Quartic Samples Suffice for Fourier Interpolation

Zhao Song* Baocheng Sun† Omri Weinstein‡ Ruizhe Zhang§

Abstract

We study the problem of interpolating a noisy Fourier-sparse signal in the time duration $[0, T]$ from noisy samples in the same range, where the ground truth signal can be any k-Fourier-sparse signal with band-limit $[-F, F]$. Our main result is an efficient Fourier Interpolation algorithm that improves the previous best algorithm by [Chen, Kane, Price, and Song, FOCS 2016] in the following three aspects:

- The sample complexity is improved from $\widetilde{O}(k^{51})$ to $\widetilde{O}(k^4)$.
- The time complexity is improved from $\widetilde{O}(k^{10\omega+40})$ to $\widetilde{O}(k^{4\omega})$.
- The output sparsity is improved from $\widetilde{O}(k^{10})$ to $\widetilde{O}(k^4)$.

Here, ω denotes the exponent of fast matrix multiplication. The state-of-the-art sample complexity of this problem is $\sim k^4$, but was only known to be achieved by an *exponential-time* algorithm. Our algorithm uses the same number of samples but has a polynomial runtime, laying the groundwork for an efficient Fourier Interpolation algorithm.

The centerpiece of our algorithm is a new sufficient condition for the frequency estimation task—a high signal-to-noise (SNR) band condition—which allows for efficient and accurate signal reconstruction. Based on this condition together with a new structural decomposition of Fourier signals (Signal Equivalent Method), we design a cheap algorithm to estimate each "significant" frequency within a narrow range, which is then combined with a signal estimation algorithm into a new Fourier Interpolation framework to reconstruct the ground-truth signal.

*zsong@adobe.com. Adobe Research.
†woafrnraetns@gmail.com. Weizmann Institute of Science.
‡omri@cs.columbia.edu. The Hebrew University and Columbia University.
§ruizhe@utexas.edu. The University of Texas at Austin.

Contents

1 Introduction

Fourier transforms are the backbone of signal processing and engineering, with profound implications to nearly every field of scientific computing and technology. This is primarily due to the discovery of the well-known Fast Fourier Transform (FFT) algorithm [CT65], which is ubiquitous in engineering applications, from image and audio processing to fast integer multiplication and optimization. The classic FFT algorithm of [CT65] computes the *Discrete Fourier Transform* (DFT) of a length-n vector x, where both the time and frequency domains are assumed to be discrete. This algorithm takes $O(n)$ samples in the time domain, and constructs $\widehat{x} = \mathrm{DFT}(x)$ in $O(n\log(n))$ time. The discrete setting of DFT limits its applicability in two main aspects: The first one is that many real-world signals are continuous (analog) by nature; Secondly, many real-world applications (such as image processing) involve signals which are *sparse* in the frequency domain (i.e., $\|\widehat{x}\|_0 = k \ll n$) [ITU92, Wat94, Rab02]. This feature underlies the *compressed sensing* paradigm [CRT06], which leverages sparsity to obtain *sublinear* algorithms for signal reconstruction, with time and sample complexity depending only on the sparsity k. Unfortunately, the continuous case cannot simply be reduced to the discrete case via standard discretization (i.e., using a sliding-window function), as it "smears out" the frequencies and blows up the sparsity, which motivates a more direct approach for the continuous problem [PS15].

The study of Fourier-sparse signals dates back to the work of Prony in 1795 [dP95], who studied the problem of exact recovery of the "ground-truth" signal x in the vanilla *noiseless* setting. By contrast, the realistic setting of reconstruction from *noisy-samples* [PS15] is a different ballgame, and exact recovery is generally impossible [Moi15]. In the *Fourier Interpolation* problem, the ground-truth signal

$$x^*(t) = \sum_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}, \quad v_j \in \mathbb{C}, f_j \in [-F, F]\ \forall j \in [k],$$

is a k-Fourier-sparse signal with bandlimit F. Given noisy access to the ground truth $x(t) = x^*(t) + g(t)$ in limited time duration $t \in [0, T]$ (which means that we need to recover $x^*(t)$ by taking samples from $x(t)$), the goal is to reconstruct a $\widetilde{k}$-Fourier-sparse signal $y(t)$ (i.e., $y(t) = \sum_{j=1}^{\widetilde{k}} \widetilde{v}_j e^{2\pi \mathbf{i} \widetilde{f}_j t}$ for some $\widetilde{v}_j \in \mathbb{C}, \widetilde{f}_j \in [-F, F]$ for all $j \in [\widetilde{k}]$) such that

$$\|y(t) - x^*(t)\|_T^2 \le c(\|g\|_T^2 + \delta\|x^*(t)\|_T^2)$$

holds for some $c = O(1)$, where the T-norm of any function $f : \mathbb{R} \to \mathbb{C}$ is defined as

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

We note that it is not necessary for $y(t)$'s frequencies and magnitudes $(\widetilde{f}_j, \widetilde{v}_j)$ being close to the ground-truth signal $x^*(t)$'s frequencies and magnitudes $(f_{j'}, v_{j'})$.

Prior to this work, the state-of-the-art algorithm for the Fourier interpolation problem was given by [CKPS16], which achieves $\widetilde{O}(k^{51})$ sample complexity, $\widetilde{O}(k^{10\omega+40})$ running time, $\widetilde{O}(k^{10})$ output sparsity, and $c \ge 2000$ approximation ratio. In [SSWZ22], the approximation ratio was improved to $\approx 1 + \sqrt{2}$, but the sample complexity remained large, and runtime remained slow. For calibration, we note that $o(k^4)$ sample complexity for Fourier interpolation is not known to be achievable even with *exponential* decoding time. In this work, we focus on improving the *efficiency* of [CKPS16]'s algorithm across all aspects: (i) runtime, (ii) sample complexity, and (iii) output-sparsity. Our main result is:

References	Samples	Time	Output Sparsity
[CKPS16]	$\widetilde{O}(k^{51})$	$\widetilde{O}(k^{10\omega+40})$	$\widetilde{O}(k^{10})$
[CP19a, SSWZ22]	$\widetilde{O}(k^4)$	$\exp(k^3)$	k
Ours (Theorem 1.1)	$\widetilde{O}(k^4)$	$\widetilde{O}(k^{4\omega})$	$\widetilde{O}(k^4)$

Table 1: Summary of the results. All the algorithms obtain $O(1)$ approximation ratio. We use ω to denote the exponent of matrix multiplication, currently $\omega \approx 2.373$ [Wil12, AW21].

Theorem 1.1 (Main Theorem). *Let $x(t) = x^*(t) + g(t)$, where $x^*(t)$ is k-Fourier-sparse signal with frequencies in $[-F, F]$. Given samples of $x(t)$ over $[0, T]$, there is an algorithm that uses*

$$k^4 \log(FT) \cdot \operatorname{poly}\log(k, 1/\delta, 1/\rho)$$

samples, runs in

$$k^{4\omega} \log(FT) \cdot \operatorname{poly}\log(k, 1/\delta, 1/\rho)$$

time, and outputs a $k^4 \cdot \operatorname{poly}\log(k/\delta)$-Fourier-sparse signal $y(t)$ s.t with probability at least $1-\rho$,

$$\|y(t) - x^*(t)\|_T \lesssim \|g(t)\|_T + \delta \|x^*(t)\|_T.$$

1.1 Related works

Sparse Fourier transform in the discrete setting The Fourier transform $\widehat{x} \in \mathbb{C}^N$ is a vector of length N. The goal of a sparse DFT algorithm is, given a bunch of samples x_i in the time domain and the sparsity parameter k, to output a k-Fourier-sparse signal x' with the ℓ_2/ℓ_2-guarantee

$$\|\widehat{x}' - \widehat{x}\|_2 \lesssim \min_{k\text{-sparse } z} \|z - \widehat{x}\|_2.$$

There are two different lines of work solving the above problem. One line [GMS05, HIKP12a, HIKP12b, IKP14, IK14, Kap16, Kap17] is carefully choosing samples (via hash function) and obtaining sublinear sample complexity and running time. The other line [CT06, RV08, Bou14, HR17, NSW19] is taking *random* samples (via RIP property [CT06] or others) and paying sublinear sample complexity but nearly linear running time.

Sparse Fourier transform in the continuous setting [PS15] defined the sparse Fourier transform in the continuous setting. It shows that as long as the sample duration T is large enough compared to the frequency gap η, then there is a sublinear time algorithm that recovers all the frequencies up to certain precision and further reconstructs the signal. [JLS23] improves and generalize several results in [PS15]. In particular, [PS15] only works for one-dimensional continuous Fourier transform, and [JLS23] generalizes it to d-dimensional Fourier transform. In order to convert the tone estimation guarantee to signal estimation guarantees, [PS15] provides a positive result which shows $T = O(\log^2(k)/\eta)$ is sufficient, and [Moi15] shows a lower bound result where $T = \Omega(1/\eta)$. [Son19] asked an open question about whether this gap can be closed. [JLS23] made positive progress on that problem by providing a new upper bound which is $T = O(\log(k)/\eta)$.

From the negative side, [Moi15] shows that in order to show tone estimation[1], we have to pay a lower bound in sample duration T. In [PS15], it shows that once we have tone estimation, we

[1]Tone refers to a (frequency, coefficient) pair in [PS15]. E.g., (f_i, v_i) is a tone of the signal $x(t) = \sum_{i=1}^k v_i e^{2\pi \mathbf{i} f_i t}$. And tone estimation means estimating each (f_i, v_i) precisely.

can obtain a signal estimation guarantee. Since [PS15] and [Moi15], there is an interesting question about whether we can reconstruct the signal without having a tone estimation guarantee, which is defined as the Fourier interpolation problem. [CKPS16] shows a positive answer to this problem. They provide a polynomial time algorithm to solve this problem. However, both sample complexity and running time in [CKPS16] have a huge polynomial factor in k. The major goal of our work is to significantly improve those polynomial factors.

2 Technical Overview

2.1 High-level approach

The high-level approach of Fourier Interpolation (also Fourier Signal reconstruction) has two steps: frequency estimation and signal estimation (also called signal recovery or Fourier set query). This work mainly contributes to the first frequency estimation step.

Filters and HashToBins The core technique in Fourier sparse recovery and interpolation algorithms is filtering. There are two kinds of filters we are using. The first filter function applied to the signal is $H(t)$ (Figure 1a), which is the bounded band limit approximation of the rectangular window function $\mathrm{rect}_T(t)$. Intuitively, since the time duration is restricted to $[0, T]$, we should view the ground truth signal as $x^*(t) \cdot \mathrm{rect}_T(t)$. However, handling $\widehat{\mathrm{rect}}_T(f)$ is not easy due to its unbounded support in the frequency domain. Therefore, we use $H(t)$ instead, which truncates the frequency domain of $\mathrm{rect}_T(t)$ and makes the analysis much easier.

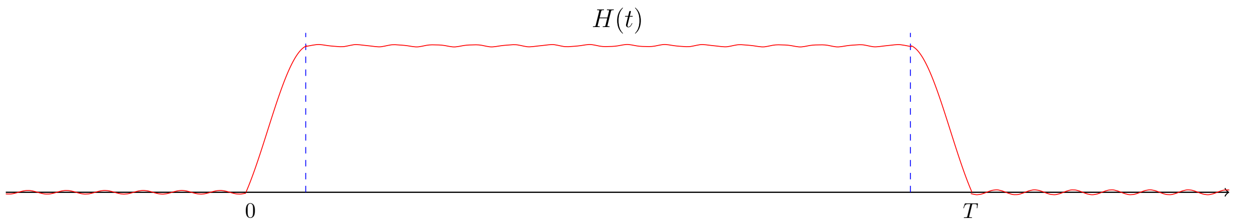
Another kind of filters we use is $G^{(j)}_{\sigma,b}(t)$ (Figure 1b), which "isolates" the signal through the procedure HashToBins and extracts the one-cluster signal in the j-th bin. More specifically, HashToBins divides the frequency domain into $B = O(k)$ bins. We can show that with high probability over the randomized hashing function, each bin contains a single cluster of frequencies. Hence, in the following frequency estimation step, we can just focus on recovering the frequency of a one-cluster *filtered signal* in each bin $j \in [B]$:

$$z_j(t) = (x \cdot H)(t) * G^{(j)}_{\sigma,b}(t).$$

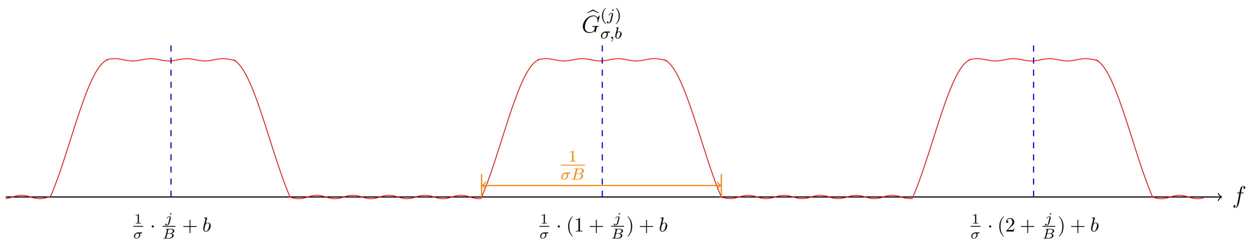
Frequency Estimation This step is the main focus on this work. To estimate the frequencies, our algorithm has two levels. The first level generates significant samples of the *local-test signal*:

$$d_z(t) = z(t)e^{2\pi \mathbf{i} f^* \beta} - z(t + \beta),$$

where $z(t) = z_j(t)$ is the filtered signal in the j-th bin and β is a perturbation parameter. A time point $\alpha \in [0, T]$ is defined to be significant with respect to the target frequency f^* if $|d_z(\alpha)|$ is small. In this case, $z(\alpha + \beta)/z(\alpha)$ is a good approximation of $e^{2\pi \mathbf{i} f^* \beta}$,which further implies the target frequency f^*. The second level is a searching algorithm that iteratively estimates the target frequency f^*. In each iteration, it calls the significant sample generation algorithm and uses the significant sample to narrow the possible range of the target frequency until reaching the desired accuracy. Based on the two-level strategy, we design an efficient, high-accuracy frequency estimation algorithm, improving the time complexity, sample complexity, and the estimation error of the frequency estimation algorithms in previous works [CKPS16, CP19b]. The theorem is stated as follows.



(a) Time domain filter $H(t)$.



(b) Frequency domain filter $\widehat{G}^{(j)}_{\sigma,b}(f)$ for the j-th bin.

Figure 1: Time and frequency domain filters.

Theorem 2.1 (Frequency estimation, Informal version of Theorem L.2)**.** *There exists an algorithm takes $O(k^2 \log(1/\delta) \log(FT))$ samples, runs in $O(k^2 \log(1/\delta) \log^2(FT))$ time, returns a set L of $O(k)$ frequencies such that with probability $1 - \rho_0$, for any "important frequency" f, there exists an $\widetilde{f} \in L$ satisfying*

$$|f - \widetilde{f}| \lesssim \Delta,$$

where $\Delta = k \cdot |\mathrm{supp}(\widehat{H})|$, where $\widehat{H}$ is the Fourier transform of H.

Signal Estimation In signal estimation, a set of estimated frequencies of $y(t)$ has been found, and it remains to interpolate the signal under these frequencies. This is often done via *set-query* techniques [Pri11]. This step is not the focus of this paper, and more discussions can be found in [CKPS16, SSWZ22]. [2]

2.2 Our techniques for frequency estimation

In the frequency estimation part, there are two central questions that need to be answered:

1. *Which frequencies or hashing bins are worth recovering?*
2. *How to recover a key frequency in a bin?*

Our answer to these questions substantially deviates from previous works, as we discuss below.

Answer to the first question: For the first question, [CKPS16]'s answer is the *heavy-cluster condition*, which is defined as follows:

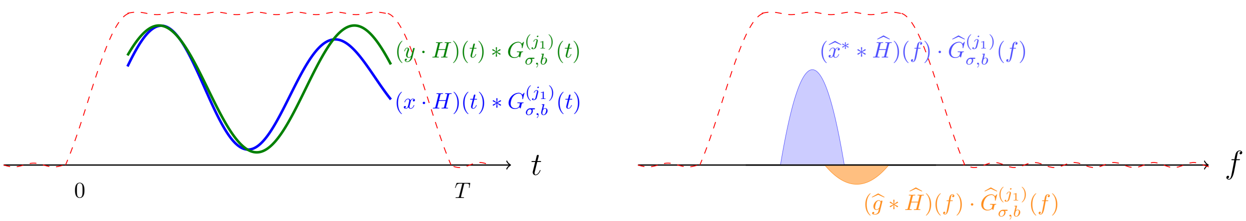
$$[f^* - \Delta, f^* + \Delta] \text{ is heavy if } \int_{f^*-\Delta}^{f^*+\Delta} |\widehat{H \cdot x^*}(f)|^2 \mathrm{d}f \geq T \cdot \mathcal{N}^2/k, \tag{1}$$

[2] We stress that this paper is self-contained and we provide all the technical details of signal estimation in Section M.

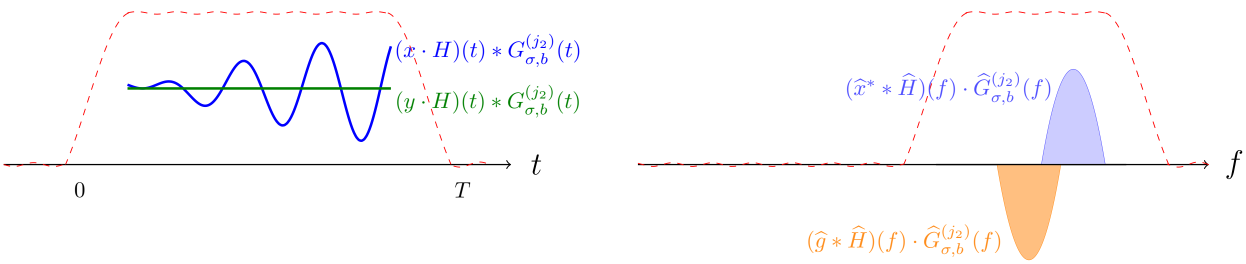
where $\mathcal{N}^2 := \|g\|_T^2 + \delta\|x^*\|_T^2$ represents the noisy-level of $x(t)$. However, only considering the energy of the ground-truth signal is not enough[3]. Indeed, their algorithm only works for "recoverable" clusters, which are defined as:

$$[f^* - \Delta, f^* + \Delta] \text{ is recoverable if } \int_{f^*-\Delta}^{f^*+\Delta} |\widehat{H\cdot x}(f)|^2 \mathrm{d}f \geq T \cdot \mathcal{N}^2/k.$$

The gap between heavy clusters and recoverable clusters is a bottleneck for improving the approximation ratio of the Fourier interpolation algorithms in [CKPS16] to an arbitrarily small constant. This gap also introduces many other technical difficulties in designing more efficient frequency estimation algorithms.



(a) Low-noise band recovery: high-accuracy frequency estimation is needed.



(b) High-noise band recovery: any frequency estimation output is acceptable.

Figure 2: The high SNR band condition. The red curves are the filters. On the left, the blue curves are the filtered noisy observation signal in the time domain, and the green curves are corresponding reconstructed signals. On the right, the light blue regions are the filtered frequencies of the ground-truth signal x^*, and the orange regions are the filtered frequencies of the noise g. Figure 2a shows a high-SNR case, where we can recover a good approximation of x^* in this band. Figure 2b shows an extremely low-SNR case, where g has almost the same energy as x^*, and a trivial signal ($y(t) =$ constant) suffices for the recovery of this band.

To overcome this gap, we introduce a new criterion for the frequency bands that need to be non-trivially reconstructed, which we call the *high signal-to-noise ratio (SNR) band condition*. Formally, we say a hashing bin $j \in [B]$ has a high SNR if the filtered signal $z_j^*(t) = (x^* \cdot H) * G_{\sigma,b}^{(j)}$ satisfies:

$$\|(g \cdot H) * G_{\sigma,b}^{(j)}(t)\|_T^2 \leq c \cdot \|z_j^*(t)\|_T^2, \tag{2}$$

where c is a universal small constant. Our frequency estimation algorithm focuses solely on recovering *heavy frequencies in high-SNR bins*. The intuition behind this condition is as follows: if the

[3] For example, consider the ground-truth signal $x^*(t) = ve^{2\pi \mathbf{i} f^* t} + ve^{2\pi\mathbf{i}(f^*+10\Delta)t}$ and the noise $g(t) = -ve^{2\pi\mathbf{i} f^* t}$. Even if $f^* \pm \Delta$ is a heavy cluster, it is impossible to recover f^* from the observation $x(t) = x^*(t) + g(t)$, since $\widehat{x}(f)$ is zero around f^*.

noise in a band (i.e., $(g \cdot H) * G^{(j)}_{\sigma,b}(t)$) is too large, then we can simply use an all-zero signal as the reconstruction of the filtered signal. We show this new condition brings many advantages for designing more efficient frequency estimation algorithms. In particular, we show that the remaining frequencies in the low-SNR bins are inconsequential for the reconstruction error, and ignoring them in the signal estimation can still achieve the approximation guarantee of Fourier interpolation.[4]

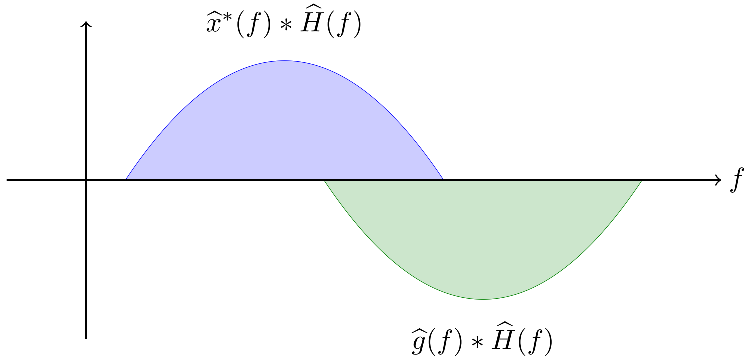


Figure 3: A case that violates our high SNR band assumption but [CKPS16] tries to recover. $\widehat{x}^*(f) * \widehat{H}(f)$ (in blue) is the filtered ground-truth signal, and $\widehat{g}(f) * \widehat{H}(f)$ (in green) is the filtered noise. This signal does not satisfy the high SNR band condition since the noise $\widehat{g}(f) * \widehat{H}(f)$ is too strong. However, the combined signal $(\widehat{x}^*(f) + \widehat{g}(f)) * \widehat{H}(f)$ still satisfies the recoverable-cluster condition since it has enough energy in the frequency domain.

Answer to the second question: As we discussed earlier, the key to answering this question is our novel "significant-samples" generation procedure (which produces samples α such that $|z(\alpha)e^{2\pi \mathbf{i} f^* \beta} - z(\alpha+\beta)|$ is small, where $z(t)$ is the filtered signal and β is a parameter). This is the content of the following lemma.

Lemma 2.2 (Significant Sample Generation, Informal version of Lemma K.4)**.** *There is a Procedure* GenerateSignificantSamples *in Algorithm 2 such that for $\beta \leq O(1/\Delta)$, it takes $\widetilde{O}(k^2)$ samples in $x(t)$ and runs in $\widetilde{O}(k^2)$ time. For each frequency f^* with $j := h_{\sigma,b}(f^*)$, if the j-th bin has "high SNR", and f^* is "heavy", then the output α_j satisfies:*

$$|z_j(\alpha_j+\beta) - z_j(\alpha_j)e^{2\pi \mathbf{i} f^* \beta}|^2 \leq 0.01|z_j(\alpha_j)|^2,$$

with a high constant probability, where $z_j(t) := (x \cdot H) * G^{(j)}_{\sigma,b}(t)$.

We first sketch the proof of Theorem 2.1 using Lemma 2.2. Intuitively, if $z(t)$ is exactly one-sparse, i.e., $z(t) = e^{2\pi \mathbf{i} f^* t}$, then we have $z(t)e^{2\pi \mathbf{i} f^* \beta} - z(t+\beta) = 0$, and $\frac{z(t+\beta)}{z(t)}$ gives the exact value of $e^{2\pi \mathbf{i} f^* \beta}$. More generally, by the guarantee of the significant sample, that ratio can well-approximate $e^{2\pi \mathbf{i} f^* \beta}$, which gives a good estimate of $f^*\beta \mod 1$ in a small constant range:

$$f^* \approx \frac{1}{2\pi\beta}\Big(\arg\Big(\frac{z(\alpha+\beta)}{z(\alpha)}\Big) + 2\pi s\Big)$$

[4]We remark our algorithm never attempts to decide whether a bin satisfies the high-SNR condition or not, but rather assumes all bins are "good". The low-SNR bins may therefore produce totally wrong frequency estimates. However, for accurate signal estimation, we only need to guarantee that all the good frequencies are reconstructed by the frequency estimation algorithm, so even if the output set contains some wrong frequencies, they can be simply ignored.

for some unknown $s \in \mathbb{Z}$. To determine s, we use a search technique to narrow down the potential range of f^* from $[-F, F]$ to $[f^* - \Delta, f^* + \Delta]$. In each iteration, we divide the region of interest into $\mathsf{num} = O(1)$ regions, and repeatedly run the Procedure GenerateSignificantSamples with several different β and pick up the heavy-hitter among all possible regions, which can exponentially increase the success probability of finding the correct interval. Now, we consider the costs of this process. The initial frequency range is $[-F, F]$, and in the last iteration, the frequency range is $[f^* - \Theta(\Delta), f^* + \Theta(\Delta)]$. Thus, we can take the number of iterations to be $O(\log(F/\Delta)) \le O(\log(FT))$. In each iteration, we call Procedure GenerateSignificantSamples for $O(\log\log(F/\Delta)) \le O(\log\log(FT))$ times. Note that each run of Procedure GenerateSignificantSamples can generate significant samples for all B bins. Therefore, by Lemma 2.2, the total time and sample complexity for frequency estimation is $\widetilde{O}(k^2) \cdot O(\log(FT)) \cdot O(\log\log(FT)) = \widetilde{O}(k^2)$.

Algorithm 1 Frequency Estimation Algorithm, Informal version of Algorithm 4, 3, and 5

procedure FrequencyEstimationX$(x, (\sigma, b))$
    for $j \leftarrow [B]$ **do**
        $\widetilde{f}_j \leftarrow$ FrequencyEstimationZ$(x, H, G^{(j)}_{\sigma,b})$ ▷ recover the heavy frequency of $z^{(j)}$
        $L \leftarrow L \cup \{\widetilde{f}_j\}$
    end for
    return L
end procedure
procedure FrequencyEstimationZ$(x, H, G^{(j)}_{\sigma,b})$
    $\mathsf{num} \leftarrow O(1)$ ▷ num-ary search in each iteration
    $D \leftarrow O(\log(\frac{FT}{\Delta}))$ ▷ number of iterations
    $\mathsf{left}_1 \leftarrow -F$, $\mathsf{len}_1 \leftarrow 2F$ ▷ initial searching interval $[\mathsf{left}_1, \mathsf{left}_1 + \mathsf{len}_1]$
    for $d \in [D]$ **do**
        $\mathsf{left}_{d+1} \leftarrow$ ArySearch$(x, H, G^{(j)}_{\sigma,b}, \mathsf{left}_d, \mathsf{len}_d, \mathsf{num})$ ▷ new searching interval's left-end
        $\mathsf{len}_{d+1} \leftarrow 5\frac{\mathsf{len}_d}{\mathsf{num}}$ ▷ new searching interval's length
    end for
    return left_{D+1}
end procedure
procedure ArySearch$(x, H, G^{(j)}_{\sigma,b}, \mathsf{left}_i, \mathsf{len}_i, \mathsf{num})$
    $I_q \leftarrow [\mathsf{left}_d + (q-1)\mathsf{len}_d/\mathsf{num}, \mathsf{left}_d + q\mathsf{len}_d/\mathsf{num}]$ for $q \in [\mathsf{num}]$ ▷ candidate regions
    $v_q \leftarrow 0$ for $q \in [\mathsf{num}]$ ▷ votes counter
    $R \leftarrow O(\log(\log(FT)))$
    for $r = 1 \rightarrow R$ **do**
        Sample $\beta \sim \mathrm{Uniform}([\frac{1}{2}\widehat{\beta}, \widehat{\beta}])$ for $\widehat{\beta} = O(\frac{\mathsf{num}}{\mathsf{len}_d})$ ▷ perturbation
        $z(\alpha + \beta), z(\alpha) \leftarrow$ GenerateSignificantSamples$(x, H, G^{(j)}_{\sigma,b})$ ▷ significant sample
        $\widetilde{S} \leftarrow \frac{1}{2\pi\beta}(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi\mathbb{Z})$ ▷ all possible frequencies
        $\widetilde{I} \leftarrow \{q \in [\mathsf{num}] \mid I_q \cap \widetilde{S} \neq \emptyset\}$ ▷ all possible regions
        $v_q \leftarrow v_q + 1$ for $q \in \widetilde{I}$ ▷ add votes to these regions
    end for
    return $\mathsf{left}_d + (q-1)\mathsf{len}_d/\mathsf{num}$ for any q such that $v_q + v_{q+1} + v_{q+2} \ge R/2$
end procedure

Then, we sketch the proof of Lemma 2.2, which contains three parts:

I. A two-level sampling procedure (see Section 2.2.1).

II. Energy estimation and Signal Equivalent Method (see Section 2.2.2).

III. Time-domain concentration of filtered signals (see Section 2.2.3).

2.2.1 Two-level sampling for significant samples generation

We may assume that in frequency domain, the energy of $\widehat{z}(f)$ is concentrated around f^*:

$$\int_{f^*-\Delta}^{f^*+\Delta} |\widehat{z}(f)|^2 \mathrm{d}f \geq 0.7 \int_{-\infty}^{+\infty} |\widehat{z}(f)|^2 \mathrm{d}f.$$

This is a very natural and necessary assumption for the frequency estimation problem. [5] Then we can show that:

$$\|z(t)e^{2\pi \mathbf{i} f^* \beta} - z(t+\beta)\|_T^2 < \gamma \|z(t)\|_T^2 \tag{3}$$

where $\gamma \in (0, 0.001)$ is a small constant. We show how to find an α such that $|z(\alpha)e^{2\pi \mathbf{i} f^* \beta} - z(\alpha + \beta)|^2 < \gamma |z(\alpha)|^2$. For ease of discussion, we scale the time domain from $[0, T]$ to $[-T, T]$.

The main idea is to use a two-level sampling procedure, which is motivated by [CP19b]. In the first level, we take a set $S = \{t_1, \ldots, t_s\}$ of $O(k \log(k))$ i.i.d. samples from the following distribution:

$$D_z(t) = \begin{cases} c \cdot (1 - |t/T|)^{-1} T^{-1} & \text{if } |t| \leq T(1 - 1/k) \\ c \cdot k T^{-1} & \text{if } |t| \in [T(1 - 1/k), T] \end{cases} \quad \forall t \in U, \tag{4}$$

where $U = \{t_0 \in \mathbb{R} \mid H(t) > 1 - \delta_1 \ \forall t \in [t_0, t_0 + \beta]\}$. Then, we assign weights $w_i := 1/(2T|S|D_z(t_i))$ for each sample $t_i \in S$.

In the second level of the sampling procedure, we sub-sample a t_i from the set S as the output according to the following distribution:

$$D_S(t_i) = \frac{w_i \cdot |z(t_i)|^2}{\sum_{j \in [s]} w_j \cdot |z(t_j)|^2} \quad \forall i \in [s].$$

Now, we explain why the two-level sampling procedure works. By the energy estimation method discussed in Section 2.2.2, we know that:

$$\|z(t)\|_T^2 \approx \|z(t)\|_{S,w}^2 := \sum_{i=1}^{s} w_i \cdot |z(t_i)|^2, \quad \text{and}$$

$$\|z(t)e^{2\pi \mathbf{i} f^* t} - z(t+\beta)\|_T^2 \approx \|z(t)e^{2\pi \mathbf{i} f^* \beta} - z(t+\beta)\|_{S,w}^2 := \sum_{i=1}^{s} w_i \cdot |z(t_i)e^{2\pi \mathbf{i} f^* \beta} - z(t_i + \beta)|^2.$$

The second level of the sampling procedure ensures that

$$\begin{aligned}\mathop{\mathbb{E}}_{t \sim D_S}\left[\frac{|z(t)e^{2\pi \mathbf{i} f^* \beta} - z(t+\beta)|^2}{|z(t)|^2}\right] &= \frac{\sum_{i=1}^{s} w_i |z(t_i)e^{2\pi \mathbf{i} f^* \beta} - z(t_i+\beta)|^2}{\sum_{j=1}^{s} w_j |z(t_j)|^2} \\ &= \frac{\|z(t)e^{2\pi \mathbf{i} f^* \beta} - z(t+\beta)\|_{S,w}^2}{\|z(t)\|_{S,w}^2}.\end{aligned}$$

[5] For the filtered signals that do not satisfy the frequency domain energy concentration assumption, it basically means that they do not contain enough information to recover f^*, and we can just ignore those "useless" clusters.

Hence, we get that

$$\mathop{\mathbb{E}}_{t\sim D_S}\left[\frac{|z(t)e^{2\pi\mathbf{i}f^*\beta}-z(t+\beta)|^2}{|z(t)|^2}\right]\approx\frac{\|z(t)e^{2\pi i f^*\beta}-z(t+\beta)\|_T^2}{\|z(t)\|_T^2}<\gamma,$$

where the last step follows from Eq. (3). Then by Markov's inequality, we get that the sample α generated by the two-level sampling procedure satisfies $|z(\alpha)e^{2\pi i f^*\beta}-z(\alpha+\beta)|^2\lesssim\gamma|z(\alpha)|^2$ with high probability.

The costs of this two-level sampling procedure are calculated as follows. In the first level, we takes $|S|=\widetilde{O}(k)$ samples from $z(t)$, where each sample $z(t_i)=((x\cdot H)*G_{\sigma,b}^{(j)})(t_i)$ can be computed by $|\mathrm{supp}(G_{\sigma,b}^{(j)}(t))|=\widetilde{O}(k)$ samples from $x(t)$ in $\widetilde{O}(k)$ time. Thus, the total time and sample complexity for the first level sampling procedure is $\widetilde{O}(k)\cdot\widetilde{O}(k)=\widetilde{O}(k^2)$. In the second level, we further select one sample from the output of the first level, which can be done in $\widetilde{O}(|S|)=\widetilde{O}(k)$ times and does not need any new sample.

We further discuss how large β we can choose in the sampling procedure since it controls the estimation accuracy of f^*.[6] We note that the range of β is determined by Eq. (3), which is an underlying assumption of our sampling procedure. To satisfy this inequality, we need to guarantee that $|e^{2\pi\mathbf{i}f^*\beta}-e^{2\pi\mathbf{i}f\beta}|\le\gamma$ for any $f\in f^*\pm\Delta$, which implies that $\beta\le O(\gamma/\Delta)$. For comparison, the upper bound of β in [CKPS16] is only $O(\gamma/(\Delta\sqrt{\Delta T}))$ due to a stronger accuracy requirement there.[7]

2.2.2 Energy estimation and Signal Equivalent Method

In this section, we show that the sampling and reweighing method we use in the significant sample generation procedure can accurately estimate the energy of $z(t)$ and $z(t)e^{2\pi\mathbf{i}f^*t}-z(t+\beta)$ with a sample complexity almost reaching the information-theoretic limit.

Lemma 2.3 (Informal version of Lemma K.1 and Lemma K.2)**.** *Suppose f^* is a heavy frequency hashed to the j-th bin which satisfies the high SNR condition. Let $z^*(t)=(x^*\cdot H)*G_{\sigma,b}^{(j)}$ and $z(t)=(x\cdot H)*G_{\sigma,b}^{(j)}$. Let $U\subseteq[0,T]$ be an interval. Let $S=\{t_1,\dots,t_s\}$ be a set of $O(k\log(k))$ i.i.d. samples from the distribution D defined by Eq. (4) with weights $w_i=1/(TsD(t_i))$. Then, with probability at least 0.8,*

$$\|z(t)\|_{S,w}^2\gtrsim\|z^*(t)\|_U^2\quad\text{and}\quad\|z(t)e^{2\pi\mathbf{i}f^*t}-z(t+\beta)\|_{S,w}^2\lesssim\|z^*(t)\|_U^2,$$

where $\|z(t)\|_U^2=(1/|U|)\cdot\int_U|z(t)|^2\mathrm{d}t$.

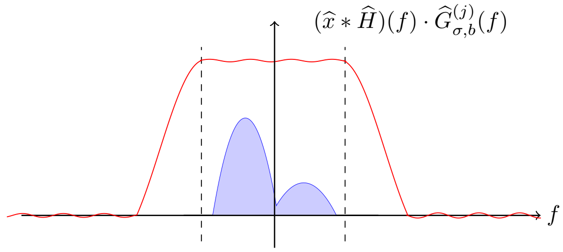
To prove Lemma 2.3, we develop a *Signal Equivalent Method.* Below, we sketch the proof of the first half of Lemma 2.3 on the energy estimation for $z(t)$. The second half follows similar ideas.

Energy estimation is also used in prior works [CKPS16, CP19a, CP19b, SSWZ22], where a key component is the following energy bound for the interested function family $\mathcal{F}$:

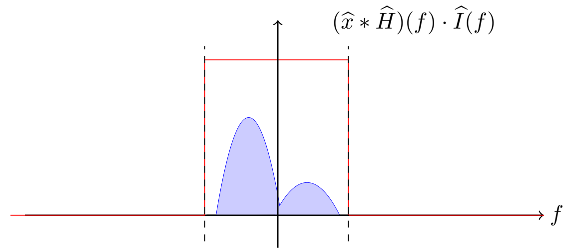
$$\sup_{f\in\mathcal{F}}\sup_{t\in[0,T]}\frac{|f(t)|^2}{\|f(t)\|_T^2}.$$

[6] By comparing $z(t+\beta)$ and $z(t)$, we get an estimate of $f^*\beta$ within some error $\pm b$, which implies an estimate of f^* within an error $\pm b/\beta$. Hence, larger β gives a higher accuracy of the frequency estimation.

[7] [CKPS16] give an ℓ_1-norm error guarantee in the frequency domain, i.e., $\int_{f^*-\Delta}^{f^*+\Delta}|e^{2\pi\mathbf{i}f^*\beta}-e^{2\pi\mathbf{i}f\beta}|\mathrm{d}f$ is small. To obtain an ℓ_2-norm guarantee (like Eq. (3)), they need to apply Cauchy-Schwarz inequality, which results in an extra $\sqrt{\Delta T}$ factor in their upper bound of β.



(a) Signal with non-ideal filter $\widehat{G}^{(j)}_{\sigma,b}(f)$.



(b) Signal with ideal filter $\widehat{I}(f)$.

Figure 4: The Signal Equivalent Method. This figure demonstrates that $(\widehat{x} * \widehat{H})(f) \cdot \widehat{G}^{(j)}_{\sigma,b}(f)$ (left) can be approximated by $(\widehat{x} * \widehat{H})(f) \cdot \widehat{I}(f)$ (right). $\widehat{I}(f)$ (red curve on the right) is the ideal filter that approximate $\widehat{G}^{(j)}_{\sigma,b}(f)$ (red curve on the left).

However, this approach is unlikely to work directly for our filtered signal $z(t)$ since it depends on the randomized hashing function. And under some hashing parameter (σ, b), there always exists some signal $x(t)$ such that $z(t) = (x \cdot H)(t) * G^{(j)}_{\sigma,b}(t)$ is in ill-condition (e.g., the frequencies are not well-isolated, or large offset events happen). As a result, bounding $\frac{|z(t)|^2}{\|z(t)\|_T^2}$ for all $z(t)$ of the form $(x \cdot H) * G^{(j)}_{\sigma,b}(t)$ by a small number is not easy. We bypass the issue by proving an energy bound only for those $z(t)$ under some well-hashed conditions (e.g. frequency is isolated and do not have a large offset), and showing that such a "refined energy bound" is still sufficient to derive the sample complexity of our algorithm.

The motivation of the Signal Equivalent Method comes from the special structure of $z(t) = (x \cdot H)(t) * G^{(j)}_{\sigma,b}(t)$ in the frequency domain. Notices that the observed signal $x(t)$'s Fourier transform $\widehat{x}(f)$ only contains some spikes (assuming small noise). By convolution with $\widehat{H}(f)$ (which corresponds to multiplying by $H(t)$ in the time domain), $(\widehat{x} * \widehat{H})(f)$ fattens the spikes in the frequency domain (and by Parseval's theorem, the area of the signal in frequency domain equals to its energy). Then, convolution with $G^{(j)}_{\sigma,b}(t)$ "zooms-in" to a narrow band around a single frequency. This construction of $z(t)$ motivates us to build a new signal $\overline{z}(t) = (x \cdot H)(t) * I(t)$, where $I(t)$ is a filter function such that $\widehat{I}(f) = 1$ when $G^{(j)}_{\sigma,b}(t) > 1/2$, and $\widehat{I}(f) = 0$ otherwise. To analyze the equivalent signal $\overline{z}(t)$, we improve the analysis of the filter $H(t)$ in [CP19b] and give a tighter bound on its value in a sub-interval of $[0, T]$. Then, we show that the equivalent signal $\overline{z}(t)$ is *almost equivalent* to $z(t)$ under some "good conditions" (i.e., the frequency is isolated and no large offset). We also prove that the ideal filter has several useful properties that can mush simplify the analysis (e.g., the function $I(t)$ is randomized, and with high probability, $I(t)$ commutes with $H(t)$).

By the Signal Equivalent Method, we can first prove an energy bound for the equivalent signal $\overline{z}(t)$, which follows from the Fourier-sparse signals' energy bounds (see Section B). Then, it remains to show that the equivalent signals' energy bound can approximate the original filtered signal $z(t)$'s energy bound. We find that the approximation error comes from two sources: the observation noise $g(t)$ and the approximation error $\overline{z}(t) - z(t)$. The first part of the error is small due to the high SNR band condition (Eq. (2)). And the second part of the error is mitigated by the tail-bound for $G^{(j)}_{\sigma,b}(t)$ and the heavy-cluster condition (Eq. (1)). More specifically, the HashToBins procedure and the filter $G^{(j)}_{\sigma,b}(t)$ can bring some interference noise from other bins to $z(t)$, which is perfectly eliminated by the ideal filter $I(t)$ in the equivalent signal $\overline{z}(t)$. Hence, we need to bound this part of noise when we transfer back from the equivalent signal to the true filtered signal. The tail bound

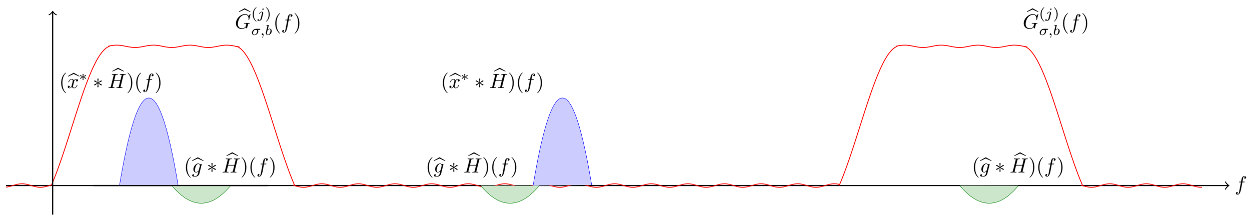


Figure 5: An illustration of a filtered noisy signal. $\widehat{G}^{(j)}_{\sigma,b}$ (the red curve) is the HashToBins filter for the j-th bin. The noise in the filtered signal comes from two parts: one is $\widehat{g}(f) * \widehat{H}(f)$ (the green signal), and another is the interference by signals outside the bin (the blue and green signals in the middle).

of $G^{(j)}_{\sigma,b}(t)$ ensures that adding small interference noise with frequencies far away from the center of the cluster will not drastically affect $z(t)$. However, by this argument, we can only bound the distance between $\overline{z}(t)$ and $z(t)$ by $\|x^*(t)\|_T$, which can be much larger than $\|z(t)\|_T$. Hence, we need to use the heavy-cluster assumption to ensure that $\|x^*(t)\|_T \lesssim \|z(t)\|_T$. Using these error-control techniques, we can prove that an energy bound for $\overline{z}(t)$ implies an energy bound for $z(t)$.

We give a comparison between ours and previous approaches for proving the energy estimation guarantee. [CKPS16] considers $z(t)$ as a generic signal that satisfies the time and frequency domains concentration properties[8]. We exploit "finer" structure of $z(t)$ and obtain a stronger energy bound and reduce the number of samples required in norm preserving. [CP19b] also proves a similar property (but only for $(x \cdot H)(t)$). However, they assume that all the frequencies of $x^*(t)$ are contained in a small interval, making the task much easier. Our filtered signal $z(t)$ does not satisfy this condition due to the interference noise caused by the HashToBins procedure.

2.2.3 Time-domain concentration of filtered signals

The proof of Lemma 2.2 relies on an underlying assumption: the most of the energy of the filtered signal is contained in the observation window $[0, T]$. That is, we need the following lemma:

Lemma 2.4 (Informal version of Lemma G.2)**.** *Let $j \in [B]$ be a bin that contains a heavy frequency. Let $z(t) = (x^* \cdot H) * G^{(j)}_{\sigma,b}$ be the filtered signal. Then, we have*

$$\int_{-\infty}^{+\infty} |z(t)|^2 \mathrm{d}t \leq 1.35 \int_0^T |z(t)|^2 \mathrm{d}t.$$

A similar concentration property is also proved in [CKPS16], using a very strict requirement on the $H(t)$ filter that it decays at an exponential rate near the boundary. More specifically, they require that $H(t)$ is exponentially small not only outside the time duration $[0, T]$, but also in the shrinking boundary $[0, T/\mathrm{poly}(k)] \cup [T - T/\mathrm{poly}(k), T]$. The additional constraint allows them to show that $x(t)$'s energy near the boundary cannot "pass" the H filter, and the energy concentration of $z(t)$ easily follows. However, it also results in a large support of H in the frequency domain, which leads to a large error in the frequency estimation, and further causes large output sparsity and time/sample complexity of their Fourier interpolation algorithm.

We resolve this issue by changing the filter function to the one defined in [CP19b], which has much smaller support and thus saves time and sample complexities. However, it is exponentially

[8] It means that most of the energy of $z(t)$ (i.e., $\|z(t)\|_2$) lies in $[0, T]$ and most of the energy of $\widehat{z}(f)$ lies in a $\mathrm{poly}(k)/T$ length interval in frequency domain.

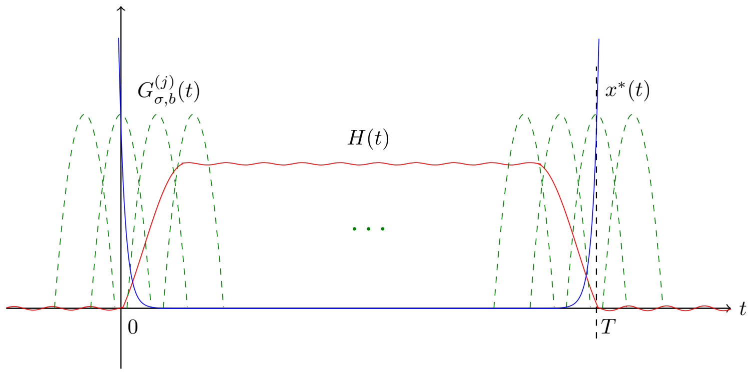


Figure 6: A bad case that may break the time domain concentration of $z(t) = (x \cdot H)(t) * G_{\sigma,b}^{(j)}(t)$ in $[0, T]$ when the filter decay slowly. $x^*(t)$ (in blue) is a k-Fourier-sparse signal. $G_{\sigma,b}^{(j)}$ (in green dashed) is the filter of frequency domain. $H(t)$ (in red) is the filter in time domain. On the one hand, since $x^*(t)$ is very small in $[T/\mathrm{poly}(k), T(1-/\mathrm{poly}(k))]$, and $H(t)$ is very small in $[0, T/\mathrm{poly}(k)] \cup [T(1-/\mathrm{poly}(k)), T]$, the filtered signal has very small energy within $[0, T]$. On the other hand, since the convolution with $G_{\sigma,b}^{(j)}(t)$ can bring some energy of $x^*(t)$ passing the boundary of $[0, T]$, and the signal $x^*(t)$ could be very large outside $[0, T]$, $(x \cdot H)(t) * G_{\sigma,b}^{(j)}(t)$ may contain very large energy in $\mathbb{R} \backslash [0, T]$. In this case, $\|z(t)\|_{L_2}^2 \gg \|z(t)\|_T^2$.

small outside $[0, T]$, but only *polynomially* small near the boundary. To prove Lemma 2.4, we use our Signal Equivalent Method again. We construct an equivalent signal $\overline{z}(t) = (x^* \cdot H) * I(t) = (x^* * I) \cdot H(t)$, where $x^*(t) * I(t)$ is a Fourier-sparse signal. Then, by some finer analysis on the $H(t)$ filter (see Lemma E.9), we can show that most of the energy of $x^*(t) * I(t)$ is preserved in $[0, T]$, i.e.,

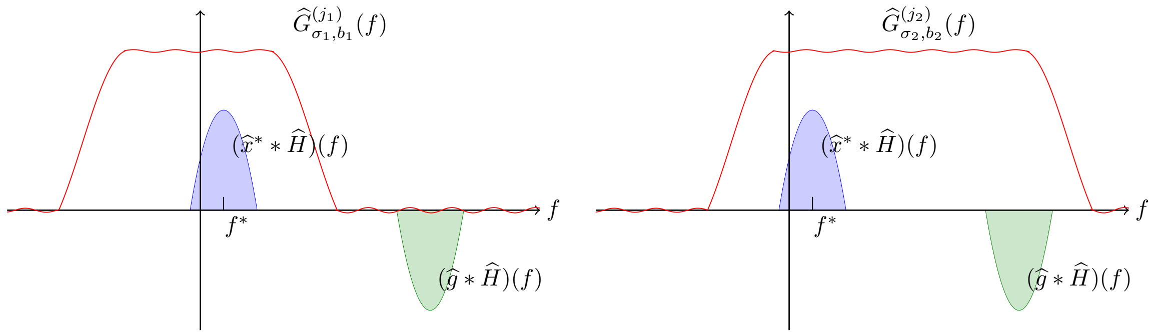
$$\int_{-\infty}^{+\infty} |\overline{z}(t)|^2 \mathrm{d}t \leq 1.1 \int_0^T |\overline{z}(t)|^2 \mathrm{d}t.$$

Finally, by the approximation guarantee of Signal Equivalent Method, we get that the energy concentration of $\overline{z}(t)$ implies the energy concentration of $z(t)$.

2.3 Our techniques for Fourier Interpolation

In this section, we discuss how to obtain a Fourier interpolation algorithm with improved efficiency and output sparsity (Theorem 1.1) based on our frequency estimation algorithm (Theorem 2.1).

We first remark that simply applying the original framework of Fourier Interpolation (e.g., [CKPS16, SSWZ22]) and combining with an existing signal estimation algorithm is still not enough to improve the previous algorithm, since the frequency estimation algorithm has a low success probability and we cannot apply the success probability boosting trick in [CKPS16] to increase it to $1-\rho$. More specifically, [CKPS16] first boosts the success probability of their frequency estimation algorithm by their merge-stage algorithm (which runs the frequency estimation algorithm for $R = \log(1/\rho)$ times, sorts all recovered frequencies, and picks every $R/2$-th entry of the sorted list), and then runs the signal estimation algorithm. It does not work here because our high SNR band condition makes frequency estimation and signal estimation "entangled". More specifically, whether a frequency is contained in a high SNR bin (which *needs* to be recovered) or not depends on the randomized hash function. However, if the outputs of multiple runs of the frequency estimation



(a) High SNR, where $j_1 = h_{\sigma_1,b_1}(f^*)$

(b) Low SNR, where $j_2 = h_{\sigma_2,b_2}(f^*)$

Figure 7: The SNR of the same signal changes with different hash functions. In (a), the noise $g \cdot H$ is hashed outside the bin and suppressed by $G^{(j_1)}_{\sigma_1,b_1}$. Thus, this bin has high SNR. In (b), by a different hash function, the noise is hashed inside the bin and $G^{(j_2)}_{\sigma_2,b_2}$ preserves its energy. Thus, the SNR of the bin becomes very low, even if the signal doesn't change.

algorithm are mixed together, it is hard to justify which frequencies are necessary, since different runs use different hash functions, resulting in different high SNR bins. In other words, if we still use [CKPS16]'s boosting strategy, we cannot guarantee the final output of the frequency estimation satisfies the requirement of the signal estimation algorithm.

We propose a new Fourier Interpolation framework that boosts the success probability after the signal estimation step. That is, in each run of the constant success probability frequency estimation algorithm, we reconstruct the signal immediately. Let $y_1, \ldots, y_{R_p}$ denote the reconstructed signals of R_p runs. Then, we boost the total success probability by outputting the signal y_{j^*}:

$$j^* = \arg\min_{j \in [R_p]} \operatorname*{median}_{i \in [R_p]} \|y_j(t) - y_i(t)\|_T^2.$$

By Chernoff bound, there are more than a half of y_i's being good approximations of the ground-truth signal $x^*(t)$. Using the median trick, we can show that y_{j^*} satisfies the recovery guarantee with an exponentially small failure probability.

It remains to estimate the distance $\|y_j(t) - y_i(t)\|_T^2$ between different reconstructed signals. Naively, it takes $O(\widetilde{k}^2)$-time since y_1, y_2 are $\widetilde{k}$-Fourier sparse, and it is enough to obtain the time complexity of our Fourier interpolation algorithm in Theorem 1.1. We further propose an $\widetilde{O}(\widetilde{k} \cdot k)$-time approximation algorithm for estimating a Fourier-sparse signal's energy, which could be of independent interest. The main idea is to use $\|y_i(t) - y_j(t)\|_{S,w}^2$ to approximate $\|y_i(t) - y_j(t)\|_T^2$, where the sample set S and weights w are defined by the significant sample generation procedure in Section 2.2.1. We show that if we take $|S| = \widetilde{O}(\widetilde{k})$, we can achieve a constant approximation ratio in $\widetilde{O}(\widetilde{k} \cdot k)$ time. In addition, we prove that even if we use the approximated distances, the output signal y_{j^*} still satisfies the recovery guarantee of Fourier interpolation.

3 Organization

In Section A, we define our notations in this paper. In Section B, we review several energy bounds for Fourier-sparse signal. In Section C, we define and show several properties of the frequency domain filters $G^{(j)}_{\sigma,b}$. In Section D, we review the HashToBins strategy and prove that bad events only happen with small probability. In Section E, we define and show some properties of the time domain filter $H(t)$.

Based on the analysis of the filters, in Section F, we study the ideal filter and develop the Signal Equivalent Method. In Section G, we show that the filtered signal satisfies some concentration properties in both time and frequency domains.

Based on the Signal Equivalent Method and the concentration properties, in Section H, we prove an energy bound for filtered Fourier-sparse signals. In Section I, we further extend the energy bound for local-test signals. Then, in Section J, we apply the energy bounds and describe how to use samples to empirically estimate the energy of filtered signals and local-test signals. In Section K, we introduce our algorithm for generating significant samples. In Section L, we use the significant samples to do frequency estimation for Fourier sparse signals. Finally, in Section M, we combine our frequency estimation algorithm with a signal estimation procedure and boost the success probability of Fourier Interpolation.

Section N presents a flowchart of the key theorems/lemmas for our Fourier interpolation algorithm.

A Preliminaries

For any positive integer n, we define $[n]$ to be the set $\{1, 2, \cdots, n\}$. We define $\mathbf{i}$ to be $\sqrt{-1}$. For a complex number $z = a + b\mathbf{i}$, we define $|z|$ to be the magnitude of z, i.e., $|z| = \sqrt{a^2 + b^2}$. For a function f, we use $\mathrm{supp}(f)$ to denote the support set of f. We use $f \lesssim g$ to denote that there exists a constant C such that $f \leq C \cdot g$. We use $f \eqsim g$ to denote that $f \lesssim g \lesssim f$. For any function f, we use $\mathrm{poly}(f)$ to denote $f^{O(1)}$, and $\widetilde{O}(f)$ to denote $f \cdot \mathrm{poly}\log(f)$. For an interval $U \subseteq \mathbb{R}$, we use $|U|$ to denote the size of the interval, and we use $\mathrm{Uniform}(U)$ to denote the uniform distribution over U.

We use ω to denote the exponent of matrix multiplication, i.e., n^ω denote the time of multiplying an $n \times n$ matrix with another $n \times n$ matrix. Currently $\omega \approx 2.373$ [Wil12, AW21].

For two functions f and g, we use $(f * g)(t) = \int_{-\infty}^{\infty} f(s)g(t-s)\mathrm{d}s$ to denote the convolution of two functions f and g. And we use f^{*l} to denote the l-fold convolution of f, i.e., $f^{*l}(t) = f(t) * f(t) * \cdots * f(t)$. For $a \in \mathbb{R}_+$, we use $\mathrm{rect}_a(t)$ to denote the box function with support set length a, i.e., $\mathrm{rect}_a(t) = \mathbf{1}_{[-a/2,a/2]}(t)$. For $a \in \mathbb{R}$, we use $\delta_a(f)$ to denote $\delta(f-a)$, where $\delta(f)$ is the Dirichlet function. We use $\mathrm{round}(x)$ to denote rounding $x \in \mathbb{R}$ to the nearest integer. For $x \in \mathbb{R}, y \in \mathbb{R}_+$, we use $x \pmod{y}$ to denote the smallest positive $z \in \mathbb{R}_+$ such that $z \in x + y\mathbb{Z}$.

We say $x(t)$ is k-Fourier-sparse if:

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

We define $\widehat{x}(f)$ to be the Fourier transform of $x(t)$:

$$\widehat{x}(f) = \int_{-\infty}^{\infty} x(t) e^{-2\pi \mathbf{i} f t} \mathrm{d}t.$$

We use $\mathcal{F}_{k,F}$ to denote the following family of signals:

$$\mathcal{F}_{k,F} := \Big\{ x(t) = \sum_{j=1}^{k} v_j \cdot e^{2\pi \mathbf{i} f_j t} \ \Big|\ f_j \in [-F, F], v_j \in \mathbb{C}\ \forall j \in [k] \Big\}.$$

Then, we define several norms for signal.

- For any discrete set $S \subseteq \mathbb{R}$, the *discrete norm* of x with respect to a set S is defined as

$$\|x(t)\|_S^2 = \frac{1}{|S|} \sum_{t \in S} |x(t)|^2,$$

  and the *weighted discrete norm* with weights $w \in \mathbb{R}^S$ is defined as

$$\|x(t)\|_{S,w}^2 = \sum_{t \in S} w_t |x(t)|^2.$$

- For any continuous interval $U \subset \mathbb{R}$, the continuous U-norm of x is defined as

$$\|x(t)\|_U^2 = \frac{1}{|U|} \int_U |x(t)|^2 \mathrm{d}t.$$

- For any $T > 0$, the *continuous T-norm* is defined as

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

- Let D be a probability distribution over $\mathbb{R}$. The *continuous D-norm* is defined as

$$\|x(t)\|_D^2 = \int_{-\infty}^{\infty} D(t)|x(t)|^2 \mathrm{d}t.$$

- The L_2-norm of $x(t)$ is defined as

$$\|x(t)\|_{L_2}^2 = \int_{-\infty}^{\infty} |x(t)|^2 \mathrm{d}t.$$

Throughout this paper, we assume that $x^*(t) \in \mathcal{F}_{k,F}$ is our ground-truth signal. And the observation signal is $x(t) = x^*(t) + g(t)$, where $g(t)$ is an arbitrary noise function. Furthermore, we assume that $x(t)$ can be observed at any point in $[0, T]$.

Lemma A.1 (Chernoff Bound [Che52])**.** *Let $X_1, X_2, \cdots, X_n$ be independent random variables. Assume that $0 \le X_i \le 1$ 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 $\varepsilon > 0$,*

$$\mathsf{Pr}[X \ge (1+\varepsilon)\mu] \le \exp(-\frac{\varepsilon^2}{2+\varepsilon}\mu) \text{ and } \mathsf{Pr}[X \le (1-\varepsilon)\mu] \le \exp(-\frac{\varepsilon^2}{2}\mu).$$

B Energy Bounds of Fourier Sparse Signals

The energy bound of a function family $\mathcal{F}$ is the largest value achieved by a function $f \in \mathcal{F}$ normalized by its norm (total energy) $\|f\|_T$. It connects the extreme value and the average value of the functions in $\mathcal{F}$, and is very useful in analyzing the concentration property.

In our setting, we take $\mathcal{F} = \mathcal{F}_{k,F}$ to be the set of F-band-limit, k-sparse Fourier signals. [Kós08] showed an energy bound that only depends on the sparsity k, without any dependence on the time point t, band-limit F, time duration T, frequency gap $\eta = \min_{i \ne j} |f_i - f_j|$:

Theorem B.1 ([Kós08])**.** *For any $t \in [0, T]$,*

$$\sup_{x \in \mathcal{F}_{k,F}} \frac{|x(t)|^2}{\|x\|_T^2} \lesssim k^2.$$

The k^2 energy bound can be further improved if we only consider the functions' value at a fixed time point t:

Theorem B.2 ([CP19a, BE06])**.** *Let $D :=$ Uniform$([-1, 1])$. For any $t \in (-1, 1)$,*

$$\sup_{x \in \mathcal{F}_{k,F}} \frac{|x(t)|^2}{\|x\|_D^2} \lesssim \frac{k}{1-|t|}.$$

C Filter in Frequency Domain

Filtering is one of the most important techniques in sparse Fourier transform literature. In this section, we introduce the frequency domain filter function $\widehat{G}^{(j)}_{\sigma,b}(f)$, which is the key to implement the HashToBins strategy. We first review the the construction given by [CKPS16] with some different parameter settings and show some known properties (see Section C.1). Then, we prove a new property of the filter functions: the frequency domain covering property (see Section C.2).

C.1 Frequency domain filter construction

In this section we review the construction and several basic properties of the frequency domain filter $G^{(j)}_{\sigma,b}(t)$, $\widehat{G}^{(j)}_{\sigma,b}(f)$.

Definition C.1 (G-filter's construction, [CKPS16])**.** *Given $B > 1$, $\delta > 0$, $\alpha > 0$. Let $l := \Theta(\log(k/\delta))$. Define $G_{B,\delta,\alpha}(t)$ and its Fourier transform $\widehat{G_{B,\delta,\alpha}}(f)$ as follows:*

$$\begin{aligned} G_{B,\delta,\alpha}(t) :=& \quad b_0 \cdot (\mathrm{rect}_{\frac{B}{(\alpha\pi)}}(t))^{\star l} \cdot \mathrm{sinc}(t\tfrac{\pi}{2B}), \\ \widehat{G_{B,\delta,\alpha}}(f) :=& \quad b_0 \cdot (\mathrm{sinc}(\tfrac{B}{\alpha\pi} f))^l * \mathrm{rect}_{\frac{\pi}{2B}}(f), \end{aligned}$$

where $b_0 = \Theta(B\sqrt{l}/\alpha)$ is the normalization factor such that $\widehat{G}(0) = 1$.

Definition C.2 (Filter for bins)**.** *Given $B > 1$, $\delta > 0$, $\alpha > 0$, let*

$$\widehat{G}(f) := \widehat{G}_{B,\delta,\alpha}(2\pi(1-\alpha)f)$$

where $G_{B,\delta,\alpha}$ is defined in Definition C.1. For any $\sigma > 0, b \in \mathbb{R}$ and $j \in [B]$, define

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

and its Fourier transformation:

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

Then, we provide several properties of G and $G^{(j)}_{\sigma,b}(t)$, which is proven by [CKPS16].

Lemma C.3 (G-filter's properties, [CKPS16])**.** *Given $B > 1$, $\delta > 0$, $\alpha > 0$, let $G := G_{B,\delta,\alpha}(t)$ be defined in Definition C.1. Then, G satisfies the following properties:*

$$\begin{aligned} \text{Property I}: &\quad \widehat{G}(f) \in [1-\delta/k, 1], \text{ if } |f| \le (1-\alpha)\frac{2\pi}{2B}. \\ \text{Property II}: &\quad \widehat{G}(f) \in [0, 1], \text{ if } (1-\alpha)\frac{2\pi}{2B} \le |f| \le \frac{2\pi}{2B}. \\ \text{Property III}: &\quad \widehat{G}(f) \in [-\delta/k, \delta/k], \text{ if } |f| > \frac{2\pi}{2B}. \\ \text{Property IV}: &\quad \mathrm{supp}(G(t)) \subset [\frac{l}{2}\cdot\frac{-B}{\pi\alpha}, \frac{l}{2}\cdot\frac{B}{\pi\alpha}]. \\ \text{Property V}: &\quad \max_t |G(t)| \lesssim \mathrm{poly}(B, l). \end{aligned}$$

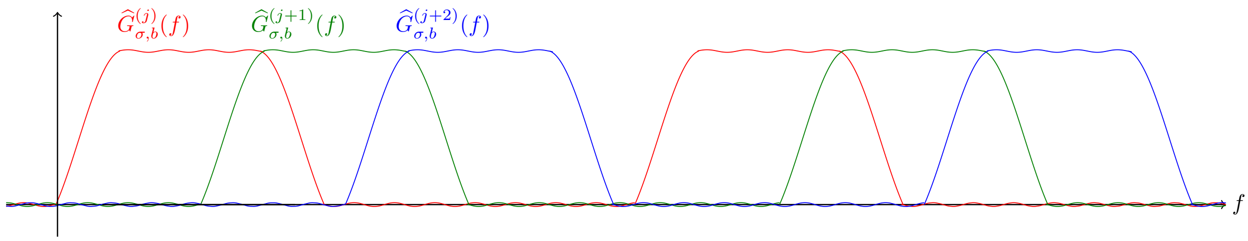


Figure 8: Filters with the frequency domain covering property. The red, green, and blue curves represent the filters $\widehat{G}^{(j)}_{\sigma,b}$, $\widehat{G}^{(j+1)}_{\sigma,b}$, and $\widehat{G}^{(j+2)}_{\sigma,b}$, respectively. The frequency domain covering property ensures that for each frequency $f \in \mathbb{R}$, there are *at least one but no more than two* filters satisfying $\widehat{G}^{(j)}_{\sigma,b}(f) \geq 1 - \delta$.

Lemma C.4. *Let $G^{(j)}_{\sigma,b}(t)$ be defined in Definition C.2. Let offset function*

$$o_{\sigma,b}(f) = |(\sigma f + \sigma b - \frac{j}{B}) - \frac{1}{2} \pmod 1| - \frac{1}{2}.$$

Then,

$$\begin{aligned}
\text{Property I}: &\quad \widehat{G}^{(b)}_{\sigma,b}(f) \in [1 - \delta/k, 1], \ \textit{if } |o_{\sigma,b}(f)| \leq (1-\alpha)\frac{2\pi}{2B}. \\
\text{Property II}: &\quad \widehat{G}^{(b)}_{\sigma,b}(f) \in [0, 1], \ \textit{if } (1-\alpha)\frac{2\pi}{2B} \leq |o_{\sigma,b}(f)| \leq \frac{2\pi}{2B}. \\
\text{Property III}: &\quad \widehat{G}^{(b)}_{\sigma,b}(f) \in [-\delta/k, \delta/k], \ \textit{if } |o_{\sigma,b}(f)| > \frac{2\pi}{2B}. \\
\text{Property IV}: &\quad \mathrm{supp}(G^{(b)}_{\sigma,b}(t)) \subset [\frac{l}{2} \cdot \frac{-B}{\pi\alpha}, \frac{l}{2} \cdot \frac{B}{\pi\alpha}]. \\
\text{Property V}: &\quad \max_t |G^{(b)}_{\sigma,b}(t)| \lesssim \mathrm{poly}(B, l).
\end{aligned}$$

C.2 Frequency domain covering

In this section, we show that the filter functions $\{\widehat{G}^{(j)}_{\sigma,b}\}_{j\in[B]}$ form a proper cover for the frequency domain. Roughly speaking, for any frequency $f \in \mathbb{R}$, we show that the sum of all filters' values (squared) at f is very close to one. This property is very important for our high SNR band assumption.

Lemma C.5. *For any $f \in \mathbb{R}$, there exists at least one $j \in [B]$ such that*

$$\widehat{G}^{(j)}_{\sigma,b}(f) \geq 1 - \frac{\delta}{k}.$$

Proof. We first prove that the lemma holds for those f^* where there exists a $j \in [B]$ such that

$$f^* \in [-b + \frac{j}{\sigma B} - \frac{1}{2\sigma B}, -b + \frac{j}{\sigma B} + \frac{1}{2\sigma B}] + \frac{1}{\sigma}\mathbb{Z}.$$

For such f^*, we have

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

$$\geq \widehat{G}((\sigma f^* + \sigma b - \frac{j}{B}) \mod 1)$$
$$\geq 1 - \frac{\delta}{k},$$

where the first step follows from the definition of $\widehat{G}^{(j)}_{\sigma,b}(f^*)$, the second step is straight forward, the third step follows from

$$\sigma f^* + \sigma b - \frac{j}{B} \mod \frac{1}{\sigma} \in [-\frac{1}{2B}, \frac{1}{2B}]$$

and Lemma C.4 Property I and Definition C.2.

It remains to show that for an arbitrary $f \in \mathbb{R}$, the condition still holds. Let

$$j := \mathrm{round}((\sigma f + \sigma b \mod 1) \cdot B).$$

We have

$$j \in [(\sigma f + \sigma b \mod 1) \cdot B - \frac{1}{2}, (\sigma f + \sigma b \mod 1) \cdot B + \frac{1}{2}],$$

which implies that

$$j \in [(\sigma f + \sigma b) \cdot B - \frac{1}{2}, (\sigma f + \sigma b) \cdot B + \frac{1}{2}] + B\mathbb{Z}.$$

Thus,

$$f \in [-b + \frac{j}{\sigma B} - \frac{1}{2\sigma B}, -b + \frac{j}{\sigma B} + \frac{1}{2\sigma B}] + \frac{1}{\sigma}\mathbb{Z}.$$

The lemma is then proved. $\square$

Lemma C.6. *For any $f \in \mathbb{R}$,*

$$\sum_{j=1}^{B} |\widehat{G}^{(j)}_{\sigma,b}(f)|^2 \eqsim 1.$$

Proof. By Lemma C.5, we have that for any $f \in \mathbb{R}$, there exist at least a $j_0 \in [B]$ such that

$$\widehat{G}^{(j_0)}_{\sigma,b}(f) \geq \frac{1}{2}. \tag{5}$$

Moreover, we have that

$$B\frac{\delta}{k} = O(\delta) \leq 0.01. \tag{6}$$

where the first step follows from $B = O(k)$, the second step follows from $\delta = o(1) \leq 0.01$.

In the followings, we give lower and upper bounds for $\sum_{j=1}^{B} |\widehat{G}^{(j)}_{\sigma,b}(f)|^2$.

Lower bound:

$$\sum_{j=1}^{B} |\widehat{G}^{(j)}_{\sigma,b}(f)|^2 \geq |\widehat{G}^{(j_0)}_{\sigma,b}(f)|^2 \gtrsim 1.$$

where the first step follows from Lemma C.4 Property I, II, and III, the second step follows from Eq. (5), the third step follows from Eq. (6) and $\delta/k \leq 1$.

Upper bound:

$$\sum_{j=1}^{B} (\widehat{G}^{(j)}_{\sigma,b}(f))^2 \leq 2 + B(\frac{\delta}{k})^2 \lesssim 1$$

where the first step follows from the definition of $\widehat{G}^{(j)}_{\sigma,b}(f)$, the second step follows from Eq. (6) and $\delta/k \leq 1$.

Combining them together, the lemma follows. □

D Hashing the Frequencies

In this section, we review the HashToBins strategy, which an important tool for Sparse Fourier Transform [HIKP12a, IKP14, PS15, CKPS16, Kap16, Kap17, JLS23]. Ideally, the HashToBins procedure randomly splits the frequency domain into B bins so that each bin contains at most one frequency. Then, the k-sparse Fourier reconstruction problem is reduced to a much easier one-sparse Fourier reconstruction problem.

We first describe the hashing strategy(see Section D.1). However, there are two kinds of bad events such that the HashToBins procedure cannot work as good as we want: two frequencies are hashed to the same bin, or some frequency lies close to the boundary of a bin. We show that these bad events only happen with small probabilities (see Sections D.2 and D.3) .

D.1 HashToBins procedure

Here, we introduce the hash function and how to compute the resulting signal of the HashToBins procedure.

We first give the definition of the hashing function:

Definition D.1 (Hash function, [CKPS16])**.** *Let* $\pi_{\sigma,b}(f) = \sigma(f+b) \pmod 1$ *and* $h_{\sigma,b}(f) = \mathrm{round}(\pi_{\sigma,b}(f) \cdot B)$ *be the hash function that maps frequency* $f \in [-F, F]$ *into bins* $\{0, \cdots, B-1\}$*.*

Intuitively, the j-th bin corresponding to f such that $\widehat{G}^{(j)}_{\sigma,b}(f) \geq 1 - \delta/k$. In general, we set $B = \Theta(k)$ and $\sigma \in [\frac{1}{B\Delta}, \frac{2}{B\Delta}]$ chosen uniformly at random, where $\Delta = k \cdot |\mathrm{supp}(\widehat{H}(f))|$.

Then, we show how to compute the HashToBins:

Lemma D.2 (Lemma 6.9 in [CKPS16])**.** *Let* $z_j(t) = x(t) * G^{(j)}_{\sigma,b}(t)$*. Let* $a := t/\sigma$ *Let* $u \in \mathbb{C}^B$ *and for* $j \in [B]$*,*

$$u_j := \sum_{i \in \mathbb{Z}} x(\sigma(a - j - iB)) e^{-2\pi \mathbf{i} \sigma b (j+iB)} G(j + iB).$$

Then, we have that for all $j \in [B]$*,*

$$\widehat{u}_j = z_j(\sigma a).$$

Note that when we apply Lemma D.2, we take $x(t) = x(t) \cdot H(t)$, where the latter $x(t)$ is the observable signal, the $H(t)$ is the filter of time domain (see Section E).

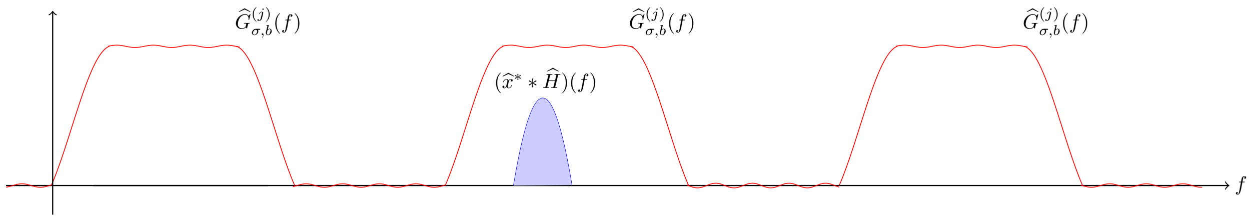


Figure 9: An example of the well-isolation event in the frequency domain. $\widehat{G}^{(j)}_{\sigma,b}$ (the red curve) is the filter of frequency domain for the j-th bin and $H(t)$ is the filter of time domain. $\widehat{x}^*(f) * \widehat{H}(f)$ (the blue curve) is the filtered ground-truth signal. Under the well-isolation event, there is one small interval that contains most of the energy. In other words, each bin only contains one-cluster of frequencies.

D.2 Frequency isolation

The goal of this section is to define and analyze the Frequency Isolation event. Frequency Isolation requires that the energy of the hashed signal in each bin is concentrated in a small band in the frequency domain. This condition is roughly equivalent to say that each bin only contains one cluster of frequencies. This condition is very useful in proving the concentration of the filtered signal in the frequency domain, which serves as one of the basic assumptions of our significant sample generation procedure.

We first introduce Claim D.3. This claim states that if two frequencies are not close to each other, with large probability, they also not hashed into the same bin.

Claim D.3 (Collision probability, [CKPS16])**.** *For any $\Delta_0 > 0$, let σ be a sample uniformly at random from $[\frac{1}{4B\Delta_0}, \frac{1}{2B\Delta_0}]$. Then, we have:*

1. *If $4\Delta_0 \leq |f^+ - f^-| < 2(B-1)\Delta_0$, then $\mathsf{Pr}[h_{\sigma,b}(f^+) = h_{\sigma,b}(f^-)] = 0$.*
2. *If $2(B-1)\Delta_0 \leq |f^+ - f^-|$, then $\mathsf{Pr}[h_{\sigma,b}(f^+) = h_{\sigma,b}(f^-)] \lesssim \frac{1}{B}$.*

We then provide the formal definition of the well-isolation event:

Definition D.4 (Well-isolation condition)**.** *We say that a frequency f^* is* well-isolated *under the hashing parameters (σ, b) if, for $j = h_{\sigma,b}(f^*)$, the hashed signal (in frequency domain) $\widehat{z}^{(j)}(f) := \widehat{x \cdot H}(f) \cdot \widehat{G}^{(j)}_{\sigma,b}(f)$ satisfies*

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

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

The following lemma shows the probability of the Frequency Isolation event under the randomized hashing functions:

Lemma D.5 (Lemma 7.19 in [CKPS16])**.** *Let f^* be any frequency. Then f^* is well-isolated by hashing parameters (σ, b) with probability ≥ 0.9.*

D.3 Large offset event

Large offset event is another kind of bad event for the HashToBins procedure, which happens when a ground-truth frequency is hashed into the changing edge of the filter $G^{(j)}_{\sigma,b}$. The large offset event breaks the guarantee of our signal equivalent method, and thus affects the performance of our significant sample generation and frequency estimation. Fortunately, this bad event only happens with a small probability.

We first state a tool for analyzing the hashing procedure, which intuitively says that the modular of a random sampling from a long interval is almost uniformly distributed:

Lemma D.6 ([PS15, CKPS16])**.** *For any $\widetilde{T}$, and $0 \leq \widetilde{\varepsilon}, \widetilde{\delta} \leq \widetilde{T}$, if we sample $\widetilde{\sigma}$ uniformly at random from $[A, 2A]$, then*

$$\frac{2\widetilde{\varepsilon}}{\widetilde{T}} - \frac{2\widetilde{\varepsilon}}{A} \leq \Pr\left[\widetilde{\sigma} \pmod{\widetilde{T}} \in [\widetilde{\delta} - \widetilde{\varepsilon}, \widetilde{\delta} + \widetilde{\varepsilon}\,]\right] \leq \frac{2\widetilde{\varepsilon}}{\widetilde{T}} + \frac{4\widetilde{\varepsilon}}{A}. \tag{7}$$

Then, we define the large offset event:

Definition D.7 (Large offset event)**.** *Given $\sigma \in \mathbb{R}_+, b \in \mathbb{R}$. Let $G^{(j)}_{\sigma,b}$ and δ be defined as in Definition C.1. For any k-Fourier-sparse signal x, we say the* Large Offset event *happens, if for any $f \in \mathrm{supp}(\widehat{x \cdot H})$ and any $j \in [B]$,*

$$\widehat{G}^{(j)}_{\sigma,b}(f) \in \Big[\frac{\delta}{k}, 1 - \frac{\delta}{k}\Big].$$

We analyze the probability of large offset event in the following lemma:

Lemma D.8. *Let $\Delta_0 = O(\Delta)$, $\widehat{\sigma} = 1/\Delta_0$. Given $b = O(\max\{F, 1/\widehat{\sigma}\})$, suppose $\sigma \sim [0.5\widehat{\sigma}, \widehat{\sigma}]$ uniformly at random. Then, with probability at least 0.99, the Large Offset event does not happen.*

Furthermore, with probability at least 0.99, for any $j \in [k]$, for any $f \in f_j + \mathrm{supp}(\widehat{H})$, it holds that $\widehat{G}^{(j)}_{\sigma,b}(f) \notin [\delta/k, 1 - \delta/k]$.

Proof. Let α be defined as in Definition C.1. Let $I_G := \{f \in \mathbb{R} \mid \widehat{G}^{(j)}_{\sigma,b}(f) \in [\delta/k, 1-\delta/k]\}$. Following from Lemma C.4 Property II, we have that $s_G := |I_G \pmod{1/\sigma}| \leq 10\alpha\Delta_0/B$.

Let $\delta_{f^*}(f)$ be the Dirichlet function at f^*. For any f_j with $j \in [k]$, let $I_{f_j} := \mathrm{supp}(\widehat{H} * \delta_{f_j})$. We also define

$$I'_{f_j} := \{f \in \mathbb{R} \mid [f - s_G, f + s_G] \cap I_{f_j} \neq \emptyset\}.$$

Since $\mathrm{supp}(\widehat{x \cdot H}) = \mathrm{supp}(\widehat{H} * \widehat{x}) \subseteq \bigcup_{j=1}^k \mathrm{supp}(\widehat{H} * \delta_{f_j})$, we know that the Large Offset event happens if

$$\Big(\bigcup_{j=1}^k I_{f_j}\Big) \cap I_G \neq \emptyset.$$

Thus, it suffices to bound $\Pr[(\cup_{j=1}^k I_{f_j}) \cap I_G \neq \emptyset]$.

First, for any $j \in [k]$, we have

$$|I'_{f_j}| \leq |I_{f_j}| + 2s_G \leq \Delta/B + 2s_G \leq O(\Delta/B) \tag{8}$$

where the first step follows from the definition of Δ, the second step follows from $s_G \leq 10\alpha\Delta_0/B$ and the setting of α.

We have that

$$
\begin{aligned}
& \mathsf{Pr}\Big[\frac{1}{2B\sigma}+\frac{j}{B\sigma}-b \pmod{1/\sigma} \in I'_{f_j} \pmod{1/\sigma}\Big] \\
= & \mathsf{Pr}\Big[\frac{1}{2B}+\frac{j}{B} \pmod 1 \in \sigma b+\sigma I'_{f_j} \pmod 1\Big] \\
= & \mathsf{Pr}\Big[\sigma b+\sigma f_j \pmod 1 \in \frac{1}{2B}+\frac{j}{B}+\sigma[-|I'_{f_j}|/2, |I'_{f_j}|/2] \pmod 1\Big] \\
\leq & \mathsf{Pr}\Big[\sigma b+\sigma f_j \pmod 1 \in \frac{1}{2B}+\frac{j}{B}+\widehat{\sigma}[-|I'_{f_j}|/2, |I'_{f_j}|/2] \pmod 1\Big] \\
\leq & \widehat{\sigma}|I'_{f_j}|+\frac{2\widehat{\sigma}|I'_{f_j}|}{0.5\widehat{\sigma}b+0.5\widehat{\sigma}f_j} \\
\leq & 2\widehat{\sigma}|I'_{f_j}| \\
\leq & 2\widehat{\sigma}\cdot O(\Delta/B) \\
\leq & O(1/B)
\end{aligned} \tag{9}
$$

where the first steps are straightforward, the second step follows from the center of I'_{f_j} is f_j, the length of the interval I'_{f_j} is $|I'_{f_j}|$, and $a \in [c-b, c+b] \Rightarrow c \in [a-b, a+b]$, the third step follows from $\sigma \leq \widehat{\sigma}$, the forth step follows by applying Lemma D.6 with the following parameters setting:

$$
\begin{aligned}
\widetilde{T} &= 1, \\
\widetilde{\delta} &= \frac{1}{2B}+\frac{j}{B}, \\
\widetilde{\varepsilon} &= \widehat{\sigma}|I'_{f_j}|/2, \\
A &= 0.5\widehat{\sigma}b+0.5\widehat{\sigma}f_j, \\
\widetilde{\sigma} &= \sigma b+\sigma f_j,
\end{aligned}
$$

the fifth step follows from $0.5b \geq F \geq f_j$ and $0.5b\widehat{\sigma} \geq 1$, the sixth step follows from Eq. (8), the last step follows from the definition of $\widehat{\sigma}$.

Similarly, we have that

$$
\mathsf{Pr}[-\frac{1}{2B\sigma}+\frac{j}{B\sigma}-b \pmod{1/\sigma} \in I'_{f_j} \pmod{1/\sigma}] \leq O(1/B) \tag{10}
$$

Note that I_G is the edge of filter $G^{(j)}_{\sigma,b}$, under the meaning of module $1/\sigma$, the center of $G^{(j)}_{\sigma,b}$ is $\frac{j}{B\sigma}-b$, the length of $G^{(j)}_{\sigma,b}$ is $\frac{1}{B\sigma}$, the length of the edge is s_G. Moreover, I_{f_j} is an interval center at f_j and length $|\mathrm{supp}(\widehat{H})|$. We can judge whether two interval have intersect $I_{f_j} \cap I_G \neq \emptyset$ by moving the length of one interval to another and judging whether $-\frac{1}{2B\sigma}+\frac{j}{B\sigma}-b \pmod{1/\sigma}$, $\frac{1}{2B\sigma}+\frac{j}{B\sigma}-b \pmod{1/\sigma}$ (the end point of I_G) contains in I'_{f_j}. By combining Eq. (9) and Eq. (10), we have that

$$
\mathsf{Pr}[I_{f_j} \cap I_G \neq \emptyset] \leq O(1/B)+O(1/B) = O(1/B). \tag{11}
$$

Therefore, by a union bound over all $j \in [k]$, we get that

$$
\mathsf{Pr}[(\cup_{j=1}^k I_{f_j}) \cap I_G \neq \emptyset] \leq \sum_{j=1}^k \mathsf{Pr}[I_{f_j} \cap I_G \neq \emptyset] \leq \sum_{j=1}^k O(1/B) \leq 0.01,
$$

where the first step is by union bound, the second step follows from Eq. (11), and the last step follows from $B = O(k)$. By the definitions of I_{f_j} and I_G, it implies that with probability at least 0.99, for any $j \in [k]$, and any $f \in f_j + \mathrm{supp}(\widehat{H})$, $\widehat{G}^{(j)}_{\sigma,b}(f) \notin [\delta/k, 1-\delta/k]$.

The proof of the lemma is then completed. $\square$

Lemma D.9. *For $x^*(t)$ be a k-Fourier-sparse signal. For frequency $f^* \in \mathrm{supp}(\widehat{x}^*)$, let $j = h_{\sigma,b}(f^*)$ be the bin that f^* hashed into. If Large Offset event not happens, then for $f \in \mathrm{supp}(\widehat{x}^* * \widehat{H})$,*

$$\widehat{G}^{(j)}_{\sigma,b}(f) \in [1-\delta/k, 1]$$

Proof. Since Large Offset event not happens, for $f \in \mathbb{R}$,

$$\widehat{G}^{(j)}_{\sigma,b}(f) \geq 1-\delta/k \text{ or } \widehat{G}^{(j)}_{\sigma,b}(f) \leq \delta/k.$$

Since Large Offset event not happens and $j = h_{\sigma,b}(f^*)$, we have that for $f \in \mathrm{supp}(\widehat{x}^* * \widehat{H})$,

$$\widehat{G}^{(j)}_{\sigma,b}(f) \geq 1-\delta/k.$$

By Lemma C.4 Property I, II, and III, we have that

$$\widehat{G}^{(j)}_{\sigma,b}(f) \in [1-\delta/k, 1].$$

$\square$

E Filter in Time Domain

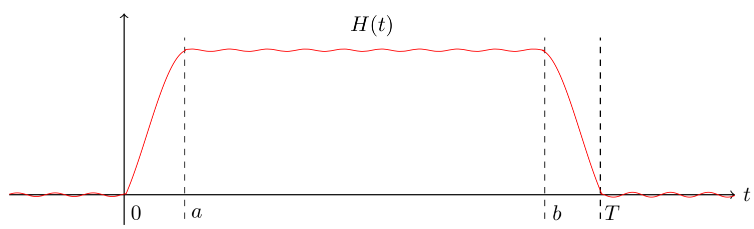


Figure 10: The filter $H(t)$ of time domain. We use decay region to refer $[0, a]$ and $[b, T]$. We use fluctuation region to refer $[a, b]$.

In this section, we discuss the filter $H(t)$ of time domain, which is an analogous of the ideal filter $\mathrm{rect}_T(t)$. In the Fourier interpolation problem, we only care about the time duration $[0, T]$. Thus, applying the filter $\mathrm{rect}_T(t)$ to the observation signal $x(t)$ can cut-off the unobservable part and much simplify the analysis. Since $\mathrm{rect}_T(t)$ have an infinite band width, for efficient computation, we need to truncate $\mathrm{rect}_T(t)$'s frequency domain to a $\mathrm{poly}(k)/T$-length interval. However, the frequency truncation loses the high frequency components of the ideal filter $\mathrm{rect}_T(t)$, and the resulting filter $H(t)$ is no longer sharp around the boundary of $[0, T]$. More specifically, $H(t)$ is exponentially close to 1 within $[T/\mathrm{poly}(k), T(1-1/\mathrm{poly}(k))]$, and exponentially close to 0 outside $[0, T]$.

We first provide the construction of $H(t)$ in [CP19b] and review some known properties (see Section E.1). Then, we discuss the normalization factor of the filter and provide a polynomial upper bound of it (see Section E.2). This bound is crucial for our fluctuation bound of the $H(t)$. Next, we bound the fluctuation of $H(t)$ during a shrinking interval $[T/\mathrm{poly}(k), T(1-1/\mathrm{poly}(k))]$ and prove that $H(t)$ is exponentially close to 1 in that range (see Section E.3). Furthermore, we prove that $H(t)$ preserve the energy of Fourier sparse signal in the duration $[0, T]$ (see Section E.4).

E.1 Time domain filter construction

We first introduce an growth rate bound from [CP19b]. This theorem bound the growth of Fourier sparse signal outside of the duration $[0,T]$ by an exponent function of base t. This bound in this theorem is high related to the size of the support set of $\widehat{H}(f)$.

Theorem E.1 ([CP19b]). *There exists $S = O(k^2 \log k)$ such that for any $|t| > T$ and $g(t) = \sum_{j=1}^{k} v_j \cdot e^{2\pi \mathbf{i} f_j t}$, $|g(t)|^2 \leq \mathrm{poly}(k) \cdot \mathop{\mathbb{E}}_{x \in [-T,T]}[|g(x)|^2] \cdot |\frac{t}{T}|^S$.*

The definition of the time domain filter $H(t)$ in [CP19b] is given in below. Intuitively, it uses some powers of $\mathrm{sinc}(t)$ to approximate $\delta_0(t) * \mathrm{rect}_1(t) = \mathrm{rect}_1(t)$, thus one can get finite band-limit and good approximation at the same time.

Definition E.2 (Definition 4.1 in [CP19b]). *Given an energy bound R satisfying*

$$|x(t)|^2 \lesssim R\|x(t)\|_T^2, \quad \forall t \in [0,T] \text{ and } k\text{-Fourier sparse signal } x(t),$$

the growth rate S a power of two, $C \in 2\mathbb{Z}$, and $C_0 \in \pi\mathbb{Z}$,we define the filter function:

$$H_1(t) = s_0 \cdot \left(\mathrm{sinc}(C_0 R \cdot t)^{C \log R} \cdot \mathrm{sinc}\left(C_0 \cdot S \cdot t\right)^C \cdot \mathrm{sinc}\left(\frac{C_0 \cdot S}{2} \cdot t\right)^{2C} \cdots \mathrm{sinc}\left(C_0 \cdot t\right)^{C \cdot S} \right) * \mathrm{rect}_1(t),$$

where $s_0 \in \mathbb{R}^+$ is a parameter to normalize $H_1(0) = 1$. Its Fourier transform is as follows:

$$\widehat{H}_1(f) = s_0 \cdot \left(\mathrm{rect}_{C_0 R}(f)^{*C \log R} * \mathrm{rect}_{C_0 \cdot S}(f)^{*C} * \mathrm{rect}_{\frac{C_0 \cdot S}{2}}(f)^{*2C} * \cdots * \mathrm{rect}_{C_0}(f)^{*CS} \right) \cdot \mathrm{sinc}(f/2).$$

We then state some basic properties of the time domain filter in [CP19b]. The following lemma bounds the support size of $\widehat{H}_1(f)$:

Lemma E.3 ([CP19b]). *Let $C_0 = \Theta(C)$. we have that*

$$|\mathrm{supp}(\widehat{H}_1(f))| = O(C^2 R \log R + C^2 S \log S).$$

The following theorem shows some time domain properties of the filter:

Theorem E.4 (Theorem 4.2 in [CP19b]). *Let $R, S > 0$, let $C \in 2\mathbb{Z}$, $C_0 \in \pi\mathbb{Z}$, $C_0 = \Theta(C)$, and define $\alpha = (\frac{1}{2} + \frac{1.2}{\pi C_0 R})$. Consider any function x satisfying the following two conditions:*

1. $\sup_{t \in [-1,1]} \left[|x(t)|^2\right] \leq R \cdot \mathop{\mathbb{E}}_{t \in [-1,1]}\left[|x(t)|^2\right]$,
2. $\mathrm{poly}(R) \cdot \mathop{\mathbb{E}}_{t \in [-1,1]}\left[|x(t)|^2\right] \cdot |t|^S$ *for* $t \notin [-1,1]$,

Then, we have that the filter function $H(t) = H_1\big(\alpha t\big)$ satisfies

- *Part 1.* $\int_{-1}^{1} |x(t) \cdot H\big(t\big)|^2 \mathrm{d}t \geq 0.9 \int_{-1}^{1} |x(t)|^2 \mathrm{d}t$,
- *Part 2.* $\int_{-1}^{1} |x(t) \cdot H\big(t\big)|^2 \mathrm{d}t \geq 0.95 \int_{-\infty}^{\infty} |x(t) \cdot H\big(t\big)|^2 \mathrm{d}t$,
- *Part 3.* $|H(t)| \leq 1.01$ *for any* t.

Throughout this paper, we denote $H(t)$ as the following re-scaling of $H_1(t)$:

Definition E.5. *Let $\alpha = (\frac{1}{2} + \frac{1.2}{\pi C_0 R})$. Let $H_1(t)$ be defined as in Definition E.2. The filter $H(t)$ is defined as:*

$$H(t) := H_1(\alpha t).$$

and setting $R = S = O(k^2)$, $R = 2^s$, where $s \in \mathbb{Z}_+$, $C = O(\log(1/\delta_1))$, $C \in 2\mathbb{Z}$, $C_0 = \Theta(C)$ and $C_0 \in \pi\mathbb{Z}$.

E.2 Normalization factor of the filter

The goal of this section is to prove Lemma E.6, an upper-bound for the normalization factor s_0. This lemma will be used later to ensure that the scaling factor will not break the exponential small fluctuation of Section E.3. We note that the same result has been proved in [CP19b], and we reprove it below for completeness.

Lemma E.6 (Lemma 7.2 in [CP19b]). *It holds that*

$$s_0 \leq O(CR\sqrt{C \log R}).$$

Proof. We first have that,

$$\begin{aligned} H_1(t) &= s_0 \cdot (\mathrm{sinc}(C_0 R \cdot t)^{C \log R} \cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,\big(\frac{C_0 \cdot S}{2^i} \cdot t\big)^{2^i \cdot C}) * \mathrm{rect}_1(t) \\ &= s_0 \cdot \int_{-\infty}^{\infty} \mathrm{sinc}(C_0 R \cdot \tau)^{C \log R} \cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,\big(\frac{C_0 \cdot S}{2^i} \cdot \tau\big)^{2^i \cdot C} \cdot \mathrm{rect}_1(t-\tau) \mathrm{d}\tau \\ &= s_0 \cdot \int_{t-0.5}^{t+0.5} \mathrm{sinc}(C_0 R \cdot \tau)^{C \log R} \cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,\big(\frac{C_0 \cdot S}{2^i} \cdot \tau\big)^{2^i \cdot C} \mathrm{d}\tau, \end{aligned}$$

where the first step follows from the definition of $H_1(t)$, the second step follows from the definition of the convolution, the third step follows from the definition of $\mathrm{rect}_2(t)$ function. Thus,

$$H_1(0) = s_0 \cdot \int_{-0.5}^{+0.5} \mathrm{sinc}(C_0 R \cdot \tau)^{C \log R} \cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,\big(\frac{C_0 \cdot S}{2^i} \cdot \tau\big)^{2^i \cdot C} \mathrm{d}\tau$$

Let $U := (2C_0 \log R)^{-1/2}$. We have that

$$\begin{aligned} &\int_{-0.5}^{+0.5} \mathrm{sinc}(C_0 R \cdot \tau)^{C \log R} \cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,\big(\frac{C_0 \cdot S}{2^i} \cdot \tau\big)^{2^i \cdot C} \mathrm{d}\tau \\ \geq & \int_{-\frac{1}{\pi C_0 R}}^{+\frac{1}{\pi C_0 R}} \mathrm{sinc}(C_0 R \cdot \tau)^{C \log R} \cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,\big(\frac{C_0 \cdot S}{2^i} \cdot \tau\big)^{2^i \cdot C} \mathrm{d}\tau \\ = & \frac{1}{\pi C_0 R} \int_{-1}^{+1} \mathrm{sinc}(\frac{\upsilon}{\pi})^{C \log R} \cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,\big(\frac{\upsilon}{2^i \pi}\big)^{2^i \cdot C} \mathrm{d}\upsilon \\ \geq & \frac{1}{\pi C_0 R} \int_{-U}^{+U} \mathrm{sinc}(\frac{\upsilon}{\pi})^{C \log R} \cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,\big(\frac{\upsilon}{2^i \pi}\big)^{2^i \cdot C} \mathrm{d}\upsilon \\ \geq & \frac{1}{\pi C_0 R} \int_{-U}^{+U} (1 - \frac{\upsilon^2}{6})^{C \log R} \cdot \prod_{i=0}^{\log(S)} (1 - \frac{\upsilon^2}{4^i 6})^{2^i \cdot C} \mathrm{d}\upsilon \\ \geq & \frac{1}{\pi C_0 R} \int_{-U}^{+U} (1 - C \log R \cdot \frac{\upsilon^2}{6} - \sum_{i=0}^{\log(S)} C \cdot \frac{\upsilon^2}{2^i 6}) \mathrm{d}\upsilon \\ \geq & \frac{1}{\pi C_0 R} \int_{-U}^{+U} (1 - C \log R \cdot \frac{\upsilon^2}{3}) \mathrm{d}\upsilon \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\pi C_0 R}(2U - 2C\log R \cdot \frac{U^3}{9}) \\
&\geq \frac{1}{\pi C_0 R\sqrt{2C_0 \log R}}, \qquad (12)
\end{aligned}
$$

where the first step follows from $R = O(k^2)$, $C_0 = O(\log(1/\delta_1))$, $0.5 \geq \frac{1}{\pi C_0 R}$, second step follows from changing the variable $\nu = \pi CR\cdot\tau$, the third step follows from $U < 1$, the forth step follows from Fact E.7, the fifth step is follows from $(1-a)(1-b) \geq 1-a-b$, the sixth step follows from $\log(R) > 2$, the seventh step is straight forward, the eighth step follows from setting $U = (2C_0 \log R)^{-1/2}$ and $C_0 = \Theta(C)$.

As a result, we have that

$$
s_0 \leq H_1(0) \cdot \pi CR\sqrt{2C\log R} = \pi CR\sqrt{2C\log R}
$$

where the first step follows from Eq. (12), the second step follows from $H_1(0) = 1$.

□

Fact E.7. *For any $t \in \mathbb{R}$,*

$$
1 - \frac{(\pi t)^2}{3!} \leq \mathrm{sinc}(t) \leq 1.
$$

E.3 Fluctuation bound

The idea filter $\mathrm{rect}_T(t)$ has a constant value 1 in the interval $[0, T]$. Due to the frequency domain truncation in $H(t)$, it deviates from $\mathrm{rect}_T(t)$ with different magnitudes in different regions. In this section, we prove the Lemma E.8, which shows that $H(t)$ is fluctuating near 1 in the "interior" of $[0, T]$ (i.e., $[0 + \frac{T}{\mathrm{poly}(k)}, T - \frac{T}{\mathrm{poly}(k)}]$). It serves as an important tool for analyzing the error in our signal equivalent method.

Lemma E.8. *For filter $H_1(t)$ defined in Definition E.2 with the parameters $C = \log(1/\delta_1)$, $C_0 = \Theta(C)$, $R = S$, and $S = 2^s$ (where $s \in \mathbb{Z}_+$), $C_0 \in \pi\mathbb{Z}$, we have that*

$$
H_1(t) \in [1 - \delta_1, 1], \forall |t| < 0.5 - \frac{\pi}{C_0 R}.
$$

Moreover, $H(t) \in [1 - \delta_1, 1]$ for any $t \in [\frac{T}{2} - \alpha^{-1}(\frac{1}{2} - \frac{\pi}{C_0 R})\frac{T}{2}, \frac{T}{2} + \alpha^{-1}(\frac{1}{2} - \frac{\pi}{C_0 R})\frac{T}{2}]$.

Proof. The proof consists of two parts: upper bound and lower bound. For the upper bound, the idea is to compare the value of $H(t)$ with $H(0)$ by analyzing the gradient of $H(t)$. And the lower bound follows from directly estimating the integral of the product of sinc functions.

Upper bound: We have that $H_1(0) = 1$ by definition. We will show $H_1(t) \leq 1$ by proving $H_1(t)$ is monotonically decreasing in t.

We have that,

$$
\begin{aligned}
H_1(t) &= s_0 \cdot (\mathrm{sinc}(C_0 R \cdot t)^{C\log R} \cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,(\frac{C_0 \cdot S}{2^i} \cdot t)^{2^i \cdot C}) * \mathrm{rect}_1(t) \\
&= s_0 \cdot \int_{-\infty}^{\infty} \mathrm{sinc}(C_0 R \cdot \tau)^{C\log R} \cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,(\frac{C_0 \cdot S}{2^i} \cdot \tau)^{2^i \cdot C} \cdot \mathrm{rect}_1(t - \tau)\mathrm{d}\tau
\end{aligned}
$$

$$= s_0 \cdot \int_{t-0.5}^{t+0.5} \mathrm{sinc}(C_0 R \cdot \tau)^{C \log R} \cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\big(\frac{C_0 \cdot S}{2^i} \cdot \tau\big)^{2^i \cdot C} \mathrm{d}\tau, \tag{13}$$

where the first step follows from the definition of $H_1(t)$, the second step follows from the definition of the convolution, the third step follows from the definition of $\mathrm{rect}_1(t)$ function.

Since $C \log R \in 2\mathbb{Z}$, $2^i \cdot C \in 2\mathbb{Z}$, we have that

$$\mathrm{sinc}(C_0 R \cdot \tau)^{C \log R} \geq 0,$$

and

$$\mathrm{sinc}\big(\frac{C_0 \cdot S}{2^i} \cdot \tau\big)^{2^i \cdot C} \geq 0.$$

Moreover, by setting $C_0 = \pi\mathbb{Z}$, we have that

$$2 \mod \frac{2\pi}{C_0} = 0. \tag{14}$$

By Eq. (14), we have that

$$\begin{aligned} \sin\big(\frac{C_0 \cdot S}{2^i} \cdot (t+0.5)\big) =&\ \sin\big(\frac{C_0 \cdot S}{2^i} \cdot (t-0.5)\big), \quad \forall i \in \{0, \cdots, \log(S)\}, \quad \text{and} \\ \sin(C_0 R \cdot (t+0.5)) =&\ \sin(C_0 R \cdot (t-0.5)). \end{aligned}$$

Furthermore, for any $t > 0$,

$$\begin{aligned} \big(\frac{C_0 \cdot S}{2^i} \cdot (t+0.5)\big)^{-1} \leq&\ \big(\frac{C_0 \cdot S}{2^i} \cdot (t-0.5)\big)^{-1}, \forall i \in \{0, \cdots, \log(S)\}, \quad \text{and} \\ (C_0 R \cdot (t+0.5))^{-1} \leq&\ (C_0 R \cdot (t-0.5))^{-1}. \end{aligned}$$

Thus,

$$\begin{aligned} |\,\mathrm{sinc}\big(\frac{C_0 \cdot S}{2^i} \cdot (t+0.5)\big)| \leq&\ |\,\mathrm{sinc}\big(\frac{C_0 \cdot S}{2^i} \cdot (t-0.5)\big)|, \quad \forall i \in \{0, \cdots, \log(S)\}, \quad \text{and} \\ |\,\mathrm{sinc}(C_0 R \cdot (t+0.5))| \leq&\ |\,\mathrm{sinc}(C_0 R \cdot (t-0.5))|. \end{aligned} \tag{15}$$

Then, we have that

$$\begin{aligned} \frac{H_1'(t)}{s_0} =&\ \mathrm{sinc}(C_0 R \cdot (t+0.5))^{C \log R} \cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\big(\frac{C_0 \cdot S}{2^i} \cdot (t+0.5)\big)^{2^i \cdot C} \\ &\quad - \mathrm{sinc}(C_0 R \cdot (t-0.5))^{C \log R} \cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\big(\frac{C_0 \cdot S}{2^i} \cdot (t-0.5)\big)^{2^i \cdot C} \\ <&\ 0, \end{aligned}$$

where the first step is straight forward, the second step follows from Eq. (15).

Thus, $H_1(t) < H_1(0) = 1$ for any $t > 0$.

Similarly, we also have that $H_1(t) < H_1(0) = 1$ for any $t \leq 0$ since $H_1(t)$ is symmetric with respect to t.

Lower bound: We have that, for any $|t| < 0.5 - \frac{\pi}{C_0 R}$,

$$\begin{aligned}
&\int_{-\infty}^{t-0.5} \mathrm{sinc}(C_0 R\cdot\tau)^{C\log R}\cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,\big(\frac{C_0\cdot S}{2^i}\cdot\tau\big)^{2^i\cdot C}\mathrm{d}\tau \\
= &\int_{0.5-t}^{\infty} \mathrm{sinc}(C_0 R\cdot\tau)^{C\log(R)}\cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,\big(\frac{C_0\cdot S}{2^i}\cdot\tau\big)^{2^i\cdot C}\mathrm{d}\tau \\
\leq &\int_{0.5-t}^{\infty} \mathrm{sinc}(C_0 R\cdot\tau)^{C\log(R)}\mathrm{d}\tau \\
\leq &\int_{0.5-t}^{\infty} (C_0 R\cdot\tau)^{-C\log(R)}\mathrm{d}\tau \\
\leq &\int_{\frac{\pi}{C_0 R}}^{\infty} (C_0 R\cdot\tau)^{-C\log(R)}\mathrm{d}\tau \\
= &\frac{1}{C_0 R}\int_{\pi}^{\infty} \upsilon^{-C\log(R)}\mathrm{d}\upsilon \\
= &\frac{1}{C_0 R}\frac{1}{C\log(R)-1}\pi^{-C\log(R)+1} \\
\lesssim &\frac{1}{C_0^2 R\log(R)}\delta_1,
\end{aligned} \tag{16}$$

where the first step is straight forward, the second step follows from $\mathrm{sinc}(x)\leq 1$, the third step is follows from $\mathrm{sinc}(x)\leq 1/x$, the forth step follows follows from $0.5 - t \geq \frac{\pi}{C_0 R}$, the fifth step follows from $\upsilon := CR\tau$, the sixth step is straight forward, the seventh step follows from $\log(R) > 1$, $C \geq \log(1/\delta_1)$.

Hence, for any $t > 0$,

$$\begin{aligned}
H_1(t) = &\, s_0\cdot\int_{t-0.5}^{t+0.5} \mathrm{sinc}(C_0 R\cdot\tau)^{C\log R}\cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,\big(\frac{C_0\cdot S}{2^i}\cdot\tau\big)^{2^i\cdot C}\mathrm{d}\tau \\
= &\, H_1(0) + s_0\cdot\int_{1}^{t+0.5} \mathrm{sinc}(C_0 R\cdot\tau)^{C\log R}\cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,\big(\frac{C_0\cdot S}{2^i}\cdot\tau\big)^{2^i\cdot C}\mathrm{d}\tau \\
&\quad - s_0\cdot\int_{-1}^{t-0.5} \mathrm{sinc}(C_0 R\cdot\tau)^{C\log R}\cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,\big(\frac{C_0\cdot S}{2^i}\cdot\tau\big)^{2^i\cdot C}\mathrm{d}\tau \\
\geq &\, H_1(0) - s_0\cdot\int_{-\infty}^{t-0.5} \mathrm{sinc}(C_0 R\cdot\tau)^{C\log R}\cdot \prod_{i=0}^{\log(S)} \mathrm{sinc}\,\big(\frac{C_0\cdot S}{2^i}\cdot\tau\big)^{2^i\cdot C}\mathrm{d}\tau \\
\geq &\, H_1(0) - s_0 O(\frac{1}{C_0^2 R\log(R)})\delta_1 \\
= &\, 1 - s_0 O(\frac{1}{C_0^2 R\log(R)})\delta_1 \\
\geq &\, 1 - O(\delta_1),
\end{aligned}$$

where the first step follows from Eq. (13), the second step is straight forward, the third step follows from

$$\text{sinc}(C_0 R\cdot\tau)^{C\log R}\cdot\prod_{i=0}^{\log(S)}\text{sinc}\left(\frac{C_0\cdot S}{2^i}\cdot\tau\right)^{2^i\cdot C}\geq 0,\quad\forall\tau\in\mathbb{R}$$

the forth step follows from Eq. (16), the fifth step follows from $H_1(0)=1$, the sixth step follows from Lemma E.6.

By re-scaling δ_1, we get that $H(t)\geq 1-\delta_1$ for any $|t|<0.5-\frac{\pi}{C_0R}$.

The lemma then follows from the upper and lower bounds.

□

E.4 Energy preserving of the time domain filter

In this section, we show the properties of $H(t)$ that we use in the rest of the paper.

We first prove Lemma E.9, which summarizes the results in above sections and prove the energy preserving property of $H(t)$.

Lemma E.9. *The filter function $(H(t),\widehat{H}(f))$ has the following properties:*

$$\begin{aligned}
\text{Property I}: &\quad |H(t)|\leq 1.01,\ \forall t\in\mathbb{R}\\
\text{Property II}: &\quad 1-\delta_1\leq H(t)\leq 1,\ \forall |t|<\alpha^{-1}\left(\frac{1}{2}-\frac{\pi}{CR}\right)\\
\text{Property III}: &\quad |\text{supp}(\widehat{H}(f))|\leq O(k^2\log^2(k)\log^2(1/\delta_1))\\
\text{Property IV}: &\quad \int_{-\infty}^{+\infty}\left|x^*(t)\cdot H(t)\cdot(1-\text{rect}_2(t))\right|^2\text{d}t<0.1\int_{-\infty}^{+\infty}|x^*(t)\cdot\text{rect}_2(t)|^2\text{d}t\\
\text{Property V}: &\quad \int_{-\infty}^{+\infty}|x^*(t)\cdot H(t)\cdot\text{rect}_2(t)|^2\text{d}t\in[0.9,1.1]\cdot\int_{-\infty}^{+\infty}|x^*(t)\cdot\text{rect}_2(t)|^2\text{d}t
\end{aligned}$$

Proof. We prove each of the five properties in below.

Property I: By Theorem E.4 Part 3, we have that

$$|H(t)|\leq 1.01.$$

Property II: By Lemma E.8, we have that

$$1-\delta_1\leq H(t)\leq 1,\ \forall|t|<\alpha^{-1}\left(\frac{1}{2}-\frac{\pi}{CR}\right).$$

Property III: By the k-Fourier-sparse signals' energy bound (Theorem B.1), we have that

$$R=O(k^2).$$

By Theorem E.1, we have that

$$S=O(k^2\log k).$$

Then, by Lemma E.3, we have that

$$\begin{aligned}
|\text{supp}(\widehat{H}(f))| &= C^2R\log R+C^2S\log S\\
&= O(\log(1/\delta_1))^2\cdot O(k^2\log k\log(k^2\log k))\\
&= O(k^2\log^2 k\log^2(1/\delta_1)).
\end{aligned}$$

Property IV: We have that

$$\begin{aligned}
&\int_{-\infty}^{+\infty} \left|x^*(t) \cdot H(t) \cdot (1 - \mathrm{rect}_1(t))\right|^2 \mathrm{d}t \\
= &\int_{-\infty}^{+\infty} \left|x^*(t) \cdot H(t)\right|^2 \mathrm{d}t - \int_{-1}^{1} \left|x^*(t) \cdot H(t)\right|^2 \mathrm{d}t \\
\leq\ & 0.06 \int_{-1}^{+1} \left|x^*(t) \cdot H(t)\right|^2 \mathrm{d}t \\
\leq\ & 0.1 \int_{-1}^{+1} \left|x^*(t)\right|^2 \mathrm{d}t \\
= &\ 0.1 \int_{-\infty}^{+\infty} |x^*(t) \cdot \mathrm{rect}_1(t)|^2 \mathrm{d}t,
\end{aligned}$$

where the first step is straight forward, the second step follows from Theorem E.4 Part 2, the third step follows from Theorem E.4 Part 3, the forth step is straight forward.

Property V: We first prove the upper bound:

$$\begin{aligned}
&\int_{-\infty}^{+\infty} |x^*(t) \cdot H(t) \cdot \mathrm{rect}_1(t)|^2 \mathrm{d}t \\
= &\int_{-1}^{+1} |x^*(t) \cdot H(t)|^2 \mathrm{d}t \\
\leq\ & 1.1 \int_{-1}^{+1} |x^*(t)|^2 \mathrm{d}t \\
= &\ 1.1 \int_{-\infty}^{+\infty} |x^*(t) \cdot \mathrm{rect}_1(t)|^2 \mathrm{d}t,
\end{aligned}$$

where the first step is straight forward, the second step follows from Theorem E.4 Part 3, the third step is straight forward.

Then, we prove the lower bound:

$$\begin{aligned}
&\int_{-\infty}^{+\infty} |x^*(t) \cdot H(t) \cdot \mathrm{rect}_1(t)|^2 \mathrm{d}t \\
= &\int_{-1}^{+1} |x^*(t) \cdot H(t)|^2 \mathrm{d}t \\
\geq\ & 0.9 \cdot \int_{-1}^{+1} |x^*(t)|^2 \mathrm{d}t \\
= &\ 0.9 \cdot \int_{-\infty}^{+\infty} |x^*(t) \cdot \mathrm{rect}_1(t)|^2 \mathrm{d}t
\end{aligned}$$

where the first step is straight forward, the second step follows from Theorem E.4 Part 1, the third step is straight forward.

□

The following lemma bounds the length of the fluctuation region (where $H(t)$ is close to 1) in the time domain.

Lemma E.10. *Let $\Delta = k|\text{supp}(\widehat{H}(f))|$, $\beta = O(1/\Delta)$, $L = \frac{T}{2} - \alpha^{-1}(\frac{1}{2} - \frac{\pi}{C_0 R})\frac{T}{2}$, $R = \frac{T}{2} + \alpha^{-1}(\frac{1}{2} - \frac{\pi}{C_0 R})\frac{T}{2} - \beta$, we have that*

$$T - k^2(T + L - R) \eqsim T,$$

and

$$R - L \eqsim T.$$

Proof. Let $U := [L, R]$. By Lemma E.8, we have that for any $t_0 \in U$,

$$H(t) > 1 - \delta_1, \forall t \in [t_0, t_0 + \beta].$$

We have that

$$\begin{aligned}
R - L &= |U| \\
&= ((\frac{1}{2} + \frac{1.2}{\pi CR})^{-1} \cdot (\frac{1}{2} - \frac{\pi}{CR}) - \beta) \cdot T \\
&\eqsim T,
\end{aligned}$$

where the first step follows from the definition of L, R, the second step follows from Lemma E.9 Property II, the third step follows from $\Delta, CR \gg 1$.

We have that

$$\begin{aligned}
T + L - R &= T - |U| \\
&= (1 - (\frac{1}{2} + \frac{1.2}{\pi CR})^{-1} \cdot (\frac{1}{2} - \frac{\pi}{CR})) \cdot T + \beta T \\
&= \frac{2\pi + 2.4/\pi}{CR + 2.4/\pi} \cdot T + \beta T, \qquad (17)
\end{aligned}$$

where the first step follows from the definition of L, R, the second step follows from Lemma E.9 Property II, the third step is straight forward.

Then, we have that

$$T - k^2(T + L - R) = T - k^2 \cdot (\frac{2\pi + 2.4/\pi}{CR + 2.4/\pi} + \beta) \cdot T \eqsim T,$$

where the first step follows from Eq. (17), the second step follows from $C = O(\log(1/\delta_1))$, $R = k^2$, $k^2\beta < 1/k$.

□

F Ideal Filter Approximation

As we discussed in previous sections, the filtered signal $z^{(j)}(t) = (x \cdot H) * G^{(j)}_{\sigma,b}(t)$ is the signal in the j-th bin by the HashToBins procedure. In this section, we consider an approximation of the frequency domain filter $G^{(j)}_{\sigma,b}$ by the *ideal filter* $I^{(j)}_{\sigma,b}(t)$ defined by its Fourier transform:

$$\widehat{I}^{(j)}_{\sigma,b}(f) := \begin{cases} 1, & \widehat{G}^{(j)}_{\sigma,b}(f) > 1 - \delta_1 \\ 0, & \text{otherwise} \end{cases} \qquad (18)$$

Intuitively, $I_{\sigma,b}^{(j)}$ is "denoising" the frequency domain filter $\widehat{G}_{\sigma,b}^{(j)}$ in the sense that it rounds the heavy Fourier coefficients of $\widehat{G}_{\sigma,b}^{(j)}$ to 1 and rounds the remaining Fourier coefficients to 0. The main purpose of this section is to show that $(x \cdot H) * I_{\sigma,b}^{(j)}$ is a good approximation of $z^{(j)}$. For simplicity, we will use I to denote $I_{\sigma,b}^{(j)}$ when σ, b, j are clear from context.

We first show a commuting property of the ideal filter (see Section F.1). Then, we derive the approximation error bound of the ideally filtered signals (see Section F.2).

F.1 Swap the order of filtering

We first prove a good property of the ideal filter that $I_{\sigma,b}^{(j)}$ "commutes" with the time domain filter H with high probability over the random hashing function.

Lemma F.1. *Let δ_1 be the δ defined in Lemma C.4.Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2. Let the ideal filter $I = I_{\sigma,b}^{(j)}$ be defined as in Eq. (18).*

Then, for any $x \in \mathcal{F}_{k,F}$, with probability 0.9 over the choice of (σ, b), for any $j \in [B]$,

$$(x \cdot H) * I(t) = (x * I)(t) \cdot H(t) \quad \forall t \in \mathbb{R}.$$

Proof. By Fourier transformation, we have

$$(x \cdot H) * I(t) = \int_{-\infty}^{\infty} (\widehat{x} * \widehat{H})(f) \cdot \widehat{I}(f) \cdot \exp(2\pi \mathbf{i} f t) \mathrm{d}f.$$

We will show that $(\widehat{x} * \widehat{H})(f) \cdot \widehat{I}(f) = ((\widehat{x} \cdot \widehat{I}) * \widehat{H})(f)$ with high probability.

On the one hand,

$$\begin{aligned}
(\widehat{x} * \widehat{H})(f) \cdot \widehat{I}(f) &= \sum_{j=1}^{k} v_j \cdot (\delta_{f_j} * \widehat{H})(f) \cdot \widehat{I}(f) \\
&= \sum_{j=1}^{k} v_j \cdot \widehat{H}(f - f_j) \cdot \widehat{I}(f) \\
&= \sum_{j=1}^{k} v_j \cdot \widehat{H}(f - f_j) \cdot \mathbf{1}_{f \in \mathrm{supp}(\widehat{I})},
\end{aligned}$$

where the first step follows from $\widehat{x}(f) = \sum_{j=1}^{k} v_j \cdot \delta_{f_j}(f)$, the second step follows from the convolution property of Delta function. By Lemma D.8, we get that with probability at least 0.9, for any $j \in [k]$ and any $f \in \mathrm{supp}(\widehat{H}) + f_j$, either $\widehat{G}_{\sigma,b}^{(j)} < \delta_1$ or $\widehat{G}_{\sigma,b}^{(j)} > 1 - \delta_1$. In other words, either $f_j + \mathrm{supp}(\widehat{H}) \subseteq \mathrm{supp}(\widehat{I})$ or $f_j + \mathrm{supp}(\widehat{H}) \cap \mathrm{supp}(\widehat{I}) = \emptyset$. Since $0 \in \mathrm{supp}(\widehat{H})$, we get that for any $f \in f_j + \mathrm{supp}(\widehat{H})$,

$$f \in \mathrm{supp}(\widehat{I}) \iff f_j \in \mathrm{supp}(\widehat{I}).$$

Hence, we have

$$(\widehat{x} * \widehat{H})(f) \cdot \widehat{I}(f) = \sum_{j \in [k]: f_j \in \mathrm{supp}(\widehat{I})} v_j \cdot \widehat{H}(f - f_j).$$

On the other hand,

$$\begin{aligned}(\widehat{x} \cdot \widehat{I}) * \widehat{H}(f) &= \sum_{j \in [k]: f_j \in \mathrm{supp}(\widehat{I})} v_j \cdot \delta_{f_j} * \widehat{H}(f) \\ &= \sum_{j \in [k]: f_j \in \mathrm{supp}(\widehat{I})} v_j \cdot \widehat{H}(f - f_j) \\ &= (\widehat{x} * \widehat{H})(f) \cdot \widehat{I}(f).\end{aligned}$$

Therefore,

$$\begin{aligned}(x \cdot H) * I(t) &= \int_{-\infty}^{\infty} (\widehat{x} * \widehat{H})(f) \cdot \widehat{I}(f) \cdot \exp(2\pi \mathbf{i} f t) \mathrm{d} f \\ &= \int_{-\infty}^{\infty} (\widehat{x} \cdot \widehat{I}) * \widehat{H}(f) \cdot \exp(2\pi \mathbf{i} f t) \mathrm{d} f \\ &= (x * I)(t) \cdot H(t),\end{aligned}$$

where the last step follows from the definition of Fourier transform.

The lemma is then proved.

□

F.2 Approximation error bounds

We analyze the approximation error due to replacing $\widehat{G}_{\sigma,b}^{(j)}$ with the ideal filter $I_{\sigma,b}^{(j)}$ defined by Eq. (18). The following lemma gives a point-wise error bound.

Lemma F.2. *Let δ_1 be defined as in Lemma C.4. Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2.*

For any $x \in \mathcal{F}_{k,F}$, we have that with probability 0.9, for any $j \in [B]$,

$$|(x \cdot H) * G_{\sigma,b}^{(j)}(t) - (x \cdot H) * I(t)| \lesssim \delta_1 \sqrt{T|S|} \cdot \|x(t)\|_T \quad \forall t \in \mathbb{R}.$$

Proof. Let $S := \mathrm{supp}(\widehat{x} * \widehat{H})$ be defined as the support set of $\widehat{x} * \widehat{H}$. Then $|S| \leq \Delta$.

We have that

$$\begin{aligned}& |(x \cdot H) * G_{\sigma,b}^{(j)}(t) - (x \cdot H) * I(t)| \qquad (19) \\ = & |(x \cdot H) * (G_{\sigma,b}^{(j)} - I)(t)| \\ = & \Big| \int_{-\infty}^{\infty} (\widehat{x} * \widehat{H})(f) \cdot (\widehat{G}_{\sigma,b}^{(j)} - \widehat{I})(f) \cdot e^{2\pi \mathbf{i} f t} \mathrm{d} f \Big| \\ \leq & \int_{-\infty}^{\infty} |(\widehat{x} * \widehat{H})(f) \cdot (\widehat{G}_{\sigma,b}^{(j)} - I)(f)| \mathrm{d} f \\ = & \int_{S} |(\widehat{x} * \widehat{H})(f) \cdot (\widehat{G}_{\sigma,b}^{(j)} - I)(f)| \mathrm{d} f \\ \leq & \int_{S} |(\widehat{x} * \widehat{H})(f) \cdot \delta_1| \mathrm{d} f \\ \leq & \delta_1 \sqrt{|S|} \cdot \sqrt{\int_{-\infty}^{\infty} |(\widehat{x} * \widehat{H})(f)|^2 \mathrm{d} f}\end{aligned}$$

$$
\begin{aligned}
&= \delta_1 \sqrt{|S|} \cdot \sqrt{\int_{-\infty}^{\infty} |(x \cdot H)(t)|^2 \mathrm{d}t} \\
&\lesssim \delta_1 \sqrt{T|S|} \cdot \|(x \cdot H)(t)\|_T \\
&\lesssim \delta_1 \sqrt{T|S|} \cdot \|x(t)\|_T
\end{aligned} \tag{20}
$$

where the first step is straight forward, the second step follows from the definition of Fourier transform, the third step follows from triangle equality, the forth step follows from the definition of S. For the fifth step, by Lemma D.8 that with probability at least 0.9, the Large Offset event does not happen (i.e., for any $f \in S$, either $\widehat{G}^{(j)}_{\sigma,b}(f) < \delta_1$ or $\widehat{G}^{(j)}_{\sigma,b}(f) > 1 - \delta_1$). Then, by Lemma C.4, we know that $-\delta_1 \leq \widehat{G}^{(j)}_{\sigma,b}(f) \leq 1$. Thus, we get that $|(\widehat{G}^{(j)}_{\sigma,b} - \widehat{I})(f)| \leq \delta_1$. The sixth step follows from Cauchy–Schwarz inequality, the seventh step follows from Parseval's theorem, the eighth step follows from Lemma E.9 Property IV and V, the last step follows from Lemma E.9 Property V. $\square$

The following lemma gives a T-norm bound for the approximation error.

Lemma F.3. *Let δ_1 be defined as in Lemma C.4. Let H be defined as in Definition E.5, $G^{(j)}_{\sigma,b}$ be defined as in Definition C.2.*

Then, for any $x \in \mathcal{F}_{k,F}$, with probability 0.9, for any $j \in [B]$,

$$
\int_{-\infty}^{\infty} |(x \cdot H) * I(t) - (x \cdot H) * G^{(j)}_{\sigma,b}(t)|^2 \mathrm{d}t \lesssim \delta_1^2 T \|x(t)\|_T^2.
$$

In particular,

$$
\|(x \cdot H) * I(t) - (x \cdot H) * G^{(j)}_{\sigma,b}(t)\|_T \lesssim \delta_1 \|x(t)\|_T.
$$

Proof. Let $S := \mathrm{supp}(\widehat{x} * \widehat{H})$ be defined as the support set of $\widehat{x} * \widehat{H}$.

We have that

$$
\begin{aligned}
& T\|(x \cdot H) * I(t) - (x \cdot H) * G^{(j)}_{\sigma,b}(t)\|_T^2 \\
&= \int_0^T |(x \cdot H) * I(t) - (x \cdot H) * G^{(j)}_{\sigma,b}(t)|^2 \mathrm{d}t \\
&\leq \int_{-\infty}^{\infty} |(x \cdot H) * I(t) - (x \cdot H) * G^{(j)}_{\sigma,b}(t)|^2 \mathrm{d}t \\
&\leq \int_{-\infty}^{\infty} |(\widehat{x} * \widehat{H})(f) \cdot (\widehat{I}(f) - \widehat{G}^{(j)}_{\sigma,b}(f))|^2 \mathrm{d}f \\
&= \int_S |(\widehat{x} * \widehat{H})(f) \cdot (I(f) - \widehat{G}^{(j)}_{\sigma,b}(f))|^2 \mathrm{d}f \\
&\leq \int_S |(\widehat{x} * \widehat{H})(f) \cdot \delta_1|^2 \mathrm{d}f \\
&\leq \int_{-\infty}^{\infty} |(\widehat{x} * \widehat{H})(f) \cdot \delta_1|^2 \mathrm{d}f \\
&= \delta_1^2 \int_{-\infty}^{\infty} |(x \cdot H)(t)|^2 \mathrm{d}t \\
&\lesssim \delta_1^2 T \|x(t)\|_T^2
\end{aligned}
$$

where the first step follows from the definition of the norm, the second step is straight forward, the third step follows from Parseval's theorem, the forth step follows from the definition of S, the fifth step follows from Lemma D.8 and Lemma C.4, the sixth step is straight forward, the seventh step follows from Parseval's theorem, the eighth step follows from Lemma E.9 Property IV and Property V.

□

G Concentration Property of the Filtered Signal

Recall that a frequency f^* is *heavy* if it satisfies the following condition:

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

In this section, we consider the filtered signal in a hashing bin that contains a heavy frequency; that is, $z(t) = (x^* \cdot H) * \widehat{G}_{\sigma,b}^{(j)}$ where $j = h_{\sigma,b}(f^*)$ is the index of the bin containing f^*. We will prove that $z(t)$ form a one-cluster signal around f^*, which means that in the frequency domain most energy are concentrated around f^*, and in the time domain, most energy are contained in the observation window $[0, T]$. The formal definition are given as follows:

Definition G.1 ((ε, Δ)-one-cluster signal)**.** *We say that a signal $z(t)$ is an (ε, Δ)-one-cluster signal around f_0 if and only if $z(t)$ and $\widehat{z}(f)$ satisfy the following two properties:*

$$\begin{aligned} \text{Property I} \quad &: \quad \int_{f_0-\Delta}^{f_0+\Delta} |\widehat{z}(f)|^2 \mathrm{d}f \geq (1-\varepsilon) \int_{-\infty}^{+\infty} |\widehat{z}(f)|^2 \mathrm{d}f \\ \text{Property II} \quad &: \quad \int_0^T |z(t)|^2 \mathrm{d}t \geq (1-\varepsilon) \int_{-\infty}^{+\infty} |z(t)|^2 \mathrm{d}t. \end{aligned}$$

We first prove the energy preservation in the time domain:

Lemma G.2 (Time domain energy preservation)**.** *Let $\Delta_h = |\mathrm{supp}(\widehat{H})|$. Let f^* satisfy*

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

and $\widehat{z} = \widehat{x^* \cdot H} \cdot \widehat{G}_{\sigma,b}^{(j)}$ where $j = h_{\sigma,b}(f^*)$. Suppose the Large Offset event does not happen. Then, we have that,

$$\int_{-\infty}^{+\infty} |z(t)|^2 \mathrm{d}t \leq 1.35 \int_0^T |z(t)|^2 \mathrm{d}t.$$

Proof. Let $I(f)$ be the ideal filter defined by Eq. (18).

We first have

$$\begin{aligned} \|z(t)\|_{L_2} &= \|(x^* \cdot H)(t) * G_{\sigma,b}^{(j)}(t)\|_{L_2} \\ &\leq \|(x^* \cdot H)(t) * I(t)\|_{L_2} + \|(x^* \cdot H)(t) * (I - G_{\sigma,b}^{(j)})(t)\|_{L_2}, \end{aligned}$$

where the second step follows from triangle inequality.

Then, we bound the two terms separately.

For the first term, by Lemma F.1, if the Large Offset event does not happen, we have that

$$(x^* \cdot H) * I(t) = (x^* * I)(t) \cdot H(t). \tag{21}$$

It implies that

$$\|(x^* \cdot H)(t) * I(t)\|_{L_2} = \|(x^* * I)(t) \cdot H(t)\|_{L_2}$$

Let $y(t) := (x^* * I)(t)$. It's easy to see that $y(y)$ is k-Fourier-sparse. Then, we have

$$\begin{aligned}
&\int_{-\infty}^{\infty} |y(t) \cdot H(t)|^2 \mathrm{d}t \\
= &\int_0^T |y(t) \cdot H(t)|^2 \mathrm{d}t + \int_{[-\infty,\infty]\setminus[0,T]} |y(t) \cdot H(t)|^2 \mathrm{d}t \\
\leq &\int_0^T |y(t) \cdot H(t)|^2 \mathrm{d}t + 0.1 \int_0^T |y(t)|^2 \mathrm{d}t \\
\leq &\ 1.1 \int_0^T |y(t)|^2 \mathrm{d}t \\
\leq &\ 1.3 \int_0^T |y(t) \cdot H(t)|^2 \mathrm{d}t, \qquad (22)
\end{aligned}$$

where the first step is straight forward, the second step follows from Lemma E.9 Property IV, the third step follows from Lemma E.9 Property V, the forth step follows from Lemma E.9 Property V. Hence,

$$\|(x^* \cdot H)(t) * I(t)\|_{L_2} = \|y(t) \cdot H(t)\|_{L_2} \leq \sqrt{1.3T} \cdot \|(x^* * I)(t) \cdot H(t)\|_T.$$

By Eq. (21) again, we can swap the order of I and H and obtain:

$$\|(x^* \cdot H)(t) * I(t)\|_{L_2} \leq \sqrt{1.3T} \cdot \|(x^* \cdot H)(t) * I(t)\|_T.$$

For the second term, by Lemma F.3, we have that

$$\|(x^* \cdot H) * I(t) - (x^* \cdot H) * G_{\sigma,b}^{(j)}(t)\|_{L_2} \lesssim \delta_1 \sqrt{T} \|x^*(t)\|_T. \qquad (23)$$

Therefore, we get that

$$\begin{aligned}
\|z(t)\|_{L_2} &\leq \sqrt{1.3T} \cdot \|(x^* \cdot H)(t) * I(t)\|_T + O(\delta_1 \sqrt{T} \|x^*(t)\|_T) \\
&\leq \sqrt{1.3T} \cdot \|(x^* \cdot H) * G_{\sigma,b}^{(j)}(t)\|_T + \sqrt{1.3T} \cdot \|(x^* \cdot H) * (I - G_{\sigma,b}^{(j)})(t)\|_T + O(\delta_1 \sqrt{T} \|x^*(t)\|_T) \\
&\leq \sqrt{1.3T} \cdot \|z(t)\|_T + O(\delta_1 \sqrt{T} \|x^*(t)\|_T),
\end{aligned}$$

where the second step follows from triangle inequality, and the last step follows from Lemma F.3 again.

We claim that the second term can be bounded by $o(1) \cdot \|z(t)\|_T$. Indeed, we have

$$\begin{aligned}
&\int_{-\infty}^{\infty} |(x^* \cdot H) * G_{\sigma,b}^{(j)}(t)|^2 \mathrm{d}t \\
= &\int_{-\infty}^{\infty} |(\widehat{x}^* * \widehat{H}) \cdot \widehat{G}_{\sigma,b}^{(j)}(f)|^2 \mathrm{d}f \\
\geq &\int_{f^*-\Delta_h}^{f^*+\Delta_h} |(\widehat{x}^* * \widehat{H}) \cdot \widehat{G}_{\sigma,b}^{(j)}(f)|^2 \mathrm{d}f \\
\gtrsim &\int_{f^*-\Delta_h}^{f^*+\Delta_h} |(\widehat{x}^* * \widehat{H})(f)|^2 \mathrm{d}f
\end{aligned}$$

$$
\begin{aligned}
&\gtrsim \int_{f^*-\Delta_h}^{f^*+\Delta_h} |(\widehat{x}^* * \widehat{H})(f)|^2 \mathrm{d}f \\
&\geq \frac{T\delta \|x^*\|_T^2}{k} \qquad (24)
\end{aligned}
$$

where the first step follows from Parseval's theorem, the second step is straightforward, the third step follows from Lemma D.8 and Lemma C.4 Property I, the forth step follows from our assumption that there exists a heavy frequency f^* hashing into the j-th bin, the fifth step follows from f^* satisfying

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

Thus, $\|x^*\|_T \leq O(\sqrt{k/(T\delta)})\|z(t)\|_{L_2}$ and we have

$$O(\delta_1\sqrt{T}\|x^*\|_T) = O\Big(\delta_1\sqrt{\frac{k}{\delta}}\|z(t)\|_{L_2}\Big) \leq O\Big(\sqrt{\frac{\delta}{k}}\Big)\|z(t)\|_{L_2} = o(1)\cdot\|z(t)\|_{L_2},$$

where the second step follows from $\delta_1 \leq \delta/k$.

Finally, we have

$$\|z(t)\|_{L_2} \leq \sqrt{1.3T}\cdot\|z(t)\|_T + o(1)\cdot\|z(t)\|_{L_2},$$

which implies that

$$\int_{-\infty}^{+\infty} |z(t)|^2 \mathrm{d}t \leq 1.35\int_0^T |z(t)|^2 \mathrm{d}t.$$

The lemma is then proved. □

We next show the frequency domain energy concentration in the following lemma. Together with Lemma G.2, we conclude that $z(t)$ is a one-cluster signal.

Lemma G.3 (Frequency domain energy concentration)**.** *Let x^* be a k-Fourier-sparse signal. Let $f^* \in [-F, F]$ satisfy the following property:*

$$\int_{f^*-\Delta_h}^{f^*+\Delta_h} |\widehat{x^* \cdot H}(f)|^2 \mathrm{d}f \geq T\mathcal{N}^2/k. \qquad (25)$$

Let σ, b be the parameter of the hashing function. Suppose that Large Offset event not happened and f^* is well-isolated. 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 $\widehat{z} = \widehat{x^* \cdot H} \cdot \widehat{G}_{\sigma,b}^{(j)}$. Then, we have

$$\int_{f^*-\Delta}^{f^*+\Delta} |\widehat{z}(f)|^2 \mathrm{d}f \geq 0.7\int_{-\infty}^{+\infty} |\widehat{z}(f)|^2 \mathrm{d}f.$$

Furthermore, $z(t)$ is a $(0.3, \Delta)$-one-cluster signal around f^*.

Proof. Define region $I_{f^*} = (f^* - \Delta, f^* + \Delta)$ with the complement $\overline{I_{f^*}} = (-\infty, \infty) \setminus I_{f^*}$. We have that

$$\int_{I_{f^*}} |\widehat{z}(f)|^2 \mathrm{d}f \geq (1 - \delta/k) \int_{I_{f^*}} |\widehat{x^* \cdot H}(f)|^2 \mathrm{d}f \geq (1 - o(1)) T\mathcal{N}^2/k$$

where the first step follows from Lemma D.9, the second step follows from Eq. (25).

On the other hand, f^* is well-isolated. Thus, by the definition of well-isolation (Definition D.4), we have that

$$\int_{\overline{I_{f^*}}} |\widehat{z}(f)|^2 \mathrm{d}f \lesssim \varepsilon \cdot T\mathcal{N}^2/k \leq 0.1 T\mathcal{N}^2/k.$$

Combining them together, we get that

$$\int_{f_0-\Delta}^{f_0+\Delta} |\widehat{z}(f)|^2 \mathrm{d}f \geq 0.7 \int_{-\infty}^{+\infty} |\widehat{z}(f)|^2 \mathrm{d}f$$

For the furthermore part, Lemma G.2 implies that

$$\int_0^T |z(t)|^2 \mathrm{d}t \geq (1 - 0.3) \int_{-\infty}^{+\infty} |z(t)|^2 \mathrm{d}t.$$

Hence, by Definition G.1, $z(t)$ is a $(0.3, \Delta)$-one-cluster. □

H Energy Bound for Filtered Fourier Sparse Signals

In this section, we prove an energy bound for the filtered signals $(x \cdot H) * G_{\sigma,b}^{(j)}$, which upper bounds the magnitude of any such signal at a point t by its energy in the time duration $[0, T]$. We first prove an energy bound for untruncated ideally filtered signals (see Section H.1). Then, we prove an energy bound for filtered signals (see Section H.2). In addition, we prove a technical claim (see Section H.3).

H.1 Energy bound for untruncated ideally filtered signals

In Section F, we show that the ideally filtered signal $(x \cdot H) * I_{\sigma,b}^{(j)}$, where $I_{\sigma,b}^{(j)}$ defined as Eq. (18) is the ideal filter, is close to the true filtered signal. Here, we further simplify the signal by ignoring the truncation filter $H(t)$, and prove an energy bound for the signals of the form $(x * I_{\sigma,b}^{(j)})(t)$:

Lemma H.1. *Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2 and the corresponding ideal filter $I = I_{\sigma,b}^{(b)}$ be defined as in Eq. (18). Let $D(t) := \mathrm{Uniform}([-1, 1])$.*

For any $x \in \mathcal{F}_{k,F}$, we have that with probability 0.6, for any $t \in (-1, 1)$

$$|(x * I)(t)|^2 \lesssim \min\Big\{\frac{k}{1 - |t|}, k^2\Big\} \cdot \|(x * I)(t)\|_D^2$$

Proof. Since $\widehat{(x * I)}(f) = \widehat{x} \cdot \widehat{I}(f)$ and x is k-Fourier-sparse, we know that $(x * \widehat{I})(t)$ is also a k-Fourier-sparse signal.

On the one hand, by the k-Fourier-sparse signal's location-dependent energy bound (Theorem B.2), we have

$$|(x * I)(t)|^2 \lesssim \frac{k}{1-|t|} \|(x * I)(t)\|_D^2 \tag{26}$$

On the other hand, by the location-independent energy bound (Theorem B.1), we have that

$$|(x * I)(t)|^2 \lesssim k^2 \|(x * I)(t)\|_D^2 \tag{27}$$

Combine Eq. (26) and Eq. (27) together, we prove the lemma:

$$|(x * I)(t)|^2 \lesssim \min\Big\{\frac{k}{1-|t|}, k^2\Big\} \cdot \|(x * I)(t)\|_D^2.$$

□

H.2 Energy bound for filtered signals

Based on Lemma H.1, we can relate the magnitude of the filtered signal with its own energy plus the original Fourier-sparse signal's energy.

Lemma H.2. *Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2 and the corresponding ideal filter $I = I_{\sigma,b}^{(b)}$ be defined as in Eq. (18).*

For any $x \in \mathcal{F}_{k,F}$, $j \in [B]$, and (σ, b) such that Large Offset event does not happen, let $z(t) = (x \cdot H) * G_{\sigma,b}^{(j)}(t)$. It holds that:

$$|z(t)|^2 \lesssim \min\Big\{\frac{k \cdot H(t)}{1-|2t/T-1|}, k^2\Big\} \cdot \|z(t)\|_T^2 + \delta_1 \|x(t)\|_T^2 \quad \forall t \in (-1,1).$$

Proof. Let $S := \mathrm{supp}(\widehat{x} * \widehat{H})$ be defined as the support set of $\widehat{x} * \widehat{H}$. Then $|S| \le \Delta$.

First, by the ideally untruncated filtered signal's energy bound (Lemma H.1), we have

$$\begin{aligned}
|(x * I)(t) \cdot H(t)|^2 &\lesssim H^2(t) \cdot \min\Big\{\frac{k}{1-|2t/T-1|}, k^2\Big\} \cdot \|(x * I)(t)\|_T^2 \\
&\lesssim \min\Big\{\frac{k \cdot H(t)}{1-|2t/T-1|}, k^2\Big\} \cdot \|(x * I)(t)\|_T^2,
\end{aligned} \tag{28}$$

where the second step follows from $H(t) \lesssim 1$ (Lemma E.9 Property I, II).

Then, we bound the magnitude of the ideal filtered signal as follows:

$$\begin{aligned}
|(x \cdot H) * I(t)|^2 &= |(x * I)(t) \cdot H(t)|^2 \\
&\lesssim \min\Big\{\frac{k \cdot H(t)}{1-|2t/T-1|}, k^2\Big\} \cdot \|(x * I)(t)\|_T^2 \\
&\lesssim \min\Big\{\frac{k \cdot H(t)}{1-|2t/T-1|}, k^2\Big\} \cdot (\|(x \cdot H) * G_{\sigma,b}^{(j)}(t)\|_T^2 + \delta_1^2 \|x(t)\|_T^2) \\
&\lesssim \min\Big\{\frac{k \cdot H(t)}{1-|2t/T-1|}, k^2\Big\} \cdot \|(x \cdot H) * G_{\sigma,b}^{(j)}(t)\|_T^2 + \delta_1 \|x(t)\|_T^2
\end{aligned} \tag{29}$$

where the first step follows from Lemma F.1, the second step follows from Eq. (28), the third step follows from Claim H.5, the forth step follows from $k^2 \delta_1 \le 1$.

Next, we consider the difference between the signals filtered by $G_{\sigma,b}^{(j)}(t)$ and $I(t)$:

$$\begin{aligned}|(x \cdot H) * G_{\sigma,b}^{(j)}(t) - (x \cdot H) * I(t)|^2 \leq & \ \delta_1^2 T|S| \cdot \|x(t)\|_T^2 \\ \leq & \ \delta_1 \cdot \|x(t)\|_T^2 \end{aligned} \tag{30}$$

where the first step follows from Lemma F.2, the second step follows from $\delta_1 T|S| \leq 1$.

Finally, we have that

$$\begin{aligned}|(x \cdot H) * G_{\sigma,b}^{(j)}(t)|^2 \leq & \ 2|(x \cdot H) * I(t)|^2 + 2|(x \cdot H) * G_{\sigma,b}^{(j)}(t) - (x \cdot H) * I(t)|^2 \\ \lesssim & \ |(x \cdot H) * I(t)|^2 + \delta_1 \|x(t)\|_T^2 \\ \lesssim & \ \min\Big\{\frac{k \cdot H(t)}{1 - |2t/T - 1|}, k^2\Big\} \cdot \|(x \cdot H) * \widehat{G}_{\sigma,b}^{(j)}(t)\|_T^2 + \delta_1 \|x(t)\|_T^2,\end{aligned}$$

where the first step follows from $(a+b)^2 \leq 2a^2 + 2b^2$, the second step follows from Eq. (30), the third step follows from Eq. (29).

The lemma is then proved.

□

The energy bound in Lemma H.2 not only depends on $\|z(t)\|_T$, but also on $\|x(t)\|_T$. The following lemma show that assuming the filtered signal contains a heavy frequency, $\|x(t)\|_T$ can be upper bounded by $\|z(t)\|_T$.

Lemma H.3. *Given $k \in \mathbb{Z}_+, F \in \mathbb{R}_+$. Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2. Let $x \in \mathcal{F}_{k,F}$ be any k-Fourier sparse signal. For $j \in [B]$ such that there exists a f^* satisfying: $j = h_{\sigma,b}(f^*)$ and*

$$\int_{f^*-\Delta_h}^{f^*+\Delta_h} |\widehat{x \cdot H}(f)|^2 \mathrm{d}f \geq T\mathcal{N}^2/k, \tag{31}$$

where $\mathcal{N}^2 \geq \delta \|x\|_T^2$ and $\Delta_h = |\mathrm{supp}(\widehat{H})|$.

For any (σ, b) that Large Offset event does not happen, we have that

$$\|(x \cdot H) * G_{\sigma,b}^{(j)}(t)\|_T^2 \gtrsim \frac{\delta \|x\|_T^2}{k}.$$

Proof. We have that

$$\begin{aligned}T\|(x \cdot H) * G_{\sigma,b}^{(j)}(t)\|_T^2 = & \ \int_0^T |(x \cdot H) * G_{\sigma,b}^{(j)}(t)|^2 \mathrm{d}t \\ \gtrsim & \ \int_{-\infty}^{\infty} |(x \cdot H) * G_{\sigma,b}^{(j)}(t)|^2 \mathrm{d}t \\ = & \ \int_{-\infty}^{\infty} |(\widehat{x} * \widehat{H}) \cdot \widehat{G}_{\sigma,b}^{(j)}(f)|^2 \mathrm{d}f \\ \geq & \ \int_{f^*-\Delta_h}^{f^*+\Delta_h} |(\widehat{x} * \widehat{H}) \cdot \widehat{G}_{\sigma,b}^{(j)}(f)|^2 \mathrm{d}f \\ \gtrsim & \ \int_{f^*-\Delta_h}^{f^*+\Delta_h} |(\widehat{x} * \widehat{H})(f)|^2 \mathrm{d}f\end{aligned}$$

$$\geq \frac{T\delta\|x\|_T^2}{k}$$

where the first step follows from the definition of norm, the second step follows from Lemma G.3, the third step follows from Parseval's theorem, the forth step is straight forward, the fifth step follows from Lemma D.9, the sixth step follows from Eq. (31).

$\square$

Lemma H.2 and Lemma H.3 implies the following energy bound:

Corollary H.4 (Energy bound for filtered signals)**.** *Given $k \in \mathbb{N}$ and $F \in \mathbb{R}_+$. Let $x \in \mathcal{F}_{k,F}$. Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2 with (σ, b) such that Large Offset event does not happen.*

For any $j \in [B]$, suppose there exists an f^* with $j = h_{\sigma,b}(f^*)$ satisfying:

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

where $\mathcal{N}^2 \geq \delta\|x\|_T^2$. Then, for $z(t) = (x \cdot H) * G_{\sigma,b}^{(j)}(t)$, it holds that:

$$|z(t)|^2 \lesssim \min\Big\{\frac{k \cdot H(t) + \delta}{1 - |2t/T - 1|}, k^2\Big\} \cdot \|z(t)\|_D^2 \quad \forall t \in (0, T).$$

Proof. We have that

$$\begin{aligned}
|z(t)|^2 &\lesssim \min\Big\{\frac{k \cdot H(t)}{1 - |2t/T - 1|}, k^2\Big\} \cdot \|z(t)\|_T^2 + \delta_1\|x(t)\|_T^2 \\
&\lesssim \min\Big\{\frac{k \cdot H(t)}{1 - |2t/T - 1|}, k^2\Big\} \cdot \|z(t)\|_T^2 + \delta^2 k^{-1}\|x(t)\|_T^2 \\
&\lesssim \min\Big\{\frac{k \cdot H(t)}{1 - |2t/T - 1|}, k^2\Big\} \cdot \|z(t)\|_T^2 + \delta\|(x \cdot H) * G_{\sigma,b}^{(j)}(t)\|_T^2 \\
&\lesssim \min\Big\{\frac{k \cdot H(t) + \delta}{1 - |2t/T - 1|}, k^2\Big\} \cdot \|z(t)\|_T^2
\end{aligned}$$

where the first step follows from Lemma H.2, the second step follows from $\delta_1 \leq \delta^2 k^{-1}$, the third step follows from Lemma H.3, the forth step is straight forward.

$\square$

H.3 Technical claim

Claim H.5. *Given $k \in \mathbb{Z}_+, F \in \mathbb{R}_+$. Let δ_1 be defined as the δ of Lemma C.4.Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2, and $I = I_{\sigma,b}^{(j)}$ be the ideal filter defined by Eq. (18).*

Then, for any $x \in \mathcal{F}_{k,F}$ and $j \in [B]$, with probability 0.6 over (σ, b), we have that

$$\|(x * I)(t)\|_T^2 \lesssim \|(x \cdot H) * \widehat{G}_{\sigma,b}^{(j)}(t)\|_T^2 + \delta_1^2\|x(t)\|_T^2.$$

Proof. We have that

$$\|(x * I)(t)\|_T^2 \lesssim \|(x * I)(t) \cdot H\|_T^2$$

$$
\begin{aligned}
&= \|(x \cdot H) * I(t)\|_T^2 \\
&\leq 2\|(x \cdot H) * G_{\sigma,b}^{(j)}(t)\|_T^2 + 2\|(x \cdot H) * G_{\sigma,b}^{(j)}(t) - (x \cdot H) * I(t)\|_T^2 \\
&\lesssim \|(x \cdot H) * G_{\sigma,b}^{(j)}(t)\|_T^2 + \delta_1^2 \|x(t)\|_T^2
\end{aligned}
$$

where the first step follows from $(x * I)(t)$ is a k-Fourier-sparse signal and Lemma E.9 Property V, the second step follows from Lemma F.1 conditioning on Large Offset event not happening, the third step follows from $(a+b)^2 \leq 2a^2 + 2b^2$, the forth step follows from Lemma F.3. $\square$

I Local-Test Signal

Recall that the filtered signal in the j-th bin of the HashToBins procedure can be written as $z(t) = (x \cdot H) * G_{\sigma,b}^{(j)}(t)$. The next step of the frequency estimation algorithm is to extract a significant frequency from $z(t)$ by considering a so-called *local-test signal*:

$$
d_z(t) := z(t) e^{2\pi f_0 \beta} - z(t+\beta), \tag{32}
$$

where $f_0 \in \mathrm{supp}(\widehat{x}^*)$, and $j = h_{\sigma,b}(f_0)$, where $\beta \in \mathbb{R}_+$ is a parameter such that $\beta \leq O(1/\Delta)$ with $\Delta = O(k \cdot |\mathrm{supp}(\widehat{H})|)$.

In this section, we will study some properties of $d_z(t)$ and its ideal versions (see Section I.1 and Section I.2) and derive an energy bound for it (See Section I.3).

I.1 Ideal local-test signal

In previous section, we've shown that ideal filter $I_{\sigma,b}^{(j)}$ can be used to approximate $G_{\sigma,b}^{(j)}$ such that the ideally filtered signal is close to the true filtered signal. We will show that under the ideal filter approximation, the *ideal local-test signal* is also close to the true local-test signal. More formally, we define the ideal filtered signal and the ideal local-test signal as follows:

$$
\begin{aligned}
z_I(t) &:= (x \cdot H) * I(t), \\
d_{I,z}(t) &:= z_I(t) e^{2\pi \mathbf{i} f_0 \beta} - z_I(t+\beta),
\end{aligned} \tag{33}
$$

The following lemma bounds the point-wise distance between $d_z(t)$ and $d_{I,z}(t)$.

Lemma I.1. *Let δ_1 be defined as in Lemma C.4. Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2 and $I = I_{\sigma,b}^{(j)}$ be the corresponding ideal filter as in Eq. (18).*

For any $x \in \mathcal{F}_{k,F}$ and (σ, b) such that Large Offset event does not happen, for any $j \in [B]$, let $z(t) = (x \cdot H) * G_{\sigma,b}^{(j)}(t)$, $d_z(t)$ be defined as Eq. (32), $z_I(t)$ and $d_{z,I}(t)$ be defined as Eq. (33).

Then, we have

$$
|d_z(t) - d_{I,z}(t)| \lesssim \delta_1 \sqrt{T|S|} \cdot \|x(t)\|_T \quad \forall t \in \mathbb{R}.
$$

Proof.

$$
\begin{aligned}
|d_z(t) - d_{I,z}(t)| &= |z(t) e^{2\pi \mathbf{i} f_0 \beta} - z_I(t) e^{2\pi \mathbf{i} f_0 \beta} - (z(t+\beta) - z_I(t+\beta))| \\
&\leq |z(t) e^{2\pi \mathbf{i} f_0 \beta} - z_I(t) e^{2\pi \mathbf{i} f_0 \beta}| + |z(t+\beta) - z_I(t+\beta)| \\
&= |z(t) - z_I(t)| + |z(t+\beta) - z_I(t+\beta)| \\
&\lesssim \delta_1 \sqrt{T|S|} \cdot \|x(t)\|_T,
\end{aligned}
$$

where the first step follows from the definition of $d_z(t)$ and $d_{I,z}(t)$, the second step follows from triangle inequality, the third step follows from $|e^{2\pi \mathbf{i} f_0 \beta}| = 1$, the forth step follows from Lemma F.2. $\square$

The following lemma bounds the L_2-distance between $d_z(t)$ and $d_{z,I}(t)$.

Lemma I.2. *Let δ_1 be defined as in Lemma C.4. Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2 and $I = I_{\sigma,b}^{(j)}$ be the corresponding ideal filter as in Eq. (18).*

For any $x \in \mathcal{F}_{k,F}$ and (σ, b) such that Large Offset event does not happen, for any $j \in [B]$, let $z(t) = (x \cdot H) * G_{\sigma,b}^{(j)}(t)$, $d_z(t)$ be defined as Eq. (32), $z_I(t)$ and $d_{z,I}(t)$ be defined as Eq. (33). Then,

$$\int_{-\infty}^{\infty} |d_{I,z}(t) - d_z(t)|^2 \mathrm{d}t \lesssim \delta_1^2 T \|x(t)\|_T^2$$

Proof. We first have that,

$$\begin{aligned}
&\int_{-\infty}^{\infty} |z_I(t) e^{2\pi \mathbf{i} f_0 \beta} - z(t) e^{2\pi \mathbf{i} f_0 \beta}|^2 \mathrm{d}t \\
= &\int_{-\infty}^{\infty} |z_I(t) - z(t)|^2 \mathrm{d}t \\
\leq &\ \delta_1^2 T \|x(t)\|_T^2,
\end{aligned} \tag{34}$$

where the first step follows from $|e^{2\pi \mathbf{i} f_0 \beta}| = 1$, the second step follows from Lemma F.3.

Then, we complete the proof as follows:

$$\begin{aligned}
&\int_{-\infty}^{\infty} |d_{I,z}(t) - d_z(t)|^2 \mathrm{d}t \\
= &\int_{-\infty}^{\infty} |z_I(t) e^{2\pi \mathbf{i} f_0 \beta} - z(t) e^{2\pi \mathbf{i} f_0 \beta} - (z_I(t+\beta) - z(t+\beta))|^2 \mathrm{d}t \\
\leq &\ 2 \int_{-\infty}^{\infty} |z_I(t) e^{2\pi \mathbf{i} f_0 \beta} - z(t) e^{2\pi \mathbf{i} f_0 \beta}|^2 \mathrm{d}t + 2 \int_{-\infty}^{\infty} |z_I(t+\beta) - z(t+\beta)|^2 \mathrm{d}t \\
\lesssim &\ \delta_1^2 T \|x(t)\|_T^2 + \int_{-\infty}^{\infty} |z_I(t+\beta) - z(t+\beta)|^2 \mathrm{d}t \\
\lesssim &\ \delta_1^2 T \|x(t)\|_T^2
\end{aligned}$$

where the first step follows from the definition of $d_{I,z}(t)$ and $d_z(t)$, the second step follows from $(a+b)^2 \leq 2a^2 + 2b^2$, the third step follows from Eq. (34), the forth step follows from Lemma F.3. $\square$

I.2 Ideal post-truncated local-test signal

It is still difficult to directly study the energy bound for $d_{z,I}(t)$. In this section, we further simplify the ideally filtered signal by removing the H filter and consider the untruncted ideally filtered signal $(x * I)(t)$. Then, in the local-test signal, we perform a post-truncation. More specifically, the untruncated ideally filtered signal and the *ideal post-truncated local-test signal* are defined as follows:

$$x_I(t) := (x * I)(t),$$

$$d_{I,x}(t) := x_I(t) \cdot H(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta) \cdot H(t+\beta). \tag{35}$$

Intuitively, $d_{I,x}(t)$ can be viewed as swapping the order of the I and H filters in $d_{I,z}(t)$.

The following lemma shows that $d_{I,z}(t)$ and $d_{I,x}(t)$ are actually the same!

Lemma I.3. *Let δ_1 be defined as in Lemma C.4. Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2 and $I = I_{\sigma,b}^{(j)}$ be the corresponding ideal filter as in Eq. (18).*

For any $x \in \mathcal{F}_{k,F}$, and (σ, b) such that Large Offset event does not happen, let $z_I(t)$ and $d_{z,I}(t)$ be defined as Eq. (33), $x_I(t)$ and $d_{x,I}(t)$ be defined as Eq. (35).

Then, we have

$$d_{I,z}(t) = d_{I,x}(t) \quad \forall t \in \mathbb{R}.$$

Proof. We have that

$$\begin{aligned}
d_{I,z}(t) &= z_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - z_I(t+\beta) \\
&= x_I(t) \cdot H(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - z_I(t+\beta) \\
&= x_I(t) \cdot H(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta) \cdot H(t+\beta) \\
&= d_{I,x}(t),
\end{aligned}$$

where the first step follows from the definition of $d_{I,z}(t)$, the second step follows from Lemma F.1, the third step follows from Lemma F.1, the last step follows from the definition of $d_{I,x}(t)$. $\square$

The structure of $d_{I,x}(t)$ makes it easy to study its magnitude at any "good point":

Lemma I.4. *Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2 and $I = I_{\sigma,b}^{(j)}$ be the corresponding ideal filter as in Eq. (18). Let $U := \{t_0 \in \mathbb{R} \mid H(t) > 1 - \delta_1 \ \forall t \in [t_0, t_0 + \beta]\}$.*

For any $x \in \mathcal{F}_{k,F}$, and (σ, b) such that Large Offset event does not happen, let $x_I(t), d_{x,I}(t)$ be defined as Eq. (35). Then, we have

$$|d_{I,x}(t)| \lesssim \left| x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta) \right| + \delta_1 k \|x_I(t)\|_T \quad \forall t \in U.$$

Proof. First, for any $t \in U$,

$$\begin{aligned}
|x_I(t) \cdot H(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta}| &= |x_I(t) \cdot H(t) - x_I(t)| \\
&= |x_I(t)| \cdot |1 - H(t)| \\
&\leq \delta_1 |x_I(t)| \\
&\lesssim \delta_1 k \|x_I(t)\|_T
\end{aligned} \tag{36}$$

where the first step follows from $|e^{2\pi \mathbf{i} f_0 \beta}| = 1$, the second step is straight forward, the third step follows from $H(t) \leq 1$ (Lemma E.9 Property I, II) and $\forall t \in U, H(t) > 1 - \delta_1$, and the last step follows from Lemma H.1.

Second, for any $t \in U$,

$$\begin{aligned}
|x_I(t+\beta) - x_I(t+\beta) \cdot H(t+\beta)| &= |x_I(t+\beta)| \cdot |1 - H(t+\beta)| \\
&\leq \delta_1 |x_I(t+\beta)| \\
&\lesssim \delta_1 k \|x_I(t)\|_T
\end{aligned} \tag{37}$$

where the first step is straight forward, the second step follows from $H(t) \leq 1$ (Lemma E.9 Property I, II) and $\forall t \in U, H(t+\beta) > 1-\delta_1$, the last step follows from Lemma H.1.

Combining them together, we have that for any $t \in U$,

$$\begin{aligned}
|d_{I,x}(t)| &= |x_I(t) \cdot H(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta) \cdot H(t+\beta)| \\
&\leq |x_I(t) \cdot H(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta}| + |x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta)| \\
&\quad + |x_I(t+\beta) - x_I(t+\beta) \cdot H(t+\beta)| \\
&\lesssim |x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta)| + |x_I(t+\beta) - x_I(t+\beta) \cdot H(t+\beta)| + \delta_1 k \|x_I(t)\|_T \\
&\lesssim |x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta)| + \delta_1 k \|x_I(t)\|_T
\end{aligned}$$

where the first step follows from the definition of $d_{I,x}(t)$, the second step follows from triangle inequality, the third step follows Eq. (36), the forth step follows from Eq. (37).

□

Furthermore, we can show that the ideal post-truncated local-test signal is close to the ideal local-test signal without truncation on most of "good points".

Lemma I.5. *Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2 and $I = I_{\sigma,b}^{(j)}$ be the corresponding ideal filter as in Eq. (18). Let $U := \{t_0 \in \mathbb{R} \mid H(t) > 1-\delta_1, \forall t \in [t_0, t_0+\beta]\}$. Let $D_U(t) := \mathrm{Uniform}(U)$ and $D_{U+\beta}(t) := \mathrm{Uniform}(U+\beta)$.*

For any $x \in \mathcal{F}_{k,F}$, and (σ, b) such that Large Offset event does not happen, let $x_I(t), d_{x,I}(t)$ be defined as Eq. (35). Then, we have

$$\|d_{I,x}(t) - (x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta))\|_{D_U} \lesssim \delta_1 \|x_I(t)\|_T.$$

Proof. First,

$$\begin{aligned}
\|x_I(t) \cdot H(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta}\|_{D_U} &= \|x_I(t) \cdot H(t) - x_I(t)\|_{D_U} \\
&\leq \max\nolimits_{t \in U}\{|1-H(t)|\} \cdot \|x_I(t)\|_{D_U} \\
&\leq \delta_1 \cdot \|x_I(t)\|_{D_U} \\
&\lesssim \delta_1 \cdot \sqrt{\frac{T}{|U|}} \|x_I(t)\|_T \\
&\lesssim \delta_1 \cdot \|x_I(t)\|_T
\end{aligned} \tag{38}$$

where the first step follows from $|e^{2\pi \mathbf{i} f_0 \beta}| = 1$, the second step is straight forward, the third step follows from $H(t) \leq 1$ (Lemma E.9 Property I, II) and $\forall t \in U, H(t) > 1-\delta_1$, the forth step follows from the definition of the norm

$$\|x(t)\|_{D_U}^2 = \frac{1}{|U|} \int_U |x(t)|^2 \mathrm{d}t \leq \frac{1}{|U|} \int_{[0,T]} |x(t)|^2 \mathrm{d}t = \frac{T}{|U|} \|x(t)\|_T^2,$$

and the last step follows from Lemma H.1.

Second,

$$\begin{aligned}
\|x_I(t+\beta) - x_I(t+\beta) \cdot H(t+\beta)\|_{D_U} &\leq \max\nolimits_{t \in U}\{|1-H(t+\beta)|\} \cdot \|x_I(t+\beta)\|_{D_U} \\
&\leq \delta_1 \cdot \|x_I(t)\|_{D_{U+\beta}} \\
&\lesssim \delta_1 \cdot \frac{1}{|U+\beta|} \|x_I(t)\|_{D_1}
\end{aligned}$$

$$\lesssim \delta_1 \cdot \|x_I(t)\|_{D_1} \tag{39}$$

where the first step is straight forward, the second step follows from $H(t) \leq 1$ (Lemma E.9 Property I, II) and $\forall t \in U, H(t+\beta) > 1-\delta_1$, the third step follows from the definition of the norm, the forth step follows from $|U+\beta| = |U| \gtrsim 1$.

Then, we have that,

$$\begin{aligned}
&\|d_{I,x}(t) - (x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta))\|_{D_U} \\
= &\|x_I(t) \cdot H(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta) \cdot H(t+\beta) - (x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta))\|_{D_U} \\
\leq &\|x_I(t) \cdot H(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta}\|_{D_U} + \|x_I(t+\beta) \cdot H(t+\beta) - x_I(t+\beta)\|_{D_U} \\
\lesssim &\delta_1 \cdot \|x_I(t)\|_{D_1} + \|x_I(t+\beta) \cdot H(t+\beta) - x_I(t+\beta)\|_{D_U} \\
\lesssim &\delta_1 \cdot \|x_I(t)\|_{D_1}
\end{aligned}$$

where the first step follows from the definition of $d_{I,x}(t)$, the second step follows from triangle inequality, the third step follows from Eq. (38), the forth step follows from Eq. (39).

□

I.3 Energy bound for local-test signals

In this section, we prove the following lemma, which gives an energy bound for local-test signals.

Lemma I.6 (Energy bound for local-test signals)**.** *Let H be defined as in Definition E.5, $G^{(j)}_{\sigma,b}$ be defined as in Definition C.2. Let U, D_U be defined as in Lemma I.5.*

For any $x \in \mathcal{F}_{k,F}$, and (σ, b) such that Large Offset event does not happen, let $z(t) = (x \cdot H) * G^{(j)}_{\sigma,b}(t)$ and $d_z(t)$ be defined as Eq. (32). Then, we have

$$|d_z(t)|^2 \lesssim \min\Big\{\frac{k}{1-|2t/T-1|}, k^2\Big\} \cdot \|d_z(t)\|^2_{D_U} + \delta_1 \|x(t)\|^2_T \quad \forall t \in U.$$

Proof. Let $I = I^{(j)}_{\sigma,b}$ be the corresponding ideal filter as in Eq. (18). Let $S := \mathrm{supp}(\widehat{x} * \widehat{H})$ be the support set of $\widehat{x} * \widehat{H}$. We have $|S| \leq \Delta$.

Let $z_I(t)$, $d_{z,I}(t)$ be defined as in Lemma I.1 and $x_I(t)$, $d_{I,x}(t)$ be defined as in Lemma I.3.

Before proving the energy bound for $|d_z(t)|$, we first consider the signal $x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta)$. By Fourier transformation, we know that its Fourier coefficient of a frequency f is:

$$\widehat{x}_I(f) e^{2\pi \mathbf{i} f_0 \beta} - \widehat{x}_I(f) e^{2\pi \mathbf{i} f \beta} = \widehat{x}(f) \cdot \widehat{I}(f) e^{2\pi \mathbf{i} f_0 \beta} - \widehat{x}(f) \cdot \widehat{I}(f) e^{2\pi \mathbf{i} f \beta}$$

Thus, $x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta)$ is at most k-Fourier-sparse.

Let $[L, R] := U$. By Fourier-sparse signals' energy bound (Theorem B.2 and Theorem B.1), we have

$$\begin{aligned}
|x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta)|^2 &\lesssim \min\Big\{\frac{k}{\min\{R-t, t-L\}}, k^2\Big\} \cdot \|x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta)\|^2_{D_U} \\
&\lesssim \min\{\frac{k}{1-|2t/T-1|}, k^2\} \cdot \|x_I(t) \cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta)\|^2_{D_U}
\end{aligned} \tag{40}$$

where the first step follows from applying Theorem B.2 with $x(t) = x(Tt/2 + T/2)$ and applying Theorem B.1 with $T = |U|, x(t) = x(t+L)$, the second step follows from $[-1 + 0.5/k, 1 - 0.5/k] \subseteq$

$[L,R]$, which implies that $k(\min\{R-t,t-L\})^{-1} \lesssim k(1-|2t/T-1|)^{-1}$ for any $|t| \in [L+1/k, R-1/k]$. Moreover, for any $|t| \in [L, L+1/k] \cup [R-1/k, R]$, $k^2 \lesssim k(1-|2t/T-1|)^{-1}$.

The RHS can be upper bounded by:

$$\begin{aligned}
\|x_I(t)\cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta)\|_{D_U}^2 &\le 2\|d_{I,x}(t)\|_{D_U}^2 + 2\|d_{I,x}(t) - (x_I(t)\cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta))\|_{D_U}^2 \\
&\lesssim \|d_{I,x}(t)\|_{D_U}^2 + \delta_1^2 \|x_I(t)\|_T^2 \\
&= \|d_{I,z}(t)\|_{D_U}^2 + \delta_1^2 \|x_I(t)\|_T^2 \\
&\lesssim \|d_z(t)\|_{D_U}^2 + \|d_{I,z}(t) - d_z(t)\|_{D_U}^2 + \delta_1^2 \|x_I(t)\|_T^2 \\
&\lesssim \|d_z(t)\|_{D_U}^2 + \|d_{I,z}(t) - d_z(t)\|_{D_U}^2 + \delta_1^2 \|x(t)\|_T^2
\end{aligned} \tag{41}$$

where the first step follows from $(a+b)^2 \le 2a^2+2b^2$, the second step follows from Lemma I.5, the third step follows from Lemma I.3, the forth step follows from $(a+b)^2 \le 2a^2+2b^2$, the last step follows from Claim I.7. For the second term, we have that

$$\begin{aligned}
&\|d_{I,z}(t) - d_z(t)\|_{D_U}^2 \\
\lesssim\ & \frac{1}{|U|}\int_U |d_{I,z}(t) - d_z(t)|^2 \mathrm{d}t \\
\lesssim\ & \frac{1}{|U|}\int_{-\infty}^{\infty} |d_{I,z}(t) - d_z(t)|^2 \mathrm{d}t \\
\lesssim\ & \frac{1}{|U|}\delta_1^2 \|x(t)\|_T^2 \\
\lesssim\ & \delta_1^2 \|x(t)\|_T^2
\end{aligned}$$

where the first step follows from the definition of the norm, second step is straight forward, the third step follows from Lemma I.2 with appropriate scaling, the forth step follows from $|U| \gtrsim 1$. Hence,

$$\|x_I(t)\cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta)\|_{D_U}^2 \lesssim \|d_z(t)\|_{D_U}^2 + \delta_1^2 \|x(t)\|_T^2. \tag{42}$$

Therefore, we have that

$$\begin{aligned}
|d_{I,z}(t)|^2 &= |d_{I,x}(t)|^2 \\
&\lesssim (|x_I(t)\cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta)| + \delta_1 k \|x_I(t)\|_T)^2 \\
&\lesssim |x_I(t)\cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta)|^2 + \delta_1^2 k^2 \|x_I(t)\|_T^2 \\
&\lesssim |x_I(t)\cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta)|^2 + \delta_1 \|x_I(t)\|_T^2 \\
&\lesssim \min\{\frac{k}{1-|2t/T-1|}, k^2\} \cdot \|x_I(t)\cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta)\|_{D_U}^2 + \delta_1 \|x_I(t)\|_T^2 \\
&\lesssim \min\{\frac{k}{1-|2t/T-1|}, k^2\} \cdot \|x_I(t)\cdot e^{2\pi \mathbf{i} f_0 \beta} - x_I(t+\beta)\|_{D_U}^2 + \delta_1 \|x(t)\|_T^2 \\
&\lesssim \min\{\frac{k}{1-|2t/T-1|}, k^2\} \cdot (\|d_z(t)\|_{D_U}^2 + \delta_1^2 \|x(t)\|_T^2) + \delta_1 \|x(t)\|_T^2 \\
&\lesssim \min\{\frac{k}{1-|2t/T-1|}, k^2\} \cdot \|d_z(t)\|_{D_U}^2 + \delta_1 \|x(t)\|_T^2,
\end{aligned} \tag{43}$$

where the first step follows from Lemma I.3, the second step follows from Lemma I.4, the third step follows from $(a+b)^2 \le 2a^2+2b^2$, the forth step follows from $\delta_1 k^2 \le 1$, the fifth step follows from Eq. (40), the six step follows from Claim I.7, the seventh step follows from Eq. (42), the last step follows from $\delta_1 k^2 \lesssim 1$.

Finally, we have

$$\begin{aligned}
|d_z(t)|^2 &\le 2|d_z(t) - d_{I,z}(t)|^2 + 2|d_{I,z}(t)|^2 \\
&\le 2\delta_1^2 T|S| \cdot \|x(t)\|_T^2 + 2|d_{I,z}(t)|^2 \\
&\le 2\delta_1 \cdot \|x(t)\|_T^2 + 2|d_{I,z}(t)|^2 \\
&\lesssim \delta_1 \cdot \|x(t)\|_T^2 + \min\{\frac{k}{1-|2t/T-1|}, k^2\} \cdot \|d_z(t)\|_{D_U}^2
\end{aligned}$$

where the first step follows from $(a+b)^2 \le 2a^2 + 2b^2$, the second step follows from Lemma I.1, the third step follows from $\delta_1 T|S| \le 1$, the forth step follows from Eq. (43).

The lemma is then proved.

□

Claim I.7 (Energy Reduction by Ideal Filter)**.** *Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2 and $I = I_{\sigma,b}^{(j)}$ be the corresponding ideal filter as in Eq. (18).*

For any $x \in \mathcal{F}_{k,F}$, for any (σ, b) such that Large Offset event does not happen, then we have

$$\|(x * I)(t)\|_T \lesssim \|x(t)\|_T$$

Proof. Let $S = \mathrm{supp}(\widehat{x} * \widehat{H})$. We have that

$$\begin{aligned}
T\|(x * I)(t)\|_T &\lesssim T\|(x * I)(t) \cdot H(t)\|_T \\
&= \int_0^T |(x * I)(t) \cdot H(t)|^2 \mathrm{d}t \\
&\le \int_{-\infty}^{\infty} |(x * I)(t) \cdot H(t)|^2 \mathrm{d}t \\
&= \int_{-\infty}^{\infty} |(\widehat{x} \cdot \widehat{I})(f) * \widehat{H}(f)|^2 \mathrm{d}f \\
&= \int_S |(\widehat{x} \cdot \widehat{I})(f) * \widehat{H}(f)|^2 \mathrm{d}f \\
&= \int_S |\widehat{x}(f) * \widehat{H}(f)|^2 \mathrm{d}f \\
&\le \int_{-\infty}^{\infty} |\widehat{x}(f) * \widehat{H}(f)|^2 \mathrm{d}f \\
&= \int_{-\infty}^{\infty} |x \cdot H(t)|^2 \mathrm{d}t \\
&\lesssim \int_0^T |x(t)|^2 \mathrm{d}t \\
&= T\|x(t)\|_T^2
\end{aligned}$$

where the first step follows from Lemma E.9 Property V, the second step follows from the definition of the norm, the third step is straight forward, the forth step follows from Parseval's theorem, the fifth and sixth steps follow from Large Offset event not happening, the seventh step is straight forward, the eighth step follows from Parseval's theorem, the ninth step follows from Lemma E.9 Property IV and VI, the last step follows from the definition of the norm.

□

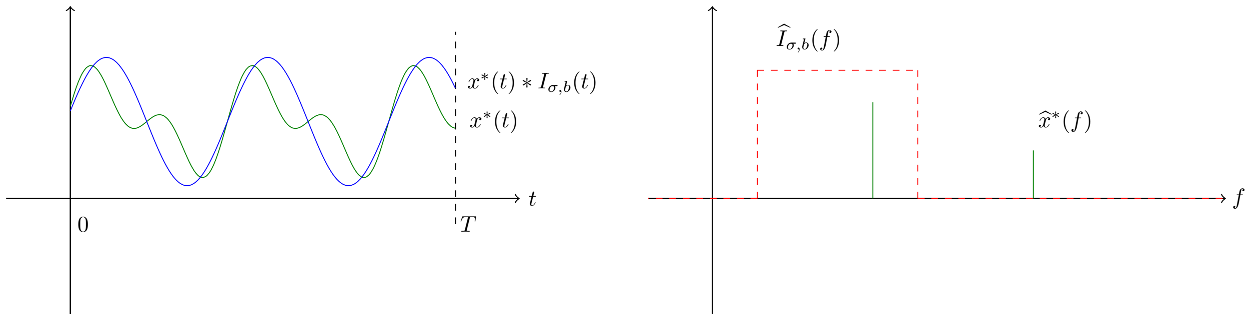


Figure 11: An illustration of the energy reduction by ideal filter. $I_{\sigma,b}$ is the ideal filter and $x^*(t)$ is a Fourier sparse signal. The energy of $x^*(t)$ in duration $[0, T]$ is reduced by applying the ideal filter, i.e., $\|x^* * I_{\sigma,b}(t)\|_T \lesssim \|x^*(t)\|_T$.

J Empirical Energy Estimation

The goal of this section is to show how to estimate a signal's energy using a few samples. We start with a general sampling and reweighing method (see Section J.1). Then, combining with the energy bounds derived in previous section, we obtain sample-efficient energy estimation methods for Fourier-sparse signals and filtered signals (see Section J.2). We further extend our methods to estimate the energy of filtered signals and local-test signals within a *sub-interval* in the time duration (see Section J.3). Finally, we prove several technical lemmas (see Section J.4).

Throughout this section, for the convenience, we use a slightly different notation for the T-norm:

$$\|z\|_T^2 := \frac{1}{2T}\int_{-T}^{T} |z(t)|^2 \mathrm{d}t.$$

This results of using this T-norm is equivalent with the result of the norm taking on $[0, T]$, since we can always re-scaling the signal and transform the result into the new T-norm result.

J.1 Sampling and reweighing

In this section, we provide a generic sample-efficient method for estimating the energy of any function using discrete samples with proper weights.

Lemma J.1. *Let $k \in \mathbb{N}_+$ and D be a probability distribution such that $\int_{-T}^{T} D(t)\mathrm{d}t = 1$. For any $\varepsilon, \rho \in (0, 1)$ and function $z : \mathbb{R} \to \mathbb{C}$, let $S_D = \{t_1, \cdots, t_s\}$ be a set of i.i.d. samples from D of size*

$$s \geq \Big(\max_{t\in[-T,T]} \frac{|z(t)|^2}{D(t)}\Big) \cdot O\Big(\frac{\log(1/\rho)}{\varepsilon^2 T \|z(t)\|_T^2}\Big).$$

Let the weight vector $w \in \mathbb{R}^s$ be defined by $w_i := 1/(2TsD(t_i))$ for $i \in [s]$.

Then with probability at least $1 - \rho$, we have

$$(1 - \varepsilon)\|z(t)\|_T^2 \leq \|z(t)\|_{S_D,w}^2 \leq (1 + \varepsilon)\|z(t)\|_T^2,$$

where $\|z\|_T^2 := \frac{1}{2T}\int_{-T}^{T} |z(t)|^2 \mathrm{d}t$.

Proof. Let $M := \max_{t\in[-T,T]} \frac{|z(t)|^2}{D(t)}$. Let $z_D(t) := \frac{1}{M}\frac{|z(t)|^2}{D(t)}$. By applying Chernoff bound (Lemma A.1) for the random variables $z_D(t_1), \ldots, z_D(t_s)$, we get that,

$$\Pr_{t_i\sim D}\Big[\Big|\sum_{i=1}^{s} z_D(t_i) - \mu\Big| \leq \varepsilon\mu\Big] \geq 1 - 2\exp(-\varepsilon^2\mu/3), \tag{44}$$

where $\mu := \sum_{i=1}^s \mathbb{E}_{t_i \sim D}[z_D(t_i)] = s \cdot \mathbb{E}_{t\sim D}[z_D(t)]$.

We first consider the expectation:

$$\begin{aligned}
\mathop{\mathbb{E}}_{t\sim D}[z_D(t)] =& \int_{-T}^{T} D(t) \cdot \frac{1}{M} \frac{|z(t)|^2}{D(t)} \mathrm{d}t \\
=& \frac{1}{M} \int_{-T}^{T} |z(t)|^2 \mathrm{d}t \\
=& \frac{2T}{M} \|z(t)\|_T^2
\end{aligned}$$

where the first step follows from the definition of expectation, the second step is straightforward, the third step follows from the definition of the norm. Thus,

$$\mu = s \cdot \mathop{\mathbb{E}}_{t\sim D}[z_D(t)] = \frac{2Ts}{M} \|z(t)\|_T^2. \tag{45}$$

Then, we consider the sum of samples:

$$\begin{aligned}
\sum_{i=1}^{s} z_D(t_i) =& \sum_{i=1}^{s} \frac{1}{M} \frac{|z(t_i)|^2}{D(t_i)} \\
=& \sum_{i=1}^{s} \frac{2w_i Ts}{M} |z(t_i)|^2 \\
=& \frac{2Ts}{M} \|z(t)\|_{S_D,w}^2
\end{aligned} \tag{46}$$

where the first step follows from the definition of z_D, the second step follows from the definition of w_i, the last step follows from the definition of the norm.

Putting Eqs. (44) - (46) together, we get that with probability at least $1 - 2\exp(-\varepsilon^2 \mu/3)$,

$$\Big| \frac{2Ts}{M} \|z(t)\|_{S_D,w}^2 - \frac{2Ts}{M} \|z(t)\|_T^2 \Big| \leq \varepsilon \cdot \frac{2Ts}{M} \|z(t)\|_T^2,$$

which can be simplified at follows:

$$|\|z(t)\|_{S_D,w}^2 - \|z(t)\|_T^2| \leq \varepsilon \cdot \|z(t)\|_T^2.$$

Finally, we need the success probability to be at least $1 - \rho$, which requires that:

$$\begin{aligned}
1 - 2\exp\Big(- \frac{\varepsilon^2}{3} \frac{2Ts}{M} \|z(t)\|_T^2 \Big) =& 1 - 2\exp\Big(- \frac{\varepsilon^2}{3} \frac{2Ts}{\cdot \max_{t\in[-T,T]} \{|z(t)|^2 / D(t)\}} \|z(t)\|_T^2 \Big) \\
\geq& 1 - \rho.
\end{aligned}$$

Hence, we need the sample complexity s to be at least

$$s \geq \Big(\max_{t\in[-T,T]} \frac{|z(t)|^2}{D(t)} \Big) \cdot O\Big(\frac{\log(1/\rho)}{\varepsilon^2 T \|z(t)\|_T^2} \Big).$$

$\square$

J.2 Energy estimation for Fourier-sparse signals and filtered signals

The goal of this section is to apply Lemma J.1 for Fourier-sparse signals and filtered signals.

The following lemma defines the sampling distribution:

Lemma J.2. *For $k \in \mathbb{N}_+$, define a probability distribution D as follows:*

$$D(t) := \begin{cases} c \cdot (1-|t/T|)^{-1} T^{-1}, & \text{for } |t| \leq T(1-1/k) \\ c \cdot k T^{-1}, & \text{for } |t| \in [T(1-1/k), T] \end{cases} \tag{47}$$

where $c = \Theta(\log(k)^{-1})$ is a normalization factor such that $\int_{-T}^{T} D(t)\mathrm{d}t = 1$. Then, D is well-defined.

Proof. We justify that D can be normalized with $c = \Theta(\log(k)^{-1}))$. By the condition $\int_{-T}^{T} D(t)\mathrm{d}t = 1$, we have

$$2\int_0^{T(1-1/k)} \frac{c}{(1-|t/T|)T}\mathrm{d}t + 2\int_{T(1-1/k)}^{T} c \cdot \frac{k}{T}\mathrm{d}t = 1,$$

which implies that

$$\begin{aligned} c^{-1} &= 2\int_0^{T(1-1/k)} \frac{1}{(1-|t/T|)T}\mathrm{d}t + 2\int_{T(1-1/k)}^{T} \frac{k}{T}\mathrm{d}t \\ &\eqsim \log(k) + 1 \\ &= \Theta(\log(k)). \end{aligned}$$

Thus, we get that $c = \Theta(\log(k)^{-1})$.

□

The following lemma gives the sampling complexity for estimating the energy of a Fourier-sparse signal. The main idea is to apply the energy bounds in Section B.

Lemma J.3 (Energy estimation for Fourier-sparse signals)**.** *Let D be the probability distribution defined as Eq. (47). Let $x \in \mathcal{F}_{k,F}$. For any $\varepsilon, \rho \in (0,1)$, let $S_D = \{t_1, \cdots, t_s\}$ be a set of i.i.d. samples from $D(t)$ of size $s \geq O(\varepsilon^{-2} k \log(k) \log(1/\rho))$. Let the weight vector $w \in \mathbb{R}^s$ be defined by $w_i := 1/(2TsD(t_i))$ for $i \in [s]$.*

Then with probability at least $1-\rho$, we have

$$(1-\varepsilon)\|x(t)\|_T^2 \leq \|x(t)\|_{S_D,w}^2 \leq (1+\varepsilon)\|x(t)\|_T^2.$$

Proof. By applying Lemma J.1, we have that the desired result satisfy when

$$s \geq \Big(\max_{t\in[-T,T]} \frac{|x(t)|^2}{D(t)}\Big) \cdot O\Big(\frac{\log(1/\rho)}{\varepsilon^2 T \|x(t)\|_T^2}\Big).$$

By Fourier-sparse signals' energy bound (Theorem B.1 and Theorem B.2 with $x(t) = x(T \cdot t)$), we have that

$$|x(t)|^2 \lesssim \min\Big\{\frac{k}{1-|t/T|}, k^2\Big\} \cdot \|x(t)\|_T^2 \quad \forall t \in [-T, T]. \tag{48}$$

Thus,

$$\max_{t\in[-T,T]} \frac{|x(t)|^2}{D(t)}$$

$$
\begin{aligned}
&\lesssim \max_{t\in[-T,T]} \min\Big\{\frac{k}{1-|t/T|}, k^2\Big\}\cdot \frac{\|x(t)\|_T^2}{D(t)}\\
&\lesssim \max_{t\in[-T,T]} \min\Big\{\frac{k}{1-|t/T|}\frac{T(1-|t/T|)}{c}, k^2\frac{T}{ck}\Big\}\cdot \|x(t)\|_T^2\\
&= kT\|x(t)\|_T^2/c\\
&\simeq k\log(k)T\|x(t)\|_T^2, \qquad (49)
\end{aligned}
$$

where the first step follows from Eq. (48), the second step follows from the definition of $D(t)$, the third step is straight forward, the forth step follows from $c = \Theta(\log(k)^{-1})$.

Hence, we get that

$$
s \geq O(k\log(k)T\|x(t)\|_T^2)\cdot O\Big(\frac{\log(1/\rho)}{\varepsilon^2 T\|x(t)\|_T^2}\Big) = O(\varepsilon^{-2}k\log(k)\log(1/\rho)).
$$

The lemma is then proved.

$\square$

Using the energy bound for filtered signals, we immediately get the following lemma.

Lemma J.4 (Energy estimation for filtered signals)**.** *Let D be the probability distribution defined as Eq. (47). Let $x \in \mathcal{F}_{k,F}$. Let H be defined as in Definition E.5. Let $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2. Let $j \in [B]$ satisfying that there exists an f^* with $h_{\sigma,b}(f^*) = j$ such that:*

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

where $\mathcal{N}^2 \geq \delta\|x\|_T^2$. Let $z(t) := (x\cdot H) * G_{\sigma,b}^{(j)}(t)$ be the filtered signal.

For any $\varepsilon,\rho \in (0,1)$, let $S_D = \{t_1,\cdots,t_s\}$ be a set of i.i.d. samples from $D(t)$ of size $s \geq O(\varepsilon^{-2}k\log(k)\log(1/\rho))$. Let the weight vector $w \in \mathbb{R}^s$ be defined by $w_i := 1/(2TsD(t_i))$ for $i \in [s]$.

Then when Large Offset event not happens, with probability at least $1-\rho$, we have

$$
(1-\varepsilon)\|z(t)\|_T^2 \leq \|z(t)\|_{S_D,w}^2 \leq (1+\varepsilon)\|z(t)\|_T^2.
$$

Proof. By applying Lemma J.1, we have that the desired result requires that

$$
s \geq \Big(\max_{t\in[-T,T]} \frac{|z(t)|^2}{D(t)}\Big)\cdot O\Big(\frac{\log(1/\rho)}{\varepsilon^2 T\|z(t)\|_T^2}\Big).
$$

By the filtered signals' energy bound (Corollary H.4), we have that

$$
\begin{aligned}
|z(t)|^2 &\lesssim \min\Big\{\frac{k\cdot H(t)+\delta}{1-|t/T|}, k^2\Big\}\cdot\|z(t)\|_T^2\\
&\lesssim \min\Big\{\frac{k}{1-|t/T|}, k^2\Big\}\cdot\|z(t)\|_T^2. \qquad (50)
\end{aligned}
$$

where the second step follows from $H(t) \lesssim 1$ (Lemma E.9 Property I, II). Then, we get that

$$
\begin{aligned}
&\max_{t\in[-T,T]} \frac{|z(t)|^2}{D(t)}\\
\lesssim\ &\max_{t\in[-T,T]} \min\Big\{\frac{k}{1-|t/T|}, k^2\Big\}\cdot\|z(t)\|_T^2
\end{aligned}
$$

$$\begin{aligned}
&\lesssim \max_{t\in[-T,T]} \min\Big\{\frac{k}{1-|t/T|}\frac{T(1-|t/T|)}{c}, k^2\frac{T}{ck}\Big\}\cdot \|z(t)\|_T^2 \\
&= kT\|z(t)\|_T^2/c \\
&\simeq k\log(k)T\|z(t)\|_T^2, \qquad (51)
\end{aligned}$$

where the first step follows from Eq. (50), the second step follows from the definition of $D(t)$, the third step is straight forward, the forth step follows from $c = \Theta(\log(k)^{-1})$.

As a result,

$$s \geq O(k\log(k)T\|z(t)\|_T^2)\cdot O\Big(\frac{\log(1/\rho)}{\varepsilon^2 T\|z(t)\|_T^2}\Big) = O(\varepsilon^{-2}k\log(k)\log(1/\rho)).$$

The lemma is then proved.

□

J.3 Partial energy estimation for filtered signals and local-test signals

In this section, we consider a variant version of energy estimation problem, which we are given a sub-interval $U \subseteq [-T,T]$ and we only want to estimate the energy within this interval.

The following lemma gives the sampling distribution with respect to U.

Lemma J.5. *Let $U = [L,R]$ such that $[-T(1-1/k), T(1-1/k)] \subseteq U \subseteq [-T,T]$. For $k \in \mathbb{N}_+$, define a probability distribution D_U as follows:*

$$D_U(t) := \begin{cases} c\cdot(1-|t/T|)^{-1}T^{-1}, & \text{for } |t| \leq T(1-1/k) \wedge t\in U \\ c\cdot kT^{-1}, & \text{for } |t| \in [T(1-1/k), T] \wedge t \in U \end{cases} \qquad (52)$$

where $c = \Theta(\log(k)^{-1})$ is a normalization factor such that $\int_{-T}^{T} D_U(t)\mathrm{d}t = 1$. Then, D_U is well-defined.

Proof. We compute the normalization factor of D_U in below. The condition that $\int_{-T}^{T} D_U(t)\mathrm{d}t = 1$ requires that

$$2\int_0^{T(1-1/k)} \frac{c}{(1-|t/T|)T}\mathrm{d}t + \int_{T(1-1/k)}^{R} c\cdot\frac{k}{T}\mathrm{d}t + \int_{L}^{-T(1-1/k)} c\cdot\frac{k}{T}\mathrm{d}t = 1,$$

which implies that

$$\begin{aligned}
c^{-1} &= 2\int_0^{T(1-1/k)} \frac{1}{(1-|t/T|)T}\mathrm{d}t + \int_{T(1-1/k)}^{R} \frac{k}{T}\mathrm{d}t + \int_{L}^{-T(1-1/k)} \frac{k}{T}\mathrm{d}t \\
&\eqsim \log(k)+1 \\
&= \Theta(\log(k)).
\end{aligned}$$

where the second step follows from $R \leq T$ and $L \geq -T$.

Thus, we get that $c = \Theta(\log(k)^{-1})$. □

Similar to Lemma J.4, we have a sample-efficient approach for estimating the partial energy of a filtered signal.

Lemma J.6 (Partial energy estimation for filtered signals)**.** *Let $U = [L, R]$ be such that $[-T(1-1/k), T(1-1/k)] \subseteq U$. For $k \in \mathbb{N}_+$, let D_U be the probability distribution defined as Eq. (52).*

Let $x \in \mathcal{F}_{k,F}$. Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2 with (σ, b) such that Large Offset event does not happen. For any $j \in [B]$, suppose there exists an f^* with $j = h_{\sigma,b}(f^*)$ satisfying:

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

where $\mathcal{N}^2 \geq \delta \|x\|_T^2$. Let $z(t) = (x \cdot H) * G_{\sigma,b}^{(j)}(t)$ be the filtered signal.

For any $\varepsilon, \rho \in (0,1)$, let $S_{D_U} = \{t_1, \cdots, t_s\}$ be a set of i.i.d. samples from D_U of size $s \geq O(\varepsilon^{-2} k \log(k) \log(1/\rho))$. Let the weight vector $w \in \mathbb{R}^s$ be defined by $w_i := 2/(TsD_U(t_i))$ for $i \in [s]$.

Then when Large Offset event not happens, with probability at least $1-\rho$, we have

$$(1-\varepsilon)\|z\|_U^2 \leq \|z\|_{S_{D_U},w}^2 \leq (1+\varepsilon)\|z\|_U^2,$$

where $\|z\|_U^2 := \frac{1}{R-L} \cdot \int_L^R |z(t)|^2 \mathrm{d}t$.

Proof. By applying Lemma J.1, we have that the desired result requires that

$$s \geq \Big(\max_{t\in U} \frac{|z(t)|^2}{D_U(t)}\Big) \cdot O\Big(\frac{\log(1/\rho)}{\varepsilon^2 T \|z(t)\|_T^2}\Big).$$

The first term can be upper bounded as follows:

$$\begin{aligned}
\max_{t\in U} \frac{|z(t)|^2}{D_U(t)} &\lesssim \max_{t\in U} \min\Big\{\frac{k}{1-|t/T|}, k^2\Big\} \cdot \frac{\|z(t)\|_T^2}{D_U(t)} \\
&\lesssim \max_{t\in U} \min\Big\{\frac{k}{1-|t/T|} \frac{T(1-|t/T|)}{c}, k^2 \frac{T}{ck}\Big\} \cdot \|z(t)\|_T^2 \\
&= kT\|z(t)\|_T^2/c \\
&\lesssim k\log(k) T\|z(t)\|_T^2 \\
&\lesssim k\log(k) \frac{R-L}{2T - k^2(2T+L-R)} \cdot T\|z(t)\|_U^2 \\
&\leq k\log(k) \cdot T\|z(t)\|_U^2,
\end{aligned} \tag{53}$$

where the first step follows from Eq. (50), the second step follows from the definition of $D_U(t)$, the third step is straight forward, the forth step follows from $c = \Theta(\log(k)^{-1})$, the fifth step follows from Lemma J.9, the sixth step follows from $R - L \leq 2T - k^2(2T + L - R)$.

Therefore, the sample complexity s should be at least:

$$s \geq O(k\log(k) \cdot T\|z(t)\|_U^2) \cdot O\Big(\frac{\log(1/\rho)}{\varepsilon^2 T\|z(t)\|_T^2}\Big) = O(\varepsilon^{-2} k \log(k) \log(1/\rho)).$$

The proof of the lemma is then completed.

□

Recall that in Section I, we study the local-test signal $d_z(t) = z(t)e^{2\pi \mathbf{i} f_0 \beta} - z(t+\beta)$. The following lemma gives a way to estimate the partial energy of a local-test signal. It can be proved by the same strategy with the energy bound in Lemma I.6.

Lemma J.7 (Partial energy estimation for local-test signals)**.** *Let $x \in \mathcal{F}_{k,F}$. Let H be defined as in Definition E.5, $G^{(j)}_{\sigma,b}$ be defined as in Definition C.2 with (σ, b) such that Large Offset event does not happen. Let $z(t) = (x \cdot H) * G^{(j)}_{\sigma,b}(t)$ be the filtered signal. Let $U := \{t_0 \in \mathbb{R} \mid H(t) > 1 - \delta_1, \forall t \in [t_0, t_0 + \beta]\}$. Let D_U be the probability distribution defined as Eq. (52). Let $D_H(t) :=$ Uniform$(\{t \in \mathbb{R} \mid H(t) > 1 - \delta_1\})$.*

For any $\varepsilon, \rho \in (0, 1)$, let $S_{D_U} = \{t_1, \cdots, t_s\}$ be a set of i.i.d. samples from D_U of size $s \geq O(k \log(k) \log(1/\rho))$. Let the weight vector $w \in \mathbb{R}^s$ be defined by $w_i := 2/(TsD_U(t_i))$ for $i \in [s]$.

Let $d_z(t) = z(t)e^{2\pi \mathbf{i} f_0 \beta} - z(t + \beta)$ be the local-test signal. Then, with probability at least $1 - \rho$, we have

$$\|d_z(t)\|^2_{S_{D_U},w} \leq 2\|d_z(t)\|^2_U + \sqrt{\delta_1}\|x(t)\|_T \cdot \|d_z(t)\|_U.$$

Proof. By Lemma J.1, we have that when

$$s \geq \Big(\max_{t\in U} \frac{|d_z(t)|^2}{D_U(t)}\Big) \cdot O\Big(\frac{\log(1/\rho)}{\xi^2|U| \cdot \|d_z(t)\|^2_U}\Big),$$

the following result holds with probability at least $1 - \rho$,

$$\|d_z(t)\|^2_{S_{D_U},w} \in (1 \pm \xi)\|d_z(t)\|^2_U, \tag{54}$$

where ξ is a parameter to be chosen later.

By the energy bound for local-test signals (Lemma I.6), we have that for any $t \in U$,

$$|d_z(t)|^2 \lesssim \min\Big\{\frac{k}{1 - |t/T|}, k^2\Big\} \cdot \|d_z(t)\|^2_U + \delta_1\|x(t)\|^2_T. \tag{55}$$

Then, we get that

$$\begin{aligned}
&\max_{t\in U} \frac{|d_z(t)|^2}{D_U(t)}\\
\lesssim\ &\max_{t\in U} \min\Big\{\frac{k}{1 - |t/T|}, k^2\Big\} \cdot \frac{\|d_z(t)\|^2_U + \delta_1\|x(t)\|^2_T}{D_U(t)}\\
\lesssim\ &\max_{t\in U} \min\Big\{\frac{k}{1 - |t/T|}\frac{T(1 - |t/T|)}{c}, k^2\frac{T}{ck}\Big\} \cdot (\|d_z(t)\|^2_U + \delta_1\|x(t)\|^2_T)\\
=\ &kTc^{-1} \cdot (\|d_z(t)\|^2_U + \delta_1\|x(t)\|^2_T)\\
\simeq\ &k\log(k)T \cdot (\|d_z(t)\|^2_{D_U} + \delta_1\|x(t)\|^2_{D_1}),
\end{aligned} \tag{56}$$

where the first step follows from Eq. (55), the second step follows from the definition of $D_U(t)$, the third step is straight forward, the forth step follows from $c = \Theta(\log(k)^{-1})$.

As a result, the sample complexity is

$$\begin{aligned}
s &\geq k\log(k)T \cdot (\|d_z(t)\|^2_{D_U} + \delta_1\|x(t)\|^2_{D_1}) \cdot O\Big(\frac{\log(1/\rho)}{\xi^2|U| \cdot \|d_z(t)\|^2_U}\Big)\\
&\simeq \xi^{-2} \cdot k\log(k) \cdot (1 + \frac{\delta_1\|x(t)\|^2_T}{\|d_z(t)\|^2_U}) \cdot \log(1/\rho)\\
&= k\log(k) \cdot \log(1/\rho),
\end{aligned}$$

where the first step follows from Eq. (56), the second step follows from $|U| \gtrsim T$, the third step follows by taking ξ to be such that

$$\xi^{-2}(1 + \frac{\delta_1 \|x(t)\|_T^2}{\|d_z(t)\|_U^2}) \simeq 1.$$

It remains to bound the estimation error. We have that

$$\begin{aligned}
\|d_z(t)\|_{S_D,w}^2 &\leq (1+\xi)\|d_z(t)\|_U^2 \\
&\simeq \Big(1 + \sqrt{1 + \frac{\delta_1 \|x(t)\|_{D_1}^2}{\|d_z(t)\|_U^2}}\Big)\|d_z(t)\|_U^2 \\
&\leq \Big(2 + \frac{\sqrt{\delta_1}\|x(t)\|_T}{\|d_z(t)\|_U}\Big)\|d_z(t)\|_U^2 \\
&\leq 2\|d_z(t)\|_U^2 + \sqrt{\delta_1}\|x(t)\|_T \cdot \|d_z(t)\|_U
\end{aligned}$$

where the first step follows from Eq. (54), the second step follows from the setting of ε, the third step follows from $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, the forth step is straight forward.

The lemma is then proved. □

J.4 Technical lemmas

We prove two technical lemmas in this section.

The following lemma bounds the energy of a Fourier-sparse signal within time duration $[L,R] \subseteq [-T,T]$ by its total energy.

Lemma J.8 (Partial energy of Fourier-sparse signal). *Given $k \in \mathbb{Z}_+, F \in \mathbb{R}_+$. For any $x \in \mathcal{F}_{k,F}$, $[L,R] \subseteq [-T(1-O(\frac{1}{k^2})), T(1-O(\frac{1}{k^2}))]$, we have that,*

$$\frac{2T - k^2(2T + L - R)}{R-L}\|x(t)\|_T^2 \lesssim \frac{1}{R-L}\int_L^R |x(t)|^2 \mathrm{d}t \leq \frac{2T}{R-L}\|x(t)\|_T^2.$$

Proof. For the upper bound, we have that

$$\frac{1}{R-L}\int_L^R |x(t)|^2 \mathrm{d}t \leq \frac{1}{R-L}\int_{-T}^T |x(t)|^2 \mathrm{d}t \leq \frac{2T}{R-L}\|x(t)\|_T^2,$$

where the first step is straight forward, the second step follows from the definition of the norm.

For the lower bound, we have that

$$\begin{aligned}
\int_L^R |x(t)|^2 \mathrm{d}t &= \int_{-T}^T |x(t)|^2 \mathrm{d}t - \int_{-T}^L |x(t)|^2 \mathrm{d}t - \int_R^T |x(t)|^2 \mathrm{d}t \\
&\geq 2T\|x(t)\|_T^2 - (L+T)\cdot \max_{t\in[-T,L]}|x(t)|^2 - (T-R)\cdot \max_{t\in[R,T]}|x(t)|^2 \\
&\gtrsim 2T\|x(t)\|_T^2 - (L+T)\cdot k^2\|x(t)\|_T^2 - (T-R)\cdot k^2\|x(t)\|_T^2 \\
&= (2T - k^2(2T+L-R))\|x(t)\|_T^2,
\end{aligned}$$

where the first step is straight forward, the second step follows from the definition of the norm, the third step follows from Theorem B.1, the forth step is straight forward.

□

By replacing the energy bound for Fourier-sparse signals with the energy bound for filtered signals (Corollary H.4), we obtain the following lemma:

Lemma J.9 (Partial energy of filtered signal)**.** *Given $k \in \mathbb{N}$ and $F \in \mathbb{R}_+$. Let $x \in \mathcal{F}_{k,F}$. Let H be defined as in Definition E.5, $G^{(j)}_{\sigma,b}$ be defined as in Definition C.2 with (σ, b) such that Large Offset event does not happen.*

For any $j \in [B]$, suppose there exists an f^* with $j = h_{\sigma,b}(f^*)$ satisfying:

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

where $\mathcal{N}^2 \geq \delta \|x\|_T^2$. Then, for $z(t) = (x \cdot H) * G^{(j)}_{\sigma,b}(t)$, we have that

$$\frac{2T - k^2(2T + L - R)}{R - L}\|z(t)\|_T^2 \lesssim \frac{1}{R-L}\int_L^R |z(t)|^2 \mathrm{d}t \leq \frac{2T}{R-L}\|z(t)\|_T^2.$$

Proof. For the upper bound, we have that

$$\frac{1}{R-L}\int_L^R |z(t)|^2 \mathrm{d}t \leq \frac{1}{R-L}\int_{-T}^T |z(t)|^2 \mathrm{d}t \leq \frac{2T}{R-L}\|z(t)\|_T^2,$$

where the first step is straight forward, the second step follows from the definition of the norm.

For the lower bound, we have that

$$\begin{aligned}
\int_L^R |z(t)|^2 \mathrm{d}t &= \int_{-T}^T |z(t)|^2 \mathrm{d}t - \int_{-T}^L |z(t)|^2 \mathrm{d}t - \int_R^T |z(t)|^2 \mathrm{d}t \\
&\geq T\|z(t)\|_T^2 - (L+T) \cdot \max_{t \in [0,L]} |z(t)|^2 - (T-R) \cdot \max_{t \in [R,T]} |z(t)|^2 \\
&\gtrsim T\|z(t)\|_T^2 - (L+T) \cdot k^2\|z(t)\|_T^2 - (T-R)k^2\|z(t)\|_T^2 \\
&= (2T - k^2(2T + L - R))\|z(t)\|_T^2
\end{aligned}$$

where the first step is straight forward, the second step follows from the definition of the norm, the third step follows from Corollary H.4, the forth step is straight forward.

$\square$

K Generate Significant Samples

In this section, we show our significant sample generation procedure for noisy signals. Recall that we use $x^*(t)$ to denote the ground-truth k-Fourier-sparse signal and $x(t) = x^*(t) + g(t)$ to denote the observation signal. We first generalize the energy estimation method in previous section to the noisy signals (see Section K.1). Then, we give our significant sample generation algorithm for a single bin (see Section K.2). Next, we show how to adapt our significant sample generation algorithm for multiple bins (see Section K.3). In addition, we provide some technical claims (see Section K.4).

K.1 Energy estimation for noisy signals

In this section, we generalize our methods in Section J to estimate the (partial) energy of the true observing signals, which contains some noise.

In the following lemma, we show that the energy of the filtered signal $z(t)$ can be estimated with a few samples, assuming it contains a small fraction of noise.

Lemma K.1. *Let $x^* \in \mathcal{F}_{k,F}$ be the ground-truth signal and $x(t) = x^*(t) + g(t)$ be the noisy observation signal. Let H be defined as in Definition E.5, $G^{(j)}_{\sigma,b}$ be defined as in Definition C.2 with (σ, b) such that Large Offset event does not happen. For any $j \in [B]$, suppose there exists an f_0 with $j = h_{\sigma,b}(f_0)$ satisfying:*

$$\int_{f_0-\Delta_h}^{f_0+\Delta_h} |\widehat{x^* \cdot H}(f)|^2 \mathrm{d}f \geq T\mathcal{N}^2/k,$$

where $\mathcal{N}^2 \geq \delta \|x^*\|_T^2$. Let $z^*(t) := (x^* \cdot H) * G^{(j)}_{\sigma,b}(t)$ and $z(t) = (x \cdot H) * G^{(j)}_{\sigma,b}(t)$. Let $g_z(t) := z(t) - z^*(t)$. Let $U = \{t_0 \in \mathbb{R} \mid H(t) > 1 - \delta_1 \ \forall t \in [t_0, t_0 + \beta]\}$. Suppose that $\|g_z(t)\|_T^2 \leq c\|z^*(t)\|_U^2$, where $c \in (0, 0.001)$ is a small universal constant.

For $s \geq O(k \log(k) \log(1/\rho))$, let $S_{D_U} = \{t_1, \ldots, t_s\}$ be a set of i.i.d. samples from the distribution D_U defined as Eq. (52). Let the weights $w_i = 1/(TsD_U(t_i))$ for $i \in [s]$.

Then, with probability at least 0.85,

$$\|z(t)\|^2_{S_{D_U},w} \geq (0.2 - 20c) \cdot \|z^*(t)\|_U^2$$

Proof. We consider the expectation of $\|g_z(t)\|^2_{S_U,w}$ first.

$$\begin{aligned}
\mathbb{E}\Big[\sum_{j=1}^s w_i |g_z(t_j)|^2\Big] &= \sum_{j=1}^s \mathop{\mathbb{E}}_{t_j \sim D_U}[w_i |g_z(t_j)|^2] \\
&= \sum_{j=1}^s \mathop{\mathbb{E}}_{t_j \sim D_U}\Big[\frac{1}{TsD_U(t_i)} |g_z(t_j)|^2\Big] \\
&\leq \mathop{\mathbb{E}}_{t \sim D_U}\Big[\frac{1}{TD_U(t)} |g_z(t)|^2\Big] \\
&\leq \int_U \frac{1}{T} |g_z(t)|^2 \mathrm{d}t \\
&\leq \frac{1}{T} \int_0^T |g_z(t)|^2 \mathrm{d}t \\
&\leq \|g_z(t)\|_T^2
\end{aligned} \tag{57}$$

where the first step is straight forward, the second step follows from the definition of w_i, the third step is straight forward, the forth step follows from the definition of expectation, the fifth step follows from $U \subseteq [0, T]$, the sixth step follows from the definition of the norm.

By Eq. (57) and Markov inequality, we have that with probability at least 0.9,

$$\sum_{j=1}^{s} w_i |g_z(t_j)|^2 \leq 20 \|g_z(t)\|_T^2. \tag{58}$$

Then, we have that

$$\begin{aligned}
\sum_{j=1}^{s} w_i |z(t_j)|^2 &\geq 0.5 \sum_{j=1}^{s} w_i |z^*(t_j)|^2 - \sum_{j=1}^{s} w_i |g_z(t_j)|^2 \\
&\geq 0.5 \sum_{j=1}^{s} w_i |z^*(t_j)|^2 - 20 \|g_z(t)\|_T^2 \\
&\geq 0.2 \|z^*(t)\|_U^2 - 20 \|g_z(t)\|_T^2 \\
&\geq (0.2 - 20c) \cdot \|z^*(t)\|_U^2
\end{aligned}$$

where the first step follows from $(a+b)^2 \geq 0.5a^2 - b^2$, the second step follows from Eq. (58), the third step follows from Lemma J.6, the forth step follows from $\|g_z(t)\|_T^2 \leq c\|z^*(t)\|_U^2$.

The total success probability follows from a union bound: $0.9 - \rho > 0.85$.

The lemma is then proved. □

The following lemma shows how to estimate the energy of a noisy local-test signal.

Lemma K.2. *Let $x^* \in \mathcal{F}_{k,F}$ be the ground-truth signal and $x(t) = x^*(t) + g(t)$ be the noisy observation signal. Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2 with (σ, b) such that Large Offset event does not happen. For any $j \in [B]$, suppose there exists an f_0 with $j = h_{\sigma,b}(f_0)$ satisfying:*

$$\int_{f_0-\Delta_h}^{f_0+\Delta_h} |\widehat{x^* \cdot H}(f)|^2 \mathrm{d}f \geq T\mathcal{N}^2/k,$$

where $\mathcal{N}^2 \geq \delta\|x^*\|_T^2$. Let $z^*(t) := (x^* \cdot H) * G_{\sigma,b}^{(j)}(t)$ and $z(t) = (x \cdot H) * G_{\sigma,b}^{(j)}(t)$. Let $g_z(t) := z(t) - z^*(t)$. Let $U = \{t_0 \in \mathbb{R} \mid H(t) > 1 - \delta_1 \ \forall t \in [t_0, t_0 + \beta]\}$. Suppose that $\|g_z(t)\|_T^2 \leq c\|z^*(t)\|_U^2$, where $c \in (0, 0.001)$ is a small universal constant.

For $s \geq O(k \log(k) \log(1/\rho))$, let $S_{D_U} = \{t_1, \ldots, t_s\}$ be a set of i.i.d. samples from the distribution D_U defined as Eq. (52). Let $w_i = 1/(T s D_U(t_i))$ for $i \in [s]$.

Then, with probability at least 0.85,

$$\|d_z(t)\|_{S_{D_U},w}^2 \lesssim (c + \sqrt{\gamma^2 + \delta_1}) \cdot \|z^*(t)\|_U^2,$$

Proof. We first consider the expectation of $\|g_z(t)e^{2\pi i f_0 \beta} - g_z(t+\beta)\|_{S_{D_U},w}^2$. We have that

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{i=1}^{s} w_i |g_z(t_i)e^{2\pi i f_0 \beta} - g_z(t_i + \beta)|^2\Big] \\
&= \sum_{i=1}^{s} \mathop{\mathbb{E}}_{t_i \sim D_U} [w_i |g_z(t_i)e^{2\pi i f_0 \beta} - g_z(t_i + \beta)|^2]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{s} \mathop{\mathbb{E}}_{t_i \sim D_U} \left[\frac{1}{TsD_U(t_i)} |g_z(t_i)e^{2\pi i f_0 \beta} - g_z(t_i+\beta)|^2 \right] \\
&= \mathop{\mathbb{E}}_{t \sim D_U} \left[\frac{1}{TD_U(t)} |g_z(t)e^{2\pi i f_0 \beta} - g_z(t+\beta)|^2 \right] \\
&= \int_U \frac{1}{T} |g_z(t)e^{2\pi i f_0 \beta} - g_z(t+\beta)|^2 \mathrm{d}t \\
&\leq \frac{4}{T} \int_U (|g_z(t)|^2 + |g_z(t+\beta)|^2) \mathrm{d}t \\
&\leq \frac{8}{T} \int_0^T |g_z(t)|^2 \mathrm{d}t \\
&\leq 10 \|g_z(t)\|_T^2, \qquad (59)
\end{aligned}
$$

where the first step is straight forward, the second step follows from the definition of w_i, the third step is straight forward, the forth step follows from the definition of expectation, the fifth step follows from $(a+b)^2 \leq 2a^2 + 2b^2$, the sixth step follows from $U \subseteq [0,T]$ and $U + \beta \subseteq [0,T]$, the seventh step follows from the definition of the norm.

By Eq. (59) and Markov inequality, we have that with probability at least 0.9,

$$
\sum_{i=1}^{s} w_i |g_z(t_i)e^{2\pi i f_0 \beta} - g_z(t_i+\beta)|^2 \leq 100 \|g_z(t)\|_T^2. \qquad (60)
$$

We have that

$$
\begin{aligned}
&\sum_{i=1}^{s} w_i |z(t_i)e^{2\pi i f_0 \beta} - z(t_i+\beta)|^2 \\
\leq\ & \sum_{i=1}^{s} (2w_i |z^*(t_i)e^{2\pi i f_0 \beta} - z^*(t_i+\beta)|^2 + 2w_i |g_z(t_i)e^{2\pi i f_0 \beta} - g_z(t_i+\beta)|^2) \\
\leq\ & 200 \|g_z(t)\|_T^2 + \sum_{i=1}^{s} 2w_i |z^*(t_i)e^{2\pi i f_0 \beta} - z^*(t_i+\beta)|^2 \\
\leq\ & 200 \|g_z(t)\|_T^2 + 4\|z^*(t+\beta) - e^{2\pi \mathbf{i} f_0 \beta} \cdot z^*(t)\|_U^2 + 2\sqrt{\delta_1} \|x(t)\|_T \|z^*(t+\beta) - e^{2\pi \mathbf{i} f_0 \beta} \cdot z^*(t)\|_U \\
\lesssim\ & (200c + 4\gamma^2 + 4\delta_1) \|z^*(t)\|_U^2 + 2\sqrt{\delta_1(\gamma^2+\delta_1)} \|x^*(t)\|_T \|z^*(t)\|_U \\
\lesssim\ & (c + \gamma^2 + \delta_1) \|z^*(t)\|_U^2 + \sqrt{\delta_1(\gamma^2+\delta_1)\frac{k}{\delta}} \|z^*(t)\|_T \|z^*(t)\|_U \\
\lesssim\ & ((c + \gamma^2 + \delta_1) + \sqrt{\delta_1(\gamma^2+\delta_1)\frac{k}{\delta}\frac{R-L}{T-k^2(T+L-R)}}) \|z^*(t)\|_U^2 \\
\lesssim\ & (c + \sqrt{\gamma^2+\delta_1}) \cdot \|z^*(t)\|_U^2
\end{aligned}
$$

where the first step follows from $(a+b)^2 \leq 2a^2 + 2b^2$, the second step follows from Eq. (60), the third step follows from the partial energy estimation for local-test signal (Lemma J.7) which holds with probability $1-\rho$, the forth step follows from the $\|g_z(t)\|_T^2 \leq c\|z^*(t)\|_U^2$ and Claim K.8, the fifth step follows from Lemma H.3, the sixth step follows from $[L,R] := U$ and Lemma J.9, the seventh step follows form $R - L \lesssim T - k^2(T+L-R)$, $\delta_1 \delta^{-1} k \lesssim 1$.

The total success probability follows from a union bound $0.9 - \rho > 0.85$.

The lemma is then proved. □

K.2 Significant sample generation for a single bin

Recall that we define a sample $t \in [0, T]$ is *significant* if the magnitude of the local-test signal at t is small, i.e., $|d_z(t)| \le O(|z(t)|)$. The following lemma shows that a significant sample can be efficiently generated, provided that the filtered noisy signal does not contain too much noise.

Lemma K.3 (Generate Significant samples for filtered noisy signals)**.** *Let $x^* \in \mathcal{F}_{k,F}$ be the ground-truth signal and $x(t) = x^*(t) + g(t)$ be the noisy observation signal. Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2 with (σ, b) such that Large Offset event does not happen. For any $j \in [B]$, suppose there exists an f_0 with $j = h_{\sigma,b}(f_0)$ satisfying:*

$$\int_{f_0-\Delta_h}^{f_0+\Delta_h} |\widehat{x^* \cdot H}(f)|^2 \mathrm{d}f \ge T\mathcal{N}^2/k,$$

where $\mathcal{N}^2 \ge \delta\|x^*\|_T^2$. Let $z^*(t) := (x^* \cdot H) * G_{\sigma,b}^{(j)}(t)$ and $z(t) = (x \cdot H) * G_{\sigma,b}^{(j)}(t)$. Let $g_z(t) := z(t) - z^*(t)$. Let $U = \{t_0 \in \mathbb{R} \mid H(t) > 1 - \delta_1 \ \forall t \in [t_0, t_0 + \beta]\}$. Suppose that $\|g_z(t)\|_T^2 \le c\|z^*(t)\|_U^2$, where $c \in (0, 0.001)$ is a small universal constant.

Then, there is an algorithm that takes $O(k \log(k))$ samples in z, runs in $O(k \log(k))$ time, and output an $\alpha \in U$ such that with probability at least 0.6,

$$|z(\alpha + \beta) - z(\alpha)e^{2\pi \mathbf{i} f_0 \beta}|^2 \le O(c + \sqrt{\gamma^2 + \delta_1})|z(\alpha)|^2 \le 0.01|z(\alpha)|^2.$$

Proof. The output α is sample in two steps:

1. For $s \ge O(k \log(k))$, generate s i.i.d. samples $S_{D_U} = \{t_1, \dots, t_s\}$ bfrom the distribution D_U defined as Eq. (52). Let $w_i = 1/(TsD_U(t_i))$ for $i \in [s]$ be the weights.

2. Define a probability distribution D_S such that

$$D_S(t_i) := \frac{w_i |z(t_i)|^2}{\sum_{i \in [s]} w_i |z(t_i)|^2} \quad \forall i \in [s]. \tag{61}$$

   And sample α according to D_S.

The sample and time complexities of this procedure are straightforward. It remains to prove that α satisfies the significance requirement stated in the lemma.

By Lemma K.1, we have that with probability at least 0.85,

$$\sum_{j=1}^{s} w_i |z(t_j)|^2 \ge (0.2 - 20c) \cdot \|z^*(t)\|_U^2 \tag{62}$$

By Lemma K.2, we have that with probability at least 0.85,

$$\sum_{i=1}^{s} w_i |z(t_i)e^{2\pi i f_0 \beta} - z(t_i + \beta)|^2 \lesssim (c + \sqrt{\gamma^2 + \delta_1}) \cdot \|z^*(t)\|_U^2 \tag{63}$$

Thus, with probability at least 0.7,

$$\frac{\sum_{i=1}^{s} w_i |z(t_i)e^{2\pi i f_0 \beta} - z(t_i + \beta)|^2}{\sum_{j=1}^{s} w_i |z(t_j)|^2}$$

$$
\begin{aligned}
&\leq \frac{O(c+\sqrt{\gamma^2+\delta_1})\cdot \|z^*(t)\|_U^2}{\sum_{j=1}^s w_i|z(t_j)|^2} \\
&\leq \frac{O(c+\sqrt{\gamma^2+\delta_1})\cdot \|z^*(t)\|_U^2}{(0.2-20c)\cdot \|z^*(t)\|_U^2} \\
&= O(c+\sqrt{\gamma^2+\delta_1}) \qquad (64)
\end{aligned}
$$

where the first step follows from Eq. (63), the second step follows from Eq. (62), the third step is straight forward.

For a random sample $\alpha \sim D_S$, we bound the following expectation:

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\alpha\sim D_S}\left[\frac{|z(\alpha)e^{2\pi i f_0\beta}-z(\alpha+\beta)|^2}{|z(\alpha)|^2}\right] \\
=& \sum_{i=1}^s \frac{w_i|z(t_i)|^2}{\sum_{j=1}^s w_j|z(t_j)|^2}\cdot\frac{|z(t_i)e^{2\pi i f_0\beta}-z(t_i+\beta)|^2}{|z(t_i)|^2} \\
=& \frac{\sum_{i=1}^s w_i|z(t_i)e^{2\pi i f_0\beta}-z(t_i+\beta)|^2}{\sum_{j=1}^s w_j|z(t_j)|^2} \\
\leq& O(c+\sqrt{\gamma^2+\delta_1}),
\end{aligned}
$$

where the first step follows from the definition of D_m, the second step is straightforward, the third step follows from Eq. (64).

Thus by Markov inequality, with probability 0.9,

$$
\frac{|z(\alpha)e^{2\pi i f_0\beta}-z(\alpha+\beta)|^2}{|z(\alpha)|^2} \leq \frac{O(c+\sqrt{\gamma^2+\delta_1})}{0.1} = O(c+\sqrt{\gamma^2+\delta_1}).
$$

The success probability follows from a union bound. And the second inequality follows from the range of the parameters c, γ, δ_1. $\square$

K.3 Significant sample generation for multiple bins

In this section, we present our significant sample generation procedure that simultaneously works for all "good bins".

We first prove the correctness of Algorithm 2.

Lemma K.4 (Generate significant samples for different bins simultaneously)**.** *Let $x^* \in \mathcal{F}_{k,F}$ be the ground-truth signal and $x(t) = x^*(t) + g(t)$ be the noisy observation signal. Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2 with (σ, b) such that Large Offset event does not happen. Let $U = \{t_0 \in \mathbb{R} \mid H(t) > 1-\delta_1\ \forall t \in [t_0, t_0+\beta]\}$.*

For $j \in [B]$, let $z_j^*(t) := (x^*\cdot H) * G_{\sigma,b}^{(j)}(t)$ and $z_j(t) = (x\cdot H) * G_{\sigma,b}^{(j)}(t)$. Let $g_j(t) := z_j(t) - z_j^*(t)$. Let

$$
S_{g1} := \left\{ j \in [B] \mid \|g_j(t)\|_T^2 \leq c\|z_j^*(t)\|_U^2 \right\},
$$

where $c \in (0, 0.001)$ is a small universal constant. Let

$$
S_{g2} := \left\{ j \in [B] \;\middle|\; \exists f_0, h_{\sigma,b}(f_0) = j \text{ and } \int_{f_0-\Delta_h}^{f_0+\Delta_h} |\widehat{x^*\cdot H}(f)|^2 \mathrm{d}f \geq T\mathcal{N}^2/k \right\},
$$

Algorithm 2 Generate Significant Samples

procedure GenerateSignificantSamples(z)
  $B \leftarrow O(k)$
  $U \leftarrow \{t_0 \in \mathbb{R} | H(t) > 1-\delta_1, \forall t \in [t_0, t_0+\beta]\}$
  $D_z \leftarrow \begin{cases} c \cdot (1-|t/T|)^{-1} T^{-1}, & \text{for } |t| \leq T(1-1/k) \wedge t \in U \\ c \cdot kT^{-1}, & \text{for } |t| \in [T(1-1/k), T] \wedge t \in U \end{cases}$
  $S \leftarrow O(k \log(k))$ i.i.d. samples from D_z
  for $t_i \in S$ **do**
    for $j \in [B]$ **do**
      $a \leftarrow t_i/\sigma$
      $u_j \leftarrow \sum_{i \in \mathbb{Z}} x \cdot H(\sigma(a-j-iB)) e^{-2\pi \mathbf{i} \sigma b(j+iB)} G(j+iB)$ ▷ $u \in \mathbb{R}^B$
      $u_j^\beta \leftarrow \sum_{i \in \mathbb{Z}} x \cdot H(\sigma(a+\beta-j-iB)) e^{-2\pi \mathbf{i} \sigma b(j+iB)} G(j+iB)$ ▷ $u^\beta \in \mathbb{R}^B$
    end for
    $\widehat{u} = \mathrm{FFT}(u)$
    $\widehat{u}^\beta = \mathrm{FFT}(u^\beta)$
    for $j \in [B]$ **do**
      $z_j(t_i) \leftarrow \widehat{u}_j$
      $z_j(t_i+\beta) \leftarrow \widehat{u}_j^\beta$
    end for
  end for
  $w_i \leftarrow D_z(t_i), \forall t_i \in S$
  $W \leftarrow \sum_{t_i \in S} w_i |z_j(t_i)|^2$
  $Z_{j,1} \leftarrow 0$, $Z_{j,2} \leftarrow 0$ ▷ $Z \in \mathbb{C}^{B \times 2}$
  for $j \in [B]$ **do**
    $D_S(t_i) \leftarrow w_i |z_j(t_i)|^2 / W, \forall t_i \in S$
    Sample $t_i \sim D_S$ ▷ $\alpha \in \mathbb{R}^B$
    $Z_{j,1} \leftarrow z_j(t_i)$, $Z_{j,2} \leftarrow z_j(t_i+\beta)$
  end for
  return Z
end procedure

where $\mathcal{N}^2 \geq \delta \|x^*\|_T^2$. Let $S_g = S_{g1} \cap S_{g2}$.

There is a Procedure GenerateSignificantSamples *(Algorithm 2) that takes $O(k^2 \log^2(k/\delta_1))$ samples in x, runs in $O(k^2 \log^3(k/\delta_1))$ time, and for each $j \in S_g$, output α_j such that with probability at least* 0.6*,*

$$|z_j(\alpha_j+\beta) - z_j(\alpha_j) e^{2\pi \mathbf{i} f_0 \beta}|^2 \leq 0.01 |z_j(\alpha_j)|^2 \quad \forall j \in S_g.$$

Proof. For $k \in \mathbb{N}_+$, define a probability distribution $D(t)$ as follows:

$$D(t) := \begin{cases} c \cdot (1-|t/T|)^{-1} T^{-1}, & \text{for } |t| \leq T(1-1/k) \wedge t \in U \\ c \cdot kT^{-1}, & \text{for } |t| \in [T(1-1/k), T] \wedge t \in U \end{cases}$$

where $c = \Theta(\log(k)^{-1})$ is a normalization factor such that $\int_{-T}^{T} D(t) \mathrm{d}t = 1$. For any $\varepsilon, \rho \in (0,1)$, let $S_D = \{t_1, \cdots, t_s\}$ be a set of i.i.d. samples from $D(t)$ of size $s \geq O(k \log(k) \log(1/\rho))$. Let the weight vector $w \in \mathbb{R}^s$ be defined by $w_i := 1/(TsD(t_i))$ for $i \in [s]$.

Suppose all the bins can access the same set of time points S_D. Then, by Lemma K.3, for any $j \in S_g$ with probability 0.6, we have that,

$$|z_j(\alpha+\beta) - z_j(\alpha)e^{2\pi \mathbf{i} f_0 \beta}|^2 \leq 0.01|z_j(\alpha)|^2.$$

Then, we show that the value of $z_j(t), j = 1, \cdots, B$ of same set of time points S_D can be compute by accessing a same set of time points in $x(t)$. By Lemma D.2 with setting $a = \alpha/\sigma$, we have that

$$z_j(\alpha) = \widehat{u}_j,$$

which is computed by the algorithm.

As a result, for each $j \in S_g$,

$$|z_j(\alpha_j+\beta) - z_j(\alpha_j)e^{2\pi \mathbf{i} f_0 \beta}|^2 \leq 0.01|z_j(\alpha_j)|^2 \quad \forall j \in S_g,$$

holds with probably 0.6.

□

We compute the time and sample complexities of Algorithm 2 in the following two lemmas.

Lemma K.5 (Running time of Procedure GenerateSignificantSamples in Algorithm 2)**.** *Procedure* GenerateSignificantSamples *in Algorithm 2 runs in* $O(k^2 \log(k) \log(k/\delta_1))$ *times.*

Proof. In each call of Procedure GenerateSignificantSamples in Algorithm 2,

- In line 5, taking $|S|$ samples runs $O(|S|)$ times.
- In line 6, the for loop repeats $|S|$ times,
  - In line 7, the for loops repeats B times and j iterate from 1 to B, in each loop line 9 and line 10, computing the summation runs in $|\{j + iB | i \in \mathbb{Z} \wedge j + iB \in \mathrm{supp}(G)\}|$ times.
  - In line 12 and 13, running Fast Fourier Transform algorithm takes $O(B \log(B))$ time, where B is the length of the vector u and u^β.
  - In line 14, the for loop repeats B times, each loop runs in $O(1)$ times.
- In line 19, assigning w_i runs in $|S|$ times.
- In line 20, computing $\sum_{t_i \in S} w_i |z_j(t_i)|^2$ runs in $|S|$ times.
- In line 22, the for loop repeats B times, each loop runs in $O(1)$ times.

Notice that

$$\sum_{j \in [B]} |\{j + iB | i \in \mathbb{Z} \wedge j + iB \in \mathrm{supp}(G)\}| \leq |\mathrm{supp}(G)|.$$

In the algorithm, we set the parameters:

$$B = \ O(k), \quad \text{and} \quad |S| = \ k \log(k). \tag{65}$$

Thus,

$$|\mathrm{supp}(G)| = O(lB/\alpha) = k \log(k/\delta_1), \tag{66}$$

where the first step follows from Lemma C.4 Property IV, the second step follows from $\alpha \approx 1$ and $l = \Theta(\log(k/\delta_1))$.

Therefore, the time complexity in total is

$$\begin{aligned}
& O(O(|S|) + |S| \cdot (|\mathrm{supp}(G)| + O(B\log(B)) + B \cdot O(1)) + |S| + |S| + B \cdot O(1)) \\
\leq & O(|S| \cdot (|\mathrm{supp}(G)| + B\log(B))) \\
\leq & O(k\log(k) \cdot (|\mathrm{supp}(G)| + k\log(k))) \\
\leq & O(k\log(k) \cdot (k\log(k/\delta_1) + k\log(k))) \\
\leq & O(k^2\log(k)\log(k/\delta_1)),
\end{aligned}$$

where the first step is straightforward, the second step follows from Eq. (65), the third step follows from Eq. (66), the forth step is straight forward. $\square$

Lemma K.6 (Sample complexity of Procedure GenerateSignificantSamples in Algorithm 2)**.** *Procedure* GenerateSignificantSamples *in Algorithm 2 takes* $O(k^2\log(k)\log(k/\delta_1))$ *samples.*

Proof. In each call of Procedure GenerateSignificantSamples in Algorithm 2,

- In line 6, the for loop repeats $|S|$ times,
- In line 7, the for loops repeats B times and j iterate from 1 to B, in each loop line 9 and line 10, computing the summation takes $O(|\{\sigma(a-j-iB)|i \in \mathbb{Z} \wedge j+iB \in \mathrm{supp}(G)\}|)$ samples.

Following from the setting in the algorithm, we have that

$$|S| = k\log(k). \tag{67}$$

Thus,

$$|\mathrm{supp}(G)| = O(lB/\alpha) = k\log(k/\delta_1). \tag{68}$$

where the first step follows from Lemma C.4 Property IV, the second step follows from $\alpha \approx 1$ and $l = \Theta(\log(k/\delta_1))$.

So, the samples complexity of Procedure GenerateSignificantSamples in Algorithm 2 is

$$\begin{aligned}
& |S| \cdot \sum_{j\in[B]} O(|\{\sigma(a-j-iB) \mid i \in \mathbb{Z} \wedge j+iB \in \mathrm{supp}(G)\}|) \\
\leq & O(|S| \cdot |\mathrm{supp}(G)|) \\
\leq & O(|S| \cdot k\log(k/\delta_1)) \\
\leq & O(k\log(k) \cdot k\log(k/\delta_1)) \\
= & O(k^2\log(k)\log(k/\delta_1))
\end{aligned}$$

where the first step is straight forward, the second step follows from Eq. (68), the third step follows from Eq. (67), the forth step is straight forward. $\square$

K.4 Technical claims

We prove two technical claims in below about the local-test signals' energy reduction.

Claim K.7 (Energy decay of local-test signals)**.** *Let $x^* \in \mathcal{F}_{k,F}$. For any (σ, b) such that Large Offset event does not happen and any $j \in [B]$, suppose there exists an f_0 with $j = h_{\sigma,b}(f_0)$ satisfying: well-isolation conditions and*

$$\int_{f_0-\Delta_h}^{f_0+\Delta_h} |\widehat{x^* \cdot H}(f)|^2 \mathrm{d}f \geq T\mathcal{N}^2/k,$$

where $\mathcal{N}^2 \geq \delta\|x^*\|_T^2$. Let $z(t) = (x^* \cdot H) * G_{\sigma,b}^{(j)}(t)$ be the filtered signal.

For $\beta \leq \gamma/\Delta_0$ with $\Delta_0 = O(\Delta)$, let $d_z(t) = z(t)e^{2\pi \mathbf{i} f_0 \beta} - z(t+\beta)$ be the local-test signal. We have that

$$\|d_z(t)\|_T^2 \lesssim (\gamma^2 + \delta_1) \cdot \|z(t)\|_T^2.$$

Proof. Let $S := \mathrm{supp}(\widehat{x}^* * \widehat{H})$ be the support set of $\widehat{x}^* * \widehat{H}$. Let

$$V := \{f \in \mathbb{R} \mid \widehat{G}_{\sigma,b}^{(j)}(f) \geq 1 - \delta_1\}.$$

Note that $\|d_z(t)\|_T^2$ can be expressed as follows (ignoring the $\frac{1}{T}$ factor):

$$\begin{aligned}
& \int_0^T |z(t+\beta) - e^{2\pi \mathbf{i} f_0 \beta} \cdot z(t)|^2 \mathrm{d}t \\
\leq & \int_{-\infty}^{\infty} |z(t+\beta) - e^{2\pi \mathbf{i} f_0 \beta} \cdot z(t)|^2 \mathrm{d}t \\
= & \int_{-\infty}^{\infty} |\widehat{z}(f)e^{2\pi \mathbf{i} f \beta} - e^{2\pi \mathbf{i} f_0 \beta} \cdot \widehat{z}(f)|^2 \mathrm{d}f \\
= & \int_{-\infty}^{\infty} |\widehat{z}(f)|^2 \cdot |e^{2\pi \mathbf{i} f \beta} - e^{2\pi \mathbf{i} f_0 \beta}|^2 \mathrm{d}f \\
= & \int_S |\widehat{z}(f)|^2 \cdot |e^{2\pi \mathbf{i} f \beta} - e^{2\pi \mathbf{i} f_0 \beta}|^2 \mathrm{d}f \\
= & \int_{S \cap V} |\widehat{z}(f)|^2 \cdot |e^{2\pi \mathbf{i} f \beta} - e^{2\pi \mathbf{i} f_0 \beta}|^2 \mathrm{d}f + \int_{S \setminus V} |\widehat{z}(f)|^2 \cdot |e^{2\pi \mathbf{i} f \beta} - e^{2\pi \mathbf{i} f_0 \beta}|^2 \mathrm{d}f,
\end{aligned}$$

where the first step is straight forward, the second step follows from Parseval's theorem, the third step is straight forward, the forth step follows from the assumption that Large Offset event does not happen, the fifth step is straight forward.

Then, for the first term, we have that

$$\begin{aligned}
\int_{S \cap V} |\widehat{z}(f)|^2 \cdot |e^{2\pi \mathbf{i} f \beta} - e^{2\pi \mathbf{i} f_0 \beta}|^2 \mathrm{d}f \leq & \int_{f_0-\Delta_0}^{f_0+\Delta_0} |\widehat{z}(f)|^2 \cdot |e^{2\pi \mathbf{i} f \beta} - e^{2\pi \mathbf{i} f_0 \beta}|^2 \mathrm{d}f \\
\lesssim & \int_{f_0-\Delta_0}^{f_0+\Delta_0} |\widehat{z}(f)|^2 \cdot \gamma^2 \mathrm{d}f \\
\leq & \gamma^2 \cdot \int_{-\infty}^{\infty} |\widehat{z}(f)|^2 \mathrm{d}f \\
\leq & \gamma^2 \cdot \int_{-\infty}^{\infty} |z(t)|^2 \mathrm{d}t \\
\leq & \gamma^2 \cdot \int_0^T |z(t)|^2 \mathrm{d}t
\end{aligned} \tag{69}$$

where the first step follows from $S \cap V \subset [f_0 - \Delta_0, f_0 + \Delta_0]$ by Claim D.3, the second step follows from

$$|e^{2\pi \mathbf{i} f\beta} - e^{2\pi \mathbf{i} f_0 \beta}| \leq 4\pi\beta |f - f_0| \leq 4\pi\beta\Delta_{h0} \lesssim \gamma,$$

the third step is straight forward, the forth step follows from Parseval's theorem, the fifth step follows from Lemma G.3.

For the second term, we have that

$$\begin{aligned}
\int_{S\setminus V} |\widehat{z}(f)|^2 \cdot |e^{2\pi \mathbf{i} f\beta} - e^{2\pi \mathbf{i} f_0\beta}|^2 \mathrm{d}f &= \int_{S\setminus V} |(\widehat{x}^* * \widehat{H})(f)|^2 \cdot |\widehat{G}^{(j)}_{\sigma,b}(f)|^2 \cdot |e^{2\pi \mathbf{i} f\beta} - e^{2\pi \mathbf{i} f_0\beta}|^2 \mathrm{d}f \\
&\leq \int_{S\setminus V} |(\widehat{x}^* * \widehat{H})(f)|^2 \cdot \delta_1^2 \cdot |e^{2\pi \mathbf{i} f\beta} - e^{2\pi \mathbf{i} f_0\beta}|^2 \mathrm{d}f \\
&\lesssim \int_{S\setminus V} |(\widehat{x}^* * \widehat{H})(f)|^2 \cdot \delta_1^2 \mathrm{d}f \\
&\leq \int_{-\infty}^{\infty} |(\widehat{x}^* * \widehat{H})(f)|^2 \cdot \delta_1^2 \mathrm{d}f \\
&= \delta_1^2 \cdot \int_{-\infty}^{\infty} |(x^* \cdot H)(t)|^2 \mathrm{d}t \\
&\lesssim \delta_1^2 \cdot \int_0^T |x^*(t)|^2 \mathrm{d}t
\end{aligned} \tag{70}$$

where the first step follows from the definition of z, the second step follows from the definition of V, the third step follows from $|e^{2\pi \mathbf{i} f\beta} - e^{2\pi \mathbf{i} f_0\beta}|^2 \lesssim 1$, the forth step is straight forward, the fifth step follows from Parseval's theorem, the sixth step follows from Lemma E.9 Property IV and V.

Putting them together, we get that

$$\begin{aligned}
\int_0^T |z(t+\beta) - e^{2\pi \mathbf{i} f_0\beta} \cdot z(t)|^2 \mathrm{d}t &\lesssim \gamma^2 \cdot \int_0^T |z(t)|^2 \mathrm{d}t + \delta_1^2 \cdot \int_0^T |x(t)|^2 \mathrm{d}t \\
&\lesssim (\gamma^2 + \delta_1^2 \delta^{-1} k) \cdot \int_0^T |z(t)|^2 \mathrm{d}t \\
&\lesssim (\gamma^2 + \delta_1) \cdot \int_0^T |z(t)|^2 \mathrm{d}t,
\end{aligned}$$

where the first step follows from Eq. (69) and Eq. (70), the second step follows from Lemma H.3, the last step follows from $\delta_1 \delta^{-1} k \lesssim 1$.

The proof of the lemma is then completed. □

Similar result also holds for the partial energy:

Claim K.8 (Partial energy decay of local-test signals)**.** *Let $x^*(t)$ be a k-Fourier-sparse signal. Let $z(t) := (x^* \cdot H) * G^{(j)}_{\sigma,b}(t)$ and $d_z(t) = z(t+\beta) - e^{2\pi \mathbf{i} f_0\beta} \cdot z(t)$. Let $U = \{t_0 \in \mathbb{R} \mid H(t) > 1 - \delta_1\ \forall t \in [t_0, t_0 + \beta]\} =: [L, R]$. For $\beta \leq \gamma/\Delta_0$ with $\Delta_0 = O(\Delta)$, we have that*

$$\int_L^R |d_z(t)|^2 \mathrm{d}t \lesssim (\gamma^2 + \delta_1) \cdot \int_L^R |z(t)|^2 \mathrm{d}t.$$

Proof. We have that

$$\int_L^R |z(t+\beta) - e^{2\pi \mathbf{i} f_0\beta} \cdot z(t)|^2 \mathrm{d}t \leq \int_0^T |z(t+\beta) - e^{2\pi \mathbf{i} f_0\beta} \cdot z(t)|^2 \mathrm{d}t$$

$$\leq (\gamma^2+\delta_1)\int_0^T |z(t)|^2 \mathrm{d}t$$
$$\leq (\gamma^2+\delta_1)\frac{T}{T-k^2(T+L-R)}\int_L^R |z(t)|^2 \mathrm{d}t$$
$$\lesssim (\gamma^2+\delta_1)\int_L^R |z(t)|^2 \mathrm{d}t,$$

where the first step is straight forward, the second step follows from Claim K.7, the third follows from Lemma J.9 with time duration changed from $[-T,T]$ to $[0,T]$, the forth step follows from $T-k^2(T+L-R)\gtrsim T$. $\square$

L Frequency Estimation

We introduce our improved frequency estimation algorithm in this section. We first show that given significant samples, we are able to estimate a specific target frequency (see Section L.1). Then, we show to generalize it to simultaneously estimate frequencies for multiple bins and give our main frequency estimation algorithm (see Section L.2). Next, we prove several technical claims on the votes distribution in the ArySearch procedure (see Section L.3).

L.1 Frequency estimation via significant samples

In this section, we show an algorithm such that for a target frequency f_0, it can use several significant samples to estimate it with high accuracy. The main idea is as follows: for a significant sample α, since $|z(\alpha+\beta)-z(\alpha)e^{2\pi\mathbf{i}f_0\beta}|$ is very small, the angle of $\frac{z(\alpha+\beta)}{z(\alpha)}$ will be close to $2\pi f_0\beta$. That is,

$$\arg\Big(\frac{z(\alpha+\beta)}{z(\alpha)}\Big)\approxeq 2\pi f_0\beta \pmod{2\pi}.$$

Solving the congruence equation gives that

$$f_0\approx\frac{1}{2\pi\beta}\Big(\arg\Big(\frac{z(\alpha+\beta)}{z(\alpha)}\Big)+2\pi s\Big)$$

for some unknown $s\in\mathbb{Z}$.

To find the unknown s, we use the same strategy as in [PS15]: perform a D-round searching procedure to narrow the possible range of f_0. More specifically, at the beginning, the possible range of f_0 (frequency interval) is $[-F,F]$. And after D rounds, f_0 is located in a frequency interval of length $O(\Delta)$, resulting in an estimate with Δ accuracy.

For $d\in[D]$, consider the d-th round of searching, where the frequency interval is:

$$[\mathsf{left}_d,\mathsf{left}_d+\mathsf{len}_d].$$

We equally partition the frequency interval into num parts and do a num-ary search. We generate R significant samples, and for each sample α, we enumerate all possible s and compute

$$\frac{1}{2\pi\beta}\Big(\arg\Big(\frac{z(\alpha+\beta)}{z(\alpha)}\Big)+2\pi s\Big).$$

Then, we find which part this quantity falls in and add a vote to that part. For robustness, we also add votes to that part's left and right neighbors. In the end, the frequency interval for the next round is the part with more than $R/2$ votes. It is easy to see that at most 5 parts can be selected in the new frequency interval. Hence, we have

$$\mathsf{len}_{d+1} \leq \frac{\mathsf{len}_d}{\mathsf{num}/5}, \tag{71}$$

i.e., the length of the possible range of f_0 decays at a constant rate.

More formally, we have the following lemma:

Lemma L.1 (significant sample to frequency estimation). *Suppose that there is an algorithm* GetSignificantSample *that*

- *takes* $z(t), \beta$ *as input where* $\beta \leq O(1/\Delta)$*,*
- *takes* $\mathcal{S}$ *samples in* $z(t)$*,*
- *runs in* $\mathcal{T}$ *time,*
- *outputs an* α *such that with probability* 0.9*,*

$$|z(\alpha+\beta) - z(\alpha)e^{2\pi \mathbf{i} f_0 \beta}|^2 \leq 0.0001|z(\alpha)|^2.$$

Then, there is an Procedure FrequencyEstimationZ *in Algorithm 3 that*

- *takes* $O(\log(FT)\log(\log(FT))\log(1/\rho_1)\mathcal{S})$ *samples,*
- *runs in* $O(\log(FT)\log(\log(FT))\log(1/\rho_1)\mathcal{T})$ *times,*
- *and outputs* $\widetilde{f}_0$ *such that with probability at least* $1-\rho_1$*,*

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

Proof. We prove the correctness, time/sample complexity of Algorithm 3 in below.

Correctness: We first compute the value of D, the number of rounds needed for the searching procedure. Note that the GetSignificantSample procedure requires that $\beta \leq O(1/\Delta)$. In our algorithm, we take $\beta_d = O(\mathsf{num}/\mathsf{len}_d)$ for the d-th round. Hence, for the last round $d = D$, we have that,

$$O(\mathsf{num}/\mathsf{len}_D) = O(1/\Delta) \implies \mathsf{len}_D \geq \mathsf{num}\Delta.$$

Then, by Eq. (71) and $\mathsf{len}_1 = F$, we get that

$$D = \log_{\mathsf{num}}(\frac{FT}{\mathsf{num}(T\Delta)}) \lesssim \log(FT)/\log(\mathsf{num}). \tag{72}$$

Then, we calculate the success probability. For $d \in [D]$, by Claim L.8, with probability at least $1-O(c+\rho)^{R/6}$, the true part containing f_0 and its left and right neighbor will get R votes in total, and the other far away parts will get at most $R/2$ votes. In this case, the new frequency interval will contain the true part, and we consider this round being success. Since the search procedure takes D rounds, by a union bound, all rounds will succeed with probability at least

$$1 - D \cdot O(c+\rho)^{R/6} \geq 1 - \frac{\log(FT)}{\log(\mathsf{num})} \cdot O(c+\rho)^{R/6} \geq 1 - \rho_1$$

where the first step follows from Eq. (72), the second step follows from

$$R \geq O\Big(\frac{\log(\log(FT)/\rho_1)}{\log(1/(c+\rho))}\Big) \geq O\Big(\frac{\log(\log(FT)/(\rho_1 \log(\mathsf{num})))}{\log(1/(c+\rho))}\Big).$$

Therefore, with probability at least $1-\rho$, the final frequency interval of length $O(\Delta)$ will contain the target frequency f_0, which means that the output $\widetilde{f}_0$ satisfies $|\widetilde{f}_0 - f_0| \lesssim \Delta$. And the correctness of the algorithm is proved.

Time complexity: We show that Procedure FrequencyEstimationZ in Algorithm 3 runs in $O(\log(FT) \cdot \log(\log(FT)/\rho_1))$ times.

In each call of the Procedure FrequencyEstimationZ in Algorithm 3,

- The for-loop repeats D times.
- In each loop, line 6 call Procedure ArySearch.

Then, in the d-th call of the Procedure ArySearch,

- In line 12, the for-loop repeats R times.
- In line 13, the Procedure GetSignificantSample is called.
- In line 14, the for-loop repeats $\beta_d \mathsf{len}_d + O(1)$ times.
- In line 16, the for-loop repeats num times.

Thus, the total time complexity is dominated by:

$$D \cdot R \cdot (\beta_d \mathsf{len}_d + O(1)) \cdot \mathsf{num} + D \cdot R \cdot \mathcal{T}.$$

By the parameter settings in Algorithm 3, we have that

$$\begin{aligned} \mathsf{num} &= O(1), \\ D &= O(\log(\frac{FT}{\Delta})/\log(\mathsf{num})), \\ R &= O(\log(\frac{\log(FT)}{\rho_1 \log(\mathsf{num})})), \\ \beta_d &= O(\frac{\mathsf{num}}{\mathsf{len}_d}), \end{aligned}$$

In particular, we have

$$D = O(\log(\frac{FT}{\Delta})/\log(\mathsf{num})) \leq O(\log(\frac{FT}{\Delta})) \leq O(\log(FT)),$$

where the first step follows from the setting of D, the second step follows from $\mathsf{num} = O(1)$, the third step follows from $\Delta = \mathrm{poly}(k)$.

Hence, the total time complexity of Algorithm 3 is

$$\begin{aligned} & O(D \cdot R \cdot (\beta_d \mathsf{len}_d + O(1)) \cdot \mathsf{num}) + D \cdot R \cdot \mathcal{T} \\ = & O(D \cdot R \cdot (O(\mathsf{num}) + O(1)) \cdot \mathsf{num}) + D \cdot R \cdot \mathcal{T} \\ = & O(D \cdot R \cdot \mathcal{T}) \\ = & O(\log(FT) \cdot \log(\log(FT)/\rho_1) \cdot \mathcal{T}), \end{aligned}$$

where the first step follows from $\beta_d = O(\frac{\mathsf{num}}{\mathsf{len}_d})$, the second step follows from $\mathsf{num} = O(1)$, the third step follows from the choices of D and R.

Sample complexity: Each call of the Procedure GetSignificantSample takes $\mathcal{S}$ samples, and it is called DR times. Thus, the total sample complexity of Algorithm 3 is

$$DR \cdot \mathcal{S} = O(\log(FT) \cdot \log(\log(FT)/\rho_1) \cdot \mathcal{S}).$$

The proof of the lemma is completed.

$\square$

Algorithm 3 Frequency Estimation of the Filtered Signal

procedure FrequencyEstimationZ$(x, H, G^{(j)}_{\sigma,b})$
    $\mathsf{num} \leftarrow O(1)$, $D \leftarrow O(\log(\frac{FT}{\Delta})/\log(\mathsf{num}))$, $R \leftarrow O(\log(\frac{\log(FT)}{\rho_1 \log(\mathsf{num})}))$
    $\mathsf{left}_1 \leftarrow -F$, $\mathsf{len}_1 \leftarrow 2F$
    for $d \in [D]$ **do**
        $\mathsf{len}_d \leftarrow 5\frac{\mathsf{len}_{d-1}}{\mathsf{num}}$
        $\mathsf{left}_{d+1} \leftarrow$ ArySearch$(x, H, G^{(j)}_{\sigma,b}, F, T, \Delta, \mathsf{left}_d, \mathsf{len}_d, \mathsf{num})$
    end for
    return left_D
end procedure
procedure ArySearch$(x, H, G^{(j)}_{\sigma,b}, F, T, \Delta, \mathsf{left}_i, \mathsf{len}_i, \mathsf{num})$
    Let $v \in \mathbb{Z}_+^{\mathsf{num}}$ and $v_q \leftarrow 0$ for $q \in [\mathsf{num}]$
    for $r = 1 \to R$ **do**
        $z(\alpha + \beta), z(\alpha) \leftarrow$ GetSignificantSample$(x, H, G^{(j)}_{\sigma,b}, r, d)$
        for $s \in [\beta\mathsf{left}_d - 10, \beta(\mathsf{left}_d + \mathsf{len}_d) + 10] \cap \mathbb{Z}$ **do**
            $\widetilde{f} = \frac{1}{2\pi\beta}(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi s)$
            for $q \in [\mathsf{num}]$ **do**
                if $\widetilde{f} \in [\mathsf{left}_d + (q-1)\mathsf{len}_d/\mathsf{num}, \mathsf{left}_d + q\mathsf{len}_d/\mathsf{num}]$ **then**
                    $v_q \leftarrow v_q + 1$
                end if
            end for
        end for
    end for
    for $q \in [\mathsf{num}]$ **do**
        if $v_q + v_{q+1} + v_{q+2} \geq R/2$ **then**
            $\mathsf{left}_{d+1} \leftarrow \mathsf{left}_d + (q-1)\mathsf{len}_d/\mathsf{num}$
            return left_{d+1}
        end if
    end for
    return $\emptyset$
end procedure

L.2 Simultaneously estimate frequencies for different bins

Combining the significant sample generation procedure discussed in Section K with Algorithm 3, we obtain the frequency estimation algorithm that improves the algorithms in [PS15] and [CKPS16].

Algorithm 4 Pre-computation of the Significant Samples

```
procedure SAMPLINGSIGNIFICANTSAMPLE(x, H, G^{(j)}_{σ,b}, F, T, Δ, num, D, R)
    len_1 ← 2F, 𝓛 ∈ ℂ^{D×R×B×2}
    for d ∈ [D] do
        len_d = 5 len_{d−1}/num
        β̂ ← O(num/len_d)
        for r ∈ [R] do
            Sample β ∈ Uniform([½β̂, β̂])
            Z ← GENERATESIGNIFICANTSAMPLES(x, H, G)      ▷ Z ∈ ℂ^{B×2}, see Algorithm 2
            for j ∈ [B] do
                𝓛_{d,r,j,1} ← Z_{j,1}                   ▷ Z_{j,1} = z^{(j)}(α + β)
                𝓛_{d,r,j,2} ← Z_{j,2}                   ▷ Z_{j,2} = z^{(j)}(α)
            end for
        end for
    end for
    return 𝓛
end procedure
procedure GETSIGNIFICANTSAMPLE(𝓛, d, r, j)
    return (𝓛_{d,r,j,1}, 𝓛_{d,r,j,2})
end procedure
```

Algorithm 5 Frequency Estimation

```
procedure FREQUENCYESTIMATIONX(x)
    𝓛 ← SAMPLINGSIGNIFICANTSAMPLE(x)
    for j ← 1, ··· , B do
        f̃_j ← FREQUENCYESTIMATIONZ(x, H, G^{(j)}_{σ,b})   ▷ z^{(j)} = (x · H) ∗ G^{(j)}_{σ,b}(t)
    end for
    L ← {f̃_1, ··· , f̃_B}
    return L
end procedure
```

Theorem L.2 (Better frequency estimation algorithm)**.** *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 the observation signal where $g(t)$ is arbitrary noise. Let $\Delta_h := O(|\mathrm{supp}(\widehat{H})|)$, $\Delta := O(k\cdot\Delta_h)$ and $\mathcal{N}^2 := \|g(t)\|_T^2 + \delta\|x^*(t)\|_T^2$. Let H be defined as in Definition E.5, $G^{(j)}_{\sigma,b}$ be defined as in Definition C.2 with (σ, b) such that Large Offset event does not happen. Let $U = \{t_0 \in \mathbb{R} \mid H(t) > 1 - \delta_1 \ \forall t \in [t_0, t_0+\beta]\}$.*

For $j \in [B]$, let $z_j^*(t) := (x^* \cdot H) * G^{(j)}_{\sigma,b}(t)$ and $z_j(t) = (x \cdot H) * G^{(j)}_{\sigma,b}(t)$. Let $g_j(t) := z_j(t) - z_j^*(t)$. Let

$$S_{g1} = \{j \in [B] \mid \|g_j(t)\|_T^2 \le c\|z_j^*(t)\|_U^2\},$$

where $c \in (0, 0.001)$ is a small universal constant. Let

$$S_{g2} = \Big\{ j \in [B] \ \Big| \ \exists f_0, h_{\sigma,b}(f_0) = j \text{ and } \int_{f_0-\Delta_h}^{f_0+\Delta_h} |\widehat{x^* \cdot H}(f)|^2 \mathrm{d}f \ge T\mathcal{N}^2/k \Big\}.$$

Let $S_g = S_{g1} \cap S_{g2}$. Let $S_f = \{f_i \mid \exists j \in S_g : h_{\sigma,b}(f_i) = j \ \forall i \in [k]\}$.

There is a Procedure FrequencyEstimationX *in Algorithm 5 such that:*

- *takes $k^2 \log(1/\delta) \log(FT)$ samples,*
- *runs in $k^2 \log(1/\delta) \log^2(FT)$ time,*
- *returns a set L of $O(k)$ frequencies such that with probability $1 - \rho_0$, for any $f \in S_f$, there exists an $\widetilde{f} \in L$ satisfying*
$$|f - \widetilde{f}| \lesssim \Delta.$$

Proof. We prove the correctness, time complexity, and sample complexity of Algorithm 5 in below.

Correctness: By Lemma K.4, we know that the Procedure GenerateSignificantSamples in Algorithm 2 takes $\mathcal{S} = O(k^2 \log^2(k/\delta_1))$ samples in x, runs in $\mathcal{T} = O(k^2 \log^3(k/\delta_1))$ time, and for each $j \in S_g$, and outputs α_j such that for each $j \in S_g$ with probability 0.6,

$$|z_j(\alpha_j + \beta) - z_j(\alpha_j) e^{2\pi \mathbf{i} f_0 \beta}|^2 \leq 0.01 |z_j(\alpha_j)|^2,$$

where f_0 satisfies

$$\int_{f_0 - \Delta}^{f_0 + \Delta} |\widehat{x^* \cdot H}(f)|^2 \mathrm{d} f \geq T \mathcal{N}^2 / k, \tag{73}$$

and $j = h_{\sigma,b}(f_0)$.

In the line 4, we call the algorithm FrequencyEstimationZ$(x, H, G_{\sigma,b}^{(j)})$. By Lemma L.1, FrequencyEstimationZ$(x, H, G_{\sigma,b}^{(j)})$ output $\widetilde{f}$ for each $f_j \in S_f$ such that with probability at least $1 - \rho_1$

$$|\widetilde{f} - f_j| \lesssim \Delta.$$

As a result, for all the $f \in S_f$, there is a $\widetilde{f} \in L$ such that

$$|\widetilde{f} - f| \lesssim \Delta$$

holds with probability at least

$$1 - B\rho_1 \geq 1 - \rho_0.$$

Time complexity: We show that the Procedure FrequencyEstimationX in Algorithm 5 runs in

$$O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1))$$

time.

In each call of the Procedure FrequencyEstimationX in Algorithm 5,

- Line 2 call Procedure SamplingSignificantSample, which runs in
$$O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1))$$
time by Lemma L.3.

- The for-loop repeats B times:
  - In each loop, line 4 call Procedure FrequencyEstimationZ, which runs in

$$O(\log(FT) \cdot \log(\log(FT)/\rho_1))$$

    time by Lemma L.1.

Thus, the total time complexity is

$$\begin{aligned}
&O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1)) + B \cdot O(\log(FT) \cdot \log(\log(FT)/\rho_1)) \\
\leq\ &O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1)) + O(k) \cdot O(\log(FT) \cdot \log(\log(FT)/\rho_1)) \\
\leq\ &O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1)),
\end{aligned}$$

where the first step follows from $B = O(k)$, the second step is straight forward.

Sample complexity: We show that the Procedure FrequencyEstimationX in Algorithm 5 takes

$$O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1))$$

samples.

In each call of the Procedure FrequencyEstimationX in Algorithm 5, Line 2 call Procedure SamplingSignificantSample, which takes $O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1))$ samples by Lemma L.4.

So, the sample complexity of Procedure FrequencyEstimationX in Algorithm 5 is

$$O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1)).$$

□

The following two lemmas shows the time complexity and sample complexity of the significant sample generation procedure in Algorithm 4.

Lemma L.3 (Running time of Procedure SamplingSignificantSample in Algorithm 4)**.** *Procedure* SamplingSignificantSample *in Algorithm 4 runs in*

$$O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1))$$

times.

Proof. In each call of Procedure SamplingSignificantSample in Algorithm 4, in line 3, the for loop repeats D times, in line 6, the for loops repeats R times,

- In line 8, by Lemma K.5, each call of Procedure GenerateSignificantSamples takes $O(k^2 \log(k) \log(k/\delta_1))$ times.

- In line 9, the for loop repeats B times, each iteration runs in $O(1)$ times.

Following from the setting in the algorithm, we have that

$$\begin{aligned}
\mathsf{num} &= O(1), \\
D &= O(\log(\frac{FT}{\Delta})/\log(\mathsf{num})), \\
R &= O(\log(\frac{\log(FT)}{\rho_1 \log(\mathsf{num})})), \\
B &= O(k).
\end{aligned} \tag{74}$$

We have that

$$D = O(\log(\frac{FT}{\Delta})/\log(\mathsf{num})) \leq O(\log(\frac{FT}{\Delta})) \leq O(\log(FT)), \tag{75}$$

where the first step follows from the setting of D, the second step follows from $\mathsf{num} = O(1)$, the third step follows from $\Delta = \mathrm{poly}(k)$.

We also have that

$$R = O(\log(\frac{\log(FT)}{\rho_1 \log(\mathsf{num})})) \leq O(\log(\log(FT)/\rho_1)), \tag{76}$$

where the first step follows from the setting of R, the second step follows from $\mathsf{num} = O(1)$.

So, the time complexity of Procedure SamplingSignificantSample in Algorithm 4 is

$$\begin{aligned}
& D \cdot R \cdot (O(k^2 \log(k) \log(k/\delta_1)) + B \cdot O(1)) \\
\leq\ & O(\log(FT)) \cdot R \cdot (O(k^2 \log(k) \log(k/\delta_1)) + B \cdot O(1)) \\
\leq\ & O(\log(FT)) \cdot O(\log(\log(FT)/\rho_1)) \cdot (O(k^2 \log(k) \log(k/\delta_1)) + B \cdot O(1)) \\
\leq\ & O(\log(FT)) \cdot O(\log(\log(FT)/\rho_1)) \cdot (O(k^2 \log(k) \log(k/\delta_1)) + O(k) \cdot O(1)) \\
=\ & O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1)),
\end{aligned}$$

where the first step follows from Eq. (75), the second step follows from Eq. (76), the third step follows from Eq. (74), the forth step is straightforward.

□

Lemma L.4 (Sample complexity of Procedure SamplingSignificantSample in Algorithm 4)**.** *Procedure* SamplingSignificantSample *in Algorithm 4 takes*

$$O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1))$$

samples.

Proof. In each call of Procedure SamplingSignificantSample in Algorithm 4, In line 3, the for loop repeats D times, in line 6, the for loops repeats R times,

- In line 8, by Lemma K.5, each call of Procedure GenerateSignificantSamples takes $O(k^2 \log(k) \log(k/\delta_1))$ samples.

Following from the setting in the algorithm, we have that

$$\begin{aligned}
\mathsf{num} &= O(1), \\
D &= O(\log(\frac{FT}{\Delta})/\log(\mathsf{num})),
\end{aligned}$$

$$R = O(\log(\frac{\log(FT)}{\rho_1 \log(\mathsf{num})})).$$

We have that

$$D = O(\log(\frac{FT}{\Delta})/\log(\mathsf{num})) \leq O(\log(\frac{FT}{\Delta})) \leq O(\log(FT)), \tag{77}$$

where the first step follows from the setting of D, the second step follows from $\mathsf{num} = O(1)$, the third step follows from $\Delta = \mathrm{poly}(k)$.

We also have that

$$R = O(\log(\frac{\log(FT)}{\rho_1 \log(\mathsf{num})})) \leq O(\log(\log(FT)/\rho_1)), \tag{78}$$

where the first step follows from the setting of R, the second step follows from $\mathsf{num} = O(1)$.

So, the sample complexity of Procedure SamplingSignificantSample in Algorithm 4 is

$$\begin{aligned}
& D \cdot R \cdot O(k^2 \log(k) \log(k/\delta_1)) \\
\leq\ & O(\log(FT)) \cdot R \cdot O(k^2 \log(k) \log(k/\delta_1)) \\
\leq\ & O(\log(FT)) \cdot O(\log(\log(FT)/\rho_1)) \cdot O(k^2 \log(k) \log(k/\delta_1)) \\
=\ & O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1)),
\end{aligned}$$

where the first step follows from Eq. (77), the second step follows from Eq. (78), the third step is straightforward.

□

L.3 Vote distributions in ArySearch

In this section, we prove several claims on the distributions of votes when we perform the num-ary search on the frequency interval.

We first consider a single voter, i.e., one significant sample. The following claim shows that, if f_0 is in the q-th part, then this part or its left neighbor or its right neighbor will get at least one vote.

Claim L.5. *For* $\mathsf{len} \in \mathbb{R}_+, \mathsf{num} \in \mathbb{Z}_+, q \in [1, \mathsf{num}]$, *let* $f_0 \in [\mathsf{left} + (q-1)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + q\frac{\mathsf{len}}{\mathsf{num}}]$. *For any* $\beta \in [\frac{c}{2} \cdot \frac{\mathsf{num}}{\mathsf{len}}, c \cdot \frac{\mathsf{num}}{\mathsf{len}}]$ *with constant* $c \in (0, 0.01)$, *for any constant* $\varepsilon \in (0, 0.01 \cdot c^2)$, *let* $\alpha \in \mathbb{R}$ *such that*

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

Let

$$\Theta = \Big\{ \frac{1}{2\pi\beta} (\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi s) \ \Big|\ s \in [\beta \mathsf{left} - 10, \beta(\mathsf{left} + \mathsf{len}) + 10] \cap \mathbb{Z} \Big\}.$$

Then, we have that

$$\Big| \Theta \cap \Big[\mathsf{left} + (q-2)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + (q+1)\frac{\mathsf{len}}{\mathsf{num}} \Big) \Big| = 1.$$

Proof. We have that

$$\Big|\frac{z(\alpha+\beta)}{z(\alpha)} - e^{2\pi \mathbf{i} f_0 \beta}\Big| \le \sqrt{\varepsilon}.$$

Since $|e^{2\pi \mathbf{i} f_0 \beta}| = 1$ and $\sin(x) \approx x$ for small x, it indicates that

$$\Big\| \arg(\frac{z(\alpha+\beta)}{z(\alpha)}) - 2\pi f_0 \beta \Big\|_{\bigcirc} \lesssim \sqrt{\varepsilon}, \tag{79}$$

where $\|a\|_{\bigcirc} = \min_{x\in\mathbb{Z}} |a + 2\pi x|$. Thus, Eq. (79) can be rewritten as:

$$\min_{x\in\mathbb{Z}} \Big| \arg(\frac{z(\alpha+\beta)}{z(\alpha)}) - 2\pi f_0 \beta + 2\pi x \Big| \lesssim \sqrt{\varepsilon}. \tag{80}$$

Let s_0 be defined as

$$s_0 := \arg\min_{x\in\mathbb{Z}} \;\; \arg(\frac{z(\alpha+\beta)}{z(\alpha)}) - 2\pi f_0 \beta + 2\pi x.$$

We first show that s_0 falls in interval in the definition of Θ. We have that

$$\begin{aligned} |2\pi s_0 - 2\pi f_0 \beta| \le & \Big| \arg(\frac{z(\alpha+\beta)}{z(\alpha)}) - 2\pi f_0 \beta + 2\pi s_0 \Big| + \Big| \arg(\frac{z(\alpha+\beta)}{z(\alpha)}) \Big| \\ \le & \Big| \arg(\frac{z(\alpha+\beta)}{z(\alpha)}) - 2\pi f_0 \beta + 2\pi s_0 \Big| + 2\pi \\ \le & \; 2\pi + O(\sqrt{\varepsilon}) \end{aligned} \tag{81}$$

where the first step follows from the triangle inequality, the second step follows from $|\arg(\frac{z(\alpha+\beta)}{z(\alpha)})| \le 2\pi$, the third step follows from Eq. (80).

As a result, s_0 has the following upper bound:

$$\begin{aligned} s_0 \le & \; f_0 \beta + 1 + O(\sqrt{\varepsilon}) \\ \le & \; \beta(\mathsf{left} + \mathsf{len}) + 1 + O(\sqrt{\varepsilon}) \\ \le & \; \beta(\mathsf{left} + \mathsf{len}) + 2 \end{aligned}$$

where the first step follows from Eq. (81), the second step follows from $f_0 \in [\mathsf{left} + (q-1)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + q\frac{\mathsf{len}}{\mathsf{num}}] \subseteq [\mathsf{left}, \mathsf{left} + \mathsf{len}]$, the third step follows from the setting of ε.

Also, s_0 has the following lower bound:

$$\begin{aligned} s_0 \ge & \; f_0 \beta - 1 - O(\sqrt{\varepsilon}) \\ \ge & \; \beta \mathsf{left} - 1 - O(\sqrt{\varepsilon}) \\ \ge & \; \beta \mathsf{left} - 2 \end{aligned}$$

where the first step follows from Eq. (81), the second step follows from $f_0 \in [\mathsf{left}, \mathsf{left} + \mathsf{len}]$, the third step follows from the setting of ε.

Combining the lower and upper bounds of s_0 together, and by the definition of the set Θ, we know that

$$\frac{1}{2\pi\beta}\Big(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi s_0 \Big) \in \Theta.$$

Then, we show that

$$\mathsf{left} + (q-2)\frac{\mathsf{len}}{\mathsf{num}} \leq \frac{1}{2\pi\beta}\Big(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi s_0\Big) < \mathsf{left} + (q+1)\frac{\mathsf{len}}{\mathsf{num}}.$$

Eq. (80) also implies that

$$\Big|\frac{1}{2\pi\beta}(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi s_0) - f_0\Big| \leq O(\frac{\sqrt{\varepsilon}}{\beta}). \tag{82}$$

Then, we have the following upper bound:

$$\begin{aligned}
\frac{1}{2\pi\beta}(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi s_0) &\leq f_0 + O(\frac{\sqrt{\varepsilon}}{\beta}) \\
&\leq \mathsf{left} + q\frac{\mathsf{len}}{\mathsf{num}} + O(\frac{\sqrt{\varepsilon}}{\beta}) \\
&\leq \mathsf{left} + q\frac{\mathsf{len}}{\mathsf{num}} + O(\frac{2\sqrt{\varepsilon}}{c}\frac{\mathsf{len}}{\mathsf{num}}) \\
&\leq \mathsf{left} + (q+1)\frac{\mathsf{len}}{\mathsf{num}},
\end{aligned}$$

where the first step follows from Eq. (82), the second step follows from $f_0 \in [\mathsf{left}+(q-1)\mathsf{len}/\mathsf{num}, \mathsf{left}+q\mathsf{len}/\mathsf{num}]$, the third step follows from $\beta \in [c\frac{\mathsf{num}}{2\mathsf{len}}, c\frac{\mathsf{num}}{\mathsf{len}}]$, the forth step follows from $O(\sqrt{\varepsilon}/c) \leq 1$.

We also have the following lower bound:

$$\begin{aligned}
\frac{1}{2\pi\beta}(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi s_0) &\geq f_0 - O(\frac{\sqrt{\varepsilon}}{\beta}) \\
&\geq \mathsf{left} + (q-1)\frac{\mathsf{len}}{\mathsf{num}} - O(\frac{\sqrt{\varepsilon}}{\beta}) \\
&\geq \mathsf{left} + (q-1)\frac{\mathsf{len}}{\mathsf{num}} - O(\frac{\sqrt{\varepsilon}}{c}\frac{\mathsf{len}}{\mathsf{num}}) \\
&> \mathsf{left} + (q-2)\frac{\mathsf{len}}{\mathsf{num}},
\end{aligned}$$

where the first step follows from Eq. (82), the second step follows from $f_0 \in [\mathsf{left}+(q-1)\mathsf{len}/\mathsf{num}, \mathsf{left}+q\mathsf{len}/\mathsf{num}]$, the third step follows from $\beta \in [c\frac{\mathsf{num}}{2\mathsf{len}}, c\frac{\mathsf{num}}{\mathsf{len}}]$, the forth step follows from $O(\sqrt{\varepsilon}/c) < 1$.

Moreover, since

$$\frac{1}{\beta} \geq \frac{\mathsf{len}}{\mathsf{num}},$$

we have that there is at most 1 element in the intersection

$$\Big|\Theta \cap \Big[\mathsf{left} + (q-2)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + (q+1)\frac{\mathsf{len}}{\mathsf{num}}\Big)\Big| \leq 1.$$

The lemma then follows. $\square$

The following claim shows that for those parts far away from the true part containing f_0, they will get no vote.

Claim L.6. *For* $\mathsf{len} \in \mathbb{R}_+, \mathsf{num} \in \mathbb{Z}_+, q \in [1, \mathsf{num}]$*, let* $f_0 \in [\mathsf{left} + (q-1)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + q\frac{\mathsf{len}}{\mathsf{num}}]$*. Let* $\beta \sim \mathit{Uniform}([\frac{c}{2} \cdot \frac{\mathsf{num}}{\mathsf{len}}, c \cdot \frac{\mathsf{num}}{\mathsf{len}}])$ *with constant* $c \in (0, 0.01)$*. For any constant* $\varepsilon \in (0, 0.01 \cdot c^2)$*, let* $\alpha \in \mathbb{R}$ *such that*

$$|z(\alpha+\beta) - z(\alpha)e^{2\pi \mathbf{i} f_0 \beta}|^2 \le \varepsilon |z(\alpha)|^2.$$

Let

$$\Theta = \Big\{ \frac{1}{2\pi\beta}(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi s) \;\Big|\; s \in [\beta\mathsf{left} - 10, \beta(\mathsf{left}+\mathsf{len}) + 10] \cap \mathbb{Z} \Big\}.$$

Then, we have that for any $q' \in [0, \mathsf{num}-1]$ *such that* $|q - q'| > 1$*, with probability at least* $1 - O(c)$*,*

$$\Theta \cap \Big[\mathsf{left} + (q'-1)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + q'\frac{\mathsf{len}}{\mathsf{num}}\Big] = \emptyset.$$

Proof. Let s_0 be defined as

$$s_0 := \arg\min_{x\in\mathbb{Z}} \arg(\frac{z(\alpha+\beta)}{z(\alpha)}) - 2\pi f_0\beta + 2\pi x.$$

By Claim L.5, we have that

$$\frac{1}{2\pi\beta}(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi s_0) \in \Big[\mathsf{left} + (q-2)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + (q+1)\frac{\mathsf{len}}{\mathsf{num}}\Big). \tag{83}$$

Then, we discuss two cases based on the range of q'.

Case 1: $1 < |q - q'| < 1/(4c)$**.**

For the ease of discussion, suppose $1 < q' - q < 1/(4c)$. We have that

$$\begin{aligned}\mathsf{left} + (q'-1)\frac{\mathsf{len}}{\mathsf{num}} &\ge \mathsf{left} + (q+1)\frac{\mathsf{len}}{\mathsf{num}} \\ &> \frac{1}{2\pi\beta}(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi s_0),\end{aligned}$$

where the first step follows from $q, q' \in \mathbb{Z}$, and the second step follows from Eq. (83).

Moreover, we also have that

$$\begin{aligned}\mathsf{left} + q'\frac{\mathsf{len}}{\mathsf{num}} &\le \mathsf{left} + (q + \frac{1}{4c} - 1)\frac{\mathsf{len}}{\mathsf{num}} \\ &\le \frac{1}{2\pi\beta}(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi s_0) + (\frac{1}{4c}+1)\frac{\mathsf{len}}{\mathsf{num}} \\ &< \frac{1}{2\pi\beta}(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi s_0) + \frac{1}{\beta} \\ &\le \frac{1}{2\pi\beta}(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi(s_0+1)),\end{aligned}$$

where the second step follows from Eq. (83), the third step follows from $(\frac{1}{4c}+1)\mathsf{len}/\mathsf{num} < 1/\beta$.

Hence, we get that $\Big[\mathsf{left} + (q'-1)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + q'\frac{\mathsf{len}}{\mathsf{num}}\Big]$ is contained in the following interval:

$$\Big(\frac{1}{2\pi\beta}(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi s_0), \frac{1}{2\pi\beta}(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi(s_0+1))\Big).$$

Since s_0 and s_0+1 are two consecutive integers, by the definition of Θ, there is no element of Θ in this open interval. Hence, we know that in this case,

$$\Theta \cap \left[\mathsf{left} + (q'-1)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + q'\frac{\mathsf{len}}{\mathsf{num}}\right] = \emptyset.$$

Case 2: $|q - q'| \geq 1/(4c)$**.**

For the ease of discussion, suppose that $q' - q \geq 1/(4c)$. We have that,

$$c(q'-q) \geq \frac{1}{4}. \tag{84}$$

Moreover, we have that

$$\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) - 2\pi\beta\left(\mathsf{left} + (q - \frac{1}{2})\frac{\mathsf{len}}{\mathsf{num}}\right) \pmod{2\pi} \in \Big[-\frac{3\pi\beta\mathsf{len}}{\mathsf{num}}, \frac{3\pi\beta\mathsf{len}}{\mathsf{num}}\Big], \tag{85}$$

which follows from Eq. (83).

We also have that,

$$\beta\frac{\mathsf{len}}{\mathsf{num}} \leq c. \tag{86}$$

Then, we have that,

$$\begin{aligned}
&\mathsf{Pr}\Big[2\pi\beta(\mathsf{left} + (q' - \frac{1}{2})\frac{\mathsf{len}}{\mathsf{num}}) - \arg(\frac{z(\alpha+\beta)}{z(\alpha)}) \pmod{2\pi} \in [-\frac{\pi\beta\mathsf{len}}{\mathsf{num}}, \frac{\pi\beta\mathsf{len}}{\mathsf{num}}]\Big] \\
\leq\ &\mathsf{Pr}\Big[2\pi\beta(q'-q)\frac{\mathsf{len}}{\mathsf{num}} \pmod{2\pi} \in [-\frac{4\pi\beta\mathsf{len}}{\mathsf{num}}, \frac{4\pi\beta\mathsf{len}}{\mathsf{num}}]\Big] \\
\leq\ &\mathsf{Pr}\Big[2\pi\beta(q'-q)\frac{\mathsf{len}}{\mathsf{num}} \pmod{2\pi} \in [-4\pi c, +4\pi c]\Big] \\
\leq\ &4c + \frac{16}{q'-q} \\
\leq\ &100c
\end{aligned} \tag{87}$$

where the first step follows from Eq. (85), the second step follows from Eq. (86), the third step follows from Lemma D.6 with the following parameters:

$$\begin{aligned}
\widetilde{T} &= 2\pi, \\
\widetilde{\sigma} &= 2\pi\beta(q'-q)\frac{\mathsf{len}}{\mathsf{num}}, \\
\widetilde{\varepsilon} &= 4\pi c, \\
\widetilde{\delta} &= 0, \\
A &= \pi c(q'-q),
\end{aligned}$$

the forth step follows from Eq. (84).

By Eq. (87), we have that

$$\mathsf{Pr}\Big[\exists s_0 \in \mathbb{Z}, \frac{1}{2\pi\beta}(\arg(\frac{z(\alpha+\beta)}{z(\alpha)}) + 2\pi s_0) \in \Big[\mathsf{left} + (q'-1)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + q'\frac{\mathsf{len}}{\mathsf{num}}\Big]\Big] \leq 100c$$

As a result, we know that in this case, with probability at least $1 - O(c)$,

$$\Theta \cap \left[\mathsf{left} + (q'-1)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + q'\frac{\mathsf{len}}{\mathsf{num}}\right] = \emptyset.$$

□

Then, we consider R independent voters, i.e., R significant samples $\alpha_1, \dots, \alpha_R$. The following claim shows that the true part and its left and right neighbors will get at least R votes. Meanwhile, those parts far away from the true part will get at most $R/2$ votes with high probability.

Claim L.7. *For For* $\mathsf{len} \in \mathbb{R}_+, \mathsf{num} \in \mathbb{Z}_+, q \in [1, \mathsf{num}]$*, let* $f_0 \in [\mathsf{left} + (q-1)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + q\frac{\mathsf{len}}{\mathsf{num}}]$*. Let* $\beta \sim$ *Uniform*$([\frac{c}{2} \cdot \frac{\mathsf{num}}{\mathsf{len}}, c \cdot \frac{\mathsf{num}}{\mathsf{len}}])$ *with constant* $c \in (0, 0.01)$*. For any constant* $\varepsilon \in (0, 0.01 \cdot c^2)$*, Let* $\alpha_1, \cdots, \alpha_R \in \mathbb{R}$ *such that for any* $i \in [R]$*,*

$$|z(\alpha_i + \beta) - z(\alpha_i)e^{2\pi \mathbf{i} f_0 \beta}|^2 \le \varepsilon |z(\alpha_i)|^2.$$

For any $i \in [R]$*, let*

$$\Theta_i = \left\{ \frac{1}{2\pi\beta}(\arg(\frac{z(\alpha_i + \beta)}{z(\alpha_i)}) + 2\pi s) \;\middle|\; s \in [\beta \mathsf{left} - 10, \beta(\mathsf{left} + \mathsf{len}) + 10] \cap \mathbb{Z} \right\}.$$

Then, it holds that:

1.

$$\sum_{i=1}^{R} \left| \Theta_i \cap \left[\mathsf{left} + (q-2)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + (q+1)\frac{\mathsf{len}}{\mathsf{num}} \right] \right| \ge R.$$

2. For any $|q' - q| \ge 3$*, with probability at least* $1 - O(c)^{R/6}$*,*

$$\sum_{i=1}^{R} \left| \Theta_i \cap \left[\mathsf{left} + (q'-2)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + (q'+1)\frac{\mathsf{len}}{\mathsf{num}} \right] \right| \le \frac{R}{2}.$$

Proof. **Part 1.**

By applying Claim L.5, we have that,

$$\left| \Theta_i \cap \left[\mathsf{left} + (q-2)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + (q+1)\frac{\mathsf{len}}{\mathsf{num}} \right] \right| \ge 1,$$

which implies that

$$\sum_{i=1}^{R} \left| \Theta_i \cap \left[\mathsf{left} + (q-2)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + (q+1)\frac{\mathsf{len}}{\mathsf{num}} \right] \right| \ge R.$$

Part 2. By applying Claim L.6, we have that, for any $|q - q'| > 1$, with probability at most $O(c)$,

$$\left| \Theta_i \cap \left[\mathsf{left} + (q'-1)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + q'\frac{\mathsf{len}}{\mathsf{num}} \right] \right| \ge 1.$$

By the setting of our parameter $\frac{1}{\beta} \ge \frac{\mathsf{len}}{\mathsf{num}}$, thus

$$\left| \Theta_i \cap \left[\mathsf{left} + (q'-1)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + q'\frac{\mathsf{len}}{\mathsf{num}} \right] \right| = 1.$$

Then, for any $|q - q'| \ge 3$, by a union bound over $q' - 1$, q', and $q' + 1$, with probability at most $O(c)$,

$$3 \ge \left| \Theta_i \cap \left[\mathsf{left} + (q'-2)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + (q'+1)\frac{\mathsf{len}}{\mathsf{num}} \right] \right| \ge 1.$$

Then, we have that

$$\begin{aligned}
&\Pr\Big[\sum_{i=1}^{R}\Big|\Theta_i \cap \Big[\mathsf{left} + (q'-2)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + (q'+1)\frac{\mathsf{len}}{\mathsf{num}}\Big]\Big| \geq \frac{R}{2}\Big] \\
\leq & \binom{R}{R/6} O(c)^{R/6} \\
\leq & (\frac{eR}{R/6})^{R/6} O(c)^{R/6} \\
\leq & O(c)^{R/6}
\end{aligned}$$

where the first step follows from there should be at least $0.5R/3 = R/6$ different $i \in [R]$ satisfying $|\Theta_i \cap [\mathsf{left} + (q'-2)\mathsf{len}/\mathsf{num}, \mathsf{left} + (q'+1)\mathsf{len}/\mathsf{num}]| \geq 1$,the second step follows from $\binom{n}{k} \leq (\frac{en}{k})^k$, the third step is straight forward.

The lemma is then proved. $\square$

Finally, we consider probabilistic voters, that is, for each sample α_i, with probability $1-\rho$, it is significant. The following claim shows the votes distribution in this case.

Claim L.8. *For* $\mathsf{len} \in \mathbb{R}_+, \mathsf{num} \in \mathbb{Z}_+, q \in [1, \mathsf{num}]$*, let* $f_0 \in [\mathsf{left} + (q-1)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + q\frac{\mathsf{len}}{\mathsf{num}}]$*. Let* $\beta \sim$ *Uniform*$([\frac{c}{2} \cdot \frac{\mathsf{num}}{\mathsf{len}}, c \cdot \frac{\mathsf{num}}{\mathsf{len}}])$ *with* $c = \Theta(1) \in (0, 0.01)$*,* $\varepsilon = \Theta(1) \in (0, 0.01 \cdot c^2)$*, let* $\alpha_1, \cdots, \alpha_R \in \mathbb{R}$ *such that for any* $i \in [R]$ *with probability at least* $1-\rho$*,*

$$|z(\alpha_i + \beta) - z(\alpha_i)e^{2\pi \mathbf{i} f_0 \beta}|^2 \leq \varepsilon |z(\alpha_i)|^2.$$

For any $i \in [R]$*, let*

$$\Theta_i = \Big\{\frac{1}{2\pi\beta}(\arg(\frac{z(\alpha_i+\beta)}{z(\alpha_i)}) + 2\pi s) \ \Big|\ s \in [\beta\mathsf{left} - 10, \beta(\mathsf{left} + \mathsf{len}) + 10] \cap \mathbb{Z}\Big\}.$$

Then, it holds that

1. *With probability at least* $1 - O(\rho)^{R/3}$*,*

$$\sum_{i=1}^{R}\Big|\Theta_i \cap \Big[\mathsf{left} + (q-2)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + (q+1)\frac{\mathsf{len}}{\mathsf{num}}\Big]\Big| \geq \frac{2R}{3}.$$

2. *For any* $|q'-q| \geq 3$*, with probability at least* $1 - O(c+\rho)^{R/6}$*,*

$$\sum_{i=1}^{R}\Big|\Theta_i \cap \Big[\mathsf{left} + (q'-2)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + (q'+1)\frac{\mathsf{len}}{\mathsf{num}}\Big]\Big| \leq \frac{R}{2}.$$

Proof. **Part 1.**

By applying Claim L.5, we have that with probability at most ρ,

$$\Big|\Theta_i \cap \Big[\mathsf{left} + (q-2)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + (q+1)\frac{\mathsf{len}}{\mathsf{num}}\Big]\Big| = 0,$$

then we have that,

$$\Pr\Big[\sum_{i=1}^{R}\Big|\Theta_i \cap [\mathsf{left} + (q-2)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + (q+1)\frac{\mathsf{len}}{\mathsf{num}}]\Big| \leq \frac{R}{3}\Big]$$

$$
\begin{aligned}
&\leq \binom{R}{R/3} O(\rho)^{R/3} \\
&\leq O(\frac{eR}{R/3})^{R/3} O(\rho)^{R/3} \\
&\leq O(\rho)^{R/3}
\end{aligned}
$$

where the first step follows from $|\Theta_i \cap [\mathsf{left} + (q-2)\mathsf{len}/\mathsf{num}, \mathsf{left} + (q+1)\mathsf{len}/\mathsf{num}]| = 0$ or 1 by our parameter setting $1/\beta > 3\mathsf{len}/\mathsf{num}$ and there should be at least $R/3$ different $i \in [R]$ satisfying $|\Theta_i \cap [\mathsf{left} + (q-2)\mathsf{len}/\mathsf{num}, \mathsf{left} + (q+1)\mathsf{len}/\mathsf{num}]| = 0$, the second step follows from $\binom{n}{k} \leq (\frac{en}{k})^k$, the third step is straight forward.

Part 2.

By applying Claim L.6, we have that, for any $|q - q'| > 1$, with probability at most $O(c) + \rho$,

$$
\left| \Theta_i \cap \left[\mathsf{left} + (q'-1)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + q'\frac{\mathsf{len}}{\mathsf{num}} \right] \right| = 1,
$$

where the probability follows from a union bound over the success of Claim L.6 and α_i being significant.

Thus, for any $|q - q'| \geq 3$, by a union bound, with probability at most $3((1-\rho)O(c) + \rho) = O(c + \rho)$,

$$
3 \geq \left| \Theta_i \cap \left[\mathsf{left} + (q'-2)\frac{\mathsf{len}}{\mathsf{num}}, \mathsf{left} + (q'+1)\frac{\mathsf{len}}{\mathsf{num}} \right] \right| \geq 1.
$$

Then, we have that

$$
\begin{aligned}
&\Pr\left[\sum_{i=1}^{R} \left| \Theta_i \cap [\mathsf{left} + (q'-2)\mathsf{len}/\mathsf{num}, \mathsf{left} + (q'+1)\mathsf{len}/\mathsf{num}] \right| \geq \frac{R}{2} \right] \\
\leq &\binom{R}{R/6} O(c+\rho)^{R/6} \\
\leq &(\frac{eR}{R/6})^{R/6} O(c+\rho)^{R/6} \\
\leq &O(c+\rho)^{R/6}
\end{aligned}
$$

where the first step follows from there should be at least $0.5R/3 = R/6$ different $i \in [R]$ satisfying $|\Theta_i \cap [\mathsf{left} + (q'-2)\mathsf{len}/\mathsf{num}, \mathsf{left} + (q'+1)\mathsf{len}/\mathsf{num}]| \geq 1$, the second step follows from $\binom{n}{k} \leq (\frac{en}{k})^k$, the third step is straight forward.

$\square$

M Signal Reconstruction

In this section, we wrap up all technical tools developed in previous sections and present our main result: a Fourier interpolation algorithm with improved time complexity, sample complexity, and output sparsity.

This section consists of two parts. The first part is devoted to the signal estimation. We first provide some tools that are useful for signal estimation (see Section M.1). Then, we formally define the heavy clusters and show their approximation property (see Section M.2). Next, we give a Fourier set query algorithm, which is a component in signal estimation (see Section M.3). We further show that it suffices to only reconstruct the signals in the bins satisfying the high SNR band condition (see Section M.4).

The second part focuses on the Fourier interpolation algorithm. Combining the frequency estimation algorithm in Section L with the signal estimation method we just developed, we obtain a Fourier interpolation algorithm with a constant success probability (see Section M.5). Then, we introduce the min-of-median signal estimator used to boost the success probability (see Section M.6). Finally, we prove our main theorem that gives a Fourier interpolation algorithm with high success probability (see Section M.7).

M.1 Preliminary

We provide some technical tools in this section.

The following two lemma shows that Fourier-polynomial mixed signals and Fourier-sparse signals can approximate each other.

Lemma M.1 ([CKPS16]). *For any $\Delta > 0$, $\delta > 0$, for any $n_1, \ldots, n_k \in \mathbb{Z}_{\geq 0}$ with $\sum_{j\in[k]} n_j = k$, let*

$$x^*(t) = \sum_{j\in[k]} e^{2\pi \mathbf{i} f_j t} \sum_{i=1}^{n_j} v_{j,i} e^{2\pi \mathbf{i} f'_{j,i} t},$$

where $|f'_{j,i}| \leq \Delta$ for each $j \in [k], i \in [n_j]$. There exist k polynomials $P_j(t)$ for $j \in [k]$ of degree at most

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

such that

$$\Big\| \sum_{j\in[k]} e^{2\pi \mathbf{i} f_j t} P_j(t) - x^*(t) \Big\|_T^2 \leq \delta \|x^*(t)\|_T^2.$$

Lemma M.2 ([CKPS16, Lemma 8.8]). *For any degree-d polynomial $Q(t) = \sum\limits_{j=0}^{d} c_j t^j$, any $T > 0$ and any $\varepsilon > 0$, there always exist $\gamma > 0$ and*

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

such that

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

The following fact shows an efficient method multi-point evaluation of a polynomial.

Fact M.3 ([VZGG99, Chapter 10]). *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 $O(d \log^2 d \log\log d)$ time to output the evaluations $\{P(t_1), P(t_2), \cdots, P(t_d)\}$.*

The following lemma shows the time complexity of evaluating a mixed polynomial.

Algorithm 6 Multipoint evaluation of a polynomial

1: **procedure** PolynomialEvaluation(P, t) ▷ Fact M.3
2: **return** $(P(t_1), P(t_2), \cdots, P(t_d))$ ▷ $t \in \mathbb{C}^d$
3: **end procedure**
4: **procedure** MixedPolynomialEvaluation$(\sum_{j=1}^{k} P_j(t)\exp(2\pi \mathbf{i} f_j t), t)$
5: **for** $j \in [k]$ **do**
6: $v_j \leftarrow (P_j(t_1), P_j(t_2), \cdots, P_j(t_d))$ ▷ $t \in \mathbb{C}^d$
7: **end for**
8: **return** $(\sum_{j=1}^{k} v_{j,1}\exp(2\pi \mathbf{i} f_j t_1), \sum_{j=1}^{k} v_{j,2}\exp(2\pi \mathbf{i} f_j t_2), \cdots, \sum_{j=1}^{k} v_{j,3}\exp(2\pi \mathbf{i} f_j t_3))$
9: **end procedure**

Lemma M.4 (Time complexity of Algorithm 6)**.** *Procedure* MixedPolynomialEvaluation *in Algorithm 6 runs*

$$O\Big(\sum_{j=1}^{k} \max\{d, \deg(P_j)\} \log^3(\max\{d, \deg(P_j)\})\Big)$$

time.

Proof. Procedure MixedPolynomialEvaluation in Algorithm 6 consists of the following steps:

- In line 5, the for loop repeats k times.
- In line 6, multipoint evaluation of a polynomial takes $d_j \log^c(d_j)$ times by Fact M.3, where $d_j = \max\{d, \deg(P_j)\}$.

Hence, the total time complexity is

$$\sum_{j=1}^{k} O(d_j \log^3(d_j)) = O\Big(\sum_{j=1}^{k} \max\{d, \deg(P_j)\} \log^3(\max\{d, \deg(P_j)\})\Big).$$

□

M.2 Heavy cluster

In this section, we formally define the heavy clusters and show that using "heavy frequencies" only yields a good approximation of the ground-truth signal.

Definition M.5 (Heavy cluster)**.** *Let* $x^*(t) = \sum\limits_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$ *and* $\mathcal{N} > 0$. *Let the filter* H *be defined as in Lemma E.9. Let* $\Delta_h = |\mathrm{supp}(\widehat{H})|$. *We say a frequency* f^* *belongs to an* $\mathcal{N}$-heavy cluster *if and only if*

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

Claim M.6. *Given* $x^*(t) = \sum\limits_{j=1}^{k} v_j e^{2\pi \mathbf{i} f_j t}$ *and any* $\mathcal{N} > 0$. *For the set of heavy frequencies:*

$$S^* = \left\{ j \in [k] \,\middle|\, \int_{f_j-\Delta_h}^{f_j+\Delta_h} |\widehat{H \cdot x^*}(f)|^2 \mathrm{d}f \geq T \cdot \mathcal{N}^2/k \right\},$$

and the signal $x_{S^*}(t) = \sum_{j \in S^*} v_j e^{2\pi \mathbf{i} f_j t}$, *it holds that*

$$\|x_{S^*} - x^*\|_T^2 \lesssim \mathcal{N}^2.$$

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

Then, we have that

$$\begin{aligned}
T\|x_{\overline{S^*}}(t)\|_T^2 = & \int_0^T |x_{\overline{S^*}}(t)|^2 \mathrm{d}t \\
\lesssim & \int_0^T |x_{\overline{S^*}}(t) \cdot H(t)|^2 \mathrm{d}t \\
\leq & \int_{-\infty}^{\infty} |x_{\overline{S^*}}(t) \cdot H(t)|^2 \mathrm{d}t \\
= & \int_{-\infty}^{\infty} |\widehat{x}_{\overline{S^*}}(f) * \widehat{H}(f)|^2 \mathrm{d}f \\
\leq & \sum_{j \in [k] \backslash S^*} \int_{f_j - \Delta_h}^{f_j + \Delta_h} |\widehat{x}_{\overline{S^*}}(f) * \widehat{H}(f)|^2 \mathrm{d}f \\
\leq & \sum_{j \in [k] \backslash S^*} T\mathcal{N}^2 / k \\
\leq & T\mathcal{N}^2,
\end{aligned}$$

where the first step follows from the definition of the norm, the second step follows from Lemma E.9 Property V, the third step is straight forward, the forth step follows from Parseval's theorem, the fifth step follows from the definition of $x_{\overline{S^*}}(t)$, the sixth step follows from the definition of heavy frequency, the seventh step is straightforward. □

M.3 Fourier set query

In this section, we present a Fourier set query algorithm such that for a Fourier-polynomial mixed signal, given all of its frequencies, the algorithm can reconstruct the signal very efficiently.

Lemma M.7. *For* $j \in [k]$, *given a* d_j*-degree polynomial* $P_j(t)$ *and a frequency* f_j. *Let* $x_S(t) = \sum_{j=1}^k P_j(t) \exp(2\pi \mathbf{i} f_j t)$. *Given observations of the form* $x(t) := x_S(t) + g(t)$ *for arbitrary noise* $g(t)$ *in time duration* $t \in [0, T]$. *Let* $D := \sum_{j=1}^k d_j$.

Then, there is an algorithm (Procedure SignalEstimation *in Algorithm 7) such that*

- *takes* $O(D \log(D))$ *samples from* $x(t)$,
- *runs* $O(D^\omega \log(D))$ *time,*
- *outputs* $y(t) = \sum_{j=1}^k P'_j(t) \exp(2\pi \mathbf{i} f_j t)$ *with* d*-degree polynomial* $P'_j(t)$, *such that with probability at least* 0.99, *we have*

$$\|y - x_S\|_T^2 \lesssim \|g\|_T^2.$$

Proof. By Lemma M.2, we have that, for all $t \in [0, T]$, there exist D-Fourier-sparse signals $y_1(t)$ and $x_{S,1}(t)$

$$|y(t) - y_1(t)| \leq \varepsilon_1, \tag{88}$$

and

$$|x_S(t) - x_{S,1}(t)| \leq \varepsilon_1. \tag{89}$$

Then, we have that

$$\begin{aligned} \|y(t) - x_S(t)\|_T^2 &\lesssim \|y(t) - y_1(t)\|_T^2 + \|x_S(t) - x_{S,1}(t)\|_T^2 + \|y_1(t) - x_{S,1}(t)\|_T^2 \\ &\lesssim 2\varepsilon_1 + \|y_1(t) - x_{S,1}(t)\|_T^2 \\ &\lesssim \|y_1(t) - x_{S,1}(t)\|_T^2 \end{aligned} \tag{90}$$

where the first step follows from $(a+b)^2 \leq 2a^2 + 2b^2$, the second step follows from Eq. (88) and Eq. (89), the third step follows from $\varepsilon_1 \lesssim \|y_1(t) - x_{S,1}(t)\|_T^2$.

We also have that

$$\begin{aligned} \|y_1(t) - x_{S,1}(t)\|_{S,w}^2 &\lesssim \|y_1(t) - y(t)\|_{S,w}^2 + \|x_{S,1}(t) - x_S(t)\|_{S,w}^2 + \|y(t) - x_S(t)\|_{S,w}^2 \\ &\lesssim 2\varepsilon_1 + \|y(t) - x_S(t)\|_{S,w}^2 \\ &\lesssim \|y(t) - x_S(t)\|_{S,w}^2 \end{aligned} \tag{91}$$

where the first step follows from $(a+b)^2 \leq 2a^2 + 2b^2$, the second step follows from Eq. (88) and Eq. (89), the third step follows from $\varepsilon_1 \lesssim \|y(t) - x_S(t)\|_{S,w}^2$.

By the definition of $y(t)$ in line 17 in Procedure SignalEstimation of Algorithm 7, we have that

$$\|y(t) - x(t)\|_{S,w}^2 \leq \|x_S(t) - x(t)\|_{S,w}^2 \tag{92}$$

We have that

$$\begin{aligned} \mathbb{E}[\|x - x_S\|_{S,w}^2] &= \mathbb{E}\Big[\sum_{i \in [|S|]} w_i |x(t_i) - x_S(t_i)|^2\Big] \\ &= \mathbb{E}\Big[\sum_{i \in [|S|]} \frac{1}{2T|S|D(t)} |x(t_i) - x_S(t_i)|^2\Big] \\ &= \sum_{i \in [|S|]} \mathop{\mathbb{E}}_{t_i \sim D(t)} \Big[\frac{1}{2T|S|D(t)} |x(t_i) - x_S(t_i)|^2\Big] \\ &= |S| \cdot \int_{-T}^{T} D(t) \frac{1}{2T|S|D(t)} |x(t) - x_S(t)|^2 \mathrm{d}t \\ &= \int_{-T}^{T} \frac{1}{2T} |x(t) - x_S(t)|^2 \mathrm{d}t \\ &= \|x(t) - x_S(t)\|_T^2 \end{aligned} \tag{93}$$

where the first step follows from the definition of the norm, the second step follows from the definition of w_i, the third step is straightforward, the forth follows from the definition of expectation, the fifth step follows from the definition of the norm.

We have that

$$\begin{aligned}
\|y - x_S\|_T^2 &\lesssim \|y_1 - x_{S,1}\|_T^2 \\
&\lesssim \|y_1 - x_{S,1}\|_{S,w}^2 \\
&\lesssim \|y - x_S\|_{S,w}^2 \\
&\lesssim \|y - x\|_{S,w}^2 + \|x - x_S\|_{S,w}^2 \\
&\lesssim \|x - x_S\|_{S,w}^2 \\
&\lesssim \|x - x_S\|_T^2,
\end{aligned}$$

where the first step follows from Eq. (90), the second step follows from Lemma J.3, the third step follows from Eq. (91), the forth step follows from $(a+b)^2 \le 2a^2 + 2b^2$, the fifth step follows from Eq. (92), the sixth step follows from Eq. (93) by Markov inequality with probability at least 0.99. $\square$

M.4 High signal-to-noise ratio band approximation

The goal of this section is to prove the following lemma, which roughly states that for the heavy frequencies, it suffices to only reconstruct those in the bins with high SNRs.

Lemma M.8. *Let $x^*(t) = \sum_{j=1}^k v_j e^{2\pi \mathbf{i} f_j t}$ be the ground-truth signal and $x(t) = x^*(t) + g(t)$ be the noisy observation signal. Let H be defined as in Definition E.5, $G_{\sigma,b}^{(j)}$ be defined as in Definition C.2 with (σ, b) such that Large Offset event does not happen. Let $U := \{t_0 \in \mathbb{R} \mid H(t) > 1 - \delta_1 \ \forall t \in [t_0, t_0 + \beta]\}$. Let*

$$S := \left\{ j \in [k] \ \middle|\ \int_{f_j - \Delta}^{f_j + \Delta} |\widehat{H \cdot x^*}(f)|^2 \mathrm{d}f \ge T\mathcal{N}^2/k \right\},$$

and $x_S(t) = \sum_{j \in S} v_j e^{2\pi \mathbf{i} f_j t}$.

For $j \in [B]$, let $z_j^*(t) := (x^* \cdot H) * G_{\sigma,b}^{(j)}(t)$ and $z_j(t) = (x \cdot H) * G_{\sigma,b}^{(j)}(t)$. Let $g_j(t) := z_j(t) - z_j^*(t)$. Let

$$S_{g1} := \{ j \in [B] \mid \|g_j(t)\|_T^2 \le c\|z_j^*(t)\|_U^2 \}, \tag{94}$$

where $c \in (0, 0.001)$ is a small universal constant. Let

$$S_{g2} := \left\{ j \in [B] \ \middle|\ \exists f_0 \in \{f_1, \ldots, f_k\}, \text{ and } h_{\sigma,b}(f_0) = j, \text{ and } \int_{f_0 - \Delta}^{f_0 + \Delta} |\widehat{x^* \cdot H}(f)|^2 \mathrm{d}f \ge T\mathcal{N}^2/k \right\}.$$

Let $S_g = S_{g1} \cap S_{g2}$. Let $S_f := \{j \in [k] \mid h_{\sigma,b}(f_j) \in S_g\} \cap S$ and $x_{S_f}(t) := \sum\limits_{j \in S_f} v_j e^{2\pi \mathbf{i} f_j t}$.

Then, we have

$$\|x_{S_f}(t) - x_S(t)\|_T^2 \lesssim \|g(t)\|_T^2.$$

Proof. By the definition of S and S_f, we have that

$$S_f \subseteq S.$$

Let $[L, R] := U$. We have that for any $f \in S \backslash S_f$, $j = h_{\sigma,b}(f)$,

$$\begin{aligned}\|(g(t) \cdot H(t)) * G^{(j)}_{\sigma,b}(t)\|^2_T &\geq c\|(x^*(t) \cdot H(t)) * G^{(j)}_{\sigma,b}(t)\|^2_U \\ &\geq c\frac{T - k^2(T + L - R)}{R - L}\|(x^*(t) \cdot H(t)) * G^{(j)}_{\sigma,b}(t)\|^2_T \\ &\geq O(c) \cdot \|(x^*(t) \cdot H(t)) * G^{(j)}_{\sigma,b}(t)\|^2_T, \end{aligned} \tag{95}$$

where the first step follows from Eq. (94), the second step follows from Lemma J.9, the third step follows from the Lemma E.10.

Let $\mathcal{T} = S \backslash S_f$. And for $j \in [B]$, let

$$\mathcal{T}_j := \begin{cases} \{i \in S \mid h_{\sigma,b}(f_i) = j\}, & \forall j \in [B] \backslash S_g, \\ \emptyset, & \text{otherwise}. \end{cases}$$

It is easy to see that

$$\mathcal{T} = \bigcup_{i=1}^{B} \mathcal{T}_i.$$

Moreover, by Lemma C.4 Property I and III, the definition of $\mathcal{T}_j$ and $\widehat{G}^{(j)}_{\sigma,b}(f)$, and the Large Offset event not happening, we have that for any $f \in \mathrm{supp}(\widehat{x}_{\mathcal{T}_j} * \widehat{H})$,

$$\widehat{G}^{(j)}_{\sigma,b}(f) \geq 1 - \frac{\delta}{k}, \tag{96}$$

where $x_{\mathcal{T}_j} = \sum_{i \in \mathcal{T}_j} v_i e^{2\pi \mathbf{i} f_i t}$ and $\widehat{x}_{\mathcal{T}_j}$ is its Fourier transform.

Then, we have that

$$\begin{aligned} & T\|(x^*(t) \cdot H(t)) * G^{(j)}_{\sigma,b}(t)\|^2_T \\ = & \int_0^T |(x^*(t) \cdot H(t)) * G^{(j)}_{\sigma,b}(t)|^2 \mathrm{d}t \\ \gtrsim & \int_{-\infty}^{\infty} |(x^*(t) \cdot H(t)) * G^{(j)}_{\sigma,b}(t)|^2 \mathrm{d}t \\ = & \int_{-\infty}^{\infty} |(\widehat{x}^*(f) * \widehat{H}(f)) \cdot \widehat{G}^{(j)}_{\sigma,b}(f)|^2 \mathrm{d}f \\ = & \int_{-\infty}^{\infty} |(\widehat{x}_{\mathcal{T}_j}(f) * \widehat{H}(f)) \cdot \widehat{G}^{(j)}_{\sigma,b}(f)|^2 \mathrm{d}f + \int_{-\infty}^{\infty} |(\widehat{x}_{[k] \backslash \mathcal{T}_j}(f) * \widehat{H}(f)) \cdot \widehat{G}^{(j)}_{\sigma,b}(f)|^2 \mathrm{d}f \\ \geq & \int_{-\infty}^{\infty} |(\widehat{x}_{\mathcal{T}_j}(f) * \widehat{H}(f)) \cdot \widehat{G}^{(j)}_{\sigma,b}(f)|^2 \mathrm{d}f \\ \gtrsim & \int_{-\infty}^{\infty} |\widehat{x}_{\mathcal{T}_j}(f) * \widehat{H}(f)|^2 \mathrm{d}f \end{aligned} \tag{97}$$

where the first step follows from the definition of the norm, the second step follows from Lemma G.3, third step follows from Parseval's theorem, the forth step follows from the Large Offset event not happening and the definition of $\mathcal{T}_j$, the fifth step is straight forward, the sixth step follows from Eq. (96).

Thus, we have that

$$\begin{aligned}
& T\|x_{S_f}(t) - x_S(t)\|_T^2 \\
= & T\|x_{\mathcal{T}}(t)\|_T^2 \\
\lesssim & T\|x_{\mathcal{T}}(t) \cdot H(t)\|_T^2 \\
= & \int_0^T |x_{\mathcal{T}}(t) \cdot H(t)|^2 \mathrm{d}t \\
\leq & \int_{-\infty}^{\infty} |x_{\mathcal{T}}(t) \cdot H(t)|^2 \mathrm{d}t \\
= & \int_{-\infty}^{\infty} |\widehat{x}_{\mathcal{T}}(f) * \widehat{H}(f)|^2 \mathrm{d}f \\
= & \sum_{j=1}^{B} \int_{-\infty}^{\infty} |\widehat{x}_{\mathcal{T}_j}(f) * \widehat{H}(f)|^2 \mathrm{d}f \\
\lesssim & \sum_{j \in [B] \setminus S_g} T\|(x^*(t) \cdot H(t)) * G_{\sigma,b}^{(j)}(t)\|_T^2 \\
\lesssim & \sum_{j \in [B] \setminus S_g} T\|(g(t) \cdot H(t)) * G_{\sigma,b}^{(j)}(t)\|_T^2 \qquad (98)
\end{aligned}$$

where the first step follows from the definition of $\mathcal{T}$, the second step follows from $x_{\mathcal{T}}$ is a k-Fourier-sparse signal and Lemma E.9 Property V, the third step follows from the definition of the norm, the forth step is straight forward, the fifth step follows from Parseval's theorem, the sixth step follows from the definition of $\mathcal{T}_j$ and the Large Offset event not happened, the seventh step follows from Eq. (97), the eighth step follows from Eq. (95).

Eq. (98) can be upper bounded by the summation over all bins, which can be further upper bounded as follows:

$$\begin{aligned}
& \sum_{j \in [B]} T \cdot \|(g(t) \cdot H(t)) * G_{\sigma,b}^{(j)}(t)\|_T^2 \\
= & \sum_{j \in [B]} \int_0^T |(g(t) \cdot H(t)) * G_{\sigma,b}^{(j)}(t)|^2 \mathrm{d}t \\
\leq & \sum_{j \in [B]} \int_{-\infty}^{\infty} |(g(t) \cdot H(t)) * G_{\sigma,b}^{(j)}(t)|^2 \mathrm{d}t \\
\leq & \sum_{j \in [B]} \int_{-\infty}^{\infty} |(\widehat{g}(f) * \widehat{H}(f)) \cdot \widehat{G}_{\sigma,b}^{(j)}(f)|^2 \mathrm{d}f \\
= & \int_{-\infty}^{\infty} |(\widehat{g}(f) * \widehat{H}(f))|^2 \cdot \sum_{j \in [B]} |\widehat{G}_{\sigma,b}^{(j)}(f)|^2 \mathrm{d}f \\
\lesssim & \int_{-\infty}^{\infty} |(\widehat{g}(f) * \widehat{H}(f))|^2 \mathrm{d}f \\
= & \int_{-\infty}^{\infty} |(g(t) \cdot H(t))|^2 \mathrm{d}t \\
= & \int_0^T |(g(t) \cdot H(t))|^2 \mathrm{d}t
\end{aligned}$$

$$\lesssim \int_0^T |g(t)|^2 \mathrm{d}t$$
$$= T\|g(t)\|_T^2 \tag{99}$$

where the first step follows from the definition of the norm, the second step is straightforward, the third step follows from Parseval's theorem, the forth step is straightforward, the fifth step follows from Lemma C.6, the sixth step follows from Parseval's theorem, the seventh step follows from $g(t) = 0, \forall t \in \mathbb{R}\backslash[0,T]$, the eighth step follows from Lemma E.9 Property I, II, the ninth step follows from the definition of the norm.

Therefore, we get that

$$\begin{aligned} & T\|x_{S_f}(t) - x_S(t)\|_T^2 \\ \lesssim & \sum_{j\in[B]\backslash S_g} T\|(g(t)\cdot H(t)) * G_{\sigma,b}^{(j)}(t)\|_T^2 \\ \leq & \sum_{j\in[B]} T\|(g(t)\cdot H(t)) * G_{\sigma,b}^{(j)}(t)\|_T^2 \\ \lesssim & \, T\|g(t)\|_T^2, \end{aligned}$$

where the first step follows from Eq. (98), the second step is straight forward, the third step follows from Eq. (99).

The lemma is then proved.

□

M.5 Fourier interpolation with constant success probability

In this section, we give an algorithm for Fourier interpolation by combining our frequency estimation algorithm with a signal estimation algorithm. However, it only succeeds with a constant probability.

Theorem M.9. *Let $x(t) = x^*(t) + g(t)$, where $x^*(t) \in \mathcal{F}_{k,F}$ and $g(t)$ is arbitrary noise. Given samples of x over $[0,T]$, there is an algorithm (Procedure* ConstantProbFourierInterpolation *in Algorithm 8) that uses*

$$O(k^4 \log^3(k) \log^2(1/\delta_1) \log(\log(1/\delta_1)) \log(FT) \log(\log(FT)))$$

samples, runs in

$$O(k^{4\omega} \log^{2\omega+1}(k) \log^{2\omega}(1/\delta_1) \log(\log(1/\delta_1)) \log(FT) \log(\log(FT)))$$

time, and outputs an $O(k^4 \log^4(k/\delta))$-Fourier-sparse signal $y(t)$ such that with probability at least 0.6*,*

$$\|y - x^*\|_T \lesssim \|g\|_T + \delta\|x^*\|_T.$$

Proof. Let $\mathcal{N}^2 := \|g(t)\|_T^2 + \delta\|x^*(t)\|_T^2$ be the noisy level of the observation signal.

Heavy-clusters approximation. Let S be the set of heavy frequencies:

$$S = \Big\{ j \in [k] \Big| \int_{f_j - \Delta_h}^{f_j + \Delta_h} |\widehat{H \cdot x^*}(f)|^2 \mathrm{d}f \geq T \cdot \mathcal{N}^2 / k \Big\},$$

where $\Delta_h = |\mathrm{supp}(\widehat{H})|$, and let $x_S(t) = \sum_{j\in S} v_j e^{2\pi \mathbf{i} f_j t}$. By Claim M.6, we have

$$\|x_S - x^*\|_T \lesssim \mathcal{N}, \tag{100}$$

which implies that it suffices to reconstruct x_S, instead of x^*.

Frequency estimation. Conditioning on Large Offset event not happening, which holds with probability at least 0.6 by Lemma D.8, let $S_f \subseteq S$ be defined as in Lemma M.8 and $x_{S_f}(t) = \sum_{j\in S_f} v_j e^{2\pi \mathbf{i} f_j t}$. By Lemma M.8, we have

$$\|x_{S_f}(t) - x_S(t)\|_T^2 \lesssim \|g(t)\|_T^2. \tag{101}$$

Furthermore, by Theorem L.2, there is an algorithm that outputs a set of frequencies $L \subset \mathbb{R}$ of size B such that with probability at least $1 - 2^{-\Omega(k)}$, for any $j \in S_f$, there exists an $\widetilde{f} \in L$ such that,

$$|f_j - \widetilde{f}| \lesssim \Delta.$$

Fourier-polynomial mixed signal approximation. We define a map $p : \mathbb{R} \to L$ as follows:

$$p(f) := \arg\min_{\widetilde{f}\in L} |f - \widetilde{f}| \quad \forall f \in \mathbb{R}.$$

Then, $x_{S_f}(t)$ can be expressed as

$$\begin{aligned} x_{S_f}(t) &= \sum_{j\in S_f} v_j e^{2\pi \mathbf{i} f_j t} \\ &= \sum_{j\in S_f} v_j e^{2\pi \mathbf{i} \cdot p(f_j)t} \cdot e^{2\pi \mathbf{i}\cdot (f_j - p(f_j))t} \\ &= \sum_{\widetilde{f}\in L} e^{2\pi \mathbf{i} \widetilde{f} t} \cdot \sum_{j\in S_f:\ p(f_j)=\widetilde{f}} v_j e^{2\pi \mathbf{i}(f_j - \widetilde{f})t}, \end{aligned}$$

where the first step follows from the definition of x_{S_f}, the last step follows from interchanging the summations.

For each $\widetilde{f}_i \in L$, by Lemma M.1 with $x^* = x_{S_f}$, there exists a degree $d = O(T\Delta + k^3 \log k + k \log 1/\delta)$ polynomial $P_i(t)$ such that,

$$\Big\|x_{S_f}(t) - \sum_{\widetilde{f}_i\in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t)\Big\|_T \le \sqrt{\delta}\|x_{S_f}(t)\|_T \tag{102}$$

Reconstructing the polynomials. Define the following function family:

$$\mathcal{F} := \mathrm{span}\Big\{e^{2\pi \mathbf{i} \widetilde{f} t} \cdot t^j \mid \widetilde{f} \in L, j \in \{0, 1, \ldots, d\}\Big\}.$$

Note that $\sum_{\widetilde{f}_i\in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t) \in \mathcal{F}$.

Let $D := d \cdot |L|$. By Lemma M.7, there is an algorithm that runs in $O(\varepsilon^{-1} D^{\omega} \log^3(D) \log(1/\rho))$-time using $O(\varepsilon^{-1} D \log^3(D) \log(1/\rho))$ samples, and outputs $y'(t) \in \mathcal{F}$ such that, with probability $1 - \rho$,

$$\Big\|y'(t) - \sum_{\widetilde{f}_i\in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t)\Big\|_T \le (1+\varepsilon)\Big\|x(t) - \sum_{\widetilde{f}_i\in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t)\Big\|_T \tag{103}$$

Thus, we have that

$$\Big\|y'(t) - \sum_{\widetilde{f}_i\in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t)\Big\|_T \lesssim \Big\|\sum_{\widetilde{f}_i\in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t) - x^*(t)\Big\|_T + \|x(t) - x^*(t)\|_T$$

$$= \Big\| \sum_{\widetilde{f}_i \in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t) - x^*(t) \Big\|_T + \|g(t)\|_T, \tag{104}$$

where the first step follows from triangle inequality, the second step follows from the definition of $g(t)$.

For the first term, we have that

$$\begin{aligned}
\Big\| \sum_{\widetilde{f}_i \in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t) - x^*(t) \Big\|_T &\lesssim \Big\| \sum_{\widetilde{f}_i \in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t) - x_{S_f}(t) \Big\|_T + \|x_{S_f}(t) - x^*(t)\|_T \\
&\lesssim \sqrt{\delta} \|x_{S_f}(t)\|_T + \|x_{S_f}(t) - x^*(t)\|_T \\
&\leq \sqrt{\delta} (\|x_{S_f}(t) - x^*(t)\|_T + \|x^*(t)\|_T) + \|x_{S_f}(t) - x^*(t)\|_T \\
&\lesssim \|x_{S_f}(t) - x^*(t)\|_T + \sqrt{\delta} \|x^*(t)\|_T \\
&\leq \|x_{S_f}(t) - x_S(t)\|_T + \|x_S(t) - x^*(t)\|_T + \sqrt{\delta} \|x^*(t)\|_T \\
&\lesssim \|x_{S_f}(t) - x_S(t)\|_T + \mathcal{N} + \sqrt{\delta} \|x^*(t)\|_T \\
&\lesssim \|g(t)\|_T + \mathcal{N} + \sqrt{\delta} \|x^*(t)\|_T,
\end{aligned} \tag{105}$$

where the first step follows from triangle inequality, the second step follows from Eq. (102), the third step follows from triangle inequality, the forth step follows is straightforward, the fifth step follows from triangle inequality, the sixth step follows from Eq. (100), and the last step follows from Eq. (101).

Hence, we get that

$$\begin{aligned}
\Big\| y'(t) - \sum_{\widetilde{f}_i \in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t) \Big\|_T &\leq \Big\| \sum_{\widetilde{f}_i \in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t) - x^*(t) \Big\|_T + \|g(t)\|_T \\
&\lesssim \|g(t)\|_T + \mathcal{N} + \sqrt{\delta} \|x^*(t)\|_T
\end{aligned} \tag{106}$$

where the first step follows from Eq. (104), the second step follows from Eq. (105).

Transforming back to Fourier-sparse signal. By Lemma M.2, we have that there is a $O(kd)$-Fourier-sparse signal $y(t)$, such that

$$\|y(t) - y'(t)\|_T \leq \delta' \tag{107}$$

where $\delta' > 0$ is any positive real number. Thus, $y(t)$ can be arbitrarily close to $y'(t)$. Moreover, the sparsity of $y(t)$ is

$$O(kd) = O(k \cdot (T\Delta + k^3 \log k + k \log 1/\delta)) = O(k^4 \log^4(k/\delta)),$$

which follows from Lemma E.9 Property III:

$$\Delta = k\Delta_h = k|\mathrm{supp}(\widehat{H})| = O(k^3 \log^2(k) \log^2(1/\delta_1)/T).$$

Moreover, we take

$$\mathcal{N} = \sqrt{\|g\|_T^2 + \delta \|x^*\|_T^2} \leq \|g\|_T + \sqrt{\delta} \|x^*\|_T. \tag{108}$$

Therefore, the total approximation error can be bounded as follows:

$$\|y(t) - x^*(t)\|_T$$

$$\begin{aligned}
&\leq \|y(t) - y'(t)\|_T + \Big\|y'(t) - \sum_{\widetilde{f}_i \in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t)\Big\|_T + \Big\| \sum_{\widetilde{f}_i \in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t) - x^*(t)\Big\|_T \\
&\lesssim \Big\|y'(t) - \sum_{\widetilde{f}_i \in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t)\Big\|_T + \Big\| \sum_{\widetilde{f}_i \in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t) - x^*(t)\Big\|_T \\
&\lesssim \Big\|y'(t) - \sum_{\widetilde{f}_i \in L} e^{2\pi \mathbf{i} \widetilde{f}_i t} P_i(t)\Big\|_T + \mathcal{N} + \|g(t)\|_T + \sqrt{\delta}\|x^*(t)\|_T \\
&\lesssim \mathcal{N} + \|g(t)\|_T + \sqrt{\delta}\|x^*(t)\|_T \\
&\lesssim \|g(t)\|_T + \sqrt{\delta}\|x^*(t)\|_T, \qquad (109)
\end{aligned}$$

where the first step follows from triangle inequality, the second step follows from Eq. (107), the third step follows from Eq. (105), the forth step follows from Eq. (106), the fifth step follows from $\mathcal{N} = \sqrt{\|g\|_T^2 + \delta\|x^*\|_T^2} \leq \|g\|_T + \sqrt{\delta}\|x^*\|_T$.

The correctness then follows by re-scaling δ.

The running time of the algorithm follows from Lemma M.10, and the sample complexity follows from Lemma M.11.

The theorem is then proved. □

Lemma M.10 (Running time of Algorithm 8)**.** *Procedure* ConstantProbFourierInterpolation *in Algorithm 8 runs in*

$$O(k^{4\omega} \log^{2\omega+1}(k) \log^{2\omega}(1/\delta_1) \log(\log(1/\delta_1)) \log(FT) \log(\log(FT)))$$

times.

Proof. Procedure ConstantProbFourierInterpolation in Algorithm 8 consists of the following two steps:

- Line 2 calls Procedure FrequencyEstimationX. By Theorem L.2, it runs in
  $$O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)))$$
  time.
- Line 3 calls Procedure SignalEstimation. By Lemma M.7, it runs in
  $$O(\varepsilon^{-1} D^{\omega} \log(D) \log(1/\rho))$$
  time, where ε, ρ are set to be universal constants and $D = B \cdot d$.

Following from the setting in the algorithm, we have that

$$\begin{aligned}
B &= O(k), \\
d &= O(\Delta T + k^3 \log k + k \log 1/\delta).
\end{aligned}$$

By Lemma E.9 Property III, we have that

$$\Delta = k\Delta_h = k|\mathrm{supp}(\widehat{H}(f))| = O(k^3 \log^2(k) \log^2(1/\delta_1)/T).$$

As a result, we have that

$$D = B \cdot d = O(k^4 \log^2(k) \log^2(1/\delta_1)) \tag{110}$$

Thus, the time complexity of Procedure ConstantProbFourierInterpolation in Algorithm 8 is

$$\begin{aligned}
& O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1)) + O(\varepsilon^{-1} D^{\omega} \log(D) \log(1/\rho)) \\
\leq\ & O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1)) \\
& \quad + O(\varepsilon^{-1} (k^4 \log^2(k) \log^2(1/\delta_1))^{\omega} \log(k^4 \log^2(k) \log^2(1/\delta_1)) \log(1/\rho)) \\
\leq\ & O(k^{4\omega} \log^{2\omega+1}(k) \log^{2\omega}(1/\delta_1) \log(\log(1/\delta_1)) \log(FT) \log(\log(FT)))
\end{aligned}$$

where the first step follows from Eq. (110), the second step follows from $\varepsilon = O(1), \rho = O(1), \rho_1 = O(1)$.

□

Lemma M.11 (Sample complexity of Algorithm 8)**.** *Procedure* ConstantProbFourierInterpolation *in Algorithm 8 takes*

$$O(k^4 \log^3(k) \log^2(1/\delta_1) \log(\log(1/\delta_1)) \log(FT) \log(\log(FT)))$$

samples.

Proof. The sample complexity of each steps of Procedure ConstantProbFourierInterpolation in Algorithm 8 is as follows:

- Line 2 calls Procedure FrequencyEstimationX. By Theorem L.2, it takes

  $$O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)))$$

  samples.

- Line 3 calls Procedure SignalEstimation. By Lemma M.7, it takes in

  $$O(\varepsilon^{-1} D \log(D) \log(1/\rho))$$

  samples, where ε, ρ are set to be a universal constant and $D = B \cdot d$.

By Eq. (110), we have

$$D = O(k^4 \log^2(k) \log^2(1/\delta_1)).$$

Thus, the sample complexity of Procedure ConstantProbFourierInterpolation in Algorithm 8 is

$$\begin{aligned}
& O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1)) + O(\varepsilon^{-1} D \log(D) \log(1/\rho)) \\
\leq\ & O(k^2 \log(k) \log(k/\delta_1) \log(FT) \log(\log(FT)/\rho_1)) \\
& \quad + O(\varepsilon^{-1} (k^4 \log^2(k) \log^2(1/\delta_1)) \log(k^4 \log^2(k) \log^2(1/\delta_1)) \log(1/\rho)) \\
\leq\ & O(k^4 \log^3(k) \log^2(1/\delta_1) \log(\log(1/\delta_1)) \log(FT) \log(\log(FT)))
\end{aligned}$$

where the first step follows from Eq. (110), the second step follows from $\varepsilon = O(1), \rho = O(1), \rho_1 = O(1)$.

□

M.6 Min-of-medians signal estimator

In this section, we propose a "min-of-medians" estimator for signals that can exponentially boost the success probability.

Lemma M.12. *Let $R_p \in \mathbb{N}$. For each $i \in [R_p]$, let $y_i(t)$ be a signal independently sampled from some distribution such that with probability at least* 0.9*,*

$$\|y_i(t) - x^*(t)\|_T^2 \lesssim \|g(t)\|_T^2.$$

Let $y(t) := y_{j^*}(t)$ where

$$j^* := \arg\min_{j \in [R_p]} \operatorname*{median}_{i \in [R_p]} \|y_j(t) - y_i(t)\|_T^2.$$

Then, with probability at least $1 - 2^{-\Omega(R_p)}$,

$$\|y(t) - x^*(t)\|_T^2 \lesssim \|g(t)\|_T^2.$$

Proof. Let $S = \{i \mid \|y_i(t) - x^*(t)\|_T^2 \lesssim \|g(t)\|_T^2\}$. By the Chernoff bound, we have that

$$\Pr[|S| \geq 3/4 R_p] \geq 1 - 2^{-\Omega(R_p)}.$$

For the ease of discussion, we suppose $|S| \geq 3/4 R_p$ holds in the following proof.

Fix any $j \in S$. Then, for any $q \in S$, we have that

$$\|y_j(t) - y_q(t)\|_T \leq \|y_j(t) - x^*(t)\|_T + \|x^*(t) - y_q(t)\|_T \lesssim \|g(t)\|_T^2, \tag{111}$$

where the first step follows from triangle inequality, the second step follows from the definition of S.

In other words, there are at least $|S| \geq (3/4)R_p$ elements such that Eq. (111) holds. By the definition of median, we get that

$$\operatorname*{median}_{i \in [R_p]} \|y_j(t) - y_i(t)\|_T^2 \lesssim \|g(t)\|_T^2. \tag{112}$$

By definition of $y(t)$, we have that,

$$\operatorname*{median}_{i \in [R_p]} \|y(t) - y_i(t)\|_T^2 \leq \operatorname*{median}_{i \in [R_p]} \|y_j(t) - y_i(t)\|_T^2 \lesssim \|g(t)\|_T^2, \tag{113}$$

where the first step follows from the definition of $y(t)$, the second step follows from Eq. (112).

By the definition of median, we know that there are $R_p/2$ elements $r \in [R_p]$ such that

$$\|y(t) - y_r(t)\|_T^2 \leq \operatorname*{median}_{i \in [R_p]} \|y(t) - y_i(t)\|_T^2 \lesssim \|g(t)\|_T^2, \tag{114}$$

where the last step follows from Eq. (113). Since $|S| \geq (3/4)R_p > (1/2)R_p$, there must exists an $r \in S$ such that Eq. (114) holds.

As a result, we have that

$$\begin{aligned}
\|y(t) - x^*(t)\|_T^2 &\leq \|y(t) - y_r(t)\|_T^2 + \|y_r(t) - x^*(t)\|_T^2 \\
&\lesssim \|g(t)\|_T^2 + \|y_r(t) - x^*(t)\|_T^2 \\
&\lesssim \|g(t)\|_T^2,
\end{aligned}$$

where the first step follows from triangle inequality, the second step follows from Eq. (113), the third step follows from the definition of S.

The lemma is then proved. $\square$

One potential issue in applying the min-of-median signal estimator is that, we may not be able to compute the distances $\|y_i(t)-y_j(t)\|_T^2$ exactly, but we can only estimate then with high accuracy. Therefore, we show that our estimator is robust with respect to approximated distances.

We first show a fact about the approximation of min and median.

Fact M.13. *Let $x_1, \cdots, x_n \in \mathbb{R}_+$, and $y_1, \cdots, y_n \in \mathbb{R}_+$ such that for any $i \in [n]$, $y_i \in [\alpha \cdot x_i, \beta \cdot x_i]$. Then, we have:*

- $\min\limits_{i\in[n]} y_i \in \left[\alpha \cdot \min\limits_{i\in[n]} x_i, \beta \cdot \min\limits_{i\in[n]} x_i\right]$.
- $\operatorname*{median}\limits_{i\in[n]} y_i \in \left[\alpha \cdot \operatorname*{median}\limits_{i\in[n]} x_i, \beta \cdot \operatorname*{median}\limits_{i\in[n]} x_i\right]$.

Proof. **Part 1:** Let $i^* = \arg\min\limits_{i\in[n]} y_i$. We have that

$$y_{i^*} \geq \alpha \cdot x_{i^*} \geq \alpha \cdot \min_{i\in[n]} x_i.$$

Let $j^* = \arg\min\limits_{j\in[n]} x_j$. We have that

$$\min_{j\in[n]} y_j \leq y_{j^*} \leq \beta \cdot x_{j^*} = \beta \cdot \min_{j\in[n]} x_j,$$

Hence,

$$\min_{i\in[n]} y_i \in \left[\alpha \cdot \min_{i\in[n]} x_i, \beta \cdot \min_{i\in[n]} x_i\right].$$

Part 2: For any $x_j \leq \operatorname*{median}\limits_{i\in[n]} x_i$, we have that

$$y_j \leq \beta \cdot x_j \leq \beta \cdot \operatorname*{median}_{i\in[n]} x_i.$$

Thus,

$$|\{j \in [n] \mid y_j \leq \beta \cdot \operatorname*{median}_{i\in[n]} x_i\}| \geq n/2,$$

which implies that

$$\operatorname*{median}_{i\in[n]} y_i \leq \beta \cdot \operatorname*{median}_{i\in[n]} x_i.$$

For any $x_j \geq \operatorname*{median}\limits_{i\in[n]} x_i$, we have that

$$y_j \geq \alpha \cdot x_j \geq \alpha \cdot \operatorname*{median}_{i\in[n]} x_i.$$

Thus,

$$|\{j \in [n] \mid y_j \geq \alpha \cdot \operatorname*{median}_{i\in[n]} x_i\}| \geq n/2,$$

which implies that

$$\operatorname*{median}_{i\in[n]} y_i \geq \alpha \cdot \operatorname*{median}_{i\in[n]} x_i.$$

As a result,

$$\operatorname*{median}_{i\in[n]} y_i \in \Big[\alpha \cdot \operatorname*{median}_{i\in[n]} x_i, \beta \cdot \operatorname*{median}_{i\in[n]} x_i\Big].$$

$\square$

The following lemma shows that our min-and-median estimator can still exponentially boost the success probability given access to approximated distances.

Lemma M.14 (Robust min-of-median signal estimator)**.** *Let $R_p \in \mathbb{N}$. For each $i \in [R_p]$, let $y_i(t)$ be a signal independently sampled from some distribution such that with probability at least* 0.9*,*

$$\|y_i(t) - x^*(t)\|_T^2 \lesssim \|g(t)\|_T^2.$$

Given $d \in \mathbb{R}_+^{R_p\times R_p}$ such that for any $i, j \in [R_p]$,

$$d_{i,j} \in \Big[\alpha \cdot \|y_i(t) - y_j(t)\|_T^2, \beta \cdot \|y_i(t) - y_j(t)\|_T^2\Big].$$

Let $y(t) := y_{j^*}(t)$ where

$$j^* := \operatorname*{arg\,min}_{j\in[R_p]} \operatorname*{median}_{i\in[R_p]} d_{j,i}.$$

Then, we have that, with probability at least $1 - 2^{-\Omega(R_p)}$,

$$\|y(t) - x^*(t)\|_T^2 \lesssim \frac{\beta}{\alpha}\|g(t)\|_T^2.$$

Proof. Let $S = \{i \mid \|y_i(t) - x^*(t)\|_T^2 \lesssim \|g(t)\|_T^2\}$. By the Chernoff bound, we have that

$$\Pr[|S| \geq 3/4R_p] \geq 1 - 2^{-\Omega(R_p)}.$$

For the ease of discussion, we suppose $|S| \geq 3/4R_p$ holds in the following proof.

Fix any $i^* \in S$. For any $q \in S$, we have that

$$\|y_{i^*}(t) - y_q(t)\|_T \leq \|y_{i^*}(t) - x^*(t)\|_T + \|x^*(t) - y_q(t)\|_T \lesssim \|g(t)\|_T^2, \tag{115}$$

where the first step follows from triangle inequality, the second step follows from the definition of S.

By the definition of median, since $|S| > R_p/2$, we know that

$$\operatorname*{median}_{i\in[R_p]} \|y_{i^*}(t) - y_i(t)\|_T^2 \lesssim \|g(t)\|_T^2. \tag{116}$$

Then, we have that

$$\begin{aligned}\operatorname*{median}_{i\in[R_p]} d_{j^*,i} &\leq \operatorname*{median}_{i\in[R_p]} d_{i^*,i}\\ &\leq \beta \cdot \operatorname*{median}_{i\in[R_p]} \|y_{i^*}(t) - y_i(t)\|_T^2\\ &\lesssim \beta \cdot \|g(t)\|_T^2, \end{aligned}\tag{117}$$

where the first step follows from the definition of j^*, the second step follows from Fact M.13, the third step follows from Eq. (116).

Since $|S| > R_p/2$, by the definition of median, there must exists an $r \in S$ such that

$$d_{j^*,r} \leq \underset{i\in[R_p]}{\text{median}}\ d_{j^*,i} \lesssim \beta \cdot \|g(t)\|_T^2. \tag{118}$$

As a result, we have that

$$\begin{aligned}
\|y(t) - x^*(t)\|_T^2 &\leq \|y(t) - y_r(t)\|_T^2 + \|y_r(t) - x^*(t)\|_T^2 \\
&\leq \frac{1}{\alpha} v_{j^*,r} + \|y_r(t) - x^*(t)\|_T^2 \\
&\lesssim \frac{\beta}{\alpha} \|g(t)\|_T^2 + \|y_r(t) - x^*(t)\|_T^2 \\
&\lesssim \frac{\beta}{\alpha} \|g(t)\|_T^2,
\end{aligned}$$

where the first step follows from triangle inequality, the second step follows from the definition of d, the third step follows from Eq. (118), the forth step follows from Eq. (117), the fifth step follows from $r \in S$.

The proof of the lemma is then completed. □

M.7 Main algorithm for Fourier interpolation

In this section, we present our main theorem—a time and sample efficient Fourier interpolation algorithm with high success probability. The pseudocode is given in Algorithm 8.

Theorem M.15 (Main Fourier interpolation algorithm). *Let $x(t) = x^*(t) + g(t)$, where x^* is k-Fourier-sparse signal with frequencies in $[-F, F]$. Given samples of x over $[0, T]$, there is an algorithm (Procedure* HighProbFourierInterpolation*) uses*

$$O(k^4 \log^6(k/\delta) \log(FT) \log(\log(FT)) \log(1/\rho))$$

samples, runs in

$$O(k^{4\omega} \log^{4\omega+2}(k/\delta) \log(FT) \log(\log(FT)) \log^5(1/\rho))$$

time, and outputs an $O(k^4 \log^4(k/\delta))$-Fourier-sparse signal $y(t)$ such that with probability at least $1 - \rho$,

$$\|y - x^*\|_T \lesssim \|g\|_T + \delta \|x^*\|_T.$$

Proof. We first prove the correctness of the algorithm.

Let

$$D(t) := \begin{cases} c \cdot (1 - |t/T|)^{-1} T^{-1}, & \text{for } |t| \leq T(1 - 1/k), \\ c \cdot k T^{-1}, & \text{for } |t| \in [T(1 - 1/k), T]. \end{cases}$$

Let $y_1(t), \cdots, y_{R_p}(t)$ be the outputs of R_p independent runs of Procedure ConstantProbFourierInterpolation in Algorithm 8. By Theorem M.9, we have that for any $j \in [R_p]$ with probability at least 0.9,

$$\|y_j(t) - x^*(t)\|_T^2 \lesssim \|g(t)\|_T^2.$$

Algorithm 7 Signal estimation algorithm

procedure WEIGHTEDSKETCH(m, k, T)
  $c \leftarrow \Theta(\log(k)^{-1})$
  $D(t)$ is defined as follows:

$$D(t) \leftarrow \begin{cases} c \cdot (1 - |t/T|)^{-1} T^{-1}, & \text{for } |t| \leq T(1 - 1/k), \\ c \cdot k T^{-1}, & \text{for } |t| \in [T(1 - 1/k), T]. \end{cases}$$

  $S_0 \leftarrow m$ i.i.d. samples from D
  for $t \in S_0$ **do**
    $w_t \leftarrow \frac{1}{2T \cdot |S_0| \cdot D(t)}$
  end for
  Set a new distribution $D'(t) \leftarrow w_t / \sum_{t' \in S_0} w_{t'}$ for all $t \in S_0$
  return D'
end procedure
procedure SIGNALESTIMATION(x, F, T, L)
  $\{f_1, f_2, \cdots, f_B\} \leftarrow L$ ▷ $L \in \mathbb{R}^B$
  $d \leftarrow O(\Delta T + k^3 \log k + k \log 1/\delta)$
  $s, \{t_1, t_2, \cdots, t_s\}, w \leftarrow$ WEIGHTEDSKETCH$(O(Bd \log(Bd)), Bd, T)$ ▷ $w \in \mathbb{R}^s$
  $A_{i, B \cdot j_2 + j_1} \leftarrow t_i^{j_2} \cdot \exp(2\pi \mathbf{i} f_{j_1} t_i)$, $A \in \mathbb{C}^{s \times B}$
  $b \leftarrow (x(t_1), x(t_2), \cdots, x(t_s))^\top$
  Solving the following weighted linear regression

$$v' \leftarrow \arg\min_{v' \in \mathbb{C}^{Bd}} \|\sqrt{w} \circ (Av' - b)\|_2.$$

  return $y(t) \leftarrow \sum_{j_1=1}^{B} \sum_{j_2=1}^{d} v'_{B \cdot j_2 + j_1} \cdot t^{j_2} \cdot \exp(2\pi \mathbf{i} f'_{j_1} t)$.
end procedure

Let $S = \{t_1, \ldots, t_s\}$ be $s = O(k \log(k) \log(R_p^2/\rho))$ i.i.d. samples from $D(t)$, and let $w_i = 1/(TsD(t_i))$ for $i \in [s]$. By Lemma J.3, for any $i, j \in [R_p]$, with probability at least $1 - \rho/R_p^2$,

$$\|y_i(t) - y_j(t)\|_{S,w}^2 \in [1/2, 3/2] \cdot \|y_i(t) - y_j(t)\|_T^2.$$

Let

$$j^* = \arg\min_{j \in [R_p]} \operatorname*{median}_{i \in [R_p]} \|y_j(t) - y_i(t)\|_T^2,$$

and let $y(t) := y_{j^*}(t)$.

By Lemma M.14, we have that with probability at least $1 - 2^{-\Omega(R_p)}$,

$$\|y_j(t) - x^*(t)\|_T^2 \lesssim \frac{3/2}{1/2} \|g(t)\|_T^2 \approx \|g(t)\|_T^2.$$

By setting $R_p = \log(1/\rho)$, we get the desired result. The correctness is then proved.

The time complexity follows from Lemma M.18. And the sample complexity follows from Lemma M.19.

The proof of the theorem is completed. □

Algorithm 8 Fourier-sparse signal interpolation

procedure CONSTANTPROBFOURIERINTERPOLATION(x, H, G, T, F)
    $L \leftarrow$ FREQUENCYESTIMATIONX(x, H, G, T, F) $\triangleright\ L \in \mathbb{R}^B$
    $y(t) \leftarrow$ SIGNALESTIMATION$(x, \varepsilon, k, F, T, L)$
    return $y(t)$
end procedure
procedure HIGHPROBFOURIERINTERPOLATION(x, H, G, T, F)
    $R_p \leftarrow \log(1/\rho)$
    for $i \in [R_p]$ **do**
        $y_i(t) \leftarrow$ CONSTANTPROBFOURIERINTERPOLATION(x, H, G, T, F)
    end for
    $y(t) \leftarrow$ MERGESIGNAL$(y_1(t), y_2(t), \cdots, y_{R_p}(t))$
    return $y(t)$
end procedure
procedure MERGESIGNAL$(y_1(t), y_2(t), \cdots, y_{R_p}(t))$
    $d \leftarrow O(\Delta T + k^3 \log k + k \log 1/\delta)$
    for $i \in [R_p]$ **do**
        for $j \in [R_p]$ **do**
            $s, \{t_1, t_2, \cdots, t_s\}, w \leftarrow$ WEIGHTEDSKETCH$(O(Bd \log(Bd) \log(R_p^2/\rho)), 2 \cdot Bd, T)$
            $\triangleright\ w \in \mathbb{R}^s$
            $S \leftarrow \{t_1, t_2, \cdots, t_s\}$
            $Y \leftarrow$ MIXEDPOLYNOMIALEVALUATION$(y_i - y_j, S)$ $\triangleright\ Y \in \mathbb{C}^s$
            $\|y_i(t) - y_j(t)\|_{S,w}^2 \leftarrow \sum_{l=1}^{s} w_l \cdot |Y_l|^2$
        end for
        $\mathsf{med}_i \leftarrow \mathrm{median}_{j \in [R_p]}\{\|y_i - y_j\|_{S,w}^2\}$
    end for
    $i^* \leftarrow \arg\min_{i \in [R_p]}\{\mathsf{med}_i\}$
    return y_{i^*}
end procedure

In the remaining of this section, we prove the time and sample complexities of Procedure HIGHPROBFOURIERINTERPOLATION in Algorithm 8.

The following two lemmas show the time complexity of Procedure MERGESIGNAL in Algorithm 8, which is used to boost the success probability of Fourier interpolation algorithm.

Lemma M.16 (Time complexity of Procedure WEIGHTEDSKETCH in Algorithm 7)**.** *Procedure* WEIGHTEDSKETCH *in Algorithm 7 runs in*

$$O(\varepsilon^{-2} k \log(k) \log(1/\rho))$$

time.

Proof. Procedure WEIGHTEDSKETCH contains the following steps:

- In line 4, sampling S_0 takes $O(\varepsilon^{-2} k \log(k) \log(1/\rho))$ times.
- In line 5, the for loop repeats $|S_0|$ times, and each takes $O(1)$ times.

Following from the setting in the algorithm, we have that

$$|S_0| = O(\varepsilon^{-2} k \log(k) \log(1/\rho)).$$

So, the time complexity of Procedure WeightedSketch in Algorithm 7 is

$$O(\varepsilon^{-2}k\log(k)\log(1/\rho)) + |S_0|\cdot O(1) = O(\varepsilon^{-2}k\log(k)\log(1/\rho)).$$

□

Lemma M.17 (Time complexity of Procedure MergeSignal in Algorithm 8)**.** *Procedure* MergeSignal *in Algorithm 8 runs in*

$$O(k^5\log^6(k)\log^4(1/\delta_1)\log^5(1/\rho))$$

time.

Proof. In each call of the Procedure MergeSignal in Algorithm 8, the for loop (Line 16) repeats R_p times, each consisting of the following steps:

- In Line 17, the for loop repeats R_p times and each iteration has the following steps:
  - Line 19 calls Procedure WeightedSketch. By Lemma M.16, it runs in

    $$O(Bd\log(Bd)\log(R_p^2/\rho))$$

    time.
  - Line 21 calls Procedure MixedPolynomialEvaluation. By Lemma M.4, it runs in

    $$O\Big(\sum_{j=1}^{k}\max\{d',\deg(P_j)\}\log^3(\max\{d',\deg(P_j)\})\Big)$$

    time, where $d' = O(Bd\log(Bd)\log(R_p^2/\rho))$ and $\deg(P_j) = d$.
- Line 26 computes the median in $R_p\log(R_p)$ time.

Following from the parameter setting in the algorithm, we have that

$$\begin{aligned}B &= O(k),\\ d &= O(\Delta T + k^3\log k + k\log 1/\delta).\end{aligned}$$

By Lemma E.9 Property III, we have that

$$\Delta = k\Delta_h = k|\mathrm{supp}(\widehat{H}(f))|/T = O(k^3\log^2(k)\log^2(1/\delta_1)/T).$$

As a result, we have that

$$B\cdot d = O(k^4\log^2(k)\log^2(1/\delta_1)) \tag{119}$$

Moreover, we have that

$$\begin{aligned}\max\{d',\deg(P_j)\} &= O(Bd\log(Bd)\log(R_p^2/\rho))\\ &= O(k^4\log^3(k)\log^2(1/\delta_1)\log(\log(1/\delta_1))\log(1/\rho)\log(\log(1/\rho)))\\ &\le O(k^4\log^3(k)\log^3(1/\delta_1)\log^2(1/\rho)).\end{aligned} \tag{120}$$

where the first step follows from the definition of d' and $\deg(P_j)$, the second step follows from Eq. (119), the third step is straight forward.

So, the time complexity of Procedure MergeSignal in Algorithm 8 is

$$\begin{aligned}
& R_p^2 \cdot \left(O(Bd\log(Bd)\log(R_p^2/\rho)) + O\Big(\sum_{j=1}^{k}\max\{d',\deg(P_j)\}\log^3(\max\{d',\deg(P_j)\})\Big)\right) \\
& + R_p \cdot O(R_p \log(R_p)) \\
& \leq O\Big(R_p^2\sum_{j=1}^{k}\max\{d',\deg(P_j)\}\log^3(\max\{d',\deg(P_j)\})\Big) \\
& \leq O(\log^2(1/\rho)\cdot k\cdot(k^4\log^3(k)\log^3(1/\delta_1)\log^2(1/\rho))\log^3(k\log(1/\delta_1)\log(1/\rho))) \\
& \leq O(k^5\log^6(k)\log^4(1/\delta_1)\log^5(1/\rho)),
\end{aligned}$$

where the first step follows from Eq. (119), the second step follows from Eq. (120), the third step follows from Eq. (120), the forth step is straight forward.

□

The following two lemmas show the time complexity and sample complexity of our main algorithm.

Lemma M.18 (Time complexity of the main algorithm)**.** *Procedure* HighProbFourierInterpolation *in Algorithm 8 runs in*

$$O(k^{4\omega}\log^{4\omega+2}(k/\delta)\log(FT)\log(\log(FT))\log^5(1/\rho))$$

times.

Proof. Procedure HighProbFourierInterpolation in Algorithm 8 consists of the following steps:

- In Line 8, the for loop repeats R_p times with the following step:
  - Line 9 calls Procedure ConstantProbFourierInterpolation. By Lemma M.10, it runs in

    $$O(k^{4\omega}\log^{2\omega+1}(k)\log^{2\omega}(1/\delta_1)\log(\log(1/\delta_1))\log(FT)\log(\log(FT)))$$

    time.
- Line 11 calls Procedure MergeSignal. By Lemma M.17, it runs in

  $$O(k^5\log^6(k)\log^4(1/\delta_1)\log^5(1/\rho))$$

  time.

Following from the setting in the algorithm, we have that

$$\delta_1 = \delta/\mathrm{poly}(k). \tag{121}$$

So, the time complexity of Procedure HighProbFourierInterpolation in Algorithm 8 in Algorithm 8 is

$$R_p \cdot O\left(k^{4\omega}\log^{2\omega+1}(k)\log^{2\omega}(1/\delta_1)\log(\log(1/\delta_1))\log(FT)\log(\log(FT))\right)$$

$$
\begin{aligned}
&+ O(k^5 \log^{c+3}(k) \log^4(1/\delta_1) \log^5(1/\rho)) \\
&= O(k^{4\omega} \log^{2\omega+1}(k) \log^{2\omega}(1/\delta_1) \log(\log(1/\delta_1)) \log(FT) \log(\log(FT)) \log^5(1/\rho)) \\
&\leq O(k^{4\omega} \log^{4\omega+2}(k/\delta) \log(FT) \log(\log(FT)) \log^5(1/\rho)),
\end{aligned}
$$

where the first step follows from $R_p = \log(1/\rho)$, the second step follows from Eq. (121).

$\square$

Lemma M.19 (Sample complexity of the main algorithm)**.** *Procedure* HighProbFourierInterpolation *in Algorithm 8 takes*

$$
O(k^4 \log^6(k/\delta) \log(FT) \log(\log(FT)) \log(1/\rho))
$$

samples.

Proof. Procedure Procedure HighProbFourierInterpolation in Algorithm 8 consists of the following steps:

- In Line 8, the for loop repeats R_p times:
  - Line 9 calls Procedure ConstantProbFourierInterpolation. By Lemma M.11, it takes

    $$
    O(k^4 \log^3(k) \log^2(1/\delta_1) \log(\log(1/\delta_1)) \log(FT) \log(\log(FT)))
    $$

    samples.

  The remaining steps do not use any new sample.

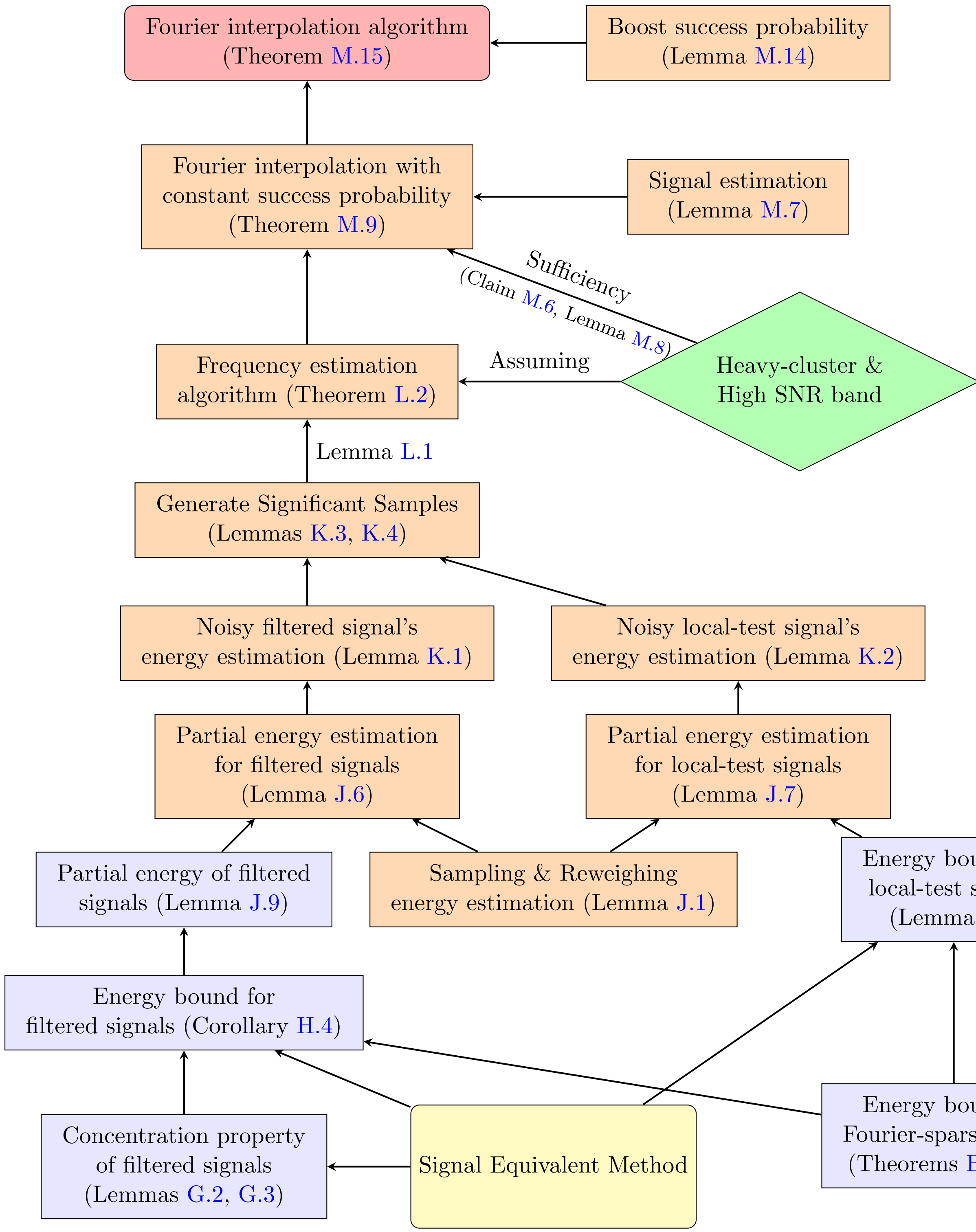
Thus, the total sample complexity is

$$
\begin{aligned}
&R_p \cdot O(k^4 \log^3(k) \log^2(1/\delta_1) \log(\log(1/\delta_1)) \log(FT) \log(\log(FT))) \\
&\leq O(k^4 \log^3(k) \log^2(1/\delta_1) \log(\log(1/\delta_1)) \log(FT) \log(\log(FT)) \log(1/\rho)) \\
&\leq O(k^4 \log^6(k/\delta) \log(FT) \log(\log(FT)) \log(1/\rho)),
\end{aligned}
$$

where the first step follows from $R_p = \log(1/\rho)$, the second step follows from Eq. (121).

$\square$

N Structure of Our Fourier Interpolation Algorithm



References

[AW21] Josh Alman and Virginia Vassilevska Williams. A refined laser method and faster matrix multiplication. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 522–539. SIAM, 2021.

[BE06] Peter Borwein and Tamás Erdélyi. Nikolskii-type inequalities for shift invariant function spaces. *Proceedings of the American Mathematical Society*, 134(11):3243–3246, 2006.

[Bou14] Jean Bourgain. An improved estimate in the restricted isometry problem. In *Geometric aspects of functional analysis*, pages 65–70. Springer, 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*, pages 493–507, 1952.

[CKPS16] Xue Chen, Daniel M Kane, Eric Price, and Zhao Song. Fourier-sparse interpolation without a frequency gap. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 741–750. IEEE, 2016.

[CP19a] Xue Chen and Eric Price. Active regression via linear-sample sparsification. In *Conference on Learning Theory (COLT)*, pages 663–695. PMLR, 2019.

[CP19b] Xue Chen and Eric Price. Estimating the frequency of a clustered signal. In *46th International Colloquium on Automata, Languages, and Programming (ICALP)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[CRT06] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8):1207–1223, 2006.

[CT65] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

[CT06] Emmanuel J Candes and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory*, 52(12):5406–5425, 2006.

[dP95] Gaspard Riche de Prony. Essai experimental et analytique: sur les lois de la dilatabilite des fluides elastique et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool, a differentes temperatures. *Journal Polytechnique ou Bulletin du Travail fait a l'Ecole Centrale des Travaux Publics*, 1795.

[GMS05] Anna C Gilbert, Shan Muthukrishnan, and Martin Strauss. Improved time bounds for near-optimal sparse fourier representations. In *Wavelets XI*, volume 5914, page 59141A. International Society for Optics and Photonics, 2005.

[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 (STOC)*, pages 563–578, 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 (SODA)*, pages 1183–1194. SIAM, 2012.

[HR17] Ishay Haviv and Oded Regev. The restricted isometry property of subsampled fourier matrices. In *Geometric aspects of functional analysis*, pages 163–179. Springer, 2017.

[IK14] Piotr Indyk and Michael Kapralov. Sample-optimal Fourier sampling in any constant dimension. In *IEEE 55th Annual Symposium onFoundations of Computer Science (FOCS)*, 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 (SODA)*, pages 480–499. SIAM, 2014.

[ITU92] ITU. Information technology - digital compression and coding of continuous - tone still images - requirements and guidelines. *CCITT, Recommendation*, 1992.

[JLS23] Yaonan Jin, Daogao Liu, and Zhao Song. Super-resolution and robust sparse continuous fourier transform in any constant dimension: Nearly linear time and sample complexity. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2023.

[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)*. https://arxiv.org/pdf/1604.00845.pdf, 2016.

[Kap17] Michael Kapralov. Sample efficient estimation and recovery in sparse FFT via isolation on average. In *58th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. https://arxiv.org/pdf/1708.04544, 2017.

[Kós08] Géza Kós. Two turán type inequalities. *Acta Mathematica Hungarica*, 119(3):219–226, 2008.

[Moi15] Ankur Moitra. Super-resolution, extremal functions and the condition number of vandermonde matrices. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 821–830, 2015.

[NSW19] Vasileios Nakos, Zhao Song, and Zhengyu Wang. (nearly) sample-optimal sparse fourier transform in any dimension; ripless and filterless. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1568–1577. IEEE, 2019.

[Pri11] Eric Price. Efficient sketches for the set query problem. In Dana Randall, editor, *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, San Francisco, California, USA, January 23-25, 2011*, pages 41–56. SIAM, 2011.

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

[Rab02] Majid Rabbani. Jpeg2000: Image compression fundamentals, standards and practice. *Journal of Electronic Imaging*, 11(2):286, 2002.

[RV08] Mark Rudelson and Roman Vershynin. On sparse reconstruction from fourier and gaussian measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 61(8):1025–1045, 2008.

[Son19] Zhao Song. *Matrix Theory: Optimization, Concentration and Algorithms*. PhD thesis, The University of Texas at Austin, 2019.

[SSWZ22] Zhao Song, Baocheng Sun, Omri Weinstein, and Ruizhe Zhang. Sparse fourier transform over lattices: A unified approach to signal reconstruction. *arXiv preprint arXiv:2205.00658*, 2022.

[VZGG99] Joachim Von Zur Gathen and Jürgen Gerhard. *Modern computer algebra*. Cambridge university press, 1999.

[Wat94] Andrew B. Watson. Image compression using the discrete cosine transform. *Mathematica Journal*, 4:81–88, 1994.

[Wil12] Virginia Vassilevska Williams. Multiplying matrices faster than coppersmith-winograd. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing (STOC)*, pages 887–898. ACM, 2012.