widetilde{f}_0 - f_0| \lesssim \Delta \cdot \sqrt{T\Delta}$$

with arbitrarily large constant probability.

Proof. For each observation, $v_{q'}$ incremented with probability at least $1-p$ and v_q is incremented with probability at most $15s + p$ for $|q - q'| > 3$. The probabilities corresponding to different observations are independent. Then after R_{loc} observations, there exists some constant $c < \frac{1}{2}$, for any q such that $|q - q'| > 3$,

$$
\begin{aligned}
&\mathsf{Pr}[\text{False region gets more than half votes}] \\
=\ &\mathsf{Pr}[v_{j,q} > R_{\mathrm{loc}}/2] \\
\leq\ &\binom{R_{\mathrm{loc}}}{R_{\mathrm{loc}}/2}(15s + p)^{R_{\mathrm{loc}}/2} \\
\leq\ &c^{\Omega(R_{\mathrm{loc}})}
\end{aligned}
$$

Similarly, on the other side,

$$
\begin{aligned}
&\mathsf{Pr}[\text{True region gets less than half votes}] \\
=\ &\mathsf{Pr}[v_{j,q'} < R_{\mathrm{loc}}/2] \\
\leq\ &\binom{R_{\mathrm{loc}}}{R_{\mathrm{loc}}/2}(p)^{R_{\mathrm{loc}}/2} \\
\leq\ &c^{\Omega(R_{\mathrm{loc}})}
\end{aligned}
$$

Taking the union bound over all the t regions, it gives with probability at least $1 - tf^{\Omega(R_{\mathrm{loc}})}$ we can find some region q such that $|q - q'| < 3$.

If we repeat the above procedure $D_{\max}$ rounds, each round we choose the "False" region with probability at most $1 - tc^{\Omega(R_{\mathrm{loc}})}$. Thus, taking the union bound over all the $D_{\max}$ rounds, we will report a region has size $\approx \Delta\sqrt{\Delta T}$ and contains f_0 with probability at least $1 - D_{\max}tc^{\Omega(R_{\mathrm{loc}})}$.

The reason for not ending up with region that has size $\approx \Delta$ is, the upper bound of the sample range of β force us to choose β is at most $\lesssim \frac{T}{(\Delta T)^{\frac{3}{2}}}$ by Claim 7.9

It remains to explain how to set $D_{\max}, t$, and R_{loc}. At the beginning of the first round, we start with frequency interval of length $2F$, at the beginning of the last round, we start with frequency interval of length $t \cdot \Delta\sqrt{T\Delta}$. Each round we do a t-ary search, thus

$$D_{\max} = \log_t\left(\frac{2F}{t\Delta\sqrt{T\Delta}}\right) \leq \log_t(F/\Delta).$$

We can set $R_{\mathrm{loc}} \approx \log_{1/c}(t/c)$ and $t > D_{\max}$, e.g. $t = \log(F/\Delta)$. Thus, the probability becomes,

$$1 - D_{\max}tc^{\Omega(R_{\mathrm{loc}})} \geq 1 - t^2 c^{\Omega(R_{\mathrm{loc}})} \geq 1 - \mathrm{poly}(1/t, c)$$

which is larger than any constant probability. $\square$

Using the same parameters setting in the proof of Lemma 7.15, we show the running time and sample complexity of Procedure LOCATE1SIGNAL,

34

Lemma 7.16. *Procedure* LOCATE1SIGNAL *in Algorithm 4 uses* $O(\mathrm{poly}(k, \log(1/\delta))) \cdot \log(FT)$ *samples and runs in* $O(\mathrm{poly}(k, \log(1/\delta))) \cdot \log^2(FT)$ *time.*

Proof. The number of "legal" observations is

$$D_{\max} R_{\mathrm{loc}} = O(\log_t(F/\Delta) \log_{1/c}(t/c)) = O(\log(F/\Delta))$$

The total number of samples is

$$R_{\mathrm{est}} + R_{\mathrm{repeat}} D_{\max} R_{\mathrm{loc}} = O(T\Delta_h)^2 + (T\Delta_h)^3 \cdot \log(FT) = \mathrm{poly}(k, \log(1/\delta)) \cdot \log(FT)$$

where the first step follows by Claim 7.11 and Lemma 7.2 and the last step follows by the setting of Δ_h in Appendix C.3.

The running time includes two parts, one is approximately computing $H(t)$ for all the samples, each sample takes $\mathrm{poly}(k, \log(1/\delta))$ time according to Lemma C.8; the other is for each legal sample we need to assign vote to some regions.

$$\mathrm{poly}(k, \log(1/\delta)) \cdot (R_{\mathrm{est}} + R_{\mathrm{repeat}} D_{\max} R_{\mathrm{loc}}) + D_{\max} R_{\mathrm{loc}} t = \mathrm{poly}(k, \log(1/\delta)) \log^2(FT)$$

$\square$

Lemma 7.17 only achieves constant success probability, using median analysis we can boost the success probability,

Lemma 7.17. *Let* $\widetilde{f}_0$ *denote the frequency output by Procedure* FREQUENCYRECOVERY1CLUSTER *in Algorithm 5, then with probability at least* $1 - 2^{-\Omega(k)}$,

$$|\widetilde{f}_0 - f_0| \lesssim \Delta\sqrt{T\Delta}$$

Proof. Because of Procedure FREQUENCYRECOVERY1CLUSTER taking the median of $O(k)$ independent results by repeating algorithm LOCATE1SIGNAL $O(k)$ times. Each sample L_r is close to $\widetilde{f}_0$ with sufficiently large probability. Thus, using the Chernoff bound will output $\widetilde{f}_0$ with probability $1 - 2^{-\Omega(k)}$ such that

$$|\widetilde{f}_0 - f_0| \lesssim \Delta\sqrt{T\Delta}.$$

$\square$

Combining Lemma 7.17 with the sample complexity and running time in Lemma 7.15, we are able to finish the proof of Theorem 7.5.

7.5 The full signal, after multiplying by H and convolving with G, is one-clustered.

The goal of this section is to prove Lemma 7.6. We fix $f^* \in [-F, F]$ satisfying (17) in this section. We first define a good hashing (σ, b) of f^* as follows.

Definition 7.18. *We say that a frequency* f^* *is* well-isolated *under the hashing* (σ, b) *if, for* $j = h_{\sigma,b}(f^*)$, *we have that the signal*

$$\widehat{z}^{(j)} = \widehat{x \cdot H} \cdot \widehat{G}^{(j)}_{\sigma,b}$$

satisfies, over the interval $\overline{I_{f^*}} = (-\infty, \infty) \setminus (f^* - \Delta, f^* + \Delta)$,

$$\int_{\overline{I_{f^*}}} |\widehat{z}^{(j)}(f)|^2 df \lesssim \epsilon \cdot T\mathcal{N}^2/k.$$

For convenience, we simplify $z^{(j)}$ by using z in the rest of this section.

Lemma 7.19. *Let f^* be any frequency. Then f^* is well-isolated by a hashing (σ, b) with probability ≥ 0.9 given $B = \Theta(k)$ and $\sigma \in [\frac{1}{B\Delta}, \frac{2}{B\Delta}]$ chosen uniformly at random.*

Proof. For any other frequency f' in x^*, its contribution in $\widehat{z}$ depends on how far it is from f^*. Either it is:

- Within Δ of f^*, f' and f^* will be mapped into the same bucket with probability at least 0.99.

- Between Δ and $1/\sigma$ far, from Claim 6.4, f' and f^* will always mapped into different buckets. Hence f' always contributes in the $\frac{\epsilon\delta}{k}$ region of Property III in Lemma 6.7 about filter function $(G(t), \widehat{G}(f))$, i.e., it contributes at most $\frac{\epsilon\delta}{k} \cdot \int_{f'-\Delta}^{f'+\Delta} |\widehat{x \cdot H}|^2 \mathrm{d}f$. Overall it will contribute

$$\frac{\epsilon\delta}{k} \cdot \int |\widehat{x \cdot H}|^2 \mathrm{d}f = \frac{\epsilon\delta}{k} \int |x \cdot H|^2 \mathrm{d}t.$$

- More than $1/\sigma$ far, in which case they contribute in the same region with probability at most $3/B$. By a union bound, it is at most $3k/B \leq 0.01$

$\square$

Without loss of generality, we assume $\mathrm{supp}(\widehat{g \cdot H}) \cap \mathrm{supp}(\widehat{x^* \cdot H}) = \emptyset$, otherwise we treat it as a part of $x^* \cdot H$. We first consider frequency $f^* \in \widehat{x^* \cdot H}$ under $G_{\sigma,b}^{(j)}$.

Lemma 7.20. *Let f^* satisfying $\int_{f^*-\Delta}^{f^*+\Delta} |\widehat{x^* \cdot H}(f)|^2 \mathrm{d}f \geq T\mathcal{N}^2/k$ and $\widehat{z} = \widehat{x^* \cdot H} \cdot \widehat{G}_{\sigma,b}^{(j)}$ where $j = h_{\sigma,b}(f^*)$. If f^* is well-isolated, then z and $\widehat{z}$ satisfying Property I(in Definition 7.1), i.e.,*

$$\int_0^T |z(t)|^2 \mathrm{d}t \geq (1 - \epsilon) \int_{-\infty}^{+\infty} |z(t)|^2 \mathrm{d}t.$$

Proof. We first notice that $z(t) = x^*(t) \cdot H(t) * G_{\sigma,b}^{(j)}(t)$ and lower bound $\int_{-\infty}^{+\infty} |z(t)|^2 \mathrm{d}t$ as follows :

$$
\begin{aligned}
&\int_{-\infty}^{+\infty} |x^*(t) \cdot H(t) * G_{\sigma,b}^{(j)}(t)|^2 \mathrm{d}t \\
=\ & \int_{-\infty}^{+\infty} |\widehat{x^* \cdot H}(f) \cdot \widehat{G}_{\sigma,b}^{(j)}(f)|^2 \mathrm{d}f && \text{by FT} \\
\geq\ & \int_{f_0-\Delta}^{f_0+\Delta} |\widehat{x^* \cdot H}(f) \cdot \widehat{G}_{\sigma,b}^{(j)}(f)|^2 \mathrm{d}f \\
\geq\ & (1 - \delta)^2 \int_{f_0-\Delta}^{f_0+\Delta} |\widehat{x^* \cdot H}(f)|^2 \mathrm{d}f \\
\geq\ & (1 - \delta)^2 T\mathcal{N}^2/k \\
\geq\ & 0.9 \frac{\delta}{k} \int_0^T |x^*(t)|^2 \mathrm{d}t && (19)
\end{aligned}
$$

We give an upper bound $\int_{-\infty}^0 |z(t)|^2 \mathrm{d}t + \int_T^{+\infty} |z(t)|^2 \mathrm{d}t \lesssim \epsilon \frac{\delta}{k} \int_0^T |x^*(t)H(t)|^2 \mathrm{d}t$ in the rest of this proof.

36

Consider the case $t < 0$, by definition of Convolution,

$$z^{(j)}(t) \;\; = \;\; x^*(t) \cdot H(t) * G_{\sigma,b}^{(j)}(t) = \int_{-\infty}^{+\infty} G_{\sigma,b}^{(j)}(t-\tau) \cdot (x^* \cdot H)(\tau)\mathrm{d}\tau$$

Without loss of generality, we can shift the original signal and $H(t)$ from $[0,T]$ to $[-T/2, T/2]$, by Property of $H(t)$, we know that if $s_3 T/2 \le |t| \le T/2$, then $H(t) \le 2^{-O\Theta(\ell)}$. Note that $G(t)$ is compact and has support DB, we also assume its compact region is $[-DB/2, DB/2]$ (Recall that $D = \frac{l}{\alpha\pi}$).

Thus, by definition of convolution,

$$z(t)$$
$$= \int_{-DB\sigma/2}^{DB\sigma/2} G_{\sigma,b}^{(j)}(s) \cdot (x \cdot H)(t-\tau)\mathrm{d}\tau$$
$$= \frac{1}{\sigma} \int_{-DB\sigma/2}^{DB\sigma/2} G(s/\sigma) e^{2\pi \mathbf{i} s(j/B - \sigma b)/\sigma} \cdot (x \cdot H)(t-\tau)\mathrm{d}\tau$$
$$\le \frac{1}{\sigma} \int_{-DB\sigma/2}^{DB\sigma/2} |G(\tau/\sigma)| \cdot |(x \cdot H)(t-\tau)|\mathrm{d}\tau$$
$$\le \left(\frac{1}{\sigma} \int_{-DB\sigma/2}^{DB\sigma/2} |G(\tau/\sigma)|\mathrm{d}\tau \right) \cdot \left(\max_{|\tau| \le DB\sigma/2} |(x \cdot H)(t-\tau)| \right)$$

So, if $t \notin [-T/2, T/2]$, then $t - s \notin [-T/2 + DB\sigma/2, T/2 - DB\sigma/2]$. By Property V of $G(t)$, $|G(t)| \le \mathrm{poly}(k, \log(1/\delta))$. Because of the parameter setting[4], we have the fact $[-Ts_3/2, Ts_3/2] \subseteq [-T/2 + DB\sigma/2, T/2 - DB\sigma/2] \subseteq [-T/2, T/2]$. Thus, we know $T(1-s_3)/2 > DB\sigma/2$, then for any $t - \tau \in [-T/2, -T/2 + DB\sigma/2] \cup [T/2 - DB\sigma/2, T/2] = S$, then

$$|z(t)|^2 \lesssim \left(DB\sigma \cdot \frac{1}{\sigma} \cdot \mathrm{poly}(k, \log(1/\delta)) \right)^2 \cdot 2^{-\Theta(\ell)} \cdot k^4 \cdot \|x^*(t)\|_T^2 \lesssim \mathrm{poly}(k, \log(1/\delta)) \cdot 2^{-\Theta(\ell)} \cdot \|x^*(t)\|_T^2.$$

Thus, taking the integral over S,

$$\int_S |z(t)|^2 \mathrm{d}t \lesssim |S| \cdot 2^{-\Theta(\ell)} \mathrm{poly}(k, \log(1/\delta)) \cdot \|x^*(t)\|^2 \lesssim 2^{-\Theta(\ell)} T \|x^*(t) \cdot H(t)\|_T^2$$

By property of filter function $H(t), \widehat{H}(f)$, we have

$$|(x \cdot H)(t)|^2 \le (\frac{t}{T})^{-\ell} \|x^*(t) \cdot H(t)\|_T^2 \text{ if } t \ge 3T$$

Thus for any constant ϵ,

$$\int_{-\infty}^{-T/2} |z(t)|^2 \mathrm{d}t + \int_{T/2}^{+\infty} |z(t)|^2 \mathrm{d}t \lesssim 2^{-\ell} T \|x^*(t) \cdot H(t)\|_T^2 \le 0.9\epsilon \cdot \frac{\delta}{k} \int_{-T/2}^{T/2} |x^*(t)|^2 \mathrm{d}t \qquad (20)$$

where the last inequality follows by $\ell \gtrsim k \log(k/\delta)$. Shifting the interval from $[-T/2, T/2]$ to $[0, T]$, the same result is still holding. Combining Equation (19) and (20) completes the proof of Property II.

$\square$

---

[4] We will set B to be $O(k)$, D to be $\mathrm{poly}(k)$ and σ to be $T/\mathrm{poly}(k)$.

We consider frequency $f^* \in \widehat{g \cdot H}$ under $G_{\sigma,b}^{(j)}$ and show the energy of noise $g(t)$ is evenly distributed over B bins on expectation.

Lemma 7.21. *Given any noise $g(t) : [0, T] \to \mathbb{C}$ and $g(t) = 0, \forall t \notin [0, T]$. We have, $\forall j \in [B]$,*

$$\mathop{\mathbb{E}}_{\sigma,b} \left[\int_{-\infty}^{+\infty} |g(t)H(t) * G_{\sigma,b}^{(j)}(t)|^2 \mathrm{d}t \right] \lesssim \frac{1}{B} \int_{-\infty}^{+\infty} |g(t)H(t)|^2 \mathrm{d}t$$

Proof. Because of Fourier Transform preserves ℓ_2 norm, it suffices to prove

$$\mathop{\mathbb{E}}_{\sigma,b} \left[\int_{-\infty}^{+\infty} |\widehat{g \cdot H}(f) \cdot \widehat{G}_{\sigma,b}^{(j)}(f)|^2 \mathrm{d}f \right] \lesssim \frac{1}{B} \int_{-\infty}^{+\infty} |\widehat{g \cdot H}(f)|^2 \mathrm{d}f$$

Since $\widehat{G}_{\sigma,b}^{(j)}(f)$ is a periodic function and outputs at most 1 on $O(1/B)$ fraction of the period, and outputs $\leq \delta$ on other part. Thus, for any frequency f, we have

$$\mathop{\mathbb{E}}_{\sigma,b} \left[|\widehat{G}_{\sigma,b}^{(j)}(f)|^2 \right] \lesssim \frac{1}{B}$$

Thus, we have

$$\mathop{\mathbb{E}}_{\sigma,b} \left[\int_{-\infty}^{+\infty} |\widehat{g \cdot H}(f) \cdot \widehat{G}_{\sigma,b}^{(j)}(f)|^2 \mathrm{d}f \right]$$
$$\leq \mathop{\mathbb{E}}_{\sigma,b} \left[\int_{-\infty}^{+\infty} |\widehat{g \cdot H}(f)|^2 \cdot |\widehat{G}_{\sigma,b}^{(j)}(f)|^2 \mathrm{d}f \right]$$
$$= \int_{-\infty}^{+\infty} |\widehat{g \cdot H}(f)|^2 \cdot \mathop{\mathbb{E}}_{\sigma,b} [|\widehat{G}_{\sigma,b}^{(j)}(f)|^2] \mathrm{d}f$$
$$\leq \int_{-\infty}^{+\infty} |\widehat{g \cdot H}(f)|^2 \mathrm{d}f \cdot \max_f \left[\mathop{\mathbb{E}}_{\sigma,b} \left| \widehat{G}_{\sigma,b}^{(j)}(f) \right|^2 \right]$$
$$\lesssim \frac{1}{B} \int_{-\infty}^{+\infty} |\widehat{g \cdot H}(f)|^2 \mathrm{d}f,$$

which completes the proof. $\qquad \square$

Proof of Lemma 7.6. Let $j = h_{\sigma,b}(f^*)$, signal

$$\widehat{z} = \widehat{x \cdot H} \cdot \widehat{G}_{\sigma,b}^{(j)}, \tag{21}$$

and region $I_{f^*} = (f^* - \Delta, f^* + \Delta)$ with complement $\overline{I_{f^*}} = (-\infty, \infty) \setminus I_{f^*}$. From Property I of G in Lemma 6.7, we have that

$$\widehat{G}_{\sigma,b}^{(l)}(f) \gtrsim 1$$

for all $f \in I_{f^*}$, so by (17)

$$\int_{I_{f^*}} |\widehat{z}(f)|^2 \mathrm{d}f \geq T\mathcal{N}^2/k.$$

On the other hand, f^* is will-isolated with probability 0.9:

$$\int_{\overline{I_{f^*}}} |\widehat{z}(f)|^2 \mathrm{d}f \lesssim \epsilon T\mathcal{N}^2/k.$$

Hence, $\widehat{z}$ satisfies the Property I(in Definition 7.1) of one-mountain recovery. Combining Lemma 7.20 and Lemma 7.21, we know that $(x^* \cdot H) * G_{\sigma,b}^{(j)}$ always satisfies Property II(in Definition 7.1) and $\int_{-\infty}^{+\infty} |g(t)H(t) * G_{\sigma,b}^{(j)}(t)|^2 \mathrm{d}t$ is less than $20T\mathcal{N}^2/B \leq \epsilon T\mathcal{N}^2/k$ with probability at least 0.95, which indicates that $z = (x^* + g) \cdot H * G_{\sigma,b}^{(j)}$ satisfies Property II(in Definition 7.1) with probability 0.95. $\quad \square$

7.6 Frequency recovery of k-clustered signals

The goal of this section is to prove that the frequencies found by Procedure FREQUENCYRECOV-ERYKCLUSTER in Algorithm 8 have some reasonable guarantee.

We first notice that Lemma 7.6 and Lemma 7.3 imply the following lemma by a union bound.

Lemma 7.22. *Let* $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$. *We observe* $x(t) = x^*(t) + g(t)$, *where* $\|g(t)\|_T^2 \leq c\|x^*(t)\|_T^2$ *for a sufficiently small constant* c *and define* $\mathcal{N}^2 := \|g(t)\|_T^2 + \delta\|x^*(t)\|_T^2$. *Then Procedure* ONESTAGE *returns a set* L *of* $O(k)$ *frequencies that covers the heavy frequencies of* x^*. *In particular, for any* f^* *with*

$$\int_{f^*-\Delta}^{f^*+\Delta} |\widehat{x \cdot H}(f)|^2 \mathrm{d}f \geq T\mathcal{N}^2/k, \tag{22}$$

there will exist an $\widetilde{f} \in L$ *satisfying* $|f^* - \widetilde{f}| \lesssim \sqrt{T\Delta} \cdot \Delta T$ *with probability* 0.99.

Lemma 7.23. *Let* $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$ *and* $R = O(k)$. *We observe* $x(t) = x^*(t) + g(t)$, *where* $\|g(t)\|_T^2 \leq c\|x^*(t)\|_T^2$ *for a sufficiently small constant* c *and choose* $\mathcal{N}^2 := \|g(t)\|_T^2 + \delta\|x^*(t)\|_T^2$. *Then Algorithm* MULTIPLESTAGES *returns a set* L *of* $O(k)$ *frequencies that approximates the heavy frequencies of* x^*. *In particular, with probability* $1 - 2^{-\Omega(k)}$, *for any* f^* *such that*

$$\int_{f^*-\Delta}^{f^*+\Delta} |\widehat{x \cdot H}(f)|^2 \mathrm{d}f \geq T\mathcal{N}^2/k, \tag{23}$$

there will exist an $\widetilde{f} \in L$ *satisfying* $|f^* - \widetilde{f}| \lesssim \sqrt{T\Delta}\Delta$.

Proof. Let $A \subset [-F, F]$ denote the set of frequencies f^* satisfying Equation (22). Let $A' \subset [-F, F]$ denote a net of A of distance 2Δ, so the intervals used in Equation (22) for each $f^* \in A'$ are disjoint. Then

$$|A'| \leq 2k + k = 3k$$

because each frequency in x^* contributes to at most two of the intervals, and the total mass of $\widehat{g}$ is at most k times the threshold $T\mathcal{N}^2$.

Let $L_1, \ldots, L_R$ be the results of R rounds of Algorithm ONESTAGE. We say that a frequency $f \in A'$ is *successfully recovered in round* r if there exists an $\widetilde{f} \in L_r$ such that $|f - \widetilde{f}| \leq \Delta_a$, where

$$\Delta_a = \Delta\sqrt{T\Delta} \lesssim \sqrt{T\Delta}\Delta.$$

By Lemma 7.22, each frequency is successfully recovered with 0.8 probability in each round. Then by the Chernoff bound, with $1 - 2^{-\Omega(k)}$ probability, every $f \in A'$ will be successfully recovered in at least $0.6R$ rounds.

Then, by Lemma 7.24, we output a set L of $O(B)$ frequencies such that every $f \in A'$ is within Δ_a of some $\widetilde{f} \in L$. Hence every $f \in A$ is within $2\Delta_a$ of some $\widetilde{f} \in L$. □

Lemma 7.24. *Let* $L_1, \ldots, L_R$ *by sets of frequencies and* f^* *be any frequency. Then* $L = $ MERGEDSTAGES$(L_1, \ldots, L_R)$ *is a set of* $2\frac{\sum |L_r|}{R}$ *frequencies satisfying*

$$\min_{\widetilde{f} \in L} |f^* - \widetilde{f}| \leq \operatorname*{median}_{r \in [R]} \min_{f \in L_r} |f^* - f|.$$

Proof. The algorithm is to take the union, sort, and take every $\frac{R}{2}$th entry of the sorted list.

Let $\Delta = \operatorname{median}_{r \in [R]} \min_{f \in L_r} |f^* - f|$. We have that at least $R/2$ different $f \in \bigcup_r L_r$ lie within Δ of f^*. This set forms a sequential subsequence of the sorted list of frequencies, so our output will include one. □

39

7.7 Time and sample complexity of frequency recovery of k-clustered signals

The goal of this section is to show that Procedure FREQUENCYRECOVERYKCLUSTER takes $\text{poly}(k, \log(1/\delta)) \log(FT)$ samples, and runs in $\text{poly}(k, \log(1/\delta)) \log^2(FT)$ time.

In order to analyze the running time and sample complexity. We need to extend the one-cluster version Procedure GETLEGAL1SAMPLE and GETEMPIRICAL1ENERGY (in Algorithm 3) to k-cluster version GETLEGALKSAMPLE and GETEMPIRICALKENERGY(in Algorithm 7)[5],

Lemma 7.25. *Procedure* GETLEGALKSAMPLE *in Algorithm 7 runs Procedure* HASHTOBINS $R_{\text{repeat}} = O((T\Delta)^3)$ *times to output two vectors* $\widehat{v}, \widehat{v}' \in \mathbb{C}^B$ *such that, for each* $j \in [B]$,

$$|\widehat{v}_j - \widehat{v}'_j e^{2\pi \mathbf{i} f_j \beta}| \leq 0.08(|\widehat{v}_j| + |\widehat{v}'_j|),$$

holds with probability at least 0.6.

Using the definition of z in Definition 7.18.

Claim 7.26. *Procedure* GETEMPIRICALKENERGY *in Algorithm 7 runs Procedure* HASHTOBINS $R_{\text{est}} O((T\Delta)^2)$ *times to output a vector* $z_{\text{emp}} \in \mathbb{R}^B$ *such that, for each* $j \in [B]$,

$$z_{\text{emp}}^j \in [0.8\|z^{(j)}\|_T, 1.2\|z^{(j)}\|_T],$$

holds with probability at least 0.9.

Claim 7.27. *Algorithm* LOCATEKSIGNAL *in Algorithm 6 uses* $O(\text{poly}(k, \log(1/\delta)) \cdot \log(FT))$, *and runs in* $O(\text{poly}(k, \log(1/\delta)) \cdot \log^2(FT))$.

Proof. We first calculate the number of samples. All the samples is basically all the Fourier samples, each time needs $B \log(k/\delta)$. In total it calls HASHTOBINS $O(R_{\text{est}} + R_{\text{repeat}} D_{\max} R_{\text{loc}})$ times where $D_{\max} R_{\text{loc}} = \Theta(\log(FT))$ by similar analysis as one-cluster frequency recovery. Thus, the total number of samples is

$$(R_{\text{est}} + R_{\text{repeat}} D_{\max} R_{\text{loc}}) B \log(k/\delta) = \text{poly}(k, \log(1/\delta)) \cdot \log(FT).$$

Then, we analyze the running time.

The expected running time includes the following parts: the first part is running Procedure HASHTOBINS $O(R_{\text{est}} + R_{\text{repeat}} D_{\max} R_{loc})$ times, each run takes $O(B \log(k/\delta) + B \log B)$ samples. For each such sample we need $\text{poly}(k, \log(1/\delta))$ time to compute $H(t)$ according to Lemma C.8 and there are $\text{poly}(k, \log(1/\delta)) \log(FT))$ many samples; the second part is updating the counter v, which takes $O(D_{\max} R_{loc} Bt)$ time. Thus, in total

$$\text{poly}(k, \log(1/\delta)) \cdot O(R_{\text{est}} + R_{\text{repeat}} D_{\max} R_{\text{loc}}) \cdot O(B \log(k/\delta) + B \log B) + O(D_{\max} R_{loc} Bt)$$
$$= \text{poly}(k, \log(1/\delta)) \cdot \log^2(FT),$$

where by similar analysis as one-cluster recovery, $t = \Theta(\log(FT))$ and $D_{\max} R_{\text{loc}} = \Theta(\log(FT))$. $\quad\square$

To boost the success probability, Procedure MULTIPLESTAGES reruns Procedure LOCATEKSIGNAL $O(k)$ times. At the end, Procedure FREQUENCYRECOVERYKCLUSTER combining Procedure MULTIPLESTAGES and MERGEDSTAGES directly, and the running time and sample complexity of MULTIPLESTAGES are dominating MERGEDSTAGES. Thus we have

Lemma 7.28. *Procedure* FREQUENCYRECOVERYKCLUSTER *in Algorithm 8 uses* $O(\text{poly}(k, \log(1/\delta)) \cdot \log(FT))$, *and runs in* $O(\text{poly}(k, \log(1/\delta)) \cdot \log^2(FT))$.

---

[5]We omitted the proofs here, because the proofs are identical to the one-cluster situation.

8 One-cluster Signal Recovery

8.1 Overview

In this section, we consider x^* whose frequencies in $\widehat{x}^*$ are in the range $[f_0 - \Delta', f_0 + \Delta']$ for some frequency f_0 and $\Delta' > 0$ and provide an algorithm to approximate it by a polynomial.

We fix T in this section and recall that $\langle f(t), g(t) \rangle_T := \frac{1}{T} \int_0^T f(t)\overline{g}(t)\mathrm{d}t$ such that $\|e^{2\pi \mathbf{i} f_i t}\|_T = \sqrt{\langle e^{2\pi \mathbf{i} f_i t}, e^{2\pi \mathbf{i} f_i t} \rangle_T} = 1$. For convenience, given $\sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$, we say the frequency gap of this signal is $\min_{i \neq j} |f_i - f_j|$.

For simplicity, we first consider frequencies clustered around 0. The main technical lemma in this section is that any signal x^* with bounded frequencies in $\widehat{x}^*$ can be approximated by a low-degree polynomial on $[0, T]$.

Lemma 2.3. *For any $\Delta > 0$ and any $\delta > 0$, let $x^*(t) = \sum_{j \in [k]} v_j e^{2\pi \mathbf{i} f_j t}$ where $|f_j| \leq \Delta$ for each $j \in [k]$. There exists a polynomial $P(t)$ of degree at most*

$$d = O(T\Delta + k^3 \log k + k \log 1/\delta)$$

such that

$$\|P(t) - x^*(t)\|_T^2 \leq \delta \|x^*\|_T^2.$$

One direct corollary is that when $\widehat{x}^*$ are in the range $[f_0 + \Delta', f_0 + \Delta']$, we can approximate x^* by $P(t) \cdot e^{2\pi \mathbf{i} f_0 t}$ for some low degree polynomial P.

We give an overview of this section first. We first show some technical tools in Section 8.2, 8.3. In Section 8.4, using those tools, we can show for any k-Fourier-sparse signal, there exists another k-Fourier-sparse signal with bounded frequency gap close to the original signal. In Section 8.5, we show that for any k-Fourier-sparse signal with bounded frequency gap, then there exists a low degree polynomial close to it. In Section 8.6, we show how to transfer low degree polynomial back to a Fourier-sparse signal. Combining all the above steps finishes the proof of Lemma 2.3.

We apply Theorem 7.5 of frequency estimation on x^* to obtain an estimation $\widetilde{f}_0$ of f_0 and use Theorem 4.5 on the approximation $Q(t)e^{2\pi \mathbf{i} \widetilde{f}_0 t}$ of x^* to recover the signal. We summarize this result as follows.

Theorem 8.1 (One-cluster Signal Recovery). *Let $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$ where $\forall j \in [k], |f_j - f_0| \leq \Delta$ and $x(t) = x^*(t) + g(t)$ be our observable signal. For any $\delta > 0$ and any $T > 0$, let $\mathcal{N}^2 := \|g\|_T^2 + \delta \|x^*\|_T^2$. Procedure $\mathrm{CFT1CULSTER}$ in Algorithm 5 finds a polynomial $P(t)$ of degree at most $d = O\left((T\Delta_h + T\Delta)^{1.5} + k^3 \log k + k \log 1/\delta\right)$ and a frequency $\widetilde{f}_0$ such that*

$$\|P(t) \cdot e^{2\pi \mathbf{i} \widetilde{f}_0 t} - x^*(t)\|_T^2 \lesssim \mathcal{N}^2 \tag{24}$$

The algorithm uses $O(kd) + \mathrm{poly}(k, \log(1/\delta)) \log(FT)$ samples, run in $O(kd^\omega) + \mathrm{poly}(k, \log(1/\delta)) \log^2(FT)$ time, and succeeds with probability at least $1 - 2^{-\Omega(k)}$.

Proof. We apply the algorithm in Theorem 7.5 to obtain an estimation $\widetilde{f}_0$ with $\mathrm{poly}(k) \log(FT)$ samples and $\mathrm{poly}(k) \log^2(FT)$ running time such that $|\widetilde{f}_0 - f_0| \lesssim (\Delta_h + \Delta)\sqrt{T(\Delta_h + \Delta)}$ holds with probability at least $1 - 2^{-\Omega(k)}$. Notice that $|f_j - \widetilde{f}_0| \leq |f_j - f_0| + |\widetilde{f}_0 - f_0| \lesssim (T(\Delta_h + \Delta))^{1.5}$.

We consider $x'(t) = e^{-2\pi \mathbf{i} \widetilde{f_0} t} x(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i}(f_j - \widetilde{f_0})t}$. By Lemma 2.3, there exists a polynomial $P(t)$ of degree at most

$$d = O\left((T\Delta_h + T\Delta)^{1.5} + k^3 \log k + k \log 1/\delta\right)$$

such that it approximates x' by

$$\|P(t) - x'(t)\|_T \leq \frac{\delta}{4}\|x'(t)\|_T = \frac{\delta}{4}\|x^*(t)\|_T.$$

which indicates $\|Q(t) - e^{-2\pi \mathbf{i} \widetilde{f_0} t} \cdot x^*(t)\|_T \leq \frac{\delta}{4}\|x^*(t)\|_T$.

Because we can sample $x(t)$, we can also sample $e^{-2\pi \mathbf{i} \widetilde{f_0} t} \cdot x(t) = Q(t) + g'(t)$ for $g'(t) = e^{-2\pi \mathbf{i} \widetilde{f_0} t} \cdot g(t) + (e^{-2\pi \mathbf{i} \widetilde{f_0} t} \cdot x^*(t) - Q(t))$. Hence we apply the algorithm in Theorem 4.5 and choose $R = O(k)$ in that proof. Then Procedure RobustPolynomialLearning$^+$ takes $O(kd)$ samples and $O(kd^\omega)$ time to find a degree d polynomial $P(t)$ approximating $Q(t)$ such that

$$\|P(t) - Q(t)\|_T \lesssim \|g'(t)\|_T,$$

holds with probability at least $1 - 2^{-\Omega(k)}$. It indicates

$$\|P(t) - e^{-2\pi \mathbf{i} \widetilde{f_0} t} \cdot x^*(t)\|_T \lesssim \|P(t) - Q(t)\|_T + \|Q(t) - x^*(t)\| \lesssim \delta\|x^*(t)\|_T + \|g(t)\|_T \approx \mathcal{N}.$$

Therefore we know $\|e^{2\pi \mathbf{i} \widetilde{f_0} t} \cdot P(t) - x^*(t)\|_T^2 \lesssim \mathcal{N}^2$. $\qquad\square$

8.2   Bounding the Gram matrix determinant

We define Gram matrix for $e^{2\pi \mathbf{i} f_1 t}, e^{2\pi \mathbf{i} f_2 t}, \cdots, e^{2\pi \mathbf{i} f_k t}$ and provide lower/upper bounds for its determinant.

Definition 8.2 (Gram matrix). *We define* $\mathrm{Gram}_{f_1, \cdots, f_k}$ *to be*

$$\begin{bmatrix} \langle e^{2\pi \mathbf{i} f_1 t}, e^{2\pi \mathbf{i} f_1 t} \rangle_T & \langle e^{2\pi \mathbf{i} f_1 t}, e^{2\pi \mathbf{i} f_2 t} \rangle_T & \cdots & \langle e^{2\pi \mathbf{i} f_1 t}, e^{2\pi \mathbf{i} f_k t} \rangle_T \\ \langle e^{2\pi \mathbf{i} f_2 t}, e^{2\pi \mathbf{i} f_1 t} \rangle_T & \langle e^{2\pi \mathbf{i} f_2 t}, e^{2\pi \mathbf{i} f_2 t} \rangle_T & \cdots & \langle e^{2\pi \mathbf{i} f_2 t}, e^{2\pi \mathbf{i} f_k t} \rangle_T \\ \cdots & \cdots & \cdots & \cdots \\ \langle e^{2\pi \mathbf{i} f_k t}, e^{2\pi \mathbf{i} f_1 t} \rangle_T & \langle e^{2\pi \mathbf{i} f_k t}, e^{2\pi \mathbf{i} f_2 t} \rangle_T & \cdots & \langle e^{2\pi \mathbf{i} f_k t}, e^{2\pi \mathbf{i} f_k t} \rangle_T \end{bmatrix}$$

Note that the above matrix is a Hermitian matrix with complex entries, thus both its determinant and all eigenvalues are in $\mathbb{R}$.

We defer the proof of the following Theorem to Appendix A.1.

Theorem 8.3. *For real numbers $\xi_1, \ldots, \xi_k$, let $G_{\xi_1, \ldots, \xi_k}$ be the matrix whose (i, j)-entry is*

$$\int_{-1}^{1} e^{2\pi \mathbf{i}(\xi_i - \xi_j)t} dt.$$

Then

$$\det(G_{\xi_1, \ldots, \xi_k}) = 2^{\tilde{O}(k^2)} \prod_{i < j} \min(|\xi_i - \xi_j|^2, 1).$$

We use the following corollary in this section.

Corollary 8.4. *There exists a universal constant $\alpha > 0$ such that, for any $T > 0$ and real numbers $f_1, \cdots, f_k$, the $k \times k$ Gram matrix of $e^{2\pi i f_1 t}, e^{2\pi i f_2 t}, \cdots, e^{2\pi i f_k t}$ whose (i,j)-entry is*

$$\mathrm{Gram}_{f_1, \cdots, f_k}(i,j) = \langle e^{2\pi i f_i t}, e^{2\pi i f_j t} \rangle_T = \frac{1}{T} \int_0^T e^{2\pi i (f_i - f_j) t} \mathrm{d}t.$$

satisfies

$$k^{-\alpha k^2} \prod_{i<j} \min((|f_i - f_j|T)^2, 1) \leq \det\left(\mathrm{Gram}_{f_1, \cdots, f_k}\right) \leq k^{\alpha k^2} \prod_{i<j} \min((|f_i - f_j|T)^2, 1).$$

Based on Corollary 8.4, we show the coefficients of a k-Fourier-sparse signal can be upper bounded by the energy $\|x\|_T^2$.

Lemma 2.2. *There exists a universal constant $c > 0$ such that for any $x(t) = \sum_{j=1}^{k} v_j e^{2\pi i f_j t}$ with frequency gap $\eta = \min_{i \neq j} |f_i - f_j|$,*

$$\|x(t)\|_T^2 \geq k^{-ck^2} \min\left((\eta T)^{2k}, 1\right) \sum_{j=1}^{k} |v_j|^2.$$

Proof. Let $\vec{v_i}$ denote the vector $e^{2\pi i f_i t}$ and $V = \{\vec{v_1}, \cdots, \vec{v_k}\}$. Notice that $\|\vec{v_i}\|_T^2 = \langle \vec{v_i}, \vec{v_i} \rangle = 1$. For each $\vec{v_i}$, we define $\vec{v_i}^{\parallel}$ to be the projection of $\vec{v_i}$ into the linear subspace $\mathrm{span}\{V \setminus \vec{v_i}\} = \mathrm{span}\{\vec{v_1}, \cdots, \vec{v_{i-1}}, \vec{v_{i+1}}, \cdots, \vec{v_k}\}$ and $\vec{v_i}^{\perp} = \vec{v_i} - \vec{v_i}^{\parallel}$ which is orthogonal to $\mathrm{span}\{V \setminus \vec{v_i}\}$ by the definition.

Therefore from the orthogonality,

$$\|x(t)\|_T^2 \geq \max_{j \in [k]} \{|v_j|^2 \cdot \|\vec{v_j}^{\perp}\|_T^2\} \geq \frac{1}{k} \sum_{j=1}^{k} |v_j|^2 \cdot \|\vec{v_j}^{\perp}\|_T^2.$$

It is enough to estimate $\|\vec{v_j}^{\perp}\|_T^2$ from Claim 3.12:

$$\|\vec{v_j}^{\perp}\|_T^2 = \frac{\det(\mathrm{Gram}(V))}{\det(\mathrm{Gram}(V \setminus \vec{v_i}))} \geq k^{-2\alpha k^2} \prod_{j \neq i} \min\left((f_j - f_i)T, 1\right)^2 \geq k^{-2\alpha k^2} (\eta T)^{2k-2},$$

where we use Corollary 8.4 to lower bound it in the last step. □

8.3 Perturbing the frequencies does not change the subspace much

We show that for a k-Fourier-sparse signal with unboundedly close frequency gap, there always exists another k-Fourier-sparse signal with slightly separated gap.

Lemma 8.5 (Slightly Shifting one Frequency). *There is a universal constant $C_0 > 0$ such that for any $x(t) = \sum_{j=1}^{k} v_j e^{2\pi i f_j t}$ and any frequency f_{k+1}, there always exists*

$$x'(t) = \sum_{j=1}^{k-1} v_j' e^{2\pi i f_j t} + v_{k+1}' e^{2\pi i f_{k+1} t}$$

with k coefficients $v_1', v_2', \cdots, v_{k-1}', v_{k+1}'$ satisfying

$$\|x'(t) - x(t)\|_T \leq k^{C_0 k^2} \cdot (|f_k - f_{k+1}|T) \cdot \|x(t)\|_T$$

Proof. We abuse the notation $e^{2\pi \mathbf{i} f_j t}$ to denote a vector in the linear subspace. We plan to shift f_k to f_{k+1} and define

$$
\begin{aligned}
V &= \{e^{2\pi \mathbf{i} f_1 t}, \cdots, e^{2\pi \mathbf{i} f_{k-1} t}, e^{2\pi \mathbf{i} f_k t}\} \\
V' &= \{e^{2\pi \mathbf{i} f_1 t}, \cdots, e^{2\pi \mathbf{i} f_{k-1} t}, e^{2\pi \mathbf{i} f_{k+1} t}\} \\
U &= \{e^{2\pi \mathbf{i} f_1 t}, \cdots, e^{2\pi \mathbf{i} f_{k-1} t}\} \\
W &= \{e^{2\pi \mathbf{i} f_1 t}, \cdots, e^{2\pi \mathbf{i} f_{k-1} t}, e^{2\pi \mathbf{i} f_k t}, e^{2\pi \mathbf{i} f_{k+1} t}\}
\end{aligned}
$$

where $f_1, f_2, \cdots, f_k$ are original frequencies in x. The idea is to show that any vector in the linear subspace span$\{V\}$ is close to some vector in the linear subspace span$\{V'\}$.

For convenience, we use $\vec{u}^{\|}$ to denote the projection of vector $e^{2\pi \mathbf{i} f_k t}$ to the linear subspace span$\{U\} = \text{span}\{e^{2\pi \mathbf{i} f_1 t}, \cdots, e^{2\pi \mathbf{i} f_{k-1} t}\}$ and $\vec{w}^{\|}$ denote the projection of vector $e^{2\pi \mathbf{i} f_{k+1} t}$ to this linear subspace span$\{U\}$. Let $\vec{u}^{\perp} = e^{2\pi \mathbf{i} f_k t} - \vec{u}^{\|}$ and $\vec{w}^{\perp} = e^{2\pi \mathbf{i} f_{k+1} t} - \vec{w}^{\|}$ be their orthogonal part to span$\{U\}$.

From the definition $e^{2\pi \mathbf{i} f_k t} = \vec{u}^{\|} + \vec{u}^{\perp}$ and $\vec{u}^{\|} \in \text{span}\{U\} = \text{span}\{e^{2\pi \mathbf{i} f_1 t}, \cdots, e^{2\pi \mathbf{i} f_{k-1} t}\}$, we rewrite the linear combination

$$
x(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t} = \sum_{j=1}^{k-1} \alpha_j e^{2\pi \mathbf{i} f_j t} + v_k \cdot \vec{u}^{\perp}
$$

for some scalars $\alpha_1, \cdots, \alpha_{k-1}$.

We will substitute $\vec{u}^{\perp}$ by $\vec{w}^{\perp}$ in the above linear combination and find a set of new coefficients. Let $\vec{w}^{\perp} = \vec{w}_1 + \vec{w}_2$ where $\vec{w}_1 = \frac{\langle \vec{u}^{\perp}, \vec{w}^{\perp} \rangle}{\|\vec{u}^{\perp}\|_T^2} \vec{u}^{\perp}$ is the projection of $\vec{w}^{\perp}$ to $\vec{u}^{\perp}$. Therefore $\vec{w}_2$ is the orthogonal part of the vector $e^{2\pi \mathbf{i} f_{k+1} t}$ to span$\{V\} = \text{span}\{e^{2\pi \mathbf{i} f_1 t}, \cdots, e^{2\pi \mathbf{i} f_{k-1} t}, e^{2\pi \mathbf{i} f_k t}\}$. We use $\delta = \frac{\|\vec{w}_2\|_T}{\|\vec{w}^{\perp}\|_T}$ for convenience.

Notice that the $\min_{\beta \in \mathbb{C}} \frac{\|\vec{u}^{\perp} - \beta \cdot \vec{w}^{\perp}\|_T}{\|\vec{u}^{\perp}\|_T} = \delta$ and $\beta^* = \frac{\langle \vec{u}^{\perp}, \vec{w}^{\perp} \rangle}{\|\vec{w}^{\perp}\|_T^2}$ is the optimal choice. Therefore we set

$$
x'(t) = \sum_{j=1}^{k-1} \beta_j e^{2\pi \mathbf{i} f_j t} + v_k \cdot \beta^* \cdot \vec{w}^{\perp} \in \text{span}\{e^{2\pi \mathbf{i} f_1 t}, \cdots, e^{2\pi \mathbf{i} f_{k-1} t}, e^{2\pi \mathbf{i} f_{k+1} t}\}
$$

where the coefficients $\beta_1, \cdots, \beta_{k-1}$ guarantee that the projection of x' onto span$\{U\}$ is as same as the projection of x onto span$\{U\}$. From the choice of β^* and the definition of x',

$$
\|x(t) - x'(t)\|_T^2 = \delta^2 \cdot |v_k|^2 \cdot \|\vec{u}^{\perp}\|_T^2 \leq \delta^2 \cdot \|x(t)\|_T^2.
$$

Eventually, we show an upper bound for δ^2 from Claim 3.12.

$$
\begin{aligned}
\delta^2 &= \frac{\|\vec{w}_2\|_T^2}{\|\vec{w}^\perp\|_T^2} \\
&= \frac{\det(\mathrm{Gram}_W)}{\det(\mathrm{Gram}_V)} \Big/ \frac{\det(\mathrm{Gram}_{V'})}{\det(\mathrm{Gram}_U)} \text{ by Claim 3.12} \\
&= \frac{\det(\mathrm{Gram}_W)}{\det(\mathrm{Gram}_V)} \cdot \frac{\det(\mathrm{Gram}_U)}{\det(\mathrm{Gram}_{V'})} \text{ by Corollary 8.4} \\
&\leq k^{4\alpha k^2} \cdot \frac{\displaystyle\prod_{i=1}^{k+1}\prod_{\substack{j=1\\j\neq i}}^{k+1}\min(|f_i-f_j|T,1)}{\displaystyle\prod_{i=1}^{k}\prod_{\substack{j=1\\j\neq i}}^{k}\min(|f_i-f_j|T,1)} \cdot \frac{\displaystyle\prod_{i=1}^{k-1}\prod_{\substack{j=1\\j\neq i}}^{k-1}\min(|f_i-f_j|T,1)}{\displaystyle\prod_{i=1}^{k-1}\prod_{\substack{j=1\\j\neq i}}^{k-1}\min(|f_i-f_j|T,1)\cdot\prod_{i=1}^{k-1}\min(|f_i-f_{k+1}|^2T^2,1)} \\
&= k^{4\alpha k^2}|f_k-f_{k+1}|^2T^2
\end{aligned}
$$

$\square$

Lemma 8.6. *For any k frequencies $f_1 < f_2 < \cdots < f_k$, there exists k frequencies $f'_1,\cdots,f'_k$ such that $\min_{i\in[k-1]} f'_{i+1} - f'_i \geq \eta$ and for all $i \in [k]$, $|f'_i - f_i| \leq k\eta$.*

Proof. We define the new frequencies f'_i as follows: $f'_1 = f_1$ and $f'_i = \max\{f'_{i-1}+\eta, f_i\}$ for $i \in \{2,3,\cdots,k\}$. $\square$

8.4 Existence of nearby k-Fourier-sparse signal with frequency gap bounded away from zero

We combine the results in the above section to finish the proof of Lemma 2.3. We first prove that for any $x^*(t) = \sum_{j=1}^k v_j e^{2\pi \mathbf{i} f_j t}$, there always exists another k-Fourier-sparse signal x' close to $x^*(t) = \sum_{j=1}^k v_j e^{2\pi \mathbf{i} f_j t}$ such that the frequency gap in x' is at least $\eta \geq 2^{-\mathrm{poly}(k)}$. Then we show how to find a low degree polynomial $P(t)$ approximating $x'(t)$.

Lemma 2.1. *There is a universal constant $C_1 > 0$ such that, for any $x^*(t) = \sum_{j=1}^k v_j e^{2\pi \mathbf{i} f_j t}$ and any $\delta > 0$, there always exist $\eta \geq \frac{\delta}{T} \cdot k^{-C_1 k^2}$ and $x'(t) = \sum_{j=1}^k v'_j e^{2\pi \mathbf{i} f'_j t}$ satisfying*

$$\|x'(t) - x^*(t)\|_T \leq \delta\|x^*(t)\|_T$$

with $\min_{i\neq j}|f'_i - f'_j| \geq \eta$ and $\max_{j\in[k]}\{|f'_j - f_j|\} \leq k\eta$.

Proof. Using Lemma 8.6 on frequencies $f_1,\cdots,f_k$, we obtain k new frequencies $f'_1,\cdots,f'_k$ such that their gap is at least η and $\max_i |f_i - f'_i| \leq k\eta$. Next we use the hybrid argument to find x'.

Let $x^{(0)}(t) = x^*(t)$. For $i = 1, \cdots, t$, we apply Lemma 8.5 to shift f_i to f_i' and obtain

$$x^{(i)}(t) = \sum_{j=i+1}^{k} v_j^{(i)} e^{2\pi \mathbf{i} f_j t} + \sum_{j=1}^{i} v_j^{(i)} e^{2\pi \mathbf{i} f_j' t}.$$

From Lemma 8.5, we know $\|x^{(i)}(t) - x^{(i-1)}(t)\|_T \leq k^{C_0 k^2}(|f_i - f_i'|T)\|x^{(i-1)}\|_T$. Thus we obtain

$$\left(1 - k^{C_0 k^2}(k\eta T)\right)^i \|x^{(0)}(t)\|_T \leq \|x^{(i)}(t)\|_T \leq \left(1 + k^{C_0 k^2}(k\eta T)\right)^i \|x^{(0)}(t)\|_T,$$

which is between $\left[\left(1 - i \cdot k^{C_0 k^2}(k\eta T)\right)\|x^{(0)}(t)\|_T, \left(1 + 2i \cdot k^{C_0 k^2}(k\eta T)\right)\|x^{(0)}(t)\|_T\right]$ for $\eta \leq \frac{1}{5T} \cdot k^{-C_1 k^2}$ with some $C_1 > C_0$.

At last, we set $x'(t) = x^{(k)}(t)$ and bound the distance between $x'(t)$ and $x^*(t)$ by

$$
\begin{aligned}
\|x^{(k)}(t) - x^{(0)}(t)\|_T &\leq \sum_{i=1}^{k} \|x^{(i)}(t) - x^{(i-1)}(t)\|_T && \text{by triangle inequality} \\
&\leq \sum_{i=1}^{k} k^{C_0 k^2}(|f_i - f_i'|T)\|x^{(i-1)}(t)\|_T && \text{by Lemma 8.5} \\
&\leq \sum_{i=1}^{k} 2k^{C_0 k^2}(k\eta T)\|x^{(i-1)}(t)\|_T && \text{by } \max_i |f_i - f_i'| \leq k\eta \\
&\leq k \cdot 2k^{C_0 k^2}(k\eta T)\|x^*(t)\|_T \\
&\leq \delta \|x^*(t)\|_T
\end{aligned}
$$

where the last inequality follows by the sufficiently small η. $\qquad\square$

8.5 Approximating k-Fourier-sparse signals by polynomials

For any k-Fourier-sparse signal with frequency gap bounded away from zero, we show that there exists a low degree polynomial which is close to the original k-Fourier-sparse signal in $\|\cdot\|_T$ distance.

Lemma 8.7 (Existence of low degree polynomial). *Let* $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$*, where* $\forall j \in [k], |f_j| \leq \Delta$ *and* $\min_{i \neq j} |f_i - f_j| \geq \eta$*. There exists a polynomial* $Q(t)$ *of degree*

$$d = O\left(T\Delta + k \log 1/(\eta T) + k^2 \log k + k \log(1/\delta)\right)$$

such that,

$$\|Q(t) - x^*(t)\|_T^2 \leq \delta \|x^*(t)\|_T^2 \tag{25}$$

Proof. For each frequency f_j, let $Q_j(t) = \sum_{k=0}^{d-1} \frac{(2\pi \mathbf{i} f_j t)^k}{k!}$ be the first d terms in the Taylor Expansion of $e^{2\pi \mathbf{i} f_j t}$. For any $t \in [0, T]$, we know the difference between $Q_j(t)$ and $e^{2\pi \mathbf{i} f_j t}$ is at most

$$|Q_j(t) - e^{2\pi \mathbf{i} f_j t}| \leq |\frac{(2\pi \mathbf{i} f_j T)^d}{d!}| \leq (\frac{2\pi T\Delta \cdot e}{d})^d.$$

We define

$$Q(t) = \sum_{j=1}^{k} v_j Q_j(t)$$

and bound the distance between Q and x^* from the above estimation:

$$
\begin{aligned}
\|Q(t) - x^*(t)\|_T^2 &= \frac{1}{T} \int_0^T |Q(t) - x^*(t)|^2 \mathrm{d}t \\
&= \frac{1}{T} \int_0^T |\sum_{j=1}^{k} v_j (Q_j(t) - e^{2\pi \mathbf{i} f_j t})|^2 \mathrm{d}t \\
&\leq 2k \sum_{j=1}^{k} \frac{1}{T} \int_0^T |v_j|^2 \cdot |Q_j(t) - e^{2\pi \mathbf{i} f_j t}|^2 \mathrm{d}t \qquad \text{by triangle inequality} \\
&\leq k \sum_{j=1}^{k} |v_j|^2 \cdot (\frac{2\pi T \Delta \cdot e}{d})^{2d} \qquad\qquad\qquad \text{by Taylor expansion}
\end{aligned}
$$

On the other hand, from Lemma 2.2, we know

$$\|x^*(t)\|_T^2 \geq (\eta T)^{2k} \cdot k^{-ck^2} \sum_j |v_j|^2.$$

Because $d = 10 \cdot \pi e (T\Delta + k \log 1/(\eta T) + k^2 \log k + k \log(1/\delta))$ is large enough, we have $k(\frac{2\pi T \Delta \cdot e}{d})^{2d} \leq \delta(\eta T)^{2k} \cdot k^{-ck^2}$, which indicates that $\|Q(t) - x^*(t)\|_T^2 \leq \delta \|x^*\|_T^2$ from all discussion above. $\qquad\square$

8.6 Transferring degree-d polynomial to $(d+1)$-Fourier-sparse signal

In this section, we show how to transfer a degree-d polynomial to $(d+1)$-Fourier-sparse signal.

Lemma 8.8. *For any degree-d polynomial $Q(t) = \sum_{j=0}^{d} c_j t^j$, any $T > 0$ and any $\epsilon > 0$, there always exist $\gamma > 0$ and*

$$x^*(t) = \sum_{i=1}^{d+1} \alpha_i e^{2\pi \mathbf{i}(\gamma i)t}$$

with some coefficients $\alpha_0, \cdots, \alpha_d$ such that

$$\forall t \in [0, T], |x^*(t) - Q(t)| \leq \epsilon.$$

Proof. We can rewrite $x^*(t)$,

$$
\begin{aligned}
x^*(t) &= \sum_{i=1}^{d+1} \alpha_i e^{2\pi \mathbf{i}\gamma it} \\
&= \sum_{i=1}^{d+1} \alpha_i \sum_{j=0}^{\infty} \frac{(2\pi \mathbf{i}\gamma it)^j}{j!} \\
&= \sum_{j=0}^{\infty} \frac{(2\pi \mathbf{i}\gamma t)^j}{j!} \sum_{i=1}^{d+1} \alpha_i \cdot i^j \\
&= \sum_{j=0}^{d} \frac{(2\pi \mathbf{i}\gamma t)^j}{j!} \sum_{i=1}^{d+1} \alpha_i \cdot i^j + \sum_{j=d+1}^{\infty} \frac{(2\pi \mathbf{i}\gamma t)^j}{j!} \sum_{i=1}^{d+1} \alpha_i \cdot i^j \\
&= Q(t) + \underbrace{\left(\sum_{j=0}^{d} \frac{(2\pi \mathbf{i}\gamma t)^j}{j!} \sum_{i=1}^{d+1} \alpha_i \cdot i^j - Q(t) \right)}_{C_1} + \underbrace{\left(\sum_{j=d+1}^{\infty} \frac{(2\pi \mathbf{i}\gamma t)^j}{j!} \sum_{i=1}^{d+1} \alpha_i \cdot i^j \right)}_{C_2}.
\end{aligned}
$$

Our goal is to show there exists some parameter γ and coefficients $\{\alpha_0, \alpha_1, \cdots, \alpha_d\}$ such that the term $C_1 = 0$ and $|C_2| \leq \epsilon$. Let's consider C_1,

$$
C_1 = \sum_{j=0}^{d} (\frac{t}{T})^j \left(\frac{(2\pi \mathbf{i}\gamma T)^j}{j!} \sum_{i=1}^{d+1} \alpha_i i^j - c_j \right)
$$

To guarantee $C_1 = 0$, we need to solve a linear system with $d+1$ unknown variables and $d+1$ constraints,

$$
\begin{aligned}
\text{Find} \quad & \alpha_1, \alpha_2, \cdots \alpha_{d+1} \\
\text{s.t.} \quad & \frac{(2\pi \mathbf{i}\gamma T)^j}{j!} \sum_{i=1}^{d+1} \alpha_i i^j - c_j = 0, \forall j \in \{0, 1, \cdots, d\}
\end{aligned}
$$

Define $c'_j = c_j j!/(2\pi \mathbf{i}\gamma)^j$, let α and c' be the length-$(d+1)$ column vectors with α_i and c'_j. Let $A \in \mathbb{R}^{d+1 \times d+1}$ denote the Vandermonde matrix where $A_{i,j} = i^j, \forall i, j \in [d+1] \times \{0, 1, \cdots, d\}$. Then we need to guarantee $A\alpha = c'$. Using the definition of determinant, $\det(A) = \prod_{i<j} |i - j| \leq 2^{O(d^2 \log d)}$.

Thus $\sigma_{\max}(A) \leq 2^{O(d^2 \log d)}$ and then

$$
\sigma_{\min}(A) = \frac{\det(A)}{\prod_{i=1}^{d-1} \sigma_i} \geq 2^{-O(d^3 \log d)}.
$$

We show how to upper bound $|\alpha_i|$,

$$
\max_{i \in [d+1]} |\alpha_i| \leq \|\alpha\|_2 = \|A^\dagger c'\|_2 \leq \|A^\dagger\|_2 \cdot \|c'\|_2 \leq \frac{1}{\sigma_{\min}(A)} \sqrt{d+1} \max_{0 \leq j \leq d} \frac{|c_j| j!}{(2\pi \gamma T)^j}
$$

Plugging the above equation into C_2, we have

$$
\begin{aligned}
|C_2| &= \left| \sum_{j=d+1}^{\infty} \frac{(2\pi\mathbf{i}\gamma t)^j}{j!} \sum_{i=1}^{d+1} \alpha_i \cdot i^j \right| \\
&\leq \sum_{j=d+1}^{\infty} \frac{(2\pi\gamma t)^j}{j!} \sum_{i=1}^{d+1} |\alpha_i| \cdot i^j \\
&\leq \sum_{j=d+1}^{\infty} \frac{(2\pi\gamma t)^j}{j!} (d+1)^{d+1} \max_{i\in[d+1]} |\alpha_i| \\
&\leq \sum_{j=d+1}^{\infty} \frac{(2\pi\gamma t)^j}{j!} (d+1)^{d+2} \frac{1}{\sigma_{\min}(A)} \frac{d!}{(2\pi\gamma T)^d} \max_{0\leq j\leq d} |c_j| \\
&\leq \epsilon
\end{aligned}
$$

where the last step follows by choosing sufficiently small

$$
\gamma \lesssim \epsilon / \left(T 2^{\Theta(d^3 \log d)} \max_{0\leq j\leq d} |c_j| \right).
$$

$\square$

9 k-cluster Signal Recovery

9.1 Overview

In this section, we prove Lemma 9.1 as the main technical lemma to finish the proof of main Theorem 1.1, which shows how to learn $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi\mathbf{i}f_j t}$ with noise.

Lemma 9.1. *Let $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi\mathbf{i}f_j t}$ and $x(t) = x^*(t) + g(t)$ be our observation. For any $\delta > 0$ and $T > 0$, let $\mathcal{N}^2 := \frac{1}{T}\int_0^T |g(t)|^2 \mathrm{d}t + \delta \cdot \frac{1}{T}\int_0^T |x^*(t)|^2 \mathrm{d}t$. For $\Delta = \mathrm{poly}(k, \log(1/\delta))/T$, Procedure* SIGNALRECOVERYKCLUSTER$^+$ *in Algorithm 8 takes $l = O(k)$ frequencies $\widetilde{f}_1, \cdots, \widetilde{f}_l$ as input and finds l polynomials $Q_1, \cdots, Q_l$ of degree $d = O((T\Delta)^{1.5} + k^3 \log k + k \log 1/\delta)$ such that*

$$
\widetilde{x}(t) = \sum_{j\in[l]} Q_j(t) e^{2\pi\mathbf{i}\widetilde{f}_j t} \text{ satisfies } \|\widetilde{x}(t) - x^*(t)\|_T^2 \lesssim \mathcal{N}^2. \tag{26}
$$

The procedure succeeds with probability at least $1 - 2^{-\Omega(k)}$, uses $\mathrm{poly}(k, \log(1/\delta)) \cdot \log(FT)$ samples, and runs in $\mathrm{poly}(k, \log(1/\delta)) \cdot \log^2(FT)$ time.

For any set $W = \{t_1, \cdots, t_m\}$ where each $t_i \in [0, T]$, we use

$$
\|\vec{v}\|_W = \sqrt{\frac{\sum_{i\in W} |\vec{v}(t_i)|^2}{|W|}} \text{ for any } \vec{v} : [0, T] \to \mathbb{C}
$$

in this section. We first show that Procedure SIGNALRECOVERYKCLUSTER succeeds with constant probability, then prove that Procedure SIGNALRECOVERYKCLUSTER$^+$ succeeds with probability at least $1 - 2^{-\Omega(k)}$.

9.2 Heavy clusters separation

Recall the definition of "heavy" clusters.

Definition 2.4. *Given $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$, any $\mathcal{N} > 0$, and a filter function $(H, \widehat{H})$ with bounded support in frequency domain. Let L_j denote the interval of $\mathrm{supp}(\widehat{e^{2\pi \mathbf{i} f_j t} \cdot H})$ for each $j \in [k]$.*

Define an equivalence relation $\sim$ on the frequencies f_i by the transitive closure of the relation $f_i \sim f_j$ if $L_i \cap L_j \neq \emptyset$. Let $S_1, \ldots, S_n$ be the equivalence classes under this relation.

Define $C_i = \bigcup_{f \in S_i} L_i$ for each $i \in [n]$. We say C_i is a "heavy" cluster iff $\int_{C_i} |\widehat{H \cdot x^*}(f)|^2 \mathrm{d}f \geq T \cdot \mathcal{N}^2 / k$.

By reordering C_i, we can assume $\{C_1, C_2, \cdots, C_l\}$ are heavy clusters, where $l \leq n \leq k$.

Claim 2.5. *Given $x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$ and any $\mathcal{N} > 0$, let H be the filter function defined in Appendix C.1 and $C_1, \cdots, C_l$ be the heavy clusters from Definition 2.4. For*

$$S = \left\{ j \in [k] \,\middle|\, f_j \in C_1 \cup \cdots C_l \right\},$$

we have $x^{(S)}(t) = \sum_{j \in S} v_j e^{2\pi \mathbf{i} f_j t}$ approximating x^* within distance $\|x^{(S)}(t) - x^*(t)\|_T^2 \lesssim \mathcal{N}^2$.

Proof. Let $x^{(\overline{S})}(t) = \sum_{j \in [k] \backslash S} v_j e^{2\pi \mathbf{i} f_j t}$. Notice that $\|x^* - x^{(S)}\|_T^2 = \|x^{(\overline{S})}\|_T^2$.

From the property VI of filter function $(H, \widehat{H})$ in Appendix C.1, we have

$$\int_{-\infty}^{+\infty} |x^{(\overline{S})}(t) \cdot H(t)|^2 \mathrm{d}t \geq 0.9 \int_0^T |x^{(\overline{S})}(t)|^2 \mathrm{d}t = 0.9 \cdot T \|x^{\overline{S}}\|_T^2.$$

From Definition 2.4, we have

$$
\begin{aligned}
\int_{-\infty}^{+\infty} |x^{(\overline{S})}(t) \cdot H(t)|^2 \mathrm{d}t &= \int_{-\infty}^{+\infty} |\widehat{x^{(\overline{S})} \cdot H}(f)|^2 \mathrm{d}f \\
&= \int_{[-\infty, +\infty] \backslash C_1 \cup \cdots \cup C_l} |\widehat{x^* \cdot H}(f)|^2 \mathrm{d}f \\
&\leq k \cdot T \mathcal{N}^2 / k.
\end{aligned}
$$

Overall, we have $\|x^{(\overline{S})}\|_T^2 \lesssim \mathcal{N}^2$. $\qquad\square$

From the guarantee of Theorem 2.6, for any $j \in S$, $\min_{i \in [l]} |f_j - \widetilde{f}_i| \leq \Delta \sqrt{\Delta T}$. From now on, we focus on the recovery of $x^{(S)}$, which is enough to approximate x^* from the above claim. Because we are looking for $\tilde{x}$ approximating $x^{(S)}$ within distance $O(\mathcal{N}^2)$, from Lemma 2.1, we can assume there is a frequency gap $\eta \geq \frac{\delta}{10T} k^{-O(k^2)}$ among $x^{(S)}$.

9.3   Approximating clusters by polynomials

In this section, we show how to approximate $x^{(S)}$ by $x'(t) = \sum_{i \in [l]} e^{2\pi \mathbf{i} \tilde{f}_i t} P_i(t)$ where $P_1, \cdots, P_l$ are low degree polynomials.

Claim 9.2. *For any $x^{(S)}(t) = \sum_{j \in S} v_j e^{2\pi \mathbf{i} f_j t}$ with a frequency gap $\eta = \min\limits_{i \neq j} |f_i - f_j|$ and l frequencies $\tilde{f}_1, \cdots, \tilde{f}_l$ with the property $\forall j \in S, \min_{i \in [l]} |f_j - \tilde{f}_i| \leq \Delta\sqrt{\Delta T}$, let*

$$d = 5\pi \left((T\Delta)^{1.5} + k^3 \log k + \log 1/\delta\right) \ \text{and} \ V = \left\{ t^j e^{2\pi \mathbf{i} \tilde{f}_i t} | i \in [l], j \in \{0, \cdots, d\} \right\}.$$

There exists $x'(t) \in \mathrm{span}\{V\}$ that approximates $x^{(S)}(t)$ as follows:

$$\forall t \in [0, T], |x'(t) - x^{(S)}(t)| \leq \delta \|x^{(S)}\|_T.$$

Proof. From Lemma 2.2, we know

$$\|x^{(S)}\|_T^2 \geq (\eta T)^{2k} \cdot k^{-ck^2} \sum_{j \in S} |v_j|^2.$$

For each frequency f_j, we use p_j to denote the index in $[l]$ such that $|f_j - \tilde{f}_{p_j}| \leq \Delta\sqrt{\Delta T}$. We rewrite

$$x^{(S)}(t) = \sum_{i=1}^l e^{2\pi \mathbf{i} \tilde{f}_i} \left(\sum_{j \in S : p_j = i} v_j e^{2\pi \mathbf{i} (f_j - \tilde{f}_i)t} \right).$$

For $d = 5\pi((T\Delta)^{1.5} + k^3 \log k + \log 1/\delta)$ and each $e^{2\pi \mathbf{i}(f_j - \tilde{f}_{p_j})t}$, let $Q_j(t) = \sum_{i=0}^{d-1} \frac{\left(2\pi \mathbf{i}(f_j - \tilde{f}_{p_j})t\right)^i}{i!}$ be the first d terms in the Taylor Expansion of $e^{2\pi \mathbf{i}(f_j - \tilde{f}_{p_j})t}$. For any $t \in [0, T]$, we know the difference between $Q_j(t)$ and $e^{2\pi \mathbf{i}(f_j - \tilde{f}_{p_j})t}$ is at most

$$\forall t \in [0, T], |Q_j(t) - e^{2\pi \mathbf{i}(f_j - \tilde{f}_{p_j})t}| \leq |\frac{(2\pi \mathbf{i}(f_j - \tilde{f}_{p_j})T)^d}{d!}| \leq (\frac{8\pi(\Delta T)^{1.5}}{d})^d.$$

Let $x' = \sum_{i=1}^l e^{2\pi \mathbf{i} \tilde{f}_i t} \left(\sum_{j \in S : p_j = i} v_j Q_j(t) \right)$. From all discussion above, we know for any $t \in [0, T]$,

$$\begin{aligned}
|x'(t) - x^{(S)}(t)|^2 &\leq \left(\sum_{j \in S} |v_j| (\frac{8\pi(T\Delta)^{1.5}}{d})^d \right)^2 \\
&\leq k (\frac{8\pi(T\Delta)^{1.5}}{d})^{2d} \sum_j |v_j|^2 \\
&\leq \frac{k(\frac{8\pi(T\Delta)^{1.5}}{d})^{2d}}{(\eta T)^{2k} \cdot k^{-ck^2}} \|x^{(S)}\|_T^2 \\
&\leq \delta^2 \|x^{(S)}\|_T^2.
\end{aligned}$$

$\square$

We provide a property of functions in $\mathrm{span}\{V\}$ such that we can use the Chernoff bound and the ϵ-net argument on vectors in $\mathrm{span}\{V\}$.

Claim 2.7. *For any $\vec{u} \in \text{span}\left\{e^{2\pi \mathbf{i} \widetilde{f}_i t} \cdot t^j \middle| j \in \{0, \cdots, d\}, i \in [l]\right\}$, there exists some universal constants $C_1 \leq 4$ and $C_2 \leq 3$ such that*

$$\max_{t \in [0,T]} \left\{|\vec{u}(t)|^2\right\} \lesssim (ld)^{C_1} \log^{C_2}(ld) \cdot \|\vec{u}\|_T^2$$

Proof. From Lemma 8.8, we can approximate each polynomial in $\vec{u}$ by a linear combination of $\{1, e^{2\pi \mathbf{i} \cdot \gamma t}, \cdots, e^{2\pi \mathbf{i} \cdot (\gamma d) t}\}$ such that we obtain $u^* \in \text{span}\left\{e^{2\pi \mathbf{i} \cdot (\gamma j) t} \cdot e^{2\pi \mathbf{i} \widetilde{f}_i t} \middle| i \in [l], j \in \{0, \cdots, d+1\}\right\}$ for some small γ such that $\forall t \in [0,T], |\vec{u}(t) - u^*(t)| \leq 0.01\|\vec{u}\|_T$.

From Lemma 5.1, we know

$$\max_{t \in [0,T]} |u^*(t)|^2 \leq C \cdot \left((ld+1)^4 \cdot \log^3(ld+1)\right) \|u^*\|_T^2.$$

For some constant C', we have

$$\max_{t \in [0,T]} |\vec{u}(t)|^2 \leq C' \left((kd)^{C_1} \log^{C_2} d\right) \|\vec{u}\|_T^2.$$

$\square$

9.4 Main result, with constant success probability

In this section, we show that the output $\widetilde{x}$ is close to x' with high probability using the ϵ-net argument, which is enough to prove $\|\widetilde{x} - x\|_T \lesssim \mathcal{N}^2$ from all discussion above. Because we can prove Lemma 9.6(which is the main goal of this section), then combining $\|x' - x^*\|_T \leq \|x' - x^{(S)}\|_T + \|x^{(S)} - x^*\|_T \lesssim \delta\|x^*\|_T$ and Lemma 9.6, we have $\|x^* - \widetilde{x}\|_T \lesssim \|g\|_T + \delta\|x^*\|_T$, which finishes the proof of Procedure SIGNALRECOVERYKCLUSTER in Algorithm 8 achieving the Equation (26) with constant success probability but not $1 - 2^{-\Omega(k)}$. We will boost the success probability in Section 9.5.

We first provide an ϵ-net $\mathcal{P}$ for the unit vectors $\mathcal{Q} = \{\vec{u} \in \text{span}\{V\} \big| \|\vec{u}\|_T^2 = 1\}$ in the linear subspace $\text{span}\{V\}$ where $V = \left\{t^j \cdot e^{2\pi \mathbf{i} \widetilde{f}_i t} \middle| j \in \{0, 1, \cdots, d\}, i \in [l]\right\}$ from the above discussion. Notice that the dimension of $\text{span}\{V\}$ is at most $l(d+1)$.

Claim 9.3. *There exists an ϵ-net $\mathcal{P} \subset \text{span}\{V\}$ such that*

1. $\forall \vec{u} \in Q, \exists \vec{w} \in \mathcal{P}, \|\vec{u} - \vec{w}\|_T \leq \epsilon$.

2. $|\mathcal{P}| \leq \left(5\frac{l(d+1)}{\epsilon}\right)^{2l(d+1)}$.

Proof. Let $\mathcal{P}'$ be an $\frac{\epsilon}{l(d+1)}$-net in the unit circle of $\mathbb{C}$ with size at most $(4\frac{l(d+1)}{\epsilon} + 1)^2$, i.e.,

$$\mathcal{P}' = \left\{\frac{\epsilon}{2l(d+1)}j_1 + \mathbf{i}\frac{\epsilon}{2l(d+1)}j_2 \middle| j_1, j_2 \in \mathbb{Z}, |j_1| \leq \frac{2l(d+1)}{\epsilon}, |j_2| \leq \frac{2l(d+1)}{\epsilon}\right\}.$$

Observe that the dimension of $\text{span}\{V\}$ is at most $l(d+1)$. Then we take an orthogonal basis $\vec{w}_1, \cdots, \vec{w}_{l(d+1)}$ in $\text{span}\{V\}$ and set

$$P = \left\{\sum_{i=1}^{l(d+1)} \alpha_i \vec{w}_i \middle| \forall i \in [l(d+1)], \alpha_i \in \mathcal{P}'\right\}.$$

Therefore $\mathcal{P}$ is an ϵ-net for Q and $|\mathcal{P}| \leq \left(5\frac{l(d+1)}{\epsilon}\right)^{2l(d+1)}$. $\square$

We first prove that W is a good estimation for all functions in the ϵ-net $\mathcal{P}$.

Claim 9.4. *For any $\epsilon > 0$, there exists a universal constant $C_3 \leq 5$ such that for a set S of i.i.d. samples chosen uniformly at random over $[0, T]$ of size $|S| \geq \frac{3(kd)^{C_3} \log^{C_3} d/\epsilon}{\epsilon^2}$, then with probability at least $1 - k^{-k}$, for all $\vec{w} \in \mathcal{P}$, we have*

$$\|\vec{w}\|_W \in \left[(1 - \epsilon)\|\vec{w}\|_T, (1 + \epsilon)\|\vec{w}\|_T\right].$$

Proof. From Claim 2.7 and Lemma 3.5, for each $\vec{w} \in \mathcal{P}$,

$$\Pr\left[\|\vec{w}(t)\|_W \notin \left[(1 - \epsilon)\|\vec{w}\|_T, (1 + \epsilon)\|\vec{w}\|_T\right]\right] \leq 2^{-\frac{|W|\epsilon^2}{3(kd)^{C_1} \log^{C_2 + 0.5} d}} \leq 2^{-kd \log^{1.5} \frac{d}{\epsilon}}.$$

From the union bound, $\|\vec{w}\|_W \in \left[(1 - \epsilon)\|\vec{w}\|_T, (1 + \epsilon)\|\vec{w}\|_T\right]$ for any $\vec{w} \in \mathcal{P}$ with probability at least $1 - (\frac{d}{\epsilon})^{-kd \log^{0.5} d} \cdot |\mathcal{P}| \geq 1 - d^{-d}$. $\qquad\square$

Then We prove that W is a good estimation for all functions in $\text{span}\{V\}$ using the property of ϵ-nets.

Claim 9.5. *For any $\epsilon > 0$, there exists a universal constant $C_3 \leq 5$ such that for a set W of i.i.d. samples chosen uniformly at random over $[0, T]$ of size $|W| \geq \frac{3(kd)^{C_3} \log^{C_3} d/\epsilon}{\epsilon^2}$, then with probability at least $1 - d^{-d}$, for all $u \in \text{span}\{V\}$, we have*

$$\|\vec{u}\|_W \in \left[(1 - 3\epsilon)\|\vec{u}\|_T, (1 + 3\epsilon)\|\vec{u}\|_T\right]$$

Proof. We assume that the above claim is true for any $\vec{w} \in \mathcal{P}$. Without loss of generality, we consider $\vec{u} \in \mathcal{Q}$ such that $\|\vec{u}\|_T = 1$.

Let $\vec{w}_0$ be the vector in $\mathcal{P}$ that minimizes $\|\vec{w} - \vec{u}\|_T$ for all $\vec{w} \in \mathcal{P}$, i.e., $\vec{w}_0 = \arg\min_{\vec{w} \in \mathcal{P}} \|\vec{w} - \vec{u}\|_T$. Define $\vec{u}_1 = \vec{u} - \vec{w}_0$ and notice that $\|\vec{u}_1\|_T \leq \epsilon$ because $\mathcal{P}$ is a ϵ-net. If $\|\vec{u}_1\|_T = 0$, then we skip the rest of this procedure. Otherwise, we define $\alpha_1 = \|\vec{u}_1\|_T$ and normalize $\widetilde{u}_1 = \vec{u}_1/\alpha_1$.

Then we choose $\vec{w}_1$ to be the vector in $\mathcal{P}$ that minimizes $\|\vec{w} - \widetilde{u}_1\|_T$ for all $\vec{w} \in \mathcal{P}$. Similarly, we set $\vec{u}_2 = \widetilde{u}_1 - \vec{w}_1$ and $\alpha_2 = \|\vec{u}_2\|_T$. Next we repeat this process for $\widetilde{u}_2 = \vec{u}_2/\alpha_2$ and so on. The recursive definition can be summarized in the following sense,

$$\begin{aligned}
\text{initial}: \quad & \widetilde{u}_0 = \vec{u} \text{ and } m = 10 \log_{1/\epsilon}(ld) + 1, \\
\text{For } i \in \{0, 1, 2, \cdots, m\}: \quad & \vec{w}_i = \arg\min_{\vec{w} \in \mathcal{P}} \|\vec{w} - \widetilde{u}_i\|_T, \\
& \vec{u}_{i+1} = \widetilde{u}_i - \vec{w}_i \text{ and } \alpha_{i+1} = \|\vec{u}_{i+1}\|_T, \\
& \text{if } \alpha_{i+1} = 0, \text{ stop.} \\
& \text{if } \alpha_{i+1} \neq 0, \widetilde{u}_{i+1} = \vec{u}_{i+1}/\alpha_{i+1} \text{ and continue,}
\end{aligned}$$

Eventually, we have $\vec{u} = \vec{w}_0 + \alpha_1 \vec{w}_1 + \alpha_1 \alpha_2 \vec{w}_2 + \cdots + \prod_{j=1}^{m} \alpha_j (\vec{w}_m + \vec{u}_{m+1})$ where each $|\alpha_i| \leq \epsilon$ and each $\vec{w}_i$ is in the ϵ-net $\mathcal{P}$. Notice that $\|\vec{u}_{m+1}\|_T \leq 1$ and $\|\vec{u}_{m+1}\|_W \leq (ld + 1)^3 \cdot \|\vec{u}_{m+1}\|_T$ from Claim 2.7. We prove a lower bound for $\|\vec{u}\|_W$,

$$\begin{aligned}
\|\vec{u}\|_W &= \|\vec{w}_0 + \alpha_1 \vec{w}_1 + \alpha_1 \alpha_2 \vec{w}_2 + \cdots + \prod_{j=1}^{m} \alpha_j (\vec{w}_m + \vec{u}_{m+1})\|_W \\
&\geq \|\vec{w}_0\|_W - \|\alpha_1 \vec{w}_1\|_W - \|\alpha_1 \alpha_2 \vec{w}_2\|_W - \cdots - \|\prod_{j=1}^{m} \alpha_j \vec{w}_m\|_W - \|\prod_{j=1}^{m} \alpha_j \vec{u}_{m+1}\|_W \\
&\geq (1 - \epsilon) - \epsilon(1 + \epsilon) - \epsilon^2(1 + \epsilon) - \cdots - \epsilon^m(1 + \epsilon) - \epsilon^m \|\vec{u}_{m+1}\|_W \\
&\geq 1 - \epsilon - \frac{(1 + \epsilon)\epsilon}{1 - \epsilon} - \epsilon^m \cdot (ld + 1)^3 \geq 1 - 3\epsilon.
\end{aligned}$$

Similarly, we have $\|\vec{u}\|_W \leq 1 + 3\epsilon$. $\qquad\square$

Lemma 9.6. *With probability at least 0.99 over the m i.i.d samples in W,*

$$\|x'(t) - \widetilde{x}(t)\|_T \leq 2200 \left(\|g(t)\|_T + \|x^{(S)}(t) - x'(t)\|_T \right).$$

Proof. Let $g'(t) = g(t) + x^*(t) - x'(t)$ such that $x(t) = x'(t) + g'(t)$. Then we choose $\epsilon = 0.03$ and bound:

$$
\begin{aligned}
&\|x'(t) - \widetilde{x}(t)\|_T \\
&\leq (1 + 3\epsilon)\|x'(t) - \widetilde{x}(t)\|_W && \text{with prob. } 1 - 2^{-\Omega(d\log d)} \text{ by Claim 9.5} \\
&= 1.09\|x'(t) - \widetilde{x}(t)\|_W && \text{by } \epsilon = 0.03 \\
&= 1.09\|x(t) - g'(t) - \widetilde{x}(t)\|_W && \text{by } x'(t) = x(t) - g'(t) \\
&\leq 1.09\|x(t) - \widetilde{x}(t)\|_W + 1.09\|g'(t)\|_W && \text{by triangle inequality} \\
&\leq 1.09\|x(t) - x'(t)\|_W + 1.09\|g'(t)\|_W && \text{by } \widetilde{x} = \argmin_{y \in \mathrm{span}\{V\}} \|x - y\|_W \\
&= 2.18\|g'(t)\|_W. && \text{by } x(t) - x'(t) = g(t)
\end{aligned}
$$

From the fact that $\mathbb{E}_W[\|g'\|_W] = \|g'\|_T$, $\|g'\|_W \leq 1000\|g'\|_T$ with probability at least .999. It indicates $\|x'(t) - \widetilde{x}(t)\|_T \leq 2200\|g'\|_T$ with probability at least 0.99 from all discussion above. $\qquad\square$

9.5 Boosting the success probability

In order to achieve $1 - 2^{-\Omega(k)}$ for the main theorem, we cannot combine Procedure SIGNALRE-COVERYKCLUSTER with FREQUENCYRECOVERYKCLUSTER directly. However, using the similar proof technique in Theorem 4.5, we are able to boost the success probability by using Procedure SIGNALRECOVERYKCLUSTER$^+$ in Algorithm 8. It runs Procedure SIGNALRECOVERYKCLUSTER $R = O(k)$ times in parallel for independent fresh samples and report R different d-Fourier-sparse signals $\widetilde{x}_i(t)$. Then, taking $m = \mathrm{poly}(k)$ new locations $\{t_1, t_2, \cdots, t_m\}$, and computing $\widetilde{A}$ as before and $\widetilde{b}_j$ by taking the median of $\{\widetilde{x}_1(t_j), \cdots, \widetilde{x}_R(t_j)\}$. At the end, solving the linear regression for matrix $\widetilde{A}$ and vector $\widetilde{b}$. Thus, we complete the proof of Lemma 9.1.

Because we can transfer a degree-d polynomial to a d-Fourier-sparse signal by Lemma 8.8, the output of Procedure CFTKCLUSTER in Algorithm 8 matches the main theorem,

Theorem 1.1. *Let $x(t) = x^*(t) + g(t)$, where x^* is k-Fourier-sparse signal with frequencies in $[-F, F]$. Given samples of x over $[0, T]$ we can output $\widetilde{x}(t)$ such that with probability at least $1 - 2^{-\Omega(k)}$,*

$$\|\widetilde{x} - x^*\|_T \lesssim \|g\|_T + \delta \|x^*\|_T.$$

Our algorithm uses $\mathrm{poly}(k, \log(1/\delta)) \cdot \log(FT)$ samples and $\mathrm{poly}(k, \log(1/\delta)) \cdot \log^2(FT)$ time. The output $\widetilde{x}$ is $\mathrm{poly}(k, \log(1/\delta))$-Fourier-sparse signal.

References

[BCG+14] Petros Boufounos, Volkan Cevher, Anna C Gilbert, Yi Li, and Martin J Strauss. What's the frequency, Kenneth?: Sublinear Fourier sampling off the grid. In *Algorithmica(A preliminary version of this paper appeared in the Proceedings of RANDOM/APPROX 2012, LNCS 7408, pp. 61-72)*, pages 1–28. Springer, 2014.

[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.

[Bou14] Jean Bourgain. An improved estimate in the restricted isometry problem. In *Geometric Aspects of Functional Analysis*, pages 65–70. Springer, 2014.

[BS12] Markus Blaser and Chandan Saha. Lecture 6: Multipoint evaluation of a polynomial. *Max-Planck-Institut für Informatik Class Notes, Computational Number Theory and Algebra*, pages 1–4, 2012.

[CDL13] Albert Cohen, Mark A Davenport, and Dany Leviatan. On the stability and accuracy of least squares approximations. *Foundations of computational mathematics*, 13(5):819–834, 2013.

[CF14] Emmanuel J Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

[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.

[CNW15] Michael B Cohen, Jelani Nelson, and David P Woodruff. Optimal approximate matrix product in terms of stable rank. *arXiv preprint arXiv:1507.02268*, 2015.

[CRT06] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.

[CW87] Don Coppersmith and Shmuel Winograd. Matrix multiplication via arithmetic progressions. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 1–6. ACM, 1987.

[CW13] Kenneth L. Clarkson and David P. Woodruff. Low rank approximation and regression in input sparsity time. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 81–90, 2013.

[DB13] Marco F Duarte and Richard G Baraniuk. Spectral compressive sensing. *Applied and Computational Harmonic Analysis*, 35(1):111–129, 2013.

[Dun10] Mark Dunster. *Legendre and Related Functions*. Handbook of Mathematical Functions, Cambridge University Press, 2010.

[FL12] Albert Fannjiang and Wenjing Liao. Coherence pattern-guided compressive sensing with unresolved grids. *SIAM Journal on Imaging Sciences*, 5(1):179–202, 2012.

[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 thiry-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.

[GZ16] Venkatesan Guruswami and David Zuckerman. Robust Fourier and polynomial curve fitting. In *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*. IEEE, 2016.

[Haz01] Michiel Hazewinkel. *Gram matrix*. Encyclopedia of Mathematics, Springer, 2001.

[HIKP12a] Haitham Hassanieh, Piotr Indyk, Dina Katabi, and Eric Price. Nearly optimal sparse Fourier transform. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*. ACM, 2012.

[HIKP12b] 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.

[HK15] Qingqing Huang and Sham M Kakade. Super-resolution off the grid. In *Advances in Neural Information Processing Systems*, pages 2647–2655, 2015.

[HR15] Ishay Haviv and Oded Regev. The restricted isometry property of subsampled Fourier matrices. *arXiv preprint arXiv:1507.01768*, 2015.

[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.

[IKP14] Piotr Indyk, Michael Kapralov, and Eric Price. (Nearly) Sample-optimal sparse Fourier transform. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 480–499. SIAM, 2014.

[Iwe13] Mark A Iwen. Improved approximation guarantees for sublinear-time Fourier algorithms. *Applied And Computational Harmonic Analysis*, 34(1):57–82, 2013.

[Kap16] Michael Kapralov. Sparse Fourier transform in any constant dimension with nearly-optimal sample complexity in sublinear time. In *Symposium on Theory of Computing Conference, STOC'16, Cambridge, MA, USA, June 19-21, 2016*, 2016.

[Mas69] James L Massey. Shift-register synthesis and BCH decoding. *Information Theory, IEEE Transactions on*, 15(1):122–127, 1969.

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

[MP14] Gregory T Minton and Eric Price. Improved concentration bounds for count-sketch. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 669–686. Society for Industrial and Applied Mathematics, 2014.

[NN13] Jelani Nelson and Huy L Nguyên. OSNAP: Faster numerical linear algebra algorithms via sparser subspace embeddings. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 117–126. IEEE, 2013.

[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.

[RPK86] Robert Roy, Arogyaswami Paulraj, and Thomas Kailath. Esprit–a subspace rotation approach to estimation of parameters of cisoids in noise. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 34(5):1340–1342, 1986.

[RV08] Mark Rudelson and Roman Vershynin. On sparse reconstruction from fourier and gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008.

[Sch81] Ralph Otto Schmidt. A signal subspace approach to multiple emitter location spectral estimation. *Ph. D. Thesis, Stanford University*, 1981.

[Str69] Volker Strassen. Gaussian elimination is not optimal. *Numerische Mathematik*, 13(4):354–356, 1969.

[Tar09] Robert E. Tarjan. Lecture 10: More chernoff bounds, sampling, and the chernoff + union bound. *Princeton Class Notes, Probability and Computing*, pages 1–9, 2009.

[TBR15] Gongguo Tang, Badri Narayan Bhaskar, and Benjamin Recht. Near minimax line spectral estimation. *Information Theory, IEEE Transactions on*, 61(1):499–512, 2015.

[TBSR13] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht. Compressed sensing off the grid. *Information Theory, IEEE Transactions on*, 59(11):7465–7490, 2013.

[Wil12] Virginia Vassilevska Williams. Multiplying matrices faster than coppersmith-winograd. In *STOC*, pages 887–898. ACM, 2012.

[Woo14] David P Woodruff. Sketching as a tool for numerical linear algebra. *arXiv preprint arXiv:1411.4357*, 2014.

[YX15] Zai Yang and Lihua Xie. Achieving high resolution for super-resolution via reweighted atomic norm minimization. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 3646–3650. IEEE, 2015.

A Technical Proofs

A.1 Proof of Theorem 8.3

We prove the following Theorem

Theorem 8.3. *For real numbers $\xi_1, \ldots, \xi_k$, let $G_{\xi_1, \ldots, \xi_k}$ be the matrix whose (i,j)-entry is*

$$\int_{-1}^{1} e^{2\pi \mathbf{i}(\xi_i - \xi_j)t} dt.$$

Then

$$\det(G_{\xi_1, \ldots, \xi_k}) = 2^{\tilde{O}(k^2)} \prod_{i<j} \min(|\xi_i - \xi_j|^2, 1).$$

First, we note by the Cauchy-Binet formula that the determinant in question is equal to

$$\int_{-1}^{1} \int_{-1}^{1} \cdots \int_{-1}^{1} \left| \det([e^{2\pi i \xi_i t_j}]_{i,j}) \right|^2 dt_1 dt_2 \ldots dt_k. \tag{27}$$

We next need to consider the integrand in the special case when $\sum |\xi_i| \leq 1/8$.

Lemma A.1. *If $\xi_i \in R$ and $t_j \in \mathbb{R}$, $\sum_i |\xi_i|(\max_i |t_i|) \leq 1/8$ then*

$$|\det([e^{2\pi i \xi_i t_j}]_{i,j})| = \Theta \left(\frac{(2\pi)^{\binom{k}{2}} \prod_{i<j} |t_i - t_j||\xi_i - \xi_j|}{1!2! \cdots k!} \right).$$

Proof. Firstly, by adding a constant to all the t_j we can make them non-negative. This multiplies the determinant by a root of unity, and at most doubles $\sum_i |\xi_i|(\max_i |t_i|)$.

By continuity, it suffices to consider the t_i to all be multiples of $1/N$ for some large integer N. By multiplying all the t_j by N and all ξ_i by $1/N$, we may assume that all of the t_j are non-negative integers with $t_1 \leq t_2 \leq \ldots \leq t_k$.

Let $z_i = \exp(2\pi i \xi_i)$. Then our determinant is

$$\det \left(\left[z_i^{t_j} \right]_{i,j} \right),$$

which is equal to the Vandermonde determinant times the Schur polynomial $s_\lambda(z_i)$ where λ is the partition $\lambda_j = t_j - (j-1)$.

Therefore, this determinant equals

$$\prod_{i<j} (z_i - z_j) s_\lambda(z_1, z_2, \ldots, z_k).$$

The absolute value of

$$\prod_{i<j} (z_i - z_j)$$

is approximately $\prod_{i<j} (2\pi i)(\xi_i - \xi_j)$, which has absolute value $(2\pi)^{\binom{k}{2}} \prod_{i<j} |\xi_i - \xi_j|$. We have left to evaluate the size of the Schur polynomial.

By standard results, s_λ is a polynomial in the z_i with non-negative coefficients, and all exponents at most $\max_j |t_j|$ in each variable. Therefore, the monomials with non-zero coefficients will all have real part at least $1/2$ and absolute value 1 when evaluated at the z_i. Therefore,

$$|s_\lambda(z_1, \ldots, z_k)| = \Theta(|s_\lambda(1, 1, \ldots, 1)|).$$

On the other hand, by the Weyl character formula

$$s_\lambda(1, 1, \ldots, 1) = \prod_{i<j} \frac{t_j - t_i}{j - i} = \frac{\prod_{i<j} |t_i - t_j|}{1! 2! \ldots k!}.$$

This completes the proof. $\qquad\square$

Next we prove our Theorem when the ξ have small total variation.

Lemma A.2. *If there exists a ξ_0 so that $\sum |\xi_i - \xi_0| < 1/8$, then*

$$\det(G_{\xi_1, \ldots, \xi_k}) = \Theta\left(\frac{2^{3k(k-1)/2} \pi^{k(k-1)} \prod_{i<j} |\xi_i - \xi_j|^2}{(k!)^3 \prod_{n=0}^{k-1} (2n)!} \right).$$

Proof. By translating the ξ_i we can assume that $\xi_0 = 0$.

By the above we have

$$\Theta\left(\frac{(2\pi)^{k(k-1)} \prod_{i<j} |\xi_i - \xi_j|^2}{(1! 2! \cdots k!)^2}\right) \int_{-1}^1 \cdots \int_{-1}^1 \prod_{i<j} |t_i - t_j|^2 dt_1 \ldots dt_k.$$

We note that by the Cauchy-Binet formula the latter term is the determinant of the matrix M with $M_{i,j} = \int_{-1}^1 t^{i+j} dt$. This is the Graham matrix associated to the polynomials t^i for $0 \le i \le k-1$. Applying Graham-Schmidt (without the renormalization step) to this set yields the basis $P_n \alpha_n$ where $\alpha_n = \frac{2^n (n!)^2}{(2n)!}$ is the inverse of the leading term of P_n. This polynomial has norm $\alpha_n^2 2/(2n+1)$. Therefore, the integral over the t_i yields

$$\prod_{n=0}^{k-1} \frac{2^{n+1} (n!)^2}{(n+1)(2n)!}.$$

This completes the proof. $\qquad\square$

Next we extend this result to the case that all the ξ are within $\text{poly}(k)$ of each other.

Proposition A.3. *If there exists a ξ_0 so that $|\xi_i - \xi_0| = \text{poly}(k)$ for all i, then*

$$\det(G_{\xi_1, \ldots, \xi_k}) = 2^{\tilde{O}(k^2)} \prod_{i<j} \min(|\xi_i - \xi_j|^2, 1).$$

Proof. We begin by proving the lower bound. We note that for $0 < x < 1$,

$$\det(G_{\xi_1, \ldots, \xi_k}) \ge \int_{-x}^x \int_{-x}^x \cdots \int_{-1}^1 \left| \det([e^{2\pi i \xi_i t_j}]_{i,j}) \right|^2 dt_1 dt_2 \ldots dt_k = x^k \det(G_{\xi_1/x, \xi_2/x, \ldots, \xi_k/x}).$$

Taking $x = 1/\text{poly}(k)$, we may apply the above Lemma to compute the determinant on the right hand side, yielding an appropriate lower bound.

To prove the lower bound, we note that we can divide our ξ_i into clusters, $\mathcal{C}_i$, where for any i, j in the same cluster $|\xi_i - \xi_j| < 1/k$ and for i and j in different clusters $|\xi_i - \xi_j| \geq 1/k^2$. We then note as a property of Graham matrices that

$$\det(G_{\xi_1,\ldots,\xi_k}) \leq \prod_{\mathcal{C}_i} \det(G_{\{\xi_j \in \mathcal{C}_i\}}) = 2^{\tilde{O}(k^2)} \prod_{i<j, \text{ in same cluster}} |\xi_i - \xi_j|^2 = 2^{\tilde{O}(k^2)} \prod_{i<j} |\xi_i - \xi_j|^2.$$

This completes the proof. $\square$

Finally, we are ready to prove our Theorem.

Proof. Let $I(t)$ be the indicator function of the interval $[-1, 1]$.

Recall that there is a function $h(t)$ so that for any function f that is a linear combination of at most k complex exponentials that $|h(t)f(t)|_2 = \Theta(|I(t)f(t)|_2)$ and so that $\hat{h}$ is supported on an interval of length $\text{poly}(k) < k^C$ about the origin.

Note that we can divide our ξ_i into clusters, $\mathcal{C}_\rangle$, so that for i and j in a cluster $|\xi_i - \xi_j| < k^{C+1}$ and for i and j in different clusters $|\xi_i - \xi_j| > k^C$.

Let $\tilde{G}_{\xi_1,\xi_2,\ldots,\xi'_k}$ be the matrix with (i, j)-entry $\int_{\mathbb{R}} |h(t)|^2 e^{(2\pi i)(\xi_i - \xi_j)t} dt$.

We claim that for any $k' \leq k$ that

$$\det(\tilde{G}_{\xi_1,\xi_2,\ldots,\xi'_k}) = 2^{O(k')} \det(G_{\xi_1,\xi_2,\ldots,\xi'_k}).$$

This is because both are Graham determinants, one for the set of functions $I(t) \exp((2\pi i)\xi_j t)$ and the other for $h(t) \exp((2\pi i)\xi_j t)$. However since any linear combination of the former has L^2 norm a constant multiple of that the same linear combination of the latter, we have that

$$\tilde{G}_{\xi_1,\xi_2,\ldots,\xi'_k} = \Theta(G_{\xi_1,\xi_2,\ldots,\xi'_k})$$

as self-adjoint matrices. This implies the appropriate bound.

Therefore, we have that
$$\det(G_{\xi_1,\ldots,\xi_k}) = 2^{O(k)} \det(\tilde{G}_{\xi_1,\ldots,\xi_k}).$$

However, note that by the Fourier support of h that

$$\int_{\mathbb{R}} |h(t)|^2 e^{(2\pi i)(\xi_i - \xi_j)t} dt = 0$$

if $|\xi_i - \xi_j| > k^C$, which happens if i and j are in different clusters. Therefore $\tilde{G}$ is block diagonal and hence its determinant equals

$$\det(\tilde{G}_{\xi_1,\ldots,\xi_k}) = \prod_{\mathcal{C}_\rangle} \det(\tilde{G}_{\{\xi_j \in \mathcal{C}_i\}}) = 2^{O(k)} \prod_{\mathcal{C}_\rangle} \det(G_{\{\xi_j \in \mathcal{C}_i\}}).$$

However the Proposition above shows that

$$\prod_{\mathcal{C}_\rangle} \det(G_{\{\xi_j \in \mathcal{C}_i\}}) = 2^{\tilde{O}(k^2)} \prod_{i<j} \min(1, |\xi_i - \xi_j|^2).$$

This completes the proof. $\square$

A.2   Proofs of Lemma 5.3 and Lemma 5.4

We fix $z_1, \cdots, z_k$ to be complex numbers on the unit circle and use $Q(z)$ to denote the degree-k polynomial $\prod_{i=1}^{k} (z - z_i)$.

Lemma 5.3. *Let $Q(z)$ be a degree k polynomial, all of whose roots are complex numbers with absolute value 1. For any integer n, let $r_{n,k}(z) = \sum_{l=0}^{k-1} r_{n,k}^{(l)} \cdot z^l$ denote the residual polynomial of*

$$r_{n,k}(z) \equiv z^n \pmod{Q(z)}.$$

Then, each coefficient of $r_{n,k}$ is bounded: $|r_{n,k}^{(l)}| \leq 2^k n^{k-1}$ for any l.

Proof. By definition, $r_{n,k}(z_i) = z_i^n$. From the polynomial interpolation, we have

$$r_{n,k}(z) = \sum_{i=1}^{k} \frac{\prod\limits_{j \in [k] \setminus i} (z - z_j) z_i^n}{\prod\limits_{j \in [k] \setminus i} (z_i - z_j)}.$$

Let $\mathrm{Sym}_{S,i}$ be the symmetry polynomial of $z_1, \cdots, z_k$ with degree i among subset $S \subseteq [k]$, i.e., $\mathrm{Sym}_{S,i} = \sum\limits_{S' \subseteq \binom{S}{i}} \prod\limits_{j \in S'} z_j$. Then

$$r_{n,k}^{(l)} = (-1)^{k-1-l} \sum_{i=1}^{k} \frac{\mathrm{Sym}_{[k] \setminus i, k-1-l} \cdot z_i^n}{\prod\limits_{j \in [k] \setminus i} (z_i - z_j)}.$$

We omit $(-1)^{k-1-l}$ in the rest of proof and use induction on n, k, and l to prove $|r_{n,k}^{(l)}| \leq \binom{k-1}{l} \binom{n}{k-1}$.

Base Case of n: For any $n < k$, from the definition, $r(z) = z^n$ and $|r_{n,k}^{(l)}| \leq 1$.

Suppose it is true for any $n < n_0$. We consider $r_{n_0,k}^l$ from now on. When $k = 1$, $r_{n,0} = z_1^n$ is bounded by 1 because z_1 is on the unit circle of $\mathbb{C}$.

Given n_0, suppose the induction hypothesis is true for any $k < k_0$ and any $l < k$. For $k = k_0$,

61

we first prove that $|r_{n_0,k_0}^{(k_0-1)}| \leq \binom{n_0}{k_0-1}$ then prove that $|r_{n_0,k_0}^{(l)}| \leq \binom{k_0-1}{l}\binom{n_0}{k_0-1}$ for $l = k_0 - 2, \cdots, 0$.

$$
\begin{aligned}
r_{n_0,k_0}^{(k_0-1)} &= \sum_{i=1}^{k_0} \frac{z_i^{n_0}}{\prod_{j\in[k_0]\backslash i}(z_i - z_j)} \\
&= \sum_{i=1}^{k_0-1} \frac{z_i^{n_0}}{\prod_{j\in[k_0]\backslash i}(z_i - z_j)} + \frac{z_{k_0}^{n_0}}{\prod_{j\in[k_0]\backslash k_0}(z_{k_0} - z_j)} \\
&= \sum_{i=1}^{k_0-1} \frac{z_i^{n_0} - z_i^{n_0-1}z_{k_0} + z_i^{n_0-1}z_{k_0}}{\prod_{j\in[k_0]\backslash i}(z_i - z_j)} + \frac{z_{k_0}^{n_0}}{\prod_{j\in k_0\backslash k_0}(z_{k_0} - z_j)} \\
&= \sum_{i=1}^{k_0-1} \left(\frac{z_i^{n_0-1}}{\prod_{j\in[k_0-1]\backslash i}(z_i - z_j)} + \frac{z_i^{n_0-1}z_{k_0}}{\prod_{j\in k_0\backslash i}(z_i - z_j)} \right) + \frac{z_{k_0}^{n_0}}{\prod_{j\in k_0\backslash k_0}(z_{k_0} - z_j)} \\
&= \left(\sum_{i=1}^{k_0-1} \frac{z_i^{n_0-1}}{\prod_{j\in[k_0-1]\backslash i}(z_i - z_j)} \right) + \left(z_{k_0}\sum_{i=1}^{k_0} \frac{z_i^{n_0-1}}{\prod_{j\in k_0\backslash i}(z_i - z_j)} \right) \\
&= r_{n_0-1,k_0-1}^{(k_0-2)} + z_{k_0} \cdot r_{n_0-1,k_0}^{(k_0-1)}
\end{aligned}
$$

Hence $|r_{n_0,k_0}^{(k_0-1)}| \leq |r_{n_0-1,[k_0-1]}^{(k_0-2)}| + |r_{n_0-1,k_0}^{(k_0-1)}| \leq \binom{n_0-2}{k_0-2} + \binom{n_0-2}{k_0-1} = \binom{n_0-1}{k_0-1}$. For $l < k_0 - 1$, we have

$$
\begin{aligned}
r_{n_0,k_0}^{(l)} &= \sum_{i=1}^{k_0} \frac{\operatorname{Sym}_{[k_0]\backslash i, k_0-1-l} \cdot z_i^{n_0}}{\prod_{j\in[k_0]\backslash i}(z_i - z_j)} \quad \text{let } l' = k_0 - 1 - l \\
&= \sum_{i=1}^{k_0-1} \frac{\left(\operatorname{Sym}_{[k_0-1]\backslash i, l'} + \operatorname{Sym}_{[k_0-1]\backslash i, l'-1}\cdot z_{k_0}\right) z_i^{n_0}}{\prod_{j\in[k_0]\backslash i}(z_i - z_j)} + \frac{\operatorname{Sym}_{[k_0-1], l'}\cdot z_{k_0}^{n_0}}{\prod_{j<k_0}(z_{k_0} - z_j)} \\
&= \sum_{i=1}^{k_0-1} \frac{\operatorname{Sym}_{[k_0-1]\backslash i, l'}\cdot(z_i - z_{k_0})z_i^{n_0-1} + \operatorname{Sym}_{[k_0-1]\backslash i, l'}\cdot z_{k_0}z_i^{n_0-1} + \operatorname{Sym}_{[k_0-1]\backslash i, l'-1}\cdot z_{k_0}z_i^{n_0}}{\prod_{j\in[k_0]\backslash i}(z_i - z_j)} \\
&\quad + \frac{\operatorname{Sym}_{[k_0-1], l'}\cdot z_{k_0}^{n_0}}{\prod_{j<k_0}(z_{k_0} - z_j)} \\
&= \sum_{i=1}^{k_0-1} \frac{\operatorname{Sym}_{[k_0-1]\backslash i, l'}\cdot(z_i - z_{k_0})z_i^{n_0-1} + \operatorname{Sym}_{[k_0-1], l'}\cdot z_{k_0}z_i^{n_0-1}}{\prod_{j\in[k_0]\backslash i}(z_i - z_j)} + \frac{\operatorname{Sym}_{[k_0-1], l'}\cdot z_{k_0}^{n_0}}{\prod_{j<k_0}(z_{k_0} - z_j)} \\
&= \sum_{i=1}^{k_0-1} \frac{\operatorname{Sym}_{[k_0-1]\backslash i, l'}\, z_i^{n_0-1}}{\prod_{j\in[k_0-1]\backslash i}(z_i - z_j)} + \sum_{i=1}^{k_0-1} \frac{\operatorname{Sym}_{[k_0-1], l'}\cdot z_{k_0}z_i^{n_0-1}}{\prod_{j\in[k_0]\backslash i}(z_i - z_j)} + \frac{\operatorname{Sym}_{[k_0-1], l'}\cdot z_{k_0}^{n_0}}{\prod_{j<k_0}(z_{k_0} - z_j)} \\
&= r_{n_0-1,k_0-1}^{(l-1)} + \operatorname{Sym}_{[k_0-1], k_0-1-l}\cdot z_{k_0} \cdot r_{n_0-1,k_0}^{(k_0-1)}
\end{aligned}
$$

By induction hypothesis, $|r_{n_0,k_0}^{(l)}| \leq \binom{k_0-2}{l-1}\binom{n_0-1}{k_0-2} + \binom{k_0-1}{l}\binom{n_0-1}{k_0-1} \leq \binom{k_0-1}{l}\binom{n_0}{k_0-1}$. □

Now we finish the proof of Lemma 5.4.

Lemma 5.4. *For any $k \in \mathbb{Z}$ and any $z_1, \cdots, z_k$ on the unit circle of $\mathbb{C}$, there always exists a degree $m = O(k^2 \log k)$ polynomial $P(z) = \sum_{j=0}^{m} c_j z^j$ with the following properties:*

$$
\begin{aligned}
\text{Property I} \quad & P(z_i) = 0, \forall i \in \{1, \cdots, k\}, \\
\text{Property II} \quad & c_0 = 1, \\
\text{Property III} \quad & |c_j| \leq 11, \forall j \in \{1, \cdots, m\}.
\end{aligned}
$$

Let $m = 10k^2 \log k$ and $\mathcal{P}$ denote a set of polynomials that has degree at most m, and all the coefficients are integers chosen from $\{-5, \cdots, -1, 0, 1, \cdots, 5\}$, i.e.,

$$
\mathcal{P} := \left\{ P(z) = \sum_{i=0}^{m} \alpha_i z^i \mid \forall i \in \{0, 1, \cdots, m\}, |\alpha_i| \leq 2 \right\}.
$$

Claim A.4. *There exists $P^*(z) = \sum_{i=0}^{m} \alpha_i z^i$ with coefficient $|\alpha_i| \leq 10$ for any $i \in \{0, 1, \cdots, m\}$, such that every coefficient of $P^*(z) \mod Q(z)$ is bounded by 2^{-m}.*

Proof. For $P(z) = \sum_{i=0}^{m} \alpha_i z^i \in \mathcal{P}$, $P(z) \mod Q(z) \equiv \sum_{i=0}^{m} \alpha_i r_{n,k}(z)$ from the definition $r_{n,k}(z)$. Hence

$$
P(z) \mod Q(z) = \sum_{i=0}^{m} \alpha_i \sum_{l=0}^{k-1} r_{i,k}^{(l)} z^l = \sum_{l=0}^{k-1} z^l \sum_{i=0}^{m} \alpha_i r_{i,k}^{(l)}.
$$

Each coefficient in $P(z) \mod Q(z)$ is bounded by $|\sum_{i=0}^{m} \alpha_i r_{i,k}^{(l)}| \leq 5 \sum_{i=0}^{m} 2^k i^{k-1} \leq 2^k m^k$.

At the same time, $|\mathcal{P}| = 11^m$. From the pigeonhole principle and $\left(\frac{2^k m^k}{(2^{-m})^2} \right)^k < 11^m$, there exists $P_1, P_2 \in \mathcal{P}$ such that for $P^*(z) = P_1(z) - P_2(z)$, $P^*(z) \mod Q(z) = \sum_{i=0}^{k-1} \gamma_i z^i$ where each coefficient $|\gamma_i| \leq 2^{-m}$. $\qquad\square$

Let $r(z) = \sum_{i=0}^{k-1} \gamma_i z^i = P^*(z) \mod Q(z)$ for convenience. If $P^*(0)$ (the constant term of P^*) is nonzero, then $|P^*(0) - r(0)| \geq 0.99$ from the above lemma. Therefore the polynomial $\frac{P^*(z) - r(z)}{P^*(0) - r(0)}$ satisfies the three properties in Lemma 5.4.

Otherwise, we assume z^l is the first term in $P^*(z)$ with a non-zero coefficient. Let

$$
r_{-l,k}(z) = z^{-l} \mod Q(z) = \sum_{i=1}^{k} \frac{\prod_{j \in [k] \setminus i} (z - z_j) z_i^{-l}}{\prod_{j \in [k] \setminus i} (z_i - z_j)}.
$$

For convenience, we use $z_S = \prod_{i \in S} z_i$ for any subset $S \subseteq [k]$. Notice that $z_i^{-l} = \frac{z_{[k] \setminus i}^l}{z_{[k]}^l}$. Hence $r_{-l,k}(z) = r_{l,k}'(z)/z_{[k]}^l$ where r' is the polynomial for k units roots $z_{[k] \setminus 1}, \cdots, z_{[k] \setminus k}$. So each coefficients of r is still bounded by $2^k l^k$, which is less than $2^{-m/2}$.

Eventually we choose $P^*(z)/z^l - r(z) \cdot r_{-l,k}(z)$ and renormalize it to satisfy the three properties in Lemma 5.4.

A.3   Proof of Lemma 4.3

Lemma 4.3. *For any degree d polynomial $P(t) : \mathbb{R} \to \mathbb{C}$ with derivative $P'(t)$, we have,*

$$\int_{-1}^{1} (1 - t^2)|P'(t)|^2 \mathrm{d}t \le 2d^2 \int_{-1}^{1} |P(t)|^2 \mathrm{d}t. \tag{28}$$

Given a degree d polynomial $P(x)$, we rewrite $P(x)$ as a linear combination of the Legendre polynomials:

$$P(x) = \sum_{i=0}^{d} \alpha_i L_i(x).$$

We use $F_i(x) = (1 - x^2)L_i'(x)$ for convenience. From the definition of the Legendre polynomials in the Equation (3), $F_i'(x) = -i(i+1) \cdot L_i(x)$ and $F_i''(x) = -i(i+1) \cdot L_i'(x)$.

Hence we have

$$
\begin{aligned}
\int_{1}^{-1} (1 - x^2)|P'(x)|^2 \mathrm{d}x = {} & \int_{1}^{-1} (1 - x^2)P'(x) \cdot \overline{P'}(x)\mathrm{d}x \\
= {} & \int_{1}^{-1} \left(\sum_{i \in [d]} \alpha_i F_i(x) \right) \cdot \left(\sum_{i \in [d]} \overline{\alpha_i} \frac{-F_i''(x)}{i(i+1)} \right) \mathrm{d}x \\
= {} & \left(\sum_{i \in [d]} \alpha_i F_i(x) \right) \cdot \left(\sum_{i \in [d]} \overline{\alpha_i} \frac{-F_i'(x)}{i(i+1)} \right) \Big|_{-1}^{1} \\
& + \int_{1}^{-1} \left(\sum_{i \in [d]} \alpha_i F_i'(x) \right) \cdot \left(\sum_{i \in [d]} \overline{\alpha_i} \frac{F_i'(x)}{i(i+1)} \right) \mathrm{d}x \\
= {} & \int_{1}^{-1} \left(\sum_{i \in [d]} \alpha_i \cdot i(i+1) \cdot L_i(x) \right) \cdot \left(\sum_{i \in [d]} \overline{\alpha_i} \frac{i(i+1) \cdot L_i(x)}{i(i+1)} \right) \mathrm{d}x \\
= {} & \sum_{i \in [d]} |\alpha_i|^2 i(i+1) \|L_i\|_T^2 \\
\le {} & d(d+1)\|P\|_T^2
\end{aligned}
$$

A.4   Proof of Lemma 6.2

Lemma 6.2. $\widehat{P_{\sigma,a,b}x}(\sigma(f - b)) = \frac{1}{\sigma}e^{-2\pi \mathbf{i}\sigma a f}\widehat{x}(f)$ *and* $\widehat{P_{\sigma,a,b}x}(f) = \frac{1}{\sigma}e^{-2\pi \mathbf{i}\sigma a (f/\sigma+b)}\widehat{x}(f/\sigma + b)$

Proof. Let's compute the Fourier Transform of $(P_{\sigma,a,b}x)(t)$,

$$\widehat{P_{\sigma,a,b}x}(f)$$

$$= \int_{-\infty}^{+\infty} (P_{\sigma,a,b}(x))(t)e^{-2\pi \mathbf{i} f t}\mathrm{d}t$$

$$= \int_{-\infty}^{+\infty} x(\sigma(t-a))e^{-2\pi \mathbf{i} \sigma b t}e^{-2\pi \mathbf{i} f t}\mathrm{d}t$$

$$= e^{-2\pi \mathbf{i}(\sigma ab + fa)}\int_{-\infty}^{+\infty} x(\sigma(t-a))e^{-2\pi \mathbf{i}\sigma b(t-a)}e^{-2\pi \mathbf{i} f(t-a)}\mathrm{d}t \qquad \text{by shifting } t \text{ by } a$$

$$= e^{-2\pi \mathbf{i}(\sigma ab + fa)}\int_{-\infty}^{+\infty} x(\sigma t)e^{-2\pi \mathbf{i}\sigma b t}e^{-2\pi \mathbf{i} f t}\mathrm{d}t \qquad \text{by replacing } t-a \text{ by } t$$

$$= \frac{1}{\sigma}e^{-2\pi \mathbf{i}(\sigma ab + fa)}\int_{-\infty}^{+\infty} x(\sigma t)e^{-2\pi \mathbf{i} b\sigma t}e^{-2\pi \mathbf{i} f\sigma t/\sigma}\mathrm{d}\sigma t \qquad$$

$$= \frac{1}{\sigma}e^{-2\pi \mathbf{i}(\sigma ab + fa)}\int_{-\infty}^{+\infty} x(t)e^{-2\pi \mathbf{i}(b + f/\sigma)t}\mathrm{d}t \qquad \text{by replacing } t\sigma \text{ by } t$$

$$= \frac{1}{\sigma}e^{-2\pi \mathbf{i} a\sigma(f/\sigma + b)}\widehat{x}(f/\sigma + b) \qquad \text{by definition of FT}$$

The first result follows immediately by replacing $f/\sigma + b$ by f', which gives

$$\widehat{P_{\sigma,a,b}x}(\sigma(f'-b)) = \frac{1}{\sigma}e^{-2\pi \mathbf{i} a\sigma f'}\widehat{x}(f').$$

Thus, we complete the proof of this Lemma. $\qquad \square$

A.5   Proof of Lemma 3.5

Lemma 3.5. *Given any function $x(t): \mathbb{R} \to \mathbb{C}$ with $\max_{t \in [0,T]} |x(t)|^2 \leq d\|x(t)\|_T^2$. Let S denote a set of points from 0 to T. If each point of S is chosen uniformly at random from $[0,T]$, we have*

$$\Pr\left[\left|\frac{1}{|S|}\sum_{i \in S}|x(t_i)|^2 - \|x(t)\|_T^2\right| \geq \epsilon\|x(t)\|_T^2\right] \leq e^{-\Omega(\epsilon^2|S|/d)}$$

Proof. Let M denote $\max_{t \in [0,T]}|x(t)|^2$. Replacing X_i by $\frac{|x(t_i)|^2}{M}$ and n by $|S|$ in Lemma B.2, we obtain that

$$\Pr[|X - \mu| > \epsilon\mu] \leq 2\exp(-\frac{\epsilon^2}{3}\mu)$$

$$\implies \Pr\left[\left|\sum_{i \in S}\frac{|x(t_i)|^2}{M} - |S|\frac{\|x(t)\|_T^2}{M}\right| > \epsilon|S|\frac{\|x(t)\|_T^2}{M}\right] \leq 2\exp(-\frac{\epsilon^2}{3}\mu)$$

$$\implies \Pr\left[\left|\frac{1}{|S|}\sum_{i \in S}|x(t_i)|^2 - \|x(t)\|_T^2\right| \geq \epsilon\|x(t)\|_T^2\right] \leq 2\exp(-\frac{\epsilon^2}{3}\mu)$$

$$\implies \Pr\left[\left|\frac{1}{|S|}\sum_{i \in S}|x(t_i)|^2 - \|x(t)\|_T^2\right| \geq \epsilon\|x(t)\|_T^2\right] \leq 2\exp(-\frac{\epsilon^2}{3}|S|\frac{\|x(t)\|_T^2}{M})$$

which is less than $2\exp(-\frac{\epsilon^2}{3}|S|/d)$, thus completes the proof. $\qquad \square$

A.6 Proof of Lemma 3.10

Lemma 3.10. *For any polynomial $P(t)$ of degree at most d from R to $\mathbb{C}$, for any interval $[S,T]$,*

$$\max_{t \in [S,T]} |P(t)|^2 \le (d+1)^2 \cdot \frac{1}{T-S} \int_S^T |P(t)|^2 \mathrm{d}x.$$

Proof. Let $t^* = \arg\max_{t \in [S,T]} |P(t)|^2$. If $t^* \in (S,T)$, then it is enough to prove that

$$|P(t^*)|^2 \le (d+1)^2 \frac{1}{t^*-S} \int_S^{t^*} |P(x)|^2 \mathrm{d}x \text{ and } |P(t^*)|^2 \le (d+1)^2 \frac{1}{T-t^*} \int_{t^*}^T |P(x)|^2 \mathrm{d}x$$

on the two intervals $[S, t^*]$ and $[t^*, T]$ separately.

Without loss of generality, we will prove the inequality for $S = -1$ and $t^* = T = 1$. We find the minimum $\|P(x)\|_T^2$ assuming $|P(1)|^2 = 1$. Because the first $(d+1)$ Legendre polynomials provide a basis of polynomials of degree at most d and their evaluation $L_n(1) = 1$ for any n, we consider:

$$\min_{\alpha_0, \alpha_1, \cdots, \alpha_d \in \mathbb{C}} \int_{-1}^1 |P(x)|^2 \mathrm{d}x$$

$$\text{s.t.} \quad P(x) = \sum_{i=0}^d \alpha_i L_i(x)$$

$$|P(1)| = |\sum_{i=0}^d \alpha_i| = 1.$$

We simplify the integration of $P(x)^2$ over $[-1,1]$ by the orthogonality of Legendre polynomials:

$$\int_{-1}^1 |P(x)|^2 \mathrm{d}x = \int_{-1}^1 \left(\sum_{i=0}^d \alpha_i L_i(x) \right) \cdot \left(\sum_{j=0}^d \overline{\alpha_j} \overline{L_j}(x) \right) \mathrm{d}x$$

$$= \int_{-1}^1 \sum_{i=0}^d |\alpha_i|^2 L_i(x)^2 + \sum_{i \ne j} \alpha_i \overline{\alpha_j} L_i(x) \overline{L_j}(x) \mathrm{d}x$$

$$= \sum_{i=0}^d |\alpha_i|^2 \frac{2}{2i+1} \text{ by Fact 3.9}$$

Using $\int_{-1}^1 |P(x)|^2 \mathrm{d}x = \sum_{i=0}^d |\alpha_i|^2 \frac{2}{2i+1}$, we simplify the optimization problem to

$$\min_{\alpha_0, \alpha_1, \cdots, \alpha_d \in \mathbb{C}} \sum_{i=0}^d |\alpha_i|^2 \frac{2}{2i+1}$$

$$\text{s.t.} \quad \left| \sum_{i=0}^d \alpha_i \right| = 1$$

From the Cauchy-Schwarz inequality, we have

$$\left| \sum_{i=0}^d \alpha_i \right|^2 \le \left(\sum_{i=0}^d |\alpha_i|^2 \frac{2}{2i+1} \right) \left(\sum_{i=0}^d \frac{2i+1}{2} \right).$$

Therefore $\sum_{i=0}^d |\alpha_i|^2 \frac{2}{2i+1} \ge \frac{2}{(d+1)^2}$ and $|P(1)|^2 \le (d+1)^2 \cdot \frac{1}{2} \int_{-1}^1 |P(x)|^2 \mathrm{d}x$. $\qquad \square$

Algorithm 1 Linear regression algorithms

1: **procedure** LINEARREGRESSION($A, b,$) — Fact B.3
2:    $x' \leftarrow \arg\min_{x} \|Ax - b\|_2$.
3:    **return** x'
4: **end procedure**
5: **procedure** LINEARREGRESSIONW(A, b, w) — Fact B.3
6:    $x' \leftarrow \arg\min_{x} \sum_{i=1}^{d} w_i |(Ax)_i - b_i|^2$.
7:    **return** x'
8: **end procedure**

B Known Facts

This section provides a list of well-known facts existing in literature.

B.1 Inequalities

We state the Hölder's inequality for complex numbers. We will use the corresponding version $p = q = 2$ of Cauchy-Schwarz inequality for complex numbers.

Lemma B.1 (Hölder's inequality). *If S is a measurable subset of $\mathbb{R}^n$ with the Lebesgue measure, and f and g are measurable complex-valued functions on S, then*

$$\int_S |f(x)g(x)| \mathrm{d}x \leq \left(\int_S |f(x)|^p \mathrm{d}x \right)^{\frac{1}{p}} \left(\int_S |g(x)|^q \mathrm{d}x \right)^{\frac{1}{q}}$$

Lemma B.2 (Chernoff Bound [Tar09],[Che52]). *Let $X_1, X_2, \cdots, X_n$ be independent random variables. Assume that $0 \leq X_i \leq 1$ always, for each $i \in [n]$. Let $X = X_1 + X_2 + \cdots + X_n$ and $\mu = \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i]$. Then for any $\epsilon > 0$,*

$$\mathsf{Pr}[X \geq (1+\epsilon)\mu] \leq \exp(-\frac{\epsilon^2}{2+\epsilon}\mu) \ and \ \mathsf{Pr}[X \geq (1-\epsilon)\mu] \leq \exp(-\frac{\epsilon^2}{2}\mu).$$

B.2 Linear regression

Given a linear subspace span$\{\vec{v}_1, \cdots, \vec{v}_d\}$ and n points, we always use ℓ_2-regression to find a vector as the linear combination of $\vec{v}_1, \cdots, \vec{v}_d$ that minimizes the distance of this vector to those n points.

Fact B.3. *Given an $n \times d$ matrix A and an $n \times 1$ column vector b , it takes $O(nd^{\omega-1})$ time to output an x' such that*

$$x' = \arg\min_{x} \|Ax - b\|_2.$$

where ω is the exponent of matrix multiplication[Wil12].

Notice that weighted linear regression can be solved by linear regression solver as a black-box.

Algorithm 2 Multipoint evaluation of a polynomial

1: **procedure** MULTIPOINTEVALUATION$(P, \{t_1, t_2, \cdots, t_d\})$ — Fact B.4
2:   **return** $P(t_1), P(t_2), \cdots, P(t_d)$
3: **end procedure**

B.3 Multipoint evaluation of a polynomial

Given a degree-d polynomial, and n locations. The naive algorithm of computing the evaluations at those n locations takes $O(nd)$. However, the running time can be improved to $O(n \operatorname{poly}(\log d))$ by using this well-known result,

Fact B.4 ([BS12]). *Given a degree-d polynomial $P(t)$, and a set of d locations $\{t_1, t_2, \cdots, t_d\}$. There exists an algorithm that takes $O(d \log^c d)$ time to output the evaluations $\{P(t_1), P(t_2), \cdots, P(t_d)\}$, for some constant c.*

C Analysis of Hash Functions and Filter Functions

C.1 Analysis of filter function $(H(t), \widehat{H}(f))$

We construct the Filter function $(H(t), \widehat{H}(f))$ in this section.

We fix the interval to be $\operatorname{supp}(\operatorname{rect}_1) = [-1/2, 1/2]$ instead of $[0, T]$ for convenience. We first define the filter function $H_1(t)$ which preserves the energy of a k-Fourier-sparse signal x^* on $[-1/2, 1/2]$ to the signal $H_1 \cdot x^*$ on $[-\infty, +\infty]$.

Definition C.1. *Let $s_1 = \Theta(k^4 \log^4 k)$, $\ell = \Omega(k \log k/\delta)$ be a even number, and $s_0 = C_0 s_1 \sqrt{\ell}$ for some constant C_0 that will normalize $H_1(0) = 1$. Recall that $\operatorname{rect}_s(t) = 1$ iff $|t| \leq s/2$ and $\widehat{\operatorname{rect}_s}(f) = \operatorname{sinc}(fs) = \frac{\sin(\pi f s)}{\pi f s}$.*

We define the filter function $H_1(t)$ and its Fourier transform $\widehat{H}_1(f)$ as follows:

$$\widehat{H}_1(f) = s_0 \cdot (\operatorname{rect}_{s_1}(f))^{*\ell} \cdot \operatorname{sinc}(f s_2),$$

$$H_1(t) = s_0 \cdot (\operatorname{sinc}(s_1 t))^{\cdot \ell} * \operatorname{rect}_{s_2}(t)$$

where s_0 is a fixed parameter s.t. $H_1(0) = 1$.

We provide some basic properties about our filter function. Notice that $\operatorname{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$ ($\operatorname{sinc}(0)$ is defined to be 1) has the following properties (shown in Figure 4):

1. $\forall t \in \mathbb{R}, 1 - \frac{(\pi t)^2}{3!} \leq \operatorname{sinc}(t) \leq 1$.

2. $\forall |t| \leq 1.2/\pi, \operatorname{sinc}(t) \leq 1 - \frac{t^2}{8}$.

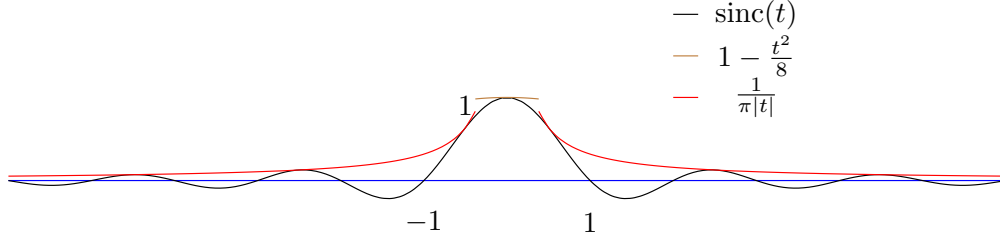3. $\forall |t| > 1.2/\pi, |\operatorname{sinc}(t)| \leq \frac{1}{\pi |t|}$.

Figure 4: The Property of $\mathrm{sinc}(t)$.

Claim C.2. $\int_{-\frac{1.2}{\pi s_1}}^{\frac{1.2}{\pi s_1}} (\mathrm{sinc}(s_1 t))^\ell \mathrm{d}t \approx \frac{1}{s_1\sqrt{\ell}}$.

Proof. We use the above properties for the sinc function to prove the upper bound:

$$
\int_{-\frac{1.2}{\pi s_1}}^{+\frac{1.2}{\pi s_1}} (\mathrm{sinc}(s_1 t))^\ell \mathrm{d}t = \frac{1}{s_1} \int_{-1.2/\pi}^{+1.2/\pi} (\mathrm{sinc}(t))^\ell \mathrm{d}t
$$

$$
= \frac{2}{s_1} \left(\int_0^{\sqrt{8/\ell}} (\mathrm{sinc}\, t)^\ell \mathrm{d}t + \int_{\sqrt{8/\ell}}^{1.2/\pi} (\mathrm{sinc}(t))^\ell \mathrm{d}t \right)
$$

$$
\leq \frac{2}{s_1} \left(\sqrt{8/\ell} + \sum_{i=1}^{\frac{1.2/\pi}{\sqrt{8/\ell}}-1} \int_{i\sqrt{8/\ell}}^{(i+1)\sqrt{8/\ell}} (1 - x^2/8)^\ell \mathrm{d}x \right)
$$

$$
\leq \frac{2}{s_1} \left(\sqrt{8/\ell} + \sum_{i=1}^{\frac{1.2/\pi}{\sqrt{8/\ell}}-1} \sqrt{8/\ell} \cdot 2^{-i^2} \right)
$$

$$
\lesssim \frac{1}{s_1\sqrt{\ell}}.
$$

We prove the lower bound:

$$
\int_{-\frac{1.2}{\pi s_1}}^{\frac{1.2}{\pi s_1}} (\mathrm{sinc}(s_1 t))^\ell \mathrm{d}t = \frac{2}{s_1} \left(\int_0^{\sqrt{8/\ell}} (\mathrm{sinc}\, t)^\ell \mathrm{d}t + \int_{\sqrt{8/\ell}}^{1.2/\pi} (\mathrm{sinc}(t))^\ell \mathrm{d}t \right)
$$

$$
\geq \frac{2}{s_1} \left(\int_0^{\sqrt{8/\ell}} (1 - \frac{\pi^2 t^2}{6})^\ell \mathrm{d}t \right)
$$

$$
\gtrsim \frac{1}{s_1\sqrt{\ell}}.
$$

$\square$

We bound the integration outside $[-\frac{1.2}{\pi s_1}, \frac{1.2}{\pi s_1}]$ from the last property of the sinc function.

Claim C.3. $\int_{\frac{1.2}{\pi s_1}}^{+\infty} (\mathrm{sinc}(s_1 t))^\ell \mathrm{d}t = O(1.2^{-\ell})$.

From these two claims, we have the existence of s_0.

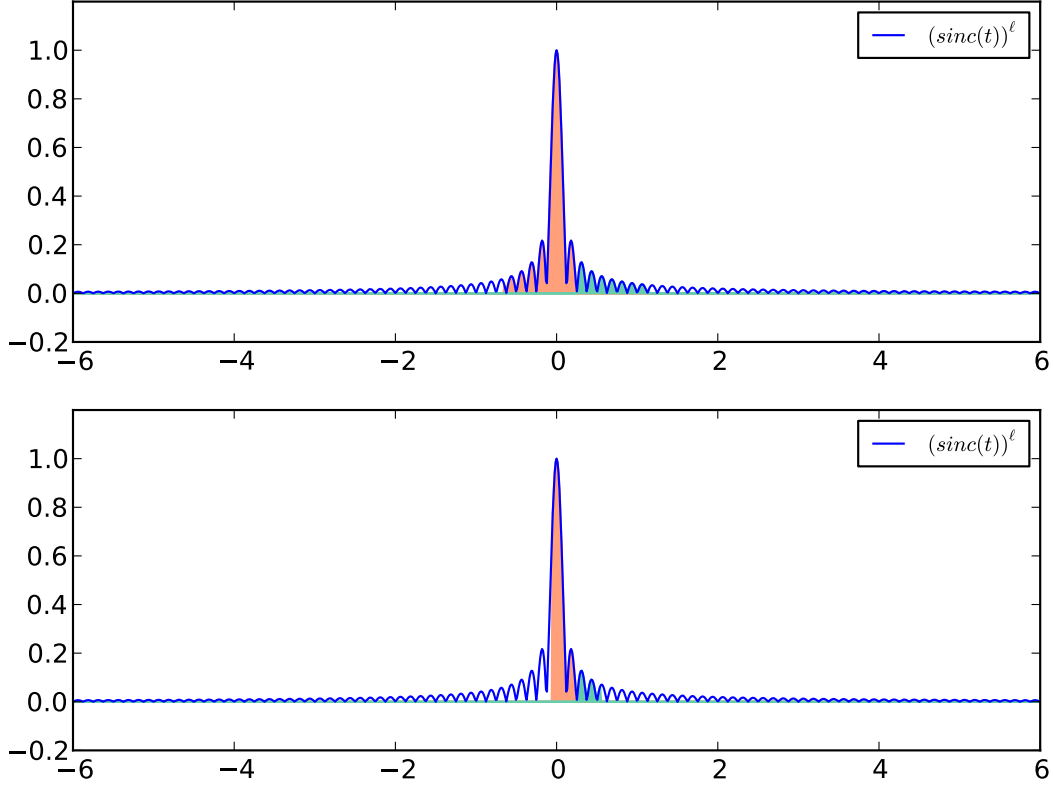**Claim C.4.** *There exists a universal constant C_0 and $s_0 = C_0 s_1 \sqrt{\ell}$ such that $H_1(0) = 1$.*

69

Figure 5: The light red area represents $\int_{(1/2-2/s_1)}^{(1/2-2/s_1)-t} s_0 \cdot \mathrm{sinc}\,(s_1(\tau))^\ell\,\mathrm{d}\tau$ and the light green area represents $\int_{1/2-2/s_1}^{1/2-2/s_1+t} s_0 \cdot \mathrm{sinc}\,(s_1(\tau))^\ell\,\mathrm{d}\tau$.

Proof. Because ℓ is a large even number, $\int_{\mathrm{rect}_{1-2/s_1}} \mathrm{sinc}(s_1 t)^\ell \mathrm{d}t \approx \frac{1}{s_1\sqrt{\ell}}$ from all discussion above. $\square$

We show several useful properties about the Filter functions $\big(H_1(t), \widehat{H_1}(f)\big)$.

Lemma C.5. *Given* s_0, s_1, s_2, ℓ, *where* $\frac{s_2}{2} + \frac{1}{s_1} \leq 1/2$ *and* $s_0 = C_0 s_1\sqrt{\ell}$ *for some constant* C_0. *The filter function* $(H_1(t), \widehat{H}_1(f))[s_0, s_1, s_2, \ell]$ *has the following properties,*

$$\text{Property I}: \quad H_1(t) \in [1 - \frac{s_0}{s_1} \cdot \frac{2\pi^{-\ell}}{\ell - 1}, 1],\ \ if\ |t| \leq \frac{s_2}{2} - \frac{1}{s_1}.$$

$$\text{Property II}: \quad H_1(t) \in [0, 1],\ if\ \frac{s_2}{2} - \frac{1}{s_1} \leq |t| \leq \frac{1}{2}$$

$$\text{Property III}: \quad H_1(t) \leq s_0 s_2 \left((s_1|t| - s_1 + 2)^2 + 1\right)^{-\ell}, \forall |t| > \frac{1}{2}$$

$$\text{Property IV}: \quad \mathrm{supp}(\widehat{H}_1(f)) \subseteq [-\frac{s_1\ell}{2}, \frac{s_1\ell}{2}]$$

Proof of Property I. First, $H_1(0) = 1$ follows by definition of s_0, then we can prove the upper bound for $H_1(t)$ by showing for any $t > 0$, $H_1(0) - H_1(t) > 0$ always holds ,

70

By definition of sinc function, we know that $\text{sinc}(s_1 t)^\ell$ reaches 0 at all the points $\{\frac{1}{s_1} + i\frac{2}{s_1} | i \in \mathbb{N}\}$. By definition of s_1, we know that $\frac{1}{s_1} \ll \frac{1}{2} - \frac{1}{s_1}$. For any $t > 0$,

$$H_1(0) - H_1(t)$$

$$= \int s_0 \cdot \text{sinc}\,(s_1(\tau))^\ell \cdot \text{rect}_{1-2/s_1}(0 - \tau)\mathrm{d}\tau - \int s_0 \cdot \text{sinc}\,(s_1(\tau))^\ell \cdot \text{rect}_{1-2/s_1}(t - \tau)\mathrm{d}\tau$$

$$= \int_{-(1/2-2/s_1)}^{1/2-2/s_1} s_0 \cdot \text{sinc}\,(s_1(\tau))^\ell\,\mathrm{d}\tau - \int_{-(1/2-2/s_1)+t}^{1/2-2/s_1+t} s_0 \cdot \text{sinc}\,(s_1(\tau))^\ell\,\mathrm{d}\tau$$

$$= \int_{-(1/2-2/s_1)}^{-(1/2-2/s_1)+t} s_0 \cdot \text{sinc}\,(s_1(\tau))^\ell\,\mathrm{d}\tau - \int_{1/2-2/s_1}^{1/2-2/s_1+t} s_0 \cdot \text{sinc}\,(s_1(\tau))^\ell\,\mathrm{d}\tau \qquad \text{shown in Figure 5}$$

$$\geq 0,$$

where the last inequality follows by choosing s_1 to be an integer. Thus, we prove an upper bound for $H_1(t)$. Third, we show the lower bound for $H_1(t)$,

$$H_1(t) = \int_{-\infty}^{+\infty} s_0 \cdot \text{sinc}(s_1\tau)^{\cdot\ell}\,\text{rect}_{s_2}(t - \tau)\mathrm{d}\tau$$

$$= \int_{t-\frac{s_2}{2}}^{t+\frac{s_2}{2}} s_0 \cdot \text{sinc}(s_1\tau)^{\cdot\ell}\mathrm{d}\tau$$

$$= 1 - \underbrace{\int_{t+\frac{s_2}{2}}^{+\infty} s_0 \cdot \text{sinc}(s_1\tau)^{\cdot\ell}\mathrm{d}\tau}_{A} - \underbrace{\int_{-\infty}^{t-\frac{s_2}{2}} s_0 \cdot \text{sinc}(s_1\tau)^{\cdot\ell}\mathrm{d}\tau}_{B}$$

Thus, as long as we can upper bound the term A and B, then we will have a lower bound for the $H_1(t)$, for any $|t| \leq \frac{s_2}{2} - \frac{1}{s_1}$.

$$A = \int_{t+\frac{s_2}{2}}^{+\infty} s_0 \cdot \text{sinc}(s_1\tau)^{\cdot\ell}\mathrm{d}\tau$$

$$\leq \int_{\frac{1}{s_1}}^{+\infty} s_0 \cdot \text{sinc}(s_1\tau)^{\cdot\ell}\mathrm{d}\tau$$

$$\leq \int_{\frac{1}{s_1}}^{+\infty} s_0 \cdot (s_1\pi\tau)^{-\ell}\mathrm{d}\tau$$

$$= s_0 \cdot (s_1\pi)^{-\ell}\frac{1}{\ell - 1}(1/s_1)^{-\ell+1}$$

$$= \frac{s_0}{s_1} \cdot (\pi)^{-\ell}\frac{1}{\ell - 1}$$

Similarly, we can bound the term B in the same way. $\qquad\square$

Proof of Property II. In the proof of Property I, we already show that $\forall t$, $H_1(t) \leq 1$. Thus, the upper bound of Property II is also holding. The lower bound follows by both $\text{sinc}(s_1 t)^{\cdot\ell}$ and $\text{rect}_{s_2}(t)$ are always nonnegative, thus the convolution of these two functions has to be nonnegative. $\qquad\square$

Proof of Property III. Let's prove the case when $t > 1$, since $H_1(t)$ is symmetric, then the case $t < -1$ will also hold. By definition of $(H_1(t), \widehat{H}_1(f))$, we have

$$
\begin{aligned}
H_1(t) &= s_0 \cdot \int_{-\infty}^{+\infty} \operatorname{sinc}(s_1(t-\tau))^{\cdot \ell} \operatorname{rect}_{s_2}(\tau) d\tau \\
&= s_0 \cdot \int_{-\frac{s_2}{2}}^{\frac{s_2}{2}} \operatorname{sinc}(s_1(t-\tau))^{\cdot \ell} d\tau \\
&= s_0 \cdot \int_{-\frac{s_2}{2}}^{\frac{s_2}{2}} \operatorname{sinc}(s_1(\tau-t))^{\cdot \ell} d\tau
\end{aligned}
$$

We'd like to choose a middle point τ_0, and then separated the interval into two parts, one is $[-\frac{s_2}{2}, \tau_0]$ and the other is $[\tau_0, \frac{s_2}{2}]$. To choose a reasonable τ_0, we need to use the following simple facts,

$$
\begin{aligned}
(\frac{\sin(x)}{x})^\ell &\le x^{-\ell} \text{ if } x \ge 1.2 \\
(\frac{\sin(x)}{x})^\ell &\le (1 - \frac{x^2}{8})^\ell \text{ if } x < 1.2
\end{aligned}
$$

Thus, $|\pi s_1(\tau_0 - t)| = 1.2$, which implies that $\tau_0^+ = t + \frac{1.2}{\pi s_1}$ or $\tau_0^- = t - \frac{1.2}{\pi s_1}$. By relationship between s_1 and s_2, we know $\tau_0^- > \frac{1}{2} - \frac{1.2}{\pi s_1} > \frac{1}{2} - \frac{1}{s_1} \ge \frac{s_2}{2}$. Thus, we can use the case $x < 1.2$ to upper bound the $H_1(t)$,

$$
\begin{aligned}
H_1&(t) \\
&\le s_0 \cdot s_2 \cdot \max_{\tau \in [-s_2/2, s_2/2]} \operatorname{sinc}(s_1(\tau - t))^{\cdot \ell} \\
&\le s_0 \cdot s_2 \max_{\tau \in [-s_2/2, s_2/2]} (1 - \frac{(s_1 \pi(\tau - t))^2}{8})^\ell \\
&= s_0 \cdot s_2 (1 - \frac{(s_1 \pi(\frac{s_2}{2} - t))^2}{8})^\ell \\
&\le s_0 \cdot s_2 \cdot (e^{-\frac{(s_1 \pi(\frac{s_2}{2} - t))^2}{8}})^\ell && \text{by } 1 - x \le e^{-x} \\
&\le s_0 \cdot s_2 \cdot (e^{(s_1(t - s_2/2))^2})^{-\ell} && \text{by } 1 < \pi^2/8 \\
&\le s_0 \cdot s_2 \cdot (1 + (s_1(t - s_2/2))^2)^{-\ell} && \text{by } 1 + x \le e^x \\
&\le s_0 \cdot (1 + (s_1(t - s_2/2))^2)^{-\ell} && \text{by } s_1 \le 1.
\end{aligned}
$$

Thus, we complete the proof. $\qquad\square$

Proof of Property IV. Because of the support of $\operatorname{rect}_{s_1}(f)$ is s_1, then the support of $(\operatorname{rect}_{s_1}(f))^{*\ell} = s_1 \ell$. Since $\widehat{H}_1(f)$ is defined to be the $(\operatorname{rect}_{s_1}(f))^{*\ell}$ multiplied by $\operatorname{sinc}(fs_2)$, thus $\operatorname{supp}(\widehat{H}_1(f)) \subseteq [-\frac{s_1 \ell}{2}, \frac{s_1 \ell}{2}]$. $\qquad\square$

Definition C.6. *Given any $0 < s_3 < 1$, $0 < \delta < 1$, we define $(H(t), \widehat{H}(f))$ to be the filter function $(H_1(t), \widehat{H}_1(f))$ by doing the following operations*

- *Setting $\ell = \Theta(k \log(k/\delta))$,*

- *Setting $s_2 = 1 - \frac{2}{s_1}$,*

- *Shrinking by a factor s_3 in time domain,*

$$
\begin{aligned}
H(t) &= H_1(t/s_3) && (29) \\
\widehat{H}(f) &= s_3 \widehat{H_1}(s_3 f) && (30)
\end{aligned}
$$

We call the "heavy cluster" around a frequency f_0 to be the support of $\delta_{f_0}(f) * \widehat{H}(f)$ in the frequency domain and use

$$
\Delta_h = |\operatorname{supp}(\widehat{H}(f))| = \frac{s_1 \cdot \ell}{s_3} \tag{31}
$$

to denote the width of the cluster.

We show several useful properties about the Filter functions $\left(H(t), \widehat{H}(f)\right)$.

Lemma 6.6. *Given $s_0, s_1, 0 < s_3 < 1, \ell > 1, 0 < \delta < 1$, where $\ell = \Theta(k \log(k/\delta))$. The filter function $(H(t), \widehat{H}(f))$ has the following properties,*

$$
\begin{aligned}
\text{Property I:} \quad & H(t) \in [1 - \delta, 1], \text{ when } |t| \leq \left(\frac{1}{2} - \frac{2}{s_1}\right)s_3. \\[2mm]
\text{Property II:} \quad & H(t) \in [0, 1], \text{ when } \left(\frac{1}{2} - \frac{2}{s_1}\right)s_3 \leq |t| \leq \frac{1}{2}s_3. \\[2mm]
\text{Property III:} \quad & H(t) \leq s_0 \cdot \left(s_1\left(\frac{|t|}{s_3} - \frac{1}{2}\right) + 2\right)^{-\ell}, \forall |t| > \frac{1}{2}s_3. \\[2mm]
\text{Property IV:} \quad & \operatorname{supp}(\widehat{H}(f)) \subseteq \left[-\frac{s_1\ell}{2s_3}, \frac{s_1\ell}{2s_3}\right].
\end{aligned}
$$

For any exact k-Fourier-sparse signal $x^*(t)$, we shift the interval from $[0, T]$ to $[-1/2, 1/2]$ and consider $x^*(t)$ for $t \in [-1/2, 1/2]$ to be our observation, which is also $x^*(t) \cdot \operatorname{rect}_1(t)$.

$$
\begin{aligned}
\text{Property V:} \quad & \int_{-\infty}^{+\infty} |x^*(t) \cdot H(t) \cdot (1 - \operatorname{rect}_1(t))|^2 \mathrm{d}t < \delta \int_{-\infty}^{+\infty} |x^*(t) \cdot \operatorname{rect}_1(t)|^2 \mathrm{d}t. \\[2mm]
\text{Property VI:} \quad & \int_{-\infty}^{+\infty} |x^*(t) \cdot H(t) \cdot \operatorname{rect}_1(t)|^2 \mathrm{d}t \in [1 - \epsilon, 1] \cdot \int_{-\infty}^{+\infty} |x^*(t) \cdot \operatorname{rect}_1(t)|^2 \mathrm{d}t.
\end{aligned}
$$

for arbitrarily small constant ϵ.

The Property I, II, III and IV follow by filter function $H(t), \widehat{H}(f)$ inheriting $H_1(t), \widehat{H}_1(f)$.

Proof of Property V. $\quad \forall t \notin [-1/2, 1/2]$, we have,

$$
\begin{aligned}
& |x^*(t) \cdot H(t)|^2 \\
\leq\ & |x^*(t)|^2 \cdot |H(t)|^2 \\
\leq\ & |x^*(t)|^2 \cdot (s_1(|t| - 1/2) + 2)^{-\ell} && \text{by Property III of } H_1(t) \\
\leq\ & k^7 \cdot (2kt)^{2.5k} \cdot \int_{-\infty}^{+\infty} |x^*(t) \cdot \operatorname{rect}_1(t)|^2 \mathrm{d}t \cdot \left(s_1\left(\frac{|t|}{s_3} - \frac{1}{2}\right) + 2\right)^{-\ell} && \text{by Lemma 5.5} \\
\leq\ & t^{O(k \log k)} \cdot \int_{-\infty}^{+\infty} |x^*(t) \cdot \operatorname{rect}_1(t)|^2 \mathrm{d}t \cdot \left(s_1\left(\frac{|t|}{s_3} - \frac{1}{2}\right) + 2\right)^{-\ell/2}. && (32)
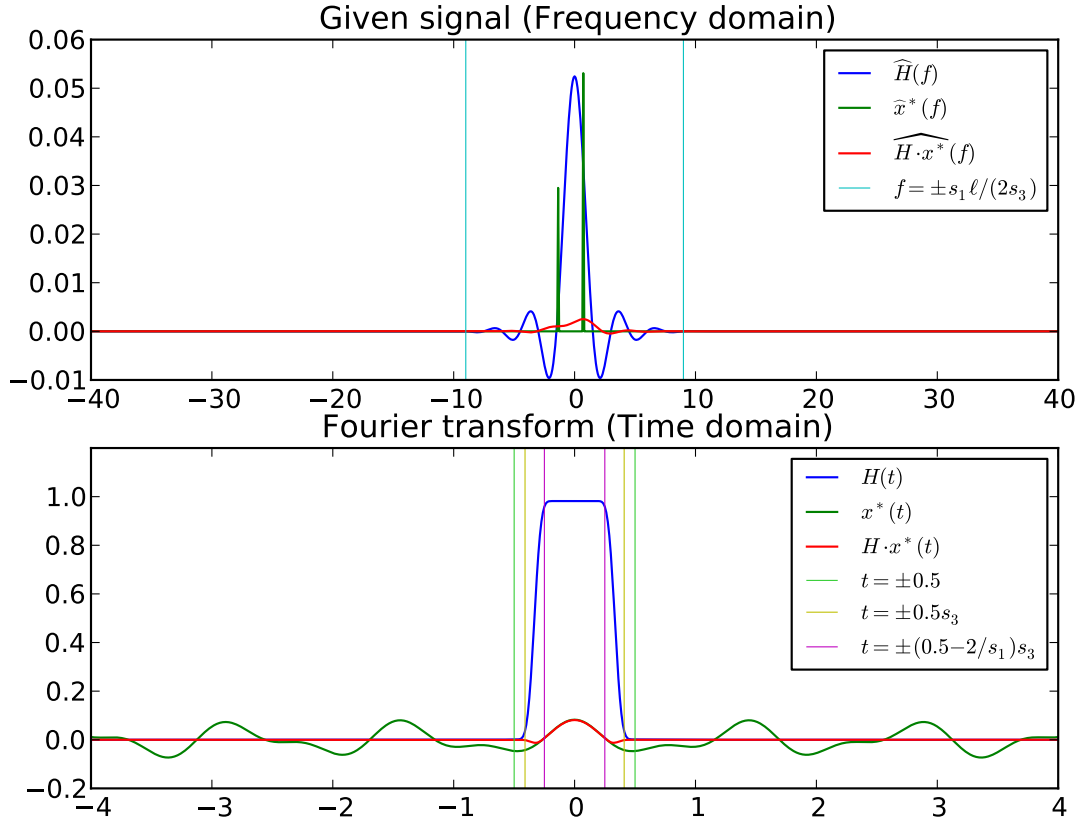\end{aligned}
$$

73

Figure 6: $\widehat{H \cdot x^*}(f)$ and $H \cdot x^*(t)$.

Thus taking the integral finishes the proof because $\ell \gtrsim k \log(k/\delta)$. □

Proof of Property VI. First, because of for any t, $|H_1(t)| \leq 1$, thus we prove the upper bound for LHS,

$$\int_{-\infty}^{+\infty} |x^*(t) \cdot H(t) \cdot \text{rect}_1(t)|^2 \mathrm{d}t \leq \int_{-\infty}^{+\infty} |x^*(t) \cdot 1 \cdot \text{rect}_1(t)|^2 \mathrm{d}t.$$

Second, as mentioned early, we need to prove the general case when $s_3 = 1 - 1/\text{poly}(k)$. Define interval $S = [-s_3(\frac{1}{2} - \frac{1}{s_1}), s_3(\frac{1}{2} - \frac{1}{s_1})]$, by definition, $S \subset [-1/2, 1/2]$. Then define $\overline{S} = [-1/2, 1/2] \backslash S$, which is $[-1/2, -s_3(\frac{1}{2} - \frac{1}{s_1})) \cup (s_3(\frac{1}{2} - \frac{1}{s_1}), 1/2]$. By Property I, we have

$$\int_S |x^*(t) \cdot H(t)|^2 \mathrm{d}t \geq (1 - \delta)^2 \int_S |x^*(t)|^2 \mathrm{d}t \tag{33}$$

74

Then we can show

$$\int_{\overline{S}} |x^*(t)|^2 \mathrm{d}t$$

$$\leq |\overline{S}| \cdot \max_{t \in [-1/2, 1/2]} |x^*(t)|^2$$

$$\leq (1 - s_3(1 - \frac{2}{s_1})) \cdot \widetilde{O}(k^4) \int_{-\frac{1}{2}}^{\frac{1}{2}} |x^*(t)|^2 \mathrm{d}t \qquad \text{by Lemma 5.1}$$

$$\lesssim \int_{-\frac{1}{2}}^{\frac{1}{2}} |x^*(t)|^2 \mathrm{d}t \qquad \text{by } \min(\frac{1}{1 - s_3}, s_1) \geq \widetilde{O}(k^4) \qquad (34)$$

Combining Equations (33) and (34) gives a lower bound for LHS,

$$\int_{-\infty}^{+\infty} |x^*(t) \cdot H(t) \cdot \mathrm{rect}_1(t)|^2 \mathrm{d}t$$

$$\geq \int_{S} |x^*(t) H(t)|^2 \mathrm{d}t$$

$$\geq (1 - 2\delta) \int_{S} |x^*(t)|^2 \mathrm{d}t \qquad \text{by Equation (33)}$$

$$\geq (1 - 2\delta) \int_{S \cup \overline{S}} |x^*(t)|^2 \mathrm{d}t - (1 - 2\delta) \int_{\overline{S}} |x^*(t)|^2 \mathrm{d}t$$

$$\geq (1 - 2\delta) \int_{S \cup \overline{S}} |x^*(t)|^2 \mathrm{d}t - (1 - 2\delta)\epsilon \int_{S \cup \overline{S}} |x^*(t)|^2 \mathrm{d}t \qquad \text{by Equation (34)}$$

$$\geq (1 - 2\delta - \epsilon) \int_{-\frac{1}{2}}^{\frac{1}{2}} |x^*(t)|^2 \mathrm{d}t$$

$$\geq (1 - 2\epsilon) \int_{-\infty}^{+\infty} |x^*(t) \cdot \mathrm{rect}_1(t)|^2 \mathrm{d}t \qquad \text{by } \epsilon \gg \delta$$

$$\square$$

Remark C.7. *To match $(H(t), \widehat{H}(f))$ on $[-1/2, 1/2]$ with signal $x(t)$ on $[0, T]$, we will scale the time domain from $[-1/2, 1/2]$ to $[-T/2, T/2]$ and shift it to $[0, T]$. For example, the rectangle function in Property V and VI will be replaced by $\mathrm{rect}_T(t - T/2)$. For the parameters $s_0, s_1, s_3, \delta, \ell$ in the definition of H, we always treat them as numbers. We assume T has seconds as unit and Δ_h has Hz as unit . For example, in time domain, the Property I becomes that given $T > 0$,*

$$H(t) \in [1 - \delta, 1] \text{ if } |t - \frac{T}{2}| \leq (\frac{1}{2} - \frac{1}{s_1}) s_3 \cdot T$$

In frequency domain, the Property IV becomes

$$\mathrm{supp}(\widehat{H}(f)) \subseteq [-\frac{\Delta_h}{2}, \frac{\Delta_h}{2}], \text{ where } \Delta_h = \frac{s_1 \ell}{s_3 T}. \qquad (35)$$

Lemma C.8. *Let $H(t)$ denote the function defined in Definition C.6. For any $t \in [-\frac{1}{2}, \frac{1}{2}]$, there exists an algorithm that takes $O(s_1 + \ell \log(s_1) + \log(1/\epsilon))$ time to output a value $\widetilde{H}(t)$ such that*

$$(1 - \epsilon)H(t) \leq \widetilde{H}(t) \leq (1 + \epsilon)H(t).$$
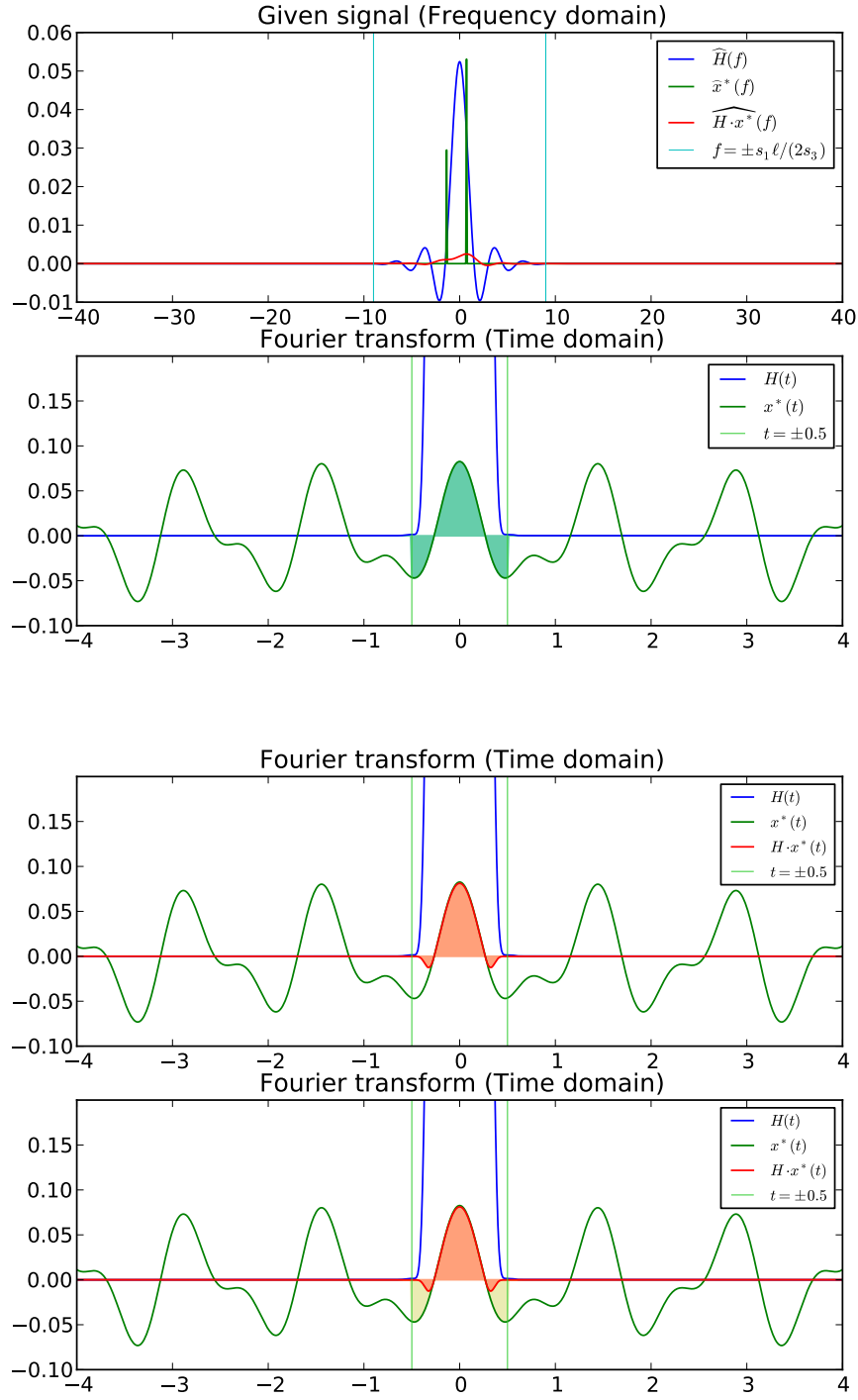
Figure 7: Property VI of filter function $H(t)$, the light green area represents RHS(without scalar) of Property VI of filter H, the light red area represents LHS of Property VI of filter H, the light yellow area represents the difference. Property VI says the light yellow area is only a small constant fraction of the light green area.

Proof. We will show that using a low degree polynomial with sufficiently large degree is able to approximate the sinc function. By definition of filter function,

$$
H(t) = s_0 \cdot \int_{-\infty}^{+\infty} \mathrm{sinc}(s_1\tau)^{\cdot\ell}\, \mathrm{rect}_{s_2}(t-\tau)\mathrm{d}\tau
$$

$$
= s_0 \cdot \int_{t-\frac{s_2}{2}}^{t+\frac{s_2}{2}} \left(\frac{\sin(\pi s_1\tau)}{\pi s_1\tau}\right)^\ell \mathrm{d}\tau
$$

$$
= \frac{s_0}{\pi s_1} \int_{(t-\frac{s_2}{2})\pi s_1}^{(t+\frac{s_2}{2})\pi s_1} \left(\frac{\sin(\tau)}{\tau}\right)^\ell \mathrm{d}\tau
$$

$$
= \frac{s_0}{\pi s_1} \int_{(t-\frac{s_2}{2})\pi s_1}^{(t+\frac{s_2}{2})\pi s_1} \left(\sum_{i=0}^{\infty}(-1)^i \frac{\tau^{2i}}{(2i+1)!}\right)^\ell \mathrm{d}\tau \qquad \text{by Taylor expansion}
$$

$$
= \frac{s_0}{\pi s_1} \int_{(t-\frac{s_2}{2})\pi s_1}^{(t+\frac{s_2}{2})\pi s_1} (A+B)^\ell \mathrm{d}\tau
$$

where the last step follows by setting $A = \sum_{i=0}^{d}(-1)^i \frac{\tau^{2i}}{(2i+1)!}$, and $B = \sum_{i=d+1}^{\infty}(-1)^i \frac{\tau^{2i}}{(2i+1)!}$.

Denote $I^+ = (t+\frac{s_2}{2})\pi s_1$ and $I^- = (t-\frac{s_2}{2})\pi s_1$. Because of $t \in [-1/2, 1/2]$, then $\max(|I^+|, |I^-|) = O(s_1)$. The goal is to show that

$$
(1-\epsilon)\int_{I^-}^{I^+} (A+B)^\ell \mathrm{d}\tau \leq \int_{I^-}^{I^+} A^\ell \mathrm{d}\tau \leq (1+\epsilon)\int_{I^-}^{I^+} (A+B)^\ell \mathrm{d}\tau
$$

Let's prove an upper first,

$$
\int_{I^-}^{I^+} (A+B-B)^\ell \mathrm{d}\tau
$$

$$
= \int_{I^-}^{I^+} (A+B)^\ell \mathrm{d}\tau + \sum_{j=1}^{\ell} \int_{I^-}^{I^+} \binom{\ell}{j}(A+B)^{\ell-j}(-B)^j \mathrm{d}\tau
$$

$$
\leq \int_{I^-}^{I^+} (A+B)^\ell \mathrm{d}\tau + \sum_{j=1}^{\ell} \int_{I^-}^{I^+} \binom{\ell}{j}|A+B|^{\ell-j}|B|^j \mathrm{d}\tau
$$

$$
\leq \int_{I^-}^{I^+} (A+B)^\ell \mathrm{d}\tau + \sum_{j=1}^{\ell} \int_{I^-}^{I^+} \binom{\ell}{j}|A+B|^{\ell-j}\mathrm{d}\tau \cdot \max_{\tau \in [I^-, I^+]} |B|^j
$$

$$
\leq \int_{I^-}^{I^+} (A+B)^\ell \mathrm{d}\tau + \ell 2^\ell \cdot \max_{\tau \in [I^-, I^+]} |B| \qquad \text{by } |H(t)| \leq 1 \text{ and } |B|^j \leq |B|
$$

$$
\leq \int_{I^-}^{I^+} (A+B)^\ell \mathrm{d}\tau + \epsilon \cdot (s_1)^{-\Theta(\ell)} \qquad \text{by Claim C.9}
$$

$$
\leq (1+\epsilon)\int_{I^-}^{I^+} (A+B)^\ell \mathrm{d}\tau \qquad \text{by Claim C.10}
$$

where all the steps by setting $d \gtrsim s_1 + \ell \log(s_1) + \log(1/\epsilon)$. Similarly, we can prove a lower bound. □

Claim C.9. *Let $B(\tau) = \sum_{i=d+1}^{+\infty}(-1)\frac{\tau^{2i}}{(2i+1)!}$, if $d \gtrsim \tau + \ell \log(s_1) + \log(1/\epsilon)$ then $|B(\tau)| \leq \epsilon(1/s_1)^{O(\ell)}$.*

Proof. We first show, for any $i \geq d+1$,

$$\frac{\tau^{2i}}{(2i+1)!}$$
$$\leq \frac{\tau^{2i}}{e((2i+1)/e)^{2i+1}} \qquad\qquad \text{by } e(n/e)^n \leq n!$$
$$\leq 2^{-2i} \qquad\qquad\qquad\qquad\qquad \text{by } i \gtrsim \tau$$
$$\leq \epsilon(1/s_1)^{O(\ell)} \qquad\qquad \text{by } i \gtrsim \ell\log(s_1) + \log(1/\epsilon)$$

Second, we can show that

$$\sum_{i=d+1}^{+\infty} (-1)\frac{\tau^{2i}}{(2i+1)!} \lesssim \frac{\tau^{2(d+1)}}{(2(d+1)+1)!} \leq \epsilon(1/s_1)^{O(\ell)}$$

Thus, we complete the proof. □

Claim C.10. $\min_{t \in [-1/2, 1/2]} |H(t)| \geq (s_1)^{-\Omega(\ell)}$.

Proof. By the property of $H(t)$,

$$\min_{\frac{1}{2}s_3 < |t| \leq \frac{1}{2}} H(t) = \min_{|t| \leq \frac{1}{2}} H(t)$$

Thus, it suffices to prove a lower bound on $H(t)$ for any t such that $\frac{1}{2}s_3 < |t| \leq \frac{1}{2}$. Because of symmetric property, we only need to prove a lower bound for one side. Let's consider $t \in [\frac{1}{2}s_3, 1/2]$,

$$H(t) \geq \frac{s_0}{\pi s_1} \int_{(t-\frac{s_2}{2})\pi s_1}^{(t+\frac{s_2}{2})\pi s_1} \left(\frac{\sin(\tau)}{\tau}\right)^\ell \mathrm{d}\tau$$
$$\geq \frac{s_0}{\pi s_1} \int_{(t+\frac{s_2}{4})\pi s_1}^{(t+\frac{s_2}{2})\pi s_1} \left(\frac{\sin(\tau)}{\tau}\right)^\ell \mathrm{d}\tau$$
$$\geq \frac{s_0}{\pi s_1} \cdot \Theta((t+\frac{s_2}{2})s_1) \cdot \frac{1}{2} \cdot \pi \cdot \Theta((t+\frac{s_2}{2})\pi s_1)^{-\ell}$$
$$\geq (s_1)^{-\Omega(\ell)}$$

□

C.2   Analysis of filter function $(G(t), \widehat{G}(f))$

We construct $(G(t), \widehat{G}(f))$ in a similar way of $(H_1(t), \widehat{H_1}(f))$ by switching the time domain and the frequency domain of $(H_1(t), \widehat{H_1}(f))$ and modify the parameters for the permutation hashing $P_{\sigma,a,b}$.

Definition C.11. *Given $B > 1$, $\delta > 0$, $\alpha > 0$, we construct $G(t), \widehat{G}(f)$ by doing the following operations,*

- $s_2 = \frac{\pi}{2B}$,

- $s_1 = \frac{B}{\alpha\pi}$,
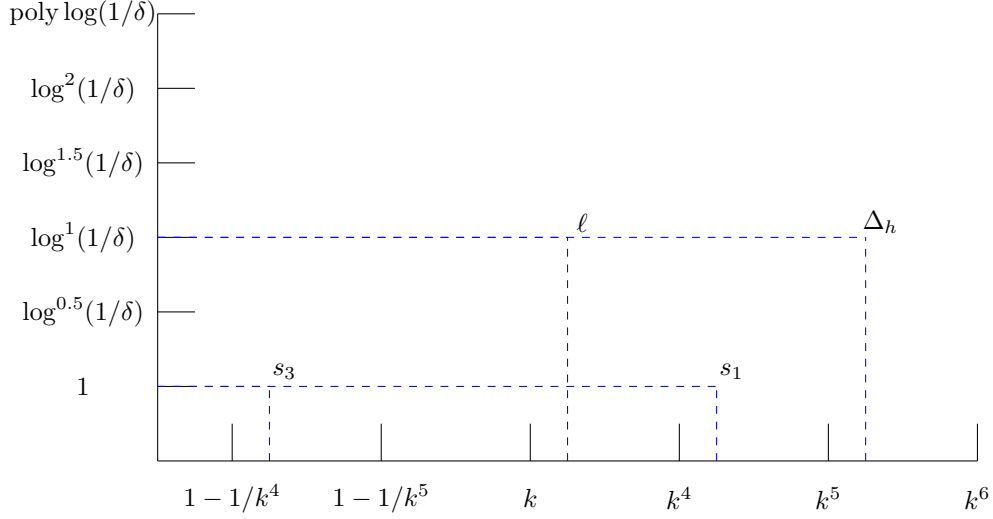
- $\ell = l = \Theta(\log(k/\delta))$.

Figure 8: Parameters for s_1, s_3 and ℓ.

Then $G(t), \widehat{G}(f)$ becomes

$$
\begin{aligned}
G(t) &= b_0 \cdot (\text{rect}_{s_1}(t))^{*l} \cdot \text{sinc}(ts_2) \\
&= b_0 \cdot (\text{rect}_{\frac{B}{(\alpha\pi)}}(t))^{*l} \cdot \text{sinc}(t\frac{\pi}{2B}), \\
\widehat{G}(f) &= b_0 \cdot (\text{sinc}(s_1 f))^{\cdot l} * \text{rect}_{s_2}(f) \\
&= b_0 \cdot (\text{sinc}(\frac{B}{\alpha\pi}f))^{\cdot l} * \text{rect}_{\frac{\pi}{2B}}(f).
\end{aligned}
$$

where the scalar $b_0 = \Theta(s_1\sqrt{l}) = \Theta(B\sqrt{l}/\alpha)$ satisfying $\widehat{G}(0) = 1$.

Lemma 6.7. *Given $B > 1$, $\delta > 0$, $\alpha > 0$, we set $l = \Omega(\log(\delta/k))$. The filter function $(G(t), \widehat{G}(f))[B, \delta, \alpha, l]$ satisfies the following properties,*

$$
\begin{aligned}
\text{Property I}: \quad & \widehat{G}(f) \in [1 - \delta/k, 1], \ \ if \ |f| \leq (1 - \alpha)\frac{2\pi}{2B}. \\
\text{Property II}: \quad & \widehat{G}(f) \in [0, 1], \ \ if \ (1 - \alpha)\frac{2\pi}{2B} \leq |f| \leq \frac{2\pi}{2B}. \\
\text{Property III}: \quad & \widehat{G}(f) \in [-\delta/k, \delta/k], \ \ if \ |f| > \frac{2\pi}{2B}. \\
\text{Property IV}: \quad & \text{supp}(G(t)) \subset [\frac{l}{2} \cdot \frac{-B}{\pi\alpha}, \frac{l}{2} \cdot \frac{B}{\pi\alpha}]. \\
\text{Property V}: \quad & \max_t |G(t)| \lesssim \text{poly}(B, l).
\end{aligned}
$$

Proof. The first five Properties follows from Lemma 6.6 directly.

$\square$

C.3 Parameters setting for filters

One-cluster Recovery. In one-cluster, we donot need filter function $(G(t), \widehat{G}(f))$.

In section C.1, by Equation (34) in the proof of Property VI of filter function $(H(t), \widehat{H}(f))$, we need $\min(\frac{1}{1-s_3}, s_1) \geq \widetilde{O}(k^4)$.

79

In section C.1, by Equation (32) in the proof of Property V of filter function $(H(t), \widehat{H}(f))$, we set $\ell \gtrsim k \log(k/\delta)$.

Δ_h is determined by the parameters of filter $(H(t), \widehat{H}(f))$ in Equation (35): $\Delta_h \approx \frac{s_1 \ell}{s_3 T}$ in section C.1. Combining the setting of s_1, s_3 ℓ, we should set $\Delta_h \geq \widetilde{O}(k^5 \log(1/\delta))/T$.

k-cluster Recovery.   Note that in the k-cluster recovery, we need to use filter function $(G(t), \widehat{G}(f))$. We choose $l = \log(k/\delta)$, $\alpha \approx 1$, $B \approx k$, and $D = l/\alpha$.

By proof of Property II of z in Lemma 7.20 from section 7.6, we need $T(1 - s_3) > \sigma B l$. By the same reason in one-cluster recovery, $1 - s_3 \leq \frac{1}{\widetilde{O}(k^4)}$. Combining $T(1 - s_3) > \sigma B l$ and $1 - s_3 \leq \frac{1}{\widetilde{O}(k^4)}$, we obtain

$$\frac{T}{\widetilde{O}(k^4)} > \sigma B l \tag{36}$$

Because in our algorithm, we will sample σ from $[\frac{1}{B\Delta_h}, \frac{2}{B\Delta_h}]$. Thus, plugging $\sigma = \Theta(\frac{1}{B\Delta_h})$ in Equation (36) we have

$$\frac{T}{\widetilde{O}(k^4)} > \frac{l}{\Delta_h}$$

which implies another lower bound for Δ_h,

$$\Delta_h \geq \widetilde{O}(k^4) l / T$$

Combining the above bound with previous lower bound in one-cluster recovery, we get

$$\Delta_h \geq \widetilde{O}(k^4 \log(1/\delta))/T + \widetilde{O}(k^5 \log(1/\delta))/T = \widetilde{O}(k^5 \log(1/\delta))/T$$

For s_1 and ℓ, we still choose the same setting as before, $s_1 \approx \widetilde{O}(k^4)$ and $\ell \approx O(k \log(k/\delta))$.

C.4   Analysis of HASHTOBINS

In this section, we explain the correctness of Procedure HASHTOBINS in Algorithm 6. Before giving the proof of that algorithm, we show how to connect CFT, DTFT and DFT.

Lemma C.12. *For any signal $W : \mathbb{R} \to \mathbb{C}$, let $A : \mathbb{Z} \to \mathbb{C}$ and $B : [n] \to \mathbb{C}$ be defined as follows:*

$$A[i] = W(i), \forall i \in \mathbb{Z} \text{ and } B[i] = \sum_{j \in \mathbb{Z}} A[i + jn], \forall i \in [n].$$

Then we consider the Fourier transform on $W, A,$ and B:

$$\begin{aligned} \text{CFT} \quad & \widehat{W} : \mathbb{R} \to \mathbb{C}, \\ \text{DTFT} \quad & \widehat{A} : [0, 1] \to \mathbb{C}, \\ \text{DFT} \quad & \widehat{B} : [n] \to \mathbb{C}. \end{aligned}$$

We have:

$$\forall f \in [0, 1), \widehat{A}(f) = \sum_{j \in \mathbb{Z}} \widehat{W}(f + j); \quad \forall i \in [n], \widehat{B}[i] = \sum_{j \in \mathbb{Z}} \widehat{W}(i/n + j).$$

Proof. Recall that $\text{Comb}_s(t) = \sum_{j \in \mathbb{Z}} \delta_{js}(t)$. First, we show $\widehat{A}(f) = \sum_{j \in \mathbb{Z}} e^{2\pi \mathbf{i} jf} A[j]$ equals to $\sum_{j \in \mathbb{Z}} \widehat{W}(f + j)$:

$$
\begin{aligned}
\widehat{A}(f) &= \sum_{j \in \mathbb{Z}} e^{2\pi \mathbf{i} jf} W[j] && \text{by } A[j] = W(j) \\
&= \int_{-\infty}^{+\infty} e^{2\pi \mathbf{i} fj} W(j) \cdot \text{Comb}_1(j) \mathrm{d}j \\
&= \widehat{W \cdot \text{Comb}_1}(f) \\
&= (\widehat{W} * \widehat{\text{Comb}_1})(f) \\
&= \sum_{j \in \mathbb{Z}} \widehat{W}(f + j). && (37)
\end{aligned}
$$

Next, we prove that $\forall i \in [n], \widehat{B}[i] = \widehat{A}(i/n)$,

$$
\begin{aligned}
\widehat{B}[i] &= \sum_{j=1}^{n} B[j] e^{\frac{2\pi \mathbf{i}}{n} ij} && \text{by DFT} \\
&= \sum_{j=1}^{n} (\sum_{k \in \mathbb{Z}} A[j+kn]) e^{\frac{2\pi \mathbf{i}}{n} ij} && \text{by } B[j] = \sum_{k \in \mathbb{Z}} A[j+kn] \\
&= \sum_{j=1}^{n} \sum_{k \in \mathbb{Z}} A[j+kn] e^{\frac{2\pi \mathbf{i}}{n} i(j+kn)} && \text{by } e^{\frac{2\pi \mathbf{i}}{n} \cdot ikn} = 1 \\
&= \sum_{j \in \mathbb{Z}} A[j] e^{2\pi \mathbf{i} j \frac{i}{n}} = \widehat{A}(i/n) && \text{by DTFT.} \quad (38)
\end{aligned}
$$

Combining Equation (38) and Equation (37), we obtain that $\widehat{B}[j] = \widehat{A}(j/n) = \sum_{i \in \mathbb{Z}} \widehat{W}(j/n + i)$ for all $j \in [n]$. □

Claim C.13. *Let $u \in \mathbb{C}^B$ and $V \in \mathbb{C}^{BD}$ such that for any $j \in B$, $u[j] = \sum_{i \in [D]} V[j + (i-1)B]$. Then*

$$
\widehat{u}[j] = \widehat{V}[jD], \forall j \in [B].
$$

Proof. We prove it through the definition of the Fourier transform:

$$
\begin{aligned}
\widehat{V}[jD] &= \sum_{i=1}^{BD} V[i] \cdot e^{\frac{2\pi \mathbf{i}}{BD} \cdot i \cdot (jD)} && \text{by definition of DFT} \\
&= \sum_{i=1}^{B} \sum_{k=1}^{D} V[i+kB] e^{\frac{2\pi \mathbf{i}}{B} \cdot (i+kB) \cdot j} && \text{by replacing } i \text{ by } i+kB \\
&= \sum_{i=1}^{B} e^{\frac{2\pi \mathbf{i}}{B} \cdot j \cdot i} \sum_{k=1}^{D} V[i+(k-1)B] && \text{by } e^{2\pi \mathbf{i} jk} = 1 \\
&= \sum_{i=1}^{B} e^{\frac{2\pi \mathbf{i}}{B} \cdot j \cdot i} u[i] = \widehat{u}[j] && \text{by definition of DFT on } u
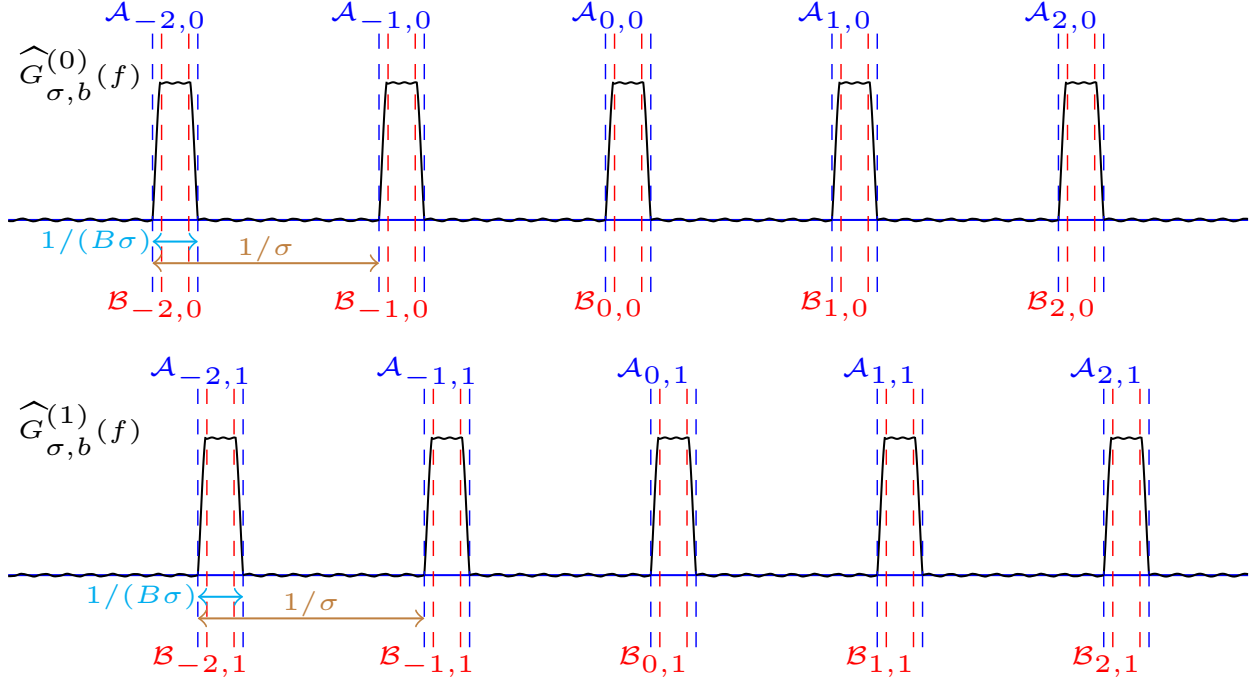\end{aligned}
$$

□

Figure 9: $\widehat{G}_{\sigma,b}^{(j)}(f)$ where the top one is $j = 0$ and the bottom one is $j = 1$, $\mathcal{A}_{i,j} = [\frac{1}{\sigma}(2\pi(i + \frac{j}{B}) - \frac{2\pi}{2B}), \frac{1}{\sigma}(2\pi(i + \frac{j}{B}) + \frac{2\pi}{2B})]$, $\mathcal{B}_{i,j} = [\frac{1}{\sigma}(2\pi(i + \frac{j}{B}) - \frac{2\pi(1-\alpha)}{2B}), \frac{1}{\sigma}(2\pi(i + \frac{j}{B}) + \frac{2\pi(1-\alpha)}{2B})]$

We use Definition 6.1 and Lemma 6.2 to generalize Lemma C.12,

Corollary C.14. *If for all $j \in [n]$, $B[j] = \sum_{i \in \mathbb{Z}} W\big((j + in)\sigma - \sigma a\big)$, then $\forall j \in [n]$,*

$$\widehat{B}[j] = \sum_{i \in \mathbb{Z}} \widehat{W}\left((\frac{j}{n} + i)/\sigma\right) \cdot \frac{1}{\sigma} e^{-2\pi \mathbf{i}(\frac{j}{n} + i)a}.$$

If for all $j \in [n]$, $B[j] = \sum_{i \in \mathbb{Z}} W\big((j + in)\sigma - \sigma a\big)e^{-2\pi \mathbf{i}\sigma b(j + in)}$, then $\forall j \in [n]$,

$$\widehat{B}[j] = \sum_{i \in \mathbb{Z}} \widehat{W}\left((\frac{j}{n} + i)/\sigma + b\right) \cdot \frac{1}{\sigma} e^{-2\pi \mathbf{i}(\frac{j}{n} + i)a - 2\pi \mathbf{i}\sigma ab}.$$

Remark C.15 (Samples of HASHTOBINS). *Procedure* HASHTOBINS *in Algorithm 6 takes BD samples in $x(t)$:*

$$x(\sigma(1 - a)), x(\sigma(2 - a)), \cdots, x(\sigma(BD - a)).$$

To analyze our algorithm, we use filter function $(G(t), \widehat{G}(f))$ and $\mathrm{Comb}_s(t) = \sum_{j \in \mathbb{Z}} \delta_{sj}(t)$ to define the discretization of G.

Definition C.16. *Define the discretization of $G(t)$ and $\widehat{G}(f)$,*

$$
\begin{aligned}
G^{\mathrm{dis}}(t) &= G(t) \cdot \mathrm{Comb}_s(t) \\
\widehat{G}^{\mathrm{dis}}(f) &= \frac{1}{s}(\widehat{G} * \mathrm{Comb}_{1/s})(f) \\
&= (\widehat{G} * \mathrm{Comb}_1)(f)
\end{aligned}
$$

where $|\operatorname{supp}(G(t))| = \frac{lB}{\pi\alpha}$, $D = \frac{l}{\pi\alpha}$, $s = |\operatorname{supp}(G(t))|/(BD) = l/(\pi\alpha D) = 1$.

Definition 6.8. $\forall \sigma > 0, b$ *and* $j \in [B]$. *Define,*

$$
\begin{aligned}
G^{(j)}_{\sigma,b}(t) &= \frac{1}{\sigma}G(t/\sigma)e^{2\pi \mathbf{i} t(j/B - \sigma b)/\sigma} \\
\widehat{G}^{(j)}_{\sigma,b}(f) &= \widehat{G}^{\mathrm{dis}}(\frac{j}{B} - \sigma f - \sigma b) = \sum_{i \in \mathbb{Z}} \widehat{G}(i + \frac{j}{B} - \sigma f - \sigma b)
\end{aligned}
$$

Lemma 6.9. *Let $u \in \mathbb{C}^B$ be the result of* HashToBins *under permutation $P_{\sigma,a,b}$, and let $j \in [B]$. Define*

$$
\widehat{z} = \widehat{x \cdot H} \cdot \widehat{G}^{(j)}_{\sigma,b},
$$

so

$$
z = (x \cdot H) * G^{(j)}_{\sigma,b}.
$$

Let vector $\widehat{u} \in \mathbb{C}^B$ denote the B-dimensional DFT of u, then $\forall j \in [B]$,

$$
\widehat{u}[j] = z_{\sigma a}.
$$

Proof. Recall B is the number of hash bins. $B \cdot D$ is the number of samples in time signal. Let $W(t) = x \cdot H(t)$, define vector $y \in \mathbb{C}^{BD}$, then $\forall j \in [BD]$, define

$$
y[j] = W(\sigma(j - a))e^{2\pi \mathbf{i}\sigma b j}
$$

Recall $G(t)$ denote the $\mathrm{rect}^{*l}_{B/\alpha}(t) \cdot \mathrm{sinc}(t/B)$, then $|\operatorname{supp}(G(t))| = \frac{lB}{\alpha}$. Let vector $G' \in \mathbb{C}^{BD}$ is the discretization of $G(t)$, where $G'[i] = G(i)$. Then, $\forall j \in [B]$,

$$
u[j] = \sum_{i \in [D]} V[j + iB]
$$

where $V[j] = y[j] \cdot G'[j]$ and $G'[j]$ is the value at the jth nonzero point of $G^{\mathrm{dis}}(t)$. Applying Claim C.13 with the definition of $u[j]$ and $V[j + iB]$, gives $\widehat{u}[j] = \widehat{V}[jD], \forall j \in [B]$.

Because of u is the result of HashToBins$(x \cdot H, P_{\sigma,a,b}, G)$ and $|\operatorname{supp}(G(t))| = BD$(choosing $D = l/\alpha$), then

$$
u[j] = \sum_{i \in \mathbb{Z}} W(\sigma(j + iB - a))e^{-2\pi \mathbf{i}\sigma b(j+iB)}G(j + iB)
$$

Then we define $G''(t) = G(t/\sigma + a)e^{-2\pi \mathbf{i}b\sigma(t/\sigma+a)}$ and $Y(t) = W(t) \cdot G''(t)$, then immediately, we have

$$
\widehat{G}''(f) = \sigma\widehat{G}(\sigma(f - b))e^{2\pi \mathbf{i}a\sigma f} \quad \text{and} \quad \widehat{Y}(f) = \widehat{W}(f) * \widehat{G}''(f)
$$

Thus, we can rewrite $u[j]$ in the following sense,

$$u[j]$$

$$= \sum_{i\in\mathbb{Z}} W(\sigma(j+iB-a))e^{-2\pi\mathbf{i}\sigma b(j+iB)}G(j+iB)$$

$$= \sum_{i\in\mathbb{Z}} W(\sigma(j+iB-a))G''(\sigma(j+iB-a)) \qquad\text{by } G''(t)=G(t/\sigma+a)e^{-2\pi\mathbf{i}b\sigma(t/\sigma+a)}$$

$$= \sum_{i\in\mathbb{Z}} Y(\sigma(j+iB-a)) \qquad\text{by } Y(t)=W(t)\cdot G''(t)$$

Then

$$\widehat{u}[j]$$

$$= \sum_{i\in\mathbb{Z}} \widehat{Y}((\tfrac{j}{B}+i)/\sigma)\cdot\frac{1}{\sigma}\cdot e^{-2\pi\mathbf{i}(\frac{j}{B}+i)a} \qquad\text{by Corollary C.14}$$

$$= \sum_{i\in\mathbb{Z}} \int_{-\infty}^{+\infty} \widehat{W}(s)\cdot\widehat{G}''(\frac{j/B+i}{\sigma}-s)\cdot\frac{1}{\sigma}\cdot e^{-2\pi\mathbf{i}\cdot(j/B+i)a}\mathrm{d}s \qquad\text{by } \widehat{Y}(f)=\widehat{W}(f)*\widehat{G}''(f)$$

$$= \sum_{i\in\mathbb{Z}} \int_{-\infty}^{+\infty} \widehat{W}(s)\cdot\widehat{G}(j/B+i-\sigma s-\sigma b)\cdot e^{-2\pi\mathbf{i}\cdot(-\sigma s)a}\mathrm{d}s \qquad\text{by } \widehat{G}''(f)=\sigma\widehat{G}(\sigma(f-b))e^{2\pi\mathbf{i}a\sigma f}$$

$$= \int_{-\infty}^{+\infty} \widehat{W}(s)\cdot\sum_{i\in\mathbb{Z}}\widehat{G}((j/B+i)-\sigma s-\sigma b)e^{2\pi\mathbf{i}a\sigma s}\mathrm{d}s$$

$$= \int_{-\infty}^{+\infty} \widehat{W}(s)\cdot\widehat{G}^{\mathrm{dis}}(\frac{j}{B}-\sigma s-\sigma b)e^{-2\pi\mathbf{i}a\sigma s}\mathrm{d}s \qquad\text{by } \widehat{G}^{\mathrm{dis}}(f)=\sum_{i\in\mathbb{Z}}\widehat{G}(f+i)$$

By definition C.16,

$$\widehat{G}_{\sigma,b}^{(j)}=\widehat{G}^{\mathrm{dis}}(\frac{j}{B}-\sigma s-\sigma b)=\sum_{i\in\mathbb{Z}}\widehat{G}(i+\frac{j}{B}-\sigma s-\sigma b)$$

By definition of $\widehat{z}$, we have

$$\widehat{z}(s)=\widehat{x\cdot H}(s)\cdot\widehat{G}^{(j)}(s)=\widehat{W}(s)\cdot\widehat{G}^{(j)}(s)$$

Then $\widehat{u}[j]$ is the $(a\sigma)^{th}$ inverse Fourier coefficients of $\widehat{z}$, basically,

$$\widehat{u}[j]=z_{a\sigma}=z(a\sigma)$$

Thus, we can conclude first computing vector $u\in\mathbb{C}^B$. Getting vector $\widehat{u}\in\mathbb{C}^B$ by using the Discrete Fourier transform $\widehat{u}=\mathrm{DFT}(u)$. This procedure allows us to sample from time domain to implicitly access the time signal's Fourier transform $\widehat{z}$. If z is one-cluster in frequency domain, then apply one-cluster recovery algorithm.

$\square$

D   Acknowledgments

E   Algorithm

This section lists the pseudocode of our algorithms.

Algorithm 3

1: **procedure** GETEMPIRICAL1ENERGY(z, T, Δ) — Claim 7.11
2:     $R_{\text{est}} \leftarrow (T\Delta)^2$
3:     **for** $i = 1 \rightarrow R_{\text{est}}$ **do**
4:         Choose $\alpha_i \in [0, T]$ uniformly at random
5:         $z_{\text{emp}} \leftarrow z_{\text{emp}} + |z(\alpha_i)|^2$
6:     **end for**
7:     $z_{\text{emp}} \leftarrow \sqrt{z_{\text{emp}}/R_{\text{est}}}$
8:     **return** z_{emp}
9: **end procedure**
10: **procedure** GETLEGAL1SAMPLE$(z, \Delta, T, \beta, z_{\text{emp}})$ — Lemma 7.2
11:     $R_{\text{repeat}} \leftarrow (T\Delta)^3, S_{\text{heavy}} \leftarrow \emptyset$
12:     **for** $i = 1 \rightarrow R_{\text{repeat}}$ **do**
13:         Choose $\alpha_i \in [0, T]$ uniformly at random
14:         **if** $|z(\alpha_i)| \geq 0.5 \cdot z_{\text{emp}}$ **then**
15:             $S_{\text{heavy}} \leftarrow S_{\text{heavy}} \cup i$
16:         **end if**
17:     **end for**
18:     **for** $i \in S_{\text{heavy}}$ **do**
19:         $w(i) \leftarrow |z(\alpha_i)|^2 + |z(\alpha_i + \beta)|^2$
20:     **end for**
21:     $\alpha \leftarrow \alpha_i$ with probability $w(i)/\sum_{j \in S_{\text{heavy}}} w(j)$ for $i \in S_{\text{heavy}}$
22:     **return** α
23: **end procedure**

Algorithm 4

1: **procedure** LOCATE1SIGNAL($z, T, F, \Delta, z_{\text{emp}}$) — Lemma 7.15
2:     Set $t \asymp \log(FT)$, $t' = t/4$, $D_{\max} \asymp \log_{t'}(FT)$, $R_{\text{loc}} \asymp \log_{1/c}(tc)$, $L^{(1)} = 2F$
3:     **for** $i \in [D_{\max}]$ **do**
4:         $l \asymp 2F/(t')^{i-1}\Delta$, $s \asymp c$, $\widehat{\beta} = \frac{ts}{2\Delta l}$
5:         **if** $\widehat{\beta} \gtrsim T/(T\Delta)^{3/2}$ **then**
6:             **break**
7:         **else**
8:             $L^{(i)} \leftarrow$ LOCATE1INNER($z, \Delta, T, \widehat{\beta}, z_{\text{emp}}, L^{(i-1)}$)
9:         **end if**
10:     **end for**
11:     **return** $L^{(i)}$
12: **end procedure**
13: **procedure** LOCATE1INNER($z, \Delta, T, \widehat{\beta}, z_{\text{emp}}, \widetilde{L}$)
14:     Let $v_q \leftarrow 0$ for $q \in [t]$
15:     **while** $r = 1 \rightarrow R_{\text{loc}}$ **do**
16:         Choose $\beta \in [\frac{1}{2}\widehat{\beta}, \widehat{\beta}]$ uniformly at random
17:         $\gamma \leftarrow$ GETLEGAL1SAMPLE($z, \Delta, T, \beta, z_{\text{emp}}$)
18:         **for** $i \in [m]$ **do**
19:             $s_i \in [\beta(\widetilde{L} - \Delta l/2), \beta(\widetilde{L} + \Delta l/2)] \cap \mathbb{Z}_+$, $\theta_i = \frac{1}{2\pi\sigma\beta}(\phi(x(\gamma)/x(\gamma + \beta)) + 2\pi s_i)$
20:             Let θ_i belong to region(q)
21:             Then add a vote to region(q) and its two neighbors, i.e., region($q-1$) and region($q+1$)
22:         **end for**
23:     **end while**
24:     $q_j^* \leftarrow \{q | v_q > \frac{R_{\text{loc}}}{2}\}$
25:     **return** $L \leftarrow$ center of region(q_j^*)
26: **end procedure**
27: **procedure** FREQUENCYRECOVERY1CLUSTER(z, T, F, Δ) — Theorem 7.5
28:     $z_{\text{emp}} \leftarrow$ GETEMPIRICAL1ENERGY(z, T, Δ)
29:     **for** $r = 1 \rightarrow O(k)$ **do**
30:         $L_r \leftarrow$ LOCATE1SIGNAL($z, T, F, \Delta, z_{\text{emp}}$)
31:     **end for**
32:     **return** $L^* \leftarrow \underset{r \in [O(k)]}{\text{median}} L_r$
33: **end procedure**

Algorithm 5 Main algorithm for one-cluster recovery

1: **procedure** CFT1CULSTER(x, H, T, F) — Theorem 8.1
2:     $\widetilde{f}_0 \leftarrow$ FREQUENCYRECOVERY1CLUSTER(x, H, T, F)
3:     $\widetilde{x} \leftarrow$ SIGNALRECOVERY1CLUSTER$(\widetilde{f}_0, \text{poly}(k)\Delta_h)$
4:     **return** $\widetilde{x}$
5: **end procedure**
6: **procedure** GENERATEINTERVALS(d)
7:     $n \leftarrow y_0 \leftarrow i \leftarrow 0, m \leftarrow \Theta(d)$
8:     **while** $y_i \leq 1 - \frac{9}{m^2}$ **do**
9:         $y_{i+1} \leftarrow y_i + \frac{\sqrt{1-y_i^2}}{m}, I_{n+1} \leftarrow [y_i, y_{i+1}], I_{n+2} \leftarrow [-y_{i+1}, -y_i]$
10:        $i \leftarrow i + 1, n \leftarrow n + 2$
11:    **end while**
12:    $I_{n+1} \leftarrow [y_i, 1], I_{n+2} \leftarrow [-y_i, -1], n \leftarrow n + 2$
13:    **return** n, I
14: **end procedure**
15: **procedure** ROBUSTPOLYNOMIALLEARNING(x, d, T) — Theorem 1.2
16:    $(n, I) \leftarrow$ GENERATEINTERVALS(d)
17:    **for** $j = 1 \rightarrow n$ **do**
18:        $w_j \leftarrow |I_j|/2$
19:        Choose t_j from I_j uniformly at random
20:        $z_j \leftarrow x(T \cdot \frac{t_j+1}{2})$
21:    **end for**
22:    $\widetilde{A}_{j,i} \leftarrow t_j^i$, for each $(j, i) \in [n] \times \{0, 1, \cdots, d\}$
23:    $\alpha \leftarrow$ LINEARREGRESSIONW$(\widetilde{A}, \widetilde{b} = z, w)$
24:    $Q(t) \leftarrow \sum_{i=0}^{d} \alpha_i t^i$
25:    **return** $\widetilde{Q}(t) = Q(T \cdot \frac{t+1}{2})$
26: **end procedure**
27: **procedure** ROBUSTPOLYNOMIALLEARNING$^+(x, d, T)$ — Theorem 4.5 — a.k.a. SIGNALRECOVERY1CLUSTER
28:    $R \leftarrow \Theta(d)$
29:    $(n, I) \leftarrow$ GENERATEINTERVALS(d)
30:    $w_j \leftarrow |I_j|/2$, for each $j \in [n]$
31:    **for** $i = 1 \rightarrow R$ **do**
32:        $Q_i \leftarrow$ ROBUSTPOLYNOMIALLEARNING(x, d, T)
33:    **end for**
34:    Choose t_j from I_j uniformly at random, for each $j \in [n]$
35:    **for** $i = 1 \rightarrow R$ **do**
36:        $Q_i(t_1), Q_i(t_2), \cdots, Q_i(t_n) \leftarrow$ MULTIPOINTEVALUATION$(Q_i, \{t_1, t_2, \cdots, t_n\})$
37:    **end for**
38:    $\widetilde{Q}_j \leftarrow \underset{i \in [R]}{\text{median}}\, Q_i(t_j)$, for each $j \in [n]$
39:    $\widetilde{A}_{j,i} \leftarrow t_j^i$, for each $(j, i) \in [n] \times \{0, 1, \cdots, d\}$
40:    $\alpha \leftarrow$ LINEARREGRESSIONW$(\widetilde{A}, \widetilde{b} = \widetilde{Q}, w)$
41:    **return** $Q(t) \leftarrow \sum_{i=0}^{d} \alpha_i t^i$
42: **end procedure**

Algorithm 6

1: **procedure** LOCATEKSIGNAL$(x, H, G, T, \Delta, \sigma, b, z_{\text{emp}})$ — Clain 7.27
2:      Set $t \asymp \log(FT)$, $t' = t/4$, $D_{\max} \asymp \log_{t'}(FT)$, $R_{\text{loc}} \asymp \log_{1/c}(tc)$, $L^{(1)} = 2F$
3:      **for** $i \in [D_{\max}]$ **do**
4:          $\Delta l \asymp 2F/(t')^{i-1}$, $s \asymp c$, $\widehat{\beta} = \frac{ts}{2\sigma \Delta l}$
5:          **if** $\sigma\widehat{\beta} \gtrsim T/(T\Delta)^{3/2}$ **then**
6:              **break**
7:          **else**
8:              $L^{(i)} \leftarrow$ LOCATEKINNER$(x, H, G, T, \Delta, \sigma, b, z_{\text{emp}}\widehat{\beta}, U, L^{(i-1)})$
9:          **end if**
10:      **end for**
11:      **return** $L^{(i)}$
12: **end procedure**
13: **procedure** LOCATEKINNER$(x, H, G, T, \Delta, \sigma, b, z_{\text{emp}}\widehat{\beta}, U, \widetilde{L})$
14:      Let $v_{j,q} \leftarrow 0$ for $(j,q) \in [B] \times [t]$
15:      **for** $r = 1 \to R_{\text{loc}}$ **do**
16:          Choose $\beta \in [\frac{1}{2}\widehat{\beta}, \widehat{\beta}]$ uniformly at random
17:          $\widehat{u}, \widehat{u}' \leftarrow$ GETLEGALKSAMPLE$(x, H, G, T, \Delta, \sigma, \beta, z_{\text{emp}})$
18:          **for** $j \in [B]$ **do**
19:              **for** $i \in [m]$ **do**
20:                  $\theta_{j,i} = \frac{1}{2\pi\sigma\beta}(\phi(\widehat{u}[j]/\widehat{u}'[j]) + 2\pi s_i)$, $s_i \in [\sigma\beta(\widetilde{L}_j - \Delta l/2), \sigma\beta(\widetilde{L}_j + \Delta l/2)] \cap \mathbb{Z}_+$
21:                  $f_{j,i} = \theta_{j,i} + b \pmod{F}$
22:                  suppose $f_{j,i}$ belongs to region(j,q),
23:                  add a vote to both region(j,q) and two neighbors nearby that region, e.g. region$(j, q-1)$ and region$(j, q+1)$
24:              **end for**
25:          **end for**
26:      **end for**
27:      **for** $j \in [B]$ **do**
28:          $q_j^* \leftarrow \{q | v_{j,q} > \frac{R_{\text{loc}}}{2}\}$
29:          $L_j \leftarrow$ center of region(j, q_j^*)
30:      **end for**
31:      **return** L
32: **end procedure**
33: **procedure** HASHTOBINS$(x, H, G, P_{\sigma,a,b})$ — Lemma 6.9
34:      Compute $u[j] = \sum_{i \in D} v[j + iB]$
35:      $\widehat{u} \leftarrow$ FFT(u)
36:      **return** $\widehat{u}$
37: **end procedure**

Algorithm 7

1: **procedure** GETEMPIRICALKENERGY$(x, H, G, T, \Delta, \sigma, b)$ — Claim 7.26
2:      $R_{\text{est}} \leftarrow (T\Delta)^2$
3:      **for** $i = 1 \rightarrow R_{\text{est}}$ **do**
4:          Choose $\alpha \in [0, T]$ uniformly at random
5:          $\widehat{u} \leftarrow$ HASHTOBINS$(x, H, G, P_{\sigma,\alpha,b})$
6:          **for** $j = 1 \rightarrow B$ **do**
7:              $z_{\text{emp}}^j \leftarrow z_{\text{emp}}^j + |\widehat{u}_j|^2$
8:          **end for**
9:      **end for**
10:     **for** $j = 1 \rightarrow B$ **do**
11:         $z_{\text{emp}}^j \leftarrow \sqrt{z_{\text{emp}}^j / R_{\text{est}}}$
12:     **end for**
13:     **return** z_{emp}.
14: **end procedure**
15: **procedure** GETLEGALKSAMPLE$(x, H, G, T, \Delta, \beta, z_{\text{emp}})$ — Lemma 7.25
16:      $R_{\text{repeat}} \leftarrow (T\Delta)^3$.
17:      $S_{\text{heavy}}^j \leftarrow \emptyset, \forall j \in [B]$
18:      **for** $i = 1 \rightarrow R_{\text{repeat}}$ **do**
19:          Choose $\alpha \in [0, T]$ uniformly at random
20:          $\widehat{u}^i \leftarrow$ HASHTOBINS$(x, H, G, P_{\sigma,\alpha,b})$
21:          $\widehat{u}'^i \leftarrow$ HASHTOBINS$(x, H, G, P_{\sigma,\alpha+\beta,b})$
22:          **for** $j = 1 \rightarrow B$ **do**
23:              **if** $|\widehat{u}_j^i| \geq 0.5 \cdot z_{\text{emp}}^j$ **then**
24:                  $S_{\text{heavy},j} \leftarrow S_{\text{heavy}}^j \cup i$
25:              **end if**
26:          **end for**
27:      **end for**
28:      **for** $j = 1 \rightarrow B$ **do**
29:          **for** $i \in S_{\text{heavy}}^j$ **do**
30:              $w(i) \leftarrow |\widehat{u}_j^i|^2 + |\widehat{u}_j'^i|^2$
31:          **end for**
32:          $(\widehat{v}_j, \widehat{v}_j') \leftarrow (\widehat{u}_j^i, \widehat{u}_j'^i)$ with probability $w(i)/\sum_{i' \in S_{\text{heavy}}^j} w(i')$ for $i \in S_{\text{heavy}}^j$
33:      **end for**
34:      **return** $\widehat{v}, \widehat{v}' \in \mathbb{C}^B$
35: **end procedure**
36: **procedure** ONESTAGE(x, H, G, σ, b) — Lemma 7.22
37:      $z_{\text{emp}} \leftarrow$ GETEMPIRICALKENERGY$(x, H, G, T, \Delta, \sigma, b)$
38:      $L \leftarrow$ LOCATEKSIGNAL$(x, H, G, T, \Delta, \sigma, b, z_{\text{emp}})$
39: **end procedure**

Algorithm 8 Main algorithm for k-cluster recovery

1: **procedure** CFTKCLUSTER(x, H, G, T, F)
2:      $\{\widetilde{f}_1, \cdots, \widetilde{f}_l\} \leftarrow$ FREQUENCYRECOVERYKCLUSTER(x, H, G, T, F)
3:      $\widetilde{x} \leftarrow$ SIGNALRECOVERYKCLUSTER$^+(\widetilde{f}_1, \cdots, \widetilde{f}_l, \Delta = \mathrm{poly}(k, \log(1/\delta))/T, T)$
4:      **return** $\widetilde{x}$ as our hypothesis
5: **end procedure**
6: **procedure** FREQUENCYRECOVERYKCLUSTER(x, H, G) — Theorem 2.6
7:      **for** $r \in [R]$ **do**
8:          Choose $\sigma \in [\frac{1}{B\Delta_h}, \frac{2}{B\Delta_h}]$ uniformly at random
9:          Choose $b \in [0, \frac{2\pi\lfloor F/\Delta_h \rfloor\rfloor}{(\sigma B)}]$ uniformly at random
10:          $L_r \leftarrow$ ONESTAGE(x, H, G, σ, b)
11:      **end for**
12:      $L^* \leftarrow$ MERGEDSTAGES$(L_1, L_2, \cdots, L_R)$
13: **end procedure**
14: **procedure** SIGNALRECOVERYKCLUSTER$(\widetilde{f}_1, \cdots, \widetilde{f}_l, \Delta, T)$
15:      $d \leftarrow 5\pi((\Delta T)^{1.5} + k^3 \log k + k \log 1/\delta)$
16:      $m \leftarrow O((kd)^{C_3} \cdot \log^{C_3} d)$ for a constant $C_3 = 5$
17:      **for** $j = 1 \to m$ **do**
18:          Sample t_j from $[0, T]$ uniformly at random
19:          $\widetilde{A}_{j, i_1 \cdot l + i_2} \leftarrow t_j^{i_1} \cdot e^{2\pi \mathbf{i} \widetilde{f}_{i_2} t_j}$ for each $(i_1, i_2) \in \{0, \cdots, d\} \times [l]$
20:          $\widetilde{b}_j \leftarrow x(t_j)$
21:      **end for**
22:      $\alpha \leftarrow$ LINEARREGRESSION$(\widetilde{A}, \widetilde{b})$
23:      **return** $\widetilde{x}(t) \leftarrow \sum_{i_1=0}^{d} \sum_{i_2=1}^{l} \alpha_{i_1 \cdot l + i_2} t^{i_1} \cdot e^{2\pi \mathbf{i} \widetilde{f}_{i_2} t}$
24: **end procedure**
25: **procedure** SIGNALRECOVERYKCLUSTER$^+(\widetilde{f}_1, \cdots, \widetilde{f}_l, \Delta, T)$ — Theorem 9.1
26:      $R \leftarrow \Theta(k)$
27:      $d \leftarrow 5\pi((\Delta T)^{1.5} + k^3 \log k + k \log 1/\delta)$
28:      $m \leftarrow O((kd)^{C_3} \cdot \log^{C_3} d)$ for a constant $C_3 = 5$
29:      **for** $i = 1 \to R$ **do**
30:          $\widetilde{x}_i(t) \leftarrow$ SIGNALRECOVERYKCLUSTER$(\widetilde{f}_1, \cdots, \widetilde{f}_l, \Delta, T)$
31:      **end for**
32:      **for** $j = 1 \to m$ **do**
33:          Sample t_j from $[0, T]$ uniformly at random
34:          $\widetilde{A}_{j, i_1 \cdot l + i_2} \leftarrow t_j^{i_1} \cdot e^{2\pi \mathbf{i} \widetilde{f}_{i_2} t_j}$ for each $(i_1, i_2) \in \{0, \cdots, d\} \times [l]$
35:          $\widetilde{b}_j \leftarrow \underset{i \in [R]}{\mathrm{median}}\ \widetilde{x}_i(t_j)$
36:      **end for**
37:      $\alpha \leftarrow$ LINEARREGRESSION$(\widetilde{A}, \widetilde{b})$
38:      **return** $\widetilde{x}(t) \leftarrow \sum_{i_1=0}^{d} \sum_{i_2=1}^{l} \alpha_{i_1 \cdot l + i_2} t^{i_1} \cdot e^{2\pi \mathbf{i} \widetilde{f}_{i_2} t}$
39: **end procedure**