

Super-resolution and Robust Sparse Continuous Fourier Transform in Any Constant Dimension: Nearly Linear Time and Sample Complexity*

Yaonan Jin[†] Daogao Liu[‡] Zhao Song[§]

Abstract

The ability to resolve detail in the object that is being imaged, named by resolution, is the core parameter of an imaging system. Super-resolution is a class of techniques that can enhance the resolution of an imaging system and even transcend the diffraction limit of systems. Despite huge success in the application, super-resolution is not well understood on the theoretical side, especially for any dimension $d \geq 2$. In particular, in order to recover a k -sparse signal, all previous results suffer from either/both $\text{poly}(k)$ samples or running time.

We design robust algorithms for any (constant) dimension under a strong noise model based on developing some new techniques in Sparse Fourier transform (Sparse FT), such as inverting a robust linear system, “eggshell” sampling schemes, and partition and voting methods in high dimension. These algorithms are the first to achieve running time and sample complexity (nearly) linear in the number of source points and logarithmic in bandwidth for any constant dimension, and we believe the techniques developed in the work can find their further applications on the Super-resolution and Sparse FT problem.

*A preliminary version of this paper appears at ACM-SIAM Symposium on Discrete Algorithms (SODA 2023).

[†]yj2552@columbia.edu. Columbia University.

[‡]dgliu@uw.com. University of Washington.

[§]zsong@adobe.com. Adobe Research.

Contents

1	Introduction	3
1.1	Our results	4
1.2	Related works	5
1.2.1	Super-resolution with a different focus	5
1.2.2	Prior works on the sparse FT problem	6
1.3	Our techniques	7
1.3.1	Hashing and sampling	7
1.3.2	Sparse recovery	10
2	Preliminaries	13
2.1	Notations	13
2.2	Fourier transform and convolution	13
2.3	An overview of previous techniques	14
2.4	Technical barriers against a better tone estimation duration	15
3	Probability tools	16
4	Filter, permutation and hashing in multiple dimensions	17
4.1	Construction of filter $(\mathcal{G}(t), \hat{\mathcal{G}}(f))$	18
4.2	Properties of filter $(\mathcal{G}(t), \hat{\mathcal{G}}(f))$	18
4.3	Proof of properties	20
4.4	Construction and properties of standard window $(\mathcal{G}'(t), \hat{\mathcal{G}}'(f))$	20
4.5	Permutation and hashing	21
4.6	HASHTOBINS: algorithm	23
4.7	HASHTOBINS: probabilities of bad events	25
4.8	HASHTOBINS: error due to noise	28
4.9	HASHTOBINS: error due to bad events	29
4.10	Performance guarantees	33
5	Locate inner	33
5.1	Definitions and algorithm	33
5.2	Sample complexity and running time	35
5.3	Voting process	37
5.4	Election process	45
5.5	Performance guarantees	48
5.6	Sampling time points	48
5.6.1	Duration requirement	48
5.6.2	Performance guarantees	50
5.7	Stronger Guarantee	51
6	Locate signal	54
6.1	Algorithm	54
6.2	Sample complexity and running time	55
6.3	Duration requirement	56
6.4	Performance guarantees	56
6.5	Stronger guarantees	57

7	Sparse recovery	58
7.1	Definitions and facts	58
7.2	ESTIMATE SIGNAL	58
7.3	ONESTAGE: algorithm, sample complexity and running time	59
7.4	ONESTAGE: performance guarantees	60
7.5	MULTISTAGE	63
7.6	MERGEDSTAGE: algorithm and running time	64
7.7	MERGEDSTAGE: performance guarantees	66
7.8	Running MERGEDSTAGE twice	69
7.9	RECOVERYSTAGE	71
8	Converting tone estimation into signal estimation	74
8.1	Improvement of signal estimation duration	74
8.2	Preliminaries and mathematical facts	75
8.3	Tone-wise errors and cross-tone errors	76
8.4	Upper bounding cross-tone errors	78
8.4.1	Both pairs are far	79
8.4.2	Both pairs are close	81
8.4.3	One pair is far and one pair is close	85
8.5	Combining tone-wise errors and cross-tone errors	86
8.6	Geometric property	88
8.7	Main result	89
9	Conclusion, future directions, other related work	90
9.1	Future directions	90
9.2	Further related works	91
A	Building-block function $(G(t), \widehat{G}(f))$ in a single dimension	92
A.1	Construction of function $(G(t), \widehat{G}(f))$	92
A.2	Properties of function $(G(t), \widehat{G}(f))$	93
A.3	Proof of properties	93
A.4	Construction and properties of standard window function $(G'(t), \widehat{G}'(f))$	100
A.5	Facts	100
B	Filter, permutation and hashing in a single dimension	103
B.1	Construction of filter $(\mathbf{G}(t), \widehat{\mathbf{G}}(f))$	103
B.2	Properties of filter $(\mathbf{G}(t), \widehat{\mathbf{G}}(f))$	103
B.3	Proof of properties	104
B.4	Construction and properties of standard window $(\mathbf{G}'(t), \widehat{\mathbf{G}}'(f))$	108
	References	110

1 Introduction

Since people began to design and study optical systems, the resolution has become the core parameter of an optical system. Roughly speaking, resolution of an imaging system is defined as its ability to distinguish two points as separate in space and resolve detail in the object being imaged. Because of the physics of diffraction, there are some fundamental limits on the resolution of an imaging system. Surprisingly, people find fantastic Super-resolution techniques, and the diffraction limit of systems is transcended. As an outstanding representative in this field, the Nobel Prize in Chemistry 2014 was awarded jointly to Eric Betzig, Stefan W. Hell, and William E. Moerner "for the development of super-resolved fluorescence microscopy."

We formalize the Super-resolution problem considered in the work here. Let $x^*(t) = \sum_{i \in [k]} v_i \cdot e^{2\pi i f_i^\top t}$ be a signal with k -point sources in d -dimensional space where $f_1, \dots, f_k \in \mathbb{R}^d$. Assume we are able to observe a complex-valued signal function $x(t) = x^*(t) + g(t) \in \mathbb{C}$ over a finite duration $t \in [0, T]^d$, where $g(t)$ captures the noise in the measurement and we do not have any assumption on $g(t)$. To access the signal $x(t) = x^*(t) + g(t)$, an algorithm can only sample $x(\tau_j)$ at a number of m time points $\{\tau_j\}_{j \in [m]}$. These $\{\tau_j\}_{j \in [m]}$ can be **arbitrarily** chosen from the duration $t \in [0, T]^d$.¹ Like the standard objective in Super-resolution, we hope to design a fast algorithm that can estimate $\{(v_i, f_i)\}_{i \in [k]}$ with few samples.

Moreover, we are more ambitious and want our algorithm should output a k -Fourier-sparse recovered signal $x'(t)$ such that, for some approximation ratio $\mathcal{C} > 1$,

$$\text{signal estimation error} := \int_{\mathbf{B}} |x'(t) - x(t)|^2 \cdot dt \leq \mathcal{C}^2 \cdot \mathcal{N}^2, \quad (1)$$

where $\int_{\mathbf{B}} := \frac{1}{T^d} \int_{[0, T]^d}$, $\mathcal{N}^2 := \int_{\mathbf{B}} |g(t)|^2 dt + \delta \cdot \sum_{i \in [k]} |v_i|^2 > 0$ be the *noise level* and $\delta > 0$ is some parameter to conclude the noiseless case (i.e. $\int_{\mathbf{B}} |g(t)|^2 dt = 0$). For simplicity, we define $\|g\|_T^2 := \int_{\mathbf{B}} |g(t)|^2 dt$.

To make the problem interesting, we assume a bounded support $\text{supp}(\hat{x}^*) = \{f_i\}_{i \in [k]} \subseteq [-F, F]^d$ for the frequencies where the parameter $F > 0$ is known, and $\{f_i\}_{i \in [k]}$ have some minimum distance $\eta > 0$, i.e. $\min_{i \neq i'} \|f_i - f_{i'}\|_2 = \eta$. The band-limited assumption on signal and separation assumption on frequencies are standard both in Super-resolution [Moi15, HK15, CM21] and Continuous Fourier Transform [PS15, CKPS16, CP19a, CP19b, SSWZ22]. As mentioned before, there is no requirement for the noise $g(t)$, which is a strong noise model compared to many previous works.

In the noise-free case, there are a variety of methods to do Super-resolution [Pis73, HS90, Sto93] when duration $T = k/\eta$. Subsequently, there are some new methods [CF14, CFG13] to solve this problem based on assumption that either it is noise-free or f_j 's are restricted to be on a grid. Moreover, there are rich literature in high dimension such as [HK15, CFG14, KPRvdO16].

We use Table 1 to give a rough comparison between previous results and ours. Because previous works are under different noise models (we are the strongest, without any restriction on the noise), different settings (some of them can only measure the input signal on grid points, which makes the problem more difficult) or with different focus (e.g., [CM21] focus on the sharp constant c for the minimum duration c/η such that one can hope to get polynomial statistical and algorithmic complexity, while [HK15] and ours result lose some logarithmic term on the duration), this is just a high-level comparison. Some more detailed discussions about related work will be given later.

One natural idea to deal with high dimension problem is to map it to one dimension. If we do transformation and project d -dimensional signal to one-dimensional, we can directly apply the

¹This assumption is standard in the Continuous Fourier Transform literature, though not standard in the Super-resolution community.

Refs	# Samples	Running time
[CC13]	$\tilde{O}(k)$	$\tilde{O}(k^d)$
[HK15]	$\tilde{O}_d(k^2)$	$\tilde{O}_d(k^2)$
[CM21]	$\tilde{O}_d(k^2) \text{ poly}(F/\eta)$	$\tilde{O}_d(k^6) \text{ poly}(F/\eta)$
Ours	$\tilde{O}_d(k \log(F/\eta))$	$\tilde{O}_d(k \log(F/\eta))$

Table 1: Rough comparison between previous works and our result for constant dimensions d .

results in one dimension such as [PS15], which loses $\text{poly}(k)$ factor in duration, sample and running time complexity. Another way is to do semi-definite programming (SDP), which is usually based on results of Candes and Fernandez-Granda [CFG13, CF14] but the sample complexity and running time can still be very large.

Despite the huge success and developments, Super-resolution in multi-dimensional cases are still not well-understood, and improving efficiency on sampling complexity and computation complexity (running time) is an important and fundamental open problem. As described in [HK15],

It remains an open problem to reduce the sample complexity ... from $O(k^2)$ to the information theoretical bound $O(k)$, while retaining the polynomial scaling of the computation complexity.

To be even more ambitious, can we achieve nearly linear computation complexity rather than being polynomial with nearly linear sample complexity? This leads to the following fundamental algorithmic and statistical problem:

How efficient a Super-resolution algorithm can be on the running time and sample complexity?

Our work makes an important step towards solving this problem.

1.1 Our results

Roughly speaking, our algorithm RECOVERYSTAGE (Algorithm 10) achieves a constant approximation to the noise level $\mathcal{N} > 0$ in any constant dimension. For the tone estimation, we have the following guarantees.

Theorem 1.1 (Informal Tone estimation, see Theorem 7.18). *When RECOVERYSTAGE observes the signal $x(t)$ over a duration² $T \gtrsim \eta^{-1} \cdot \log(k/\delta)$, it outputs $k \geq 1$ recovered tones $\{(v'_i, f'_i)\}_{i \in [k]} \subseteq \mathbb{C} \times \mathbb{R}^d$ that approximate the true tones $\{(v_i, f_i)\}_{i \in [k]}$ up to an error proportional to the noise level $\mathcal{N} > 0$, with high probability. The algorithm RECOVERYSTAGE takes $k \cdot (\log k)^{d+O(1)} \cdot \log(F/\eta) \cdot 2^{O(d \log d)}$ samples and time.*

As for the signal estimation, we have the following guarantee, which to our knowledge is a new guarantee in the Super-resolution literature.

Theorem 1.2 (Informal Signal reconstruction, see Theorem 8.12). *When RECOVERYSTAGE observes the signal $x(t)$ over a duration $T \gtrsim \eta^{-1} \cdot k^{1-1/d} \cdot \log(k/\delta)$, the signal estimation error of the k -Fourier-sparse recovered signal $x'(t) := \sum_{i \in [k]} x'_i(t)$ against the observed signal $x(t) = x^*(t) + g(t)$ is bounded as follows:*

$$\int_{\mathbf{B}} |x'(t) - x(t)|^2 \cdot dt \lesssim \mathcal{N}^2. \quad (2)$$

²We often denote $f \gtrsim g$ when $f \geq C_0 \cdot g$ for some universal constant $C_0 > 0$, and the notation $f \lesssim g$ has a similar meaning. Also, we denote $f \approx g$ when both equations $f \gtrsim g$ and $f \lesssim g$ hold. In this page, these notations hide the dependence on dimension d .

Remark 1.3. In one dimension $d = 1$, our algorithm works when the duration $T \gtrsim \eta^{-1} \cdot \log(k/\delta)$, but the state-of-art result [PS15] requires $T \gtrsim \eta^{-1} \cdot \log^2(k/\delta)$ for the signal estimation. (For more details about this improvement, see Section 8.1.) Indeed, the duration is an equally important optimization goal as the sample complexity and the running time.

Preliminary Discussion: For any constant dimensions, we succeed to get an algorithm with both nearly optimal sample complexity and run-time, which is the goal in most of the literature on sparse Fourier transforms. However, due to the exponential dependence on the dimension in our result, this is not the end of story.

Up to the iterated logarithmic factors, our algorithm RECOVERYSTAGE takes $k \cdot (\log k)^{d+O(1)} \cdot \log(F/\eta) \cdot 2^{O(d \log d)}$ samples/running time. Merely extending the filter functions into high dimensions requires some very non-trivial efforts, but it already leads to an exponential loss in the dimension. This is a consequence of our “*precise*” filter function, seems to be unavoidable using current filtering techniques since even if the one-dimensional filter’s support size is off by a constant factor, it would lead to an exponential loss in the dimension anyways. As quoted:

[Kap16, Kap17] “in the discrete settings ... the price to pay for the precision of the filter, however, is that each hashing becomes a $\log^d k$ factor more costly in terms of sample complexity and running time than in the idealized case ...”

To shave the $\log^d k$ term in the discrete model, the past works [IK14, Kap16] randomize the noise by using the “crude” filters. However, randomizing the noise does not work in the continuous model, since two noise frequencies f, f' can be arbitrarily close and, no matter how we randomized the noise, the errors can accumulate in the estimation. The exponential dependence on dimension seems to be intrinsic to the current sampling methods, and avoiding it could need completely new methods.

1.2 Related works

1.2.1 Super-resolution with a different focus

The previous results [Moi15, CM21] are focused on finding the minimum possible separations between source points for fixed cutoff frequency (denoted by duration in this paper), such that there exists an algorithm with polynomial running time by using a polynomial number of samples. As a result, their algorithms are not efficient in running time and sample complexity.

In the following, we compare our work with [CM21] in more detail. Chen and Moitra [CM21] investigate a two-dimensional Super-resolution problem which they reduce to the problem of continuous Sparse FT. The main difference between their model and our model, is the way how the noise hampers the frequency recovery. Recall that we consider a signal $x(t) = x^*(t) + g(t) \in \mathbb{C}$ over a duration $t \in [0, T]^d$, where $x^*(t) \in \mathbb{C}$ is the actual signal that we aim to recover, and the noise $g(t) \in \mathbb{C}$ has a small enough *constant-proportional* energy compared to $x^*(t)$, that is, $\|g\|_T \leq 10^{-3} \cdot \|x^*\|_T$.³ In particular, the noise magnitude $|g(t_0)|$ at a certain time point $t_0 \in [0, T]^d$ has no requirement, and can even be much larger than the signal magnitude $|x^*(t_0)|$. In contrast, [CM21] make a stronger assumption on the noise $g(t) \in \mathbb{C}$. At *any* time point $t_0 \in [0, T]^d$, they need the noise magnitude $|g(t_0)|$ is always *inverse-polynomially* small, compared to the corresponding average signal energy $\|x^*\|_T$.

For their model, Chen and Moitra focus on the two-dimensional case, and their primary emphasis is on refining the duration requirement in the two-dimensional case, i.e., on the exact constant in front of η^{-1} for constant $d = 2$. For general constant d , Chen and Moitra can (via tensor

³Recall that the average energy, e.g., of the noise $g(t)$ over duration $t \in [0, T]^d$, is defined as $\|g\|_T = \frac{1}{T^d} \cdot \int_0^T |g(t)|^2 dt$.

decomposition) get sample complexity $\tilde{O}_d(k^2) \text{poly}(F/\eta)$ and running time $\tilde{O}_d(k^6) \text{poly}(F/\eta)$,⁴ while our running time and sample complexity are $\tilde{O}_d(k \log(F/\eta) \log^{O(1)}(k))$. As a trade off, their duration is $T \gtrsim 1/\eta$ while ours is $T \gtrsim \log(k)/\eta$.

Note that in sparse Fourier transform/sparse recovery literature, the major goal is to get nearly linear in k sample complexity, and $k^{1+\Omega(1)}$ is not allowed (see Table 1 in [NS19] and Table 1 in [NSW19]). It is well-known that in many cases, k^2 or even $k^{1+\Omega(1)}$ samples can make the problem subsequently easier. Also, it is worth mentioning that, [CM21] considers the “tone recovery” problem only, without studying the “signal recovery” problem, whereas our paper investigates the both problems.

1.2.2 Prior works on the sparse FT problem

As our technology originates from Fourier Transform, in this and next sub-subsection, we briefly review several previous works for classic prior works on the discrete FT (DFT) and continuous FT (CFT) separately. For a more detailed overview, the reader can refer to Section 2.3.

The discrete model. In any dimension $d \geq 1$, the Fourier transform $\hat{x} \in \mathbb{C}^N$ is a vector of length $N = n^d$. 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

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

Following the framework of [GMS05, HIKP12a, IKP14, IK14, Kap16], the idea is to take, multiple times, a set of $\mathcal{B} = B^d = \Theta_d(k)$ ⁵ linear measurements of the form $u_j := \sum_{i: h(i)=j} x_i \cdot s(i)$, where $h : [N] \mapsto [\mathcal{B}]$ are random hash functions and $s : [N] \mapsto \{\pm 1\}$ are random sign functions. This means “hashing into \mathcal{B} bins”. If the linear measurements are ideal, then $O(\log(N/k))$ hashes are enough for sparse recovery and the sample complexity is $O(k \log(N/k))$.

Based on the linear combinations of the samples x_i , the sparse DFT algorithms will approximate the u_j ’s. That is, we first permute the samples x_i via a pseudorandom affine permutation \mathcal{P} . Then, the permuted samples $(\mathcal{P}x)_i$ are respectively scaled by coefficients $\mathcal{G}(l_i)$, i.e., the values of a filter function $\mathcal{G} : \mathbb{R}^d \mapsto \mathbb{R}$ at a bunch of lattice points $l_i \in \mathbb{R}^d$. Hence, we use a modified combination

$$u_j = \sum_{i: h(i)=j} (\mathcal{P}x)_i \cdot \mathcal{G}(l_i). \quad (3)$$

Different from the *binary-valued* sign functions, the filter functions \mathcal{G} shall be “imperfect” to reduce the sample complexity. Namely, every coordinate $i \in [n]$ not only contributes $\approx 100\%$ fraction to a target bin, but also “leak” a small fraction to each other bin. (And to balance the trade-off between the sample complexity and the running time, the past works like [HIKP12a, IK14, Kap16, Kap17] use different leakage levels.)

The above approach “isolates” most of the head frequencies $\{f_i\}_{i \in [k]}$ (i.e., the top- k coordinates of \hat{x}). In precise, most $\{f_i\}_{i \in [k]}$ are hashed to unique bins, and the “tail” frequencies $[n]^d \setminus \{f_i\}_{i \in [k]}$ contribute very little to those bins. So the algorithm can exactly identify the head frequencies and approximately evaluate the magnitudes $\hat{x}(f_i)$, producing a k -sparse estimation $\hat{x}' \approx \hat{x}$.

Also, notice that the DFT preserves the ℓ_2 -norm of a Fourier spectrum, namely $\|\hat{z}\|_2 = \|z\|_2$ for any $z \in [n]^d$, so the ℓ_2/ℓ_2 -guarantees in the frequency/time domains are equivalent.

The continuous model. The one-dimensional sparse CFT problem is introduced by [PS15], and our formulation is a natural multi-dimensional extension. Different from the discrete model, we

⁴The notation $\tilde{O}_d(f)$ assumes a constant dimension $d \geq 1$ and hides the term $\text{poly}(\log f)$; similar for $\tilde{\Omega}_d$ and $\tilde{\Theta}_d$.

⁵Here the notation $\Theta_d(f)$ assumes a constant dimension $d \geq 1$; similar for O_d and Ω_d .

cannot recover the exact head frequencies $\{f_i\}_{i \in [k]}$ in the continuous model. The current frequencies are off-the-grid, so (i) any two frequencies $f \neq f'$ can be too close to distinguish [Moi15]; and (ii) even if a head frequency f_i is well separated from the others, we can only recover it up to some precision that depends on the duration $T > 0$.

As the frequency recovery is not exact, we cannot hope for the best k -sparse Fourier spectrum. For this reason, [PS15] considers the tone/signal estimations under the ℓ_2/ℓ_2 -guarantee in the time domain. In addition to the approximation guarantee, sample complexity and running time, we have one more optimization goal – minimizing the duration $t \in [0, T]^d$ for the sampling.

1.3 Our techniques

Similar to the previous works, our main task is to recover the head frequencies $\{f_i\}_{i \in [k]}$. As if we promise a good approximation $\{f'_i\}_{i \in [k]} \approx \{f_i\}_{i \in [k]}$, then the magnitudes $\{v_i\}_{i \in [k]}$ can be easily recovered.

To deal with the continuous model, the overall ideas in [PS15] are to translate the hash functions, filter functions and estimation algorithms from the DFT setting to the CFT setting, and we adopt the similar framework. However, extension one/two-dimensional ([PS15]/[CM21]) cases to the multi-dimensional continuous case presents a number of challenges, which are addressed in this paper by some interesting techniques. Among these, there are three most remarkable ones.

- Our hashing scheme is specifically designed for the multi-dimensional continuous model, and the “eggshell” sampling scheme (for time points) which differs from all the previous ones.
- To learn the frequencies f_i ’s more accurately (while ensuring a *logarithmic* algorithm in F/η which also means beating $\text{poly}(F/\eta)$ sample/time in [CM21]), we apply (i) a coarse-grained location procedure, for which we employ technical ingredients from high-dimensional geometry; and then (ii) a fine-grained location procedure, which is built upon a robust linear-system solver.
- The duration bound required by a recovery algorithm is an equally important optimization goal as the sample complexity and the running time in the super-resolution. To improve the duration bound against the previous algorithm [PS15], we provide a better analysis by leveraging Parseval’s theorem and the convolution theorem in a different manner.

1.3.1 Hashing and sampling

The obstacles. As mentioned, we assume the head frequencies (defined by $\text{HEAD} := \{f_i\}_{i \in [k]}$) locate within the hypercube $[-F, F]^d$ and are separated by $\eta = \min_{i \neq i' \in [k]} \|f_i - f_{i'}\|_2 > 0$, corresponding to the k -Fourier-sparse signal $x^*(t)$. The other tail frequencies TAIL correspond to the noise $g(t)$.

To recover the head frequencies, a direct attempt is to handle all dimensions $r \in [d]$ separately, through the one-dimensional hashing scheme in [PS15]. Unfortunately, this approach fails to work. For example, suppose two frequencies are equal in the first dimension, i.e., $f_{i,1} = f_{i',1}$ for $i \neq i' \in [k]$ (but the overall ℓ_2 -distance in the other dimensions is $\geq \eta$). Then regarding the first dimension, no hashing scheme can distinguish these two scenarios: (i) the desired tones (f_i, v_i) and $(f_{i'}, v_{i'})$; and (ii) a single tone (f, v) given that $f_1 = f_{i,1} = f_{i',1}$ and $v = v_i + v_{i'}$. Thus, an algorithm can *miscount* the tones, and recover the top- $(k+1)$ or even more magnitudes. Also, when the miscount happens (in one or more dimensions), an algorithm cannot match the dimension-wise frequencies correctly. For these reasons, the multi-dimensional model requires a “not-very-naive” hashing scheme.

Our approach. Similar to Eq. (3), we will leverage the measurements $u_j = \sum_{i:h(i)=j} \mathcal{P}x(\tau_i) \cdot \mathcal{G}(l_i)$, where $\tau_i \in [0, T]^d$ are the sampling time points. To define permutation, we introduce three notations : $\Sigma \in \mathbb{R}^{d \times d}$ is scaling frequency domain, and $b \in \mathbb{R}^d$ is shifting frequency domain and $a \in \mathbb{R}^d$ is shifting time domain. We explain how to select them later. Now, let us present the formal permutation:

$$\widehat{\mathcal{P}x}(\text{frac}(\Sigma f - b)) = \widehat{x}(f) \cdot \det(\Sigma)^{-1} \cdot e^{-2\pi i \cdot f^\top a}, \quad (4)$$

where the function $\text{frac} : \mathbb{R}^d \mapsto [0, 1]^d$ computes the coordinate-wise *fractional part* of the input.

There are two requirements for the random matrix Σ : (i) it must be invertible; and (ii) makes any two different head frequencies $\xi_i \neq \xi_{i'} \in \text{HEAD}$ hashed into the same bin with probability at most $0.01 \cdot k^{-1}$ (i.e., the collision probability). To these ends, we construct the Σ in three steps.

- Step I. We first sample an interim matrix $\Sigma' \sim \text{Unif}(\mathbf{SO}(d))$ uniformly at random from the d -dimensional *rotation group*, leading to a rotation matrix Σ' with determinant $|\det(\Sigma')| = 1$. Clearly, such an interim matrix $\Sigma' \in \mathbb{R}^{d \times d}$ is invertible.
- Step II. Let us explain what the bins stand for in the continuous model. Given the transformation $\text{frac}(\Sigma f - b)$ in Eq. (4), we are interested in the codomain $[0, 1]^d$. We partition this unit hypercube into $\mathcal{B} = B^d = \Theta_d(k)$ isomorphic sub-hypercubes, with the volume $1/\mathcal{B}$ each. These sub-hypercubes are exactly the bins in the continuous setting.
- Step III. We sample a random scaling factor $\beta \sim \text{Unif}[\widehat{\beta}, 2\widehat{\beta}]$, where the parameter $\widehat{\beta} > 0$ is sufficiently large, and derive the ultimate random matrix by letting $\Sigma := \beta \Sigma'$. Clearly, $\Sigma \in \mathbb{R}^{d \times d}$ is invertible. Below We will explain why this Σ gives a small collision probability.

According to Eq. (4), whether two head frequencies $f_i \neq f_{i'} \in \text{HEAD}$ collides or not relies on the difference vector $\Sigma(f_i - f_{i'}) \in \mathbb{R}^d$. Since Σ is a random *rotation matrix* scaled by $\beta \sim \text{Unif}[\widehat{\beta}, 2\widehat{\beta}]$, this difference vector is distributed *almost uniformly* within the ℓ_2 -norm “eggshell”

$$\{z \in \mathbb{R}^d : \widehat{\beta} \cdot \|f_i - f_{i'}\|_2 \leq \|z\|_2 \leq 2\widehat{\beta} \cdot \|f_i - f_{i'}\|_2\}.$$

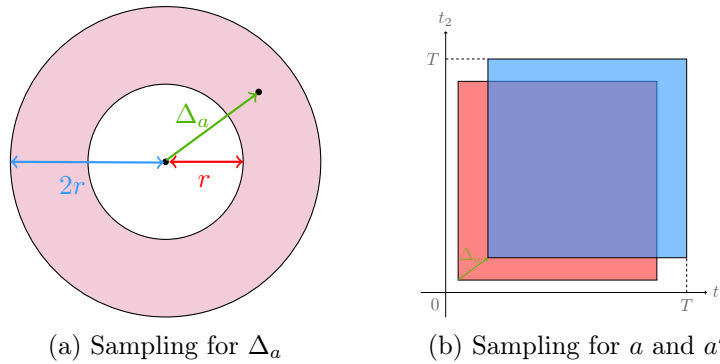


Figure 1: Demonstration of the sampling scheme. The time difference $\Delta_a = (a' - a)$ is sampled from the pink region in Figure (a), which looks like an “eggshell”. The time points a, a' are sampled respectively from the red region and the blue region in Figure (b).

The concerning frequencies $f_i \neq f_{i'}$ have an ℓ_2 -distance $\|f_i - f_{i'}\|_2 \geq \eta$ and thus, the above “eggshell” is thick enough. That is, the random difference vector $\Sigma(f_i - f_{i'})$ is distributed on a large enough support. After rounding, the $\text{frac}(\Sigma(f_i - f_{i'}))$ is distributed *almost uniformly* within the unit

hypercube $[0, 1]^d$, and the collision probability roughly equals the volume $1/\mathcal{B} = \Theta_d(k^{-1})$ of a single bin. The parameter $\hat{\beta}$ is set carefully, to ensure a small collision probability $\leq 0.01 \cdot k^{-1}$. Hence, the matrix Σ is likely to isolate at least 90% head frequencies.

The vector b serves as the “anchor point” of the hashing scheme \mathcal{h} . Independent of Σ , we just sample a uniform $b \sim \text{Unif}[0, 1]^d$ from the unit hypercube. Then due to Eq. (4), a certain frequency $f \in [-F, F]^d$ is equally likely to be hashed into one of the $\mathcal{B} = B^d = \Theta_d(k)$ bins.

The “eggshell” sampling scheme. As Figure 1 shows, the vector $a \in \mathbb{R}^d$ is sampled *non-uniformly*, which differs from all the previous sampling schemes [HIKP12a, IK14, PS15, Kap16, Kap17, CKPS16, NSW19]. Recall that this vector a rotates any magnitude $\hat{x}(f) \in \mathbb{C}$ by a certain angle $-2\pi \cdot f^\top a \in \mathbb{R}$ (see Eq. (4)). Let $S = \{f \in \text{TAIL} : \mathcal{h}(f) = j\}$ be the tail frequencies hashed into a certain bin $j \in [B]^d$, then we hope a small total rotated magnitude

$$|\int_{f \in S} \hat{x}(f) \cdot e^{-2\pi i \cdot f^\top a} \cdot df| \ll (\int_{f \in S} |\hat{x}(f)|^2 \cdot df)^{1/2}.$$

In the continuous model, the vector $a \in \mathbb{R}^d$ represents a sampling time point $t \in [0, T]^d$. We must sample this time point *almost* (but not exactly) uniformly from a *constant proportion* of the duration, such as $a \sim \text{Unif}[\frac{0.01}{d} \cdot T, (1 - \frac{0.01}{d}) \cdot T]^d$. This is due to the following two reasons.

- Recall that the noise level \mathcal{N}^2 involves the term $\|g\|_T^2 = \frac{1}{T^d} \cdot \int_{t \in [0, T]^d} |g(t)|^2 \cdot dt$, but we have no guarantee on the noise $g(t)$ at a specific time point $t \in [0, T]^d$. If the sampling range $A \ni a$ is too small (namely $|A| \ll T^d$), the average noise $\frac{1}{|A|} \cdot \int_{t \in A} |g(t)|^2 \cdot dt \gg \|g\|_T^2$ can be intolerably large, and makes the samples $a \in A$ useless.
- Unlike the discrete case, where the *on-the-grid* frequencies are perfectly separated, two “continuous” frequencies $f \neq f' \in [-F, F]^d$ can be arbitrarily close (when not both of f, f' are head frequencies). If $\|f - f'\|_2 \ll 1/(\sqrt{d} \cdot T)$ and $\hat{x}(f) = \hat{x}(f')$, then over the whole duration $t \in [0, T]^d$ (i.e., $\|t\|_2 \leq \sqrt{d} \cdot T$) the two signals are always close $\hat{x}(f) \cdot e^{-2\pi i \cdot f^\top t} \approx \hat{x}(f') \cdot e^{-2\pi i \cdot f'^\top t}$. To distinguish the frequencies $f \neq f'$, sampling the a nearly from the whole duration achieves the best we can.

We often sample a pair of $a, a' \in [0, T]^d$ and consider their difference $\Delta_a := (a' - a)$ rather than a, a' themselves. Over the difference vector $\Delta_a \in \mathbb{R}^d$, a signal with frequency $f \in \mathbb{R}^d$ rotates by an angle $2\pi \cdot f^\top \Delta_a \in \mathbb{R}$. Denote by $\|\theta\|_\circ := \min_{z \in \mathbb{Z}} |\theta + 2\pi z|$ the “circular distance”. Our actual observation would be the circular distance $\|2\pi \cdot f^\top \Delta_a\|_\circ \in [0, \pi]$.

To distinguish this frequency $f \in \mathbb{R}^d$ from the others, and to recover $f \in \mathbb{R}^d$ more accurately, we need a largest possible ℓ_2 -norm $\|\Delta_a\|_2$. Moreover, because we do not know the direction of the frequency $f \in \mathbb{R}^d$ (or the direction of the difference between f and the interim estimation of it), the sampled $\Delta_a \in \mathbb{R}^d$ must have a uniformly random direction.

The above two requirements for the time difference $\Delta_a = a' - a$ can violate our previous requirement that, both time points a, a' shall be sampled almost uniformly from a constant proportion of the duration $t \in [0, T]^d$. In particular, the dimensionality $d \geq 2$ incurs many technical issues. To overcome these challenges, we sample a, a' in a *coupling* fashion. We first determine the time difference Δ_a , making it have a uniform random direction. Moreover, the ℓ_2 -norm $\|\Delta_a\|_2$ cannot be too large; otherwise, we cannot ensure that the sampling ranges $A \ni a$ and $A' \ni a'$ are large enough, namely $|A| \approx |A'| \approx T^d$. Both the sampling range of the ℓ_2 -norm $\|\Delta_a\|_2$, and the sampling scheme for $a, a' \in [0, T]^d$ (given a specific Δ_a) are carefully chosen.

In contrast, suppose we sample two uniform random $a, a' \sim \text{Unif}[0, T]^d$, then the time difference Δ_a has a *non-uniform direction*. So the observed circular distance $\|2\pi \cdot f^\top \Delta_a\|_\circ$ will follow a more

complicated distribution, being hard to analyze. More importantly, both the true observations $\|2\pi \cdot f^\top \Delta_a\|_\circ$ and the “fake” observations $\|2\pi \cdot f'^\top \Delta_a\|_\circ$ (due to other frequencies $f' \neq f$) may concentrate in a small range like $[0, \frac{\pi}{100}]$. Then, we can’t distinguish $f \neq f'$. This issue does not exist in the one-dimensional continuous case or the discrete case:

- In the one-dimensional continuous case, Δ_a is just a random number instead of a vector. We need not concern the *direction* of Δ_a , let alone whether this direction is uniform random.
- In the multi-dimensional discrete case, the frequencies are on-the-grid. Thereby, the observed circular distance just has finite possibilities, e.g., $\{0, \frac{1}{N} \cdot \pi, \dots, \frac{N-1}{N} \cdot \pi, \pi\}$. It turns out that we can easily distinguish true observations from fake observations.

For more details about the sampling scheme, the reader can refer to Section 5.6.

1.3.2 Sparse recovery

The obstacles. Using the hash functions and the filters, several kinds of recovery algorithms have been developed in the literature. Again, the main task is to recover the head frequencies $\{f_i\}_{i \in [k]}$, and the continuous model is harder since the estimations $f'_i \approx f_i$ are limited to some precision.

Similar to the past work [PS15], we use a *voting-based* algorithm. Roughly speaking, [PS15] handles the one-dimensional case as follows: twist the frequency domain $[-F, F]$, partition it into $\Theta(k)$ sub-regions, and vote for the *probably approximately correct* sub-region(s). Although simple in spirit, generalizing this idea to a higher dimension $d \geq 2$ incurs many new challenges.⁶ For example, the twist of a hypercube $[-F, F]^d$ is complex (but the twist of $[-F, F]$ is just an interval), so a more sophisticated partition scheme is required. Moreover, since we consider the ℓ_2 -distances among f_i ’s but the domain $[-F, F]^d$ is a ℓ_∞ -ball, switching between the ℓ_2 -/ ℓ_∞ -norms raises more technical difficulties. (However, this switch follows automatically in one dimension $d = 1$.)

En route to the final algorithm, we will address some of these challenges.

Our approach. For ease of presentation, we will restrict our attention to a tone $(v_i, f_i) \in \mathbb{C} \times \mathbb{R}^d$ that is isolated by the permutation \mathcal{P} and hashing \mathcal{h} . According to Eq. (4), a sampling time point $a \in [0, T]^d$ gives a measurement $y_i(a) \in \mathbb{C}$ such that $y_i(a) \approx v_i \cdot \det(\Sigma)^{-1} \cdot e^{-2\pi i \cdot f_i^\top a}$. Here, the “ \approx ” notation hides a small error, which stems from the noise frequencies (i.e., $g(t) \in \mathbb{C}$) hashed into the same bin $j := \mathcal{h}(f_i) \in [B]^d$. To recover the frequency f_i , the idea is to leverage the difference $\Delta_a := (a' - a)$ between two time points $a, a' \in [0, T]^d$ and the relative phase

$$\psi_i(a, a') := \arg(y_i(a)/y_i(a')) \approx \arg(e^{2\pi i \cdot f_i^\top \Delta_a}) = 2\pi \cdot f_i^\top \Delta_a. \quad (5)$$

The above “ \approx ” notation hides an error phase of, say, $\pm(2\pi)/10^3$.

We recover the frequency $f'_i \approx f_i$ in two steps. First, the “coarse-grained” location (Algorithm 2) keeps track of a hypothesis region $\mathcal{H}_i \ni f_i$ for the frequency (e.g., at the beginning $\mathcal{H}_i = [-F, F]^d$) and shrinks \mathcal{H}_i round by round, and get the rough location of the frequencies in the end. Second, after receiving the “coarse-grained” location \mathcal{H}_i , the “fine-grained” locating (Algorithm 4) carefully derives d linear equations of the form $2\pi \cdot f_i'^\top \Delta_a^r = \psi_i^r$ (for all $r \in [d]$) based on d time differences $\Delta_a^r \in \mathbb{R}^d$, and solves these linear equations to find $f'_i \approx f_i$ within the hypothesis region \mathcal{H}_i .

Coarse-grained location via partition and voting in high dimension. Suppose that a frequency f_i locates in some hypothesis region \mathcal{H}_i . We carefully divide $\mathcal{H}_i = \bigcup_{q \in Q} \mathcal{H}_{i,q}$ into smaller

⁶Some of these challenges do not exist (or are less severe) in the discrete model [HIKP12a, IK14], because the twist of the discrete frequency domain $[n]^d$, under an appropriate modulo operation, is still itself.

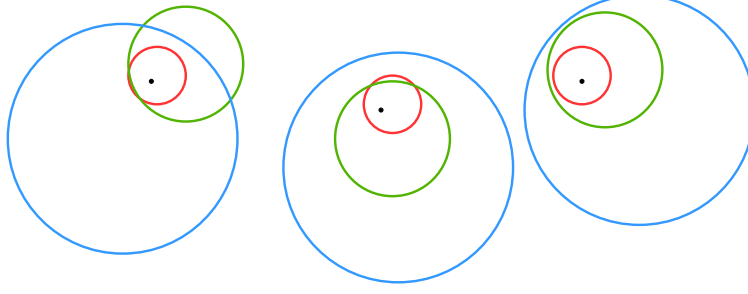


Figure 2: Demonstration for the **coarse-grained location** in two dimensions $d = 2$. The black points refer to the true frequencies. The blue/green/red circles show that we gradually shrink the hypothesis regions for the frequencies.

sub-regions and pick a candidate frequency ξ_q for each sub-region. The frequency f_i locates in a unique *true* sub-region \mathcal{H}_{i,q^*} . Based on the measurements, we can prune some of the *wrong* sub-regions $\mathcal{H}_{i,q} \not\ni f_i$ and get a *smaller* new hypothesis region. As Figure 2 shows, the coarse-grained location repeats this pruning process.

Given a pair of sampling time points $a, a' \in [0, T]^d$, in view of Eq. (5), we will vote for every candidates frequency ξ_q that satisfies

$$\|2\pi \cdot \xi_q^\top \Delta_a - \psi_i(a, a')\|_\circ \leq (2\pi)/50, \quad (6)$$

where the RHS can be other suitable thresholds. By doing so, (i) the *true* candidate frequency ξ_{q^*} (for which $\mathcal{H}_{i,q^*} \ni f_i$) gets a vote with probability 90%, since ξ_{q^*} is close enough to f_i . In contrast, (ii) if a *wrong* candidate frequency ξ_q (for which $\mathcal{H}_{i,q} \not\ni f_i$) is too far from f_i , then we hope ξ_q to get a vote with probability $< 50\%$. Given Eq. (5) and (6), the wrong candidate frequency ξ_q loses a vote when $\|2\pi \cdot (\xi_q - f_i)^\top \Delta_a\|_\circ \geq (2\pi)/40$. Namely, with probability $> 50\%$, we hope the gap between $(\xi_q - f_i)^\top \Delta_a \in \mathbb{R}$ and its closest integer to be at least

$$\min_{z \in \mathbb{Z}} |(\xi_q - f_i)^\top \Delta_a - z| \geq 1/40, \quad (7)$$

To this end, the time difference $\Delta_a \in \mathbb{R}^d$ is sampled to have a uniform random direction and a random ℓ_2 -norm $\|\Delta_a\|_2 \sim \text{Unif}[w, 2w]$, for some $w > 0$. In any dimension $d \geq 2$, we have

$$(\xi_q - f_i)^\top \Delta_a = \|\xi_q - f_i\|_2 \cdot \|\Delta_a\|_2 \cdot \cos(\gamma), \quad (8)$$

where the random angle $\gamma := \langle \xi_q - f_i, \Delta_a \rangle$. Clearly, when a fixed $|\cos(\gamma)| \in [0, 1]$ (namely a fixed direction of Δ_a) is not too small, a large enough sampling range for the ℓ_2 -norm $\|\Delta_a\|_2 \sim \text{Unif}[w, 2w]$ ensures Eq. (7) with probability $> 50\%$. This is exactly what we desire.

Nonetheless, the coarse-grained location recovers the frequencies by at most $\|\xi_{q^*} - f_i\|_2 \lesssim d/T$ (instead of $\|\xi_{q^*} - f_i\|_2 \lesssim 1/T$). When the difference $\Delta_a = (a' - a)$ has a uniform random direction, the angle $\gamma \in [0, \pi]$ concentrates within the range $\pi/2 \pm \pi/(2\sqrt{d})$, so with high probability we have $|\cos(\gamma)| \lesssim 1/\sqrt{d}$. Given Eq. (7) and (8), in order to vote for a wrong candidate frequency ξ_q with probability $< 50\%$, we require $\|\xi_q - f_i\|_2 \cdot \|\Delta_a\|_2 \gtrsim \sqrt{d}$.

Given a specific $\Delta_a \in \mathbb{R}^d$, the largest possible range from which we sample the two time points $A \ni a, a'$, has the volume $|A| = T^d \cdot (1 - \|\Delta_a\|_1/T)$. As mentioned (Section 1.3.1), this range A must be a constant proportion of the whole duration $t \in [0, T]^d$, which requires $\|\Delta_a\|_1 \lesssim T$. However, when $\Delta_a \in \mathbb{R}^d$ has a uniform random direction, with high probability we have $\|\Delta_a\|_1 \approx \sqrt{d} \cdot \|\Delta_a\|_2$. Thus, it is required that $\|\Delta_a\|_2 \lesssim T/\sqrt{d}$.

Putting the above arguments together gives $\|\xi_q - f_i\|_2 \gtrsim \sqrt{d}/\|\Delta_a\|_2 \gtrsim d/T$. Namely, we can not recover the frequency $f_i \in [-F, F]^d$ too well by the coarse-grained location, but it can provide some rough estimations.

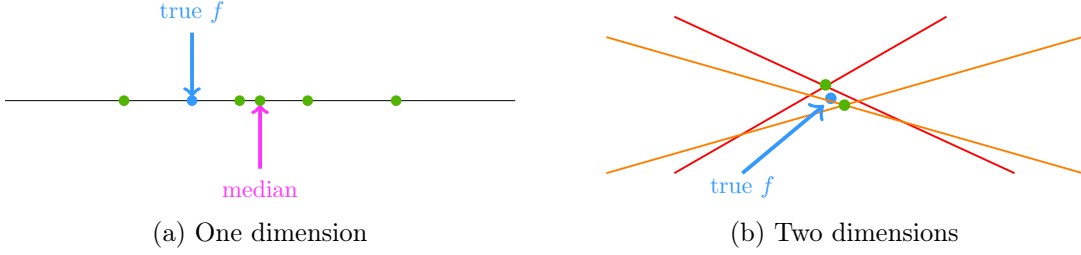


Figure 3: Demonstration of the **fine-grained location** in one or two dimension(s). When $d = 2$, each observation $\varphi^r \in \mathbb{R}$ gives a hypothesis line (i.e., a hypothesis one-dimensional hyper-plane) that is close to the true frequency f , so a pair of observations/lines determines an estimation $f' \approx f$. Given a bunch of estimations f' , we will find a small ℓ_∞ ball to cover a sufficient amount of estimations. Since this ℓ_∞ ball is small yet (with high probability) contains the true frequency f , its center must be a good enough approximation to f . (In one dimension, we just take the median of the estimations f' .)

Fine-grained location via inverting robust linear system. The coarse-grained location recovers the frequencies up to an ℓ_2 -distance $\lesssim_d 1/T$.⁷ Then the fine-grained location improves this precision to $\lesssim_d 1/(\rho T)$, where $\rho \gg 1$ is the *signal-to-noise ratio* (Definition 7.1). In one dimension $d = 1$, the past work [PS15] easily achieves so by first deriving a bunch of candidates $\xi_{q^*} \approx f_i$ from getting a few of the coarse-grained locations, and then taking the median of these ξ_{q^*} 's (as Figure 3(a) suggests). However, this idea fails in the multi-dimensional case when $d \geq 2$, and the fine-grained location becomes far more complicated.

Roughly speaking, based on a time difference $\Delta_a^r = (a' - a) \in \mathbb{R}^d$, we get an observation $\psi^r \in \mathbb{R}$ such that $\mathbf{E}[|\psi^r - f_i^\top \Delta_a^r|^2] = 0.1/\rho^2$. Due to Markov's inequality,

$$\Pr[|\psi^r - f_i^\top \Delta_a^r| \leq \sqrt{d}/\rho] \geq 1 - 0.1/d. \quad (9)$$

When $d = 1$, we can just take ψ^r/Δ_a^r as an approximation of f_i . It suffices to get a good estimation $f'_i \approx f_i$ via a small number of samples. However, when $d \geq 2$, we cannot extract enough information from the inner product $f_i^\top \Delta_a^r \in \mathbb{R}$ of the two vectors. To handle this issue, as Figure 3(b) illustrates, we will use d random vectors to form a random matrix $\Delta = [\Delta_a^r]_{r \in [d]} \in \mathbb{R}^{d \times d}$ that has a bounded spectral norm,⁸ and d observations $\psi = (\psi_r)_{r \in [d]}$. Then Eq. (9) implies that

$$\Pr[\|\psi - \Delta f_i\|_2 \leq \sqrt{d} \cdot \sqrt{d}/\rho] \geq 1 - 0.1 = 0.9.$$

This gives a good estimation $\Delta^{-1}\psi \approx f_i$ with $\|\Delta^{-1}\psi - f_i\|_2 \leq \|\Delta^{-1}\| \cdot \|\psi - \Delta f_i\|_2 \leq \|\Delta^{-1}\| \cdot (d/\rho)$. (For a illustration, see Figure 14 in Section 5.7.) This approach needs d observations, and the estimation error must be amplified by a \sqrt{d} factor to enable the union bound.

To get a more accurate estimation, our new sampling method discussed before ensures that Δ_a^r $\|\Delta_a^r\|_2 \gtrsim T/d$. One additional issue is how to analyze the random matrix $\Delta = [\Delta_a^r]_{r \in [d]} \in \mathbb{R}^{d \times d}$. Fortunately, one can show the vectors $\Delta_a^r \cdot \sqrt{d}/\|\Delta_a^r\|_2$ are sub-Gaussian isotropic, so we can upper bound the spectral norm $\|\Delta^{-1}\|$. Combining everything and solving the robust linear systems gives $\|f'_i - f_i\|_2 = \|\Delta^{-1}\psi - f_i\|_2 \lesssim_d 1/(\rho T)$.

⁷We say $a \lesssim_d b$ if $a \leq \text{poly}(d) \cdot b$.

⁸For a matrix Δ , we use $\|\Delta\|$ to denote the spectral norm of Δ .

Roadmap

Section 2 provides some basic notations and definitions. Section 3 provides a list of probability tools. Filter, permutation and hashing in one dimension are given in Appendix A and B for completeness, which can be skipped if readers are familiar with them. Section 4 presents the counterpart filter, permutation and hashing in the multi-dimensional setting. In Section 5 and 6, we show how to give accurate estimations of the frequencies. In Section 7, we present our sparse recovery algorithm. In Section 8, we show how to obtain the signal estimation by paying a slightly longer duration. Finally, in Section 9, we give a short discussion on some bottlenecks of current methods, and some interesting future directions.

2 Preliminaries

2.1 Notations

We denote by $[n]$ the set $\{0, 1, 2, \dots, n-1\}$, by \mathbb{R} the set of real numbers, by \mathbb{Z} the set of integers, and by \mathbb{C} the set of complex numbers. Also, $\mathbb{N}_{\geq a}$ refers to the set of integers no less than $a \geq 0$. Let $\text{supp}(f)$ denote the support of a function or vector f , and let $\|f\|_0 = |\text{supp}(f)|$ be the cardinality. For a random variable X , for convenience we may abuse the notation $\text{supp}(X)$ to denote the support of X 's probability density function (PDF).

We use $\max\{a, b\}$ or $\max(a, b)$ (resp. $\min\{a, b\}$ or $\min(a, b)$) to denote the maximum (resp. the minimum) between $a, b \in \mathbb{R}$. Given any $p \geq 1$, a vector $x = (x_i)_{i \in [n]} \in \mathbb{R}^n$ has the ℓ_p -norm $\|x\|_p := (\sum_{i \in [n]} |x_i|^p)^{1/p}$; in the case that $p = \infty$, we define $\|x\|_\infty := \max_{i \in [n]} |x_i|$.

We use the notations $\mathbf{i} := \sqrt{-1}$ and $e^{i\theta} := \cos(\theta) + \mathbf{i} \cdot \sin(\theta)$ for any phase $\arg(e^{i\theta}) = \theta \in \mathbb{R}$. For a complex number $z = a + \mathbf{i} \cdot b \in \mathbb{C}$, let $a \in \mathbb{R}$ be the real part and let $b \in \mathbb{R}$ be the imaginary part. Also, $\bar{z} := a - \mathbf{i}b \in \mathbb{C}$ denotes the conjugate, and $|z| := \sqrt{z\bar{z}} = \sqrt{a^2 + b^2} \geq 0$ denotes the norm.

2.2 Fourier transform and convolution

For convenience, throughout this paper we use the shorthand CFT (the continuous Fourier transform), DFT (the discrete Fourier transform), DTFT (the discrete-time Fourier transform) and FFT (the fast Fourier transform).

- In the time domain, we often use the notations t and τ .
- In the frequency domain, we often use the notations f and ξ .

Given a d -variate function $x(t)$ for $t = (t_s)_{s \in [d]} \in \mathbb{R}^d$, we have the CFT $\widehat{x}(f)$ for $f = (f_r)_{r \in [d]} \in \mathbb{R}^d$ and the inverse CFT $x(t)$ for $t \in \mathbb{R}^d$:

$$\widehat{x}(f) := \int_{\tau \in \mathbb{R}^d} x(\tau) \cdot e^{-2\pi \mathbf{i} \cdot f^\top \tau} \cdot d\tau \quad \text{and} \quad x(t) := \int_{\xi \in \mathbb{R}^d} \widehat{x}(\xi) \cdot e^{2\pi \mathbf{i} \cdot t^\top \xi} \cdot d\xi.$$

Definition 2.1 (k -Fourier-sparse signal). Given any k -Fourier-sparse signal $x^*(t)$ with the tones $\{(v_i, f_i)\}_{i \in [k]} \subseteq \mathbb{C} \times \mathbb{R}^d$, the corresponding CFT $\widehat{x^*}(f)$ is the combination of $k \geq 1$ many (scaled) d -dimensional Dirac delta functions, each of which has a point mass (i.e. the involved magnitude) $v_i \in \mathbb{C}$ at the corresponding frequency $f_i \in \text{supp}(\widehat{x^*})$. Without ambiguity, we denote $\widehat{x^*}[f_i] := v_i \in \mathbb{C}$ for convenience. Then the k -sparse Fourier spectrum $\widehat{x^*}(f)$ for $f \in \mathbb{R}^d$ can be formulated as

$$\widehat{x^*}(f) := \sum_{i \in [k]} v_i \cdot \text{Delta}_{=f_i}(f) = \sum_{i \in [k]} \widehat{x^*}[f_i] \cdot \text{Delta}_{=f_i}(f).$$

Definition 2.2 (Convolution). The convolution $(f * g)(t)$ for $t \in \mathbb{R}^d$ of two d -variate continuous function $f(t)$ and $g(t)$ is given by

$$(f * g)(t) := \int_{\tau \in \mathbb{R}^d} f(\tau) \cdot g(t - \tau) \cdot d\tau,$$

And the discrete convolution $(f * g)[i]$ for $i \in \mathbb{Z}$ of two same-length vectors f and g is given by⁹

$$(f * g)[i] = \sum_{j \in \mathbb{Z}} f[j] \cdot g[i - j].$$

2.3 An overview of previous techniques

The Sparse FT problem falls into the “*sparse recovery*” paradigm. Among such problems, an exemplar is to learn an *approximately k -sparse* length- N vector $\hat{y} \in \mathbb{R}^N$, by just accessing the length- N' *measurements* $y := \Phi \hat{y}$ resulted from an amount of N' -to- N *sensing matrices* $\Phi \in \mathbb{R}^{N' \times N}$, for some $N' \ll N$. Based on the measurements, an algorithm should output a k -sparse vector $\hat{y}' \in \mathbb{R}^N$ that approximates the vector \hat{y} . E.g., under the ℓ_2/ℓ_2 guarantee, we aim at achieving

$$\|\hat{y}' - \hat{y}\|_2 \lesssim \min_{k\text{-sparse } z} \|z - \hat{y}\|_2.$$

Given the flexibility of designing the Φ 's, the above problem is known as *compressed sensing*, and the optimization goals are threefold: (i) to access the fewest measurements, i.e., sample complexity¹⁰; (ii) to fast extract the k -sparse approximation $\hat{y}' \approx \hat{y}$, i.e., decoding time; and (iii) to use column-sparsest possible Φ 's, hence a faster encoding time.¹¹

We instead face the (discrete) sparse Fourier transform problem, if the above vector $\hat{y} \in \mathbb{R}^N$ is replaced by a length- N Fourier spectrum $\hat{x} \in \mathbb{C}^N$ (of any dimension $d \geq 1$) and the measurements y are replaced by the signal samples $x \in \mathbb{C}$. Again, the Fourier spectrum $\hat{x} \in \mathbb{C}^N$ is unknown, and we can only leverage the signal samples $x \in \mathbb{C}$. Now our optimization goals are to reduce the sample complexity and the decoding/running time.

Compressed sensing. To leverage the measurements, several past works on compressed sensing [GLPS10, DBIPW10, IP11, IPW11, BIP⁺16, NS19] first get a bunch of pseudorandom *hash functions* $h : [N] \mapsto [\mathcal{B}]$, where $\mathcal{B} = \Theta_d(k)$ is the number of bins. Such a “hashing” is associated with a random sign function $s : [m] \mapsto \{\pm 1\}$.¹² In one hashing, we derive the linear combination of the form

$$u_j := \sum_{i \in [m] : h(i)=j} y_i \cdot s(i), \quad (10)$$

for every bin $j \in [\mathcal{B}]$, based on a certain amount of $m = o(N)$ measurements $\{y_i\}_{i \in [m]} \subseteq \mathbb{R}^{N'}$. This scheme is known as “*hashing into \mathcal{B} bins*”. Following such ideas, $O(k \log(n/k))$ samples suffice to get a desired k -sparse approximation $\hat{y}' \approx \hat{y}$ [GLPS10, NS19].

Discrete Fourier transform. The very first obstacle to adopting a compressed sensing algorithm to the discrete Sparse FT problem is, how to implement the “hashing into \mathcal{B} bins” scheme by using the Fourier samples. Now we observe the signal x in the time domain, but aim to recover its Fourier spectrum $\hat{x} \in \mathbb{C}^N$ in the frequency domain.

⁹We define $(f * g)[i] := 0$ in the case that $i \notin \text{supp}(f) = \text{supp}(g)$.

¹⁰Only in the literature on compressed sensing, sample complexity is often called the *number of measurements*.

¹¹Optimizing encoding time only makes sense when we are allowed to design the sensing matrix, for more details of encoding time, we refer the readers to [NS19].

¹²Some previous works use the random Gaussian instead of the random sign functions.

The approach in the past works [HIKP12a, IK14, Kap16, Kap17] is to mimic the transformation in Eq. (10). That is, we first permute a bunch of $m = o(N)$ signal samples $\{x_i\}_{i \in [m]}$ via a pseudo-random affine permutation \mathcal{P} . Then, the permuted samples $\{(\mathcal{P}x)_i\}_{i \in [m]}$ are respectively scaled by coefficients $\{\mathcal{G}(l_i)\}_{i \in [m]}$, i.e., the values of a filter function $\mathcal{G} : \mathbb{R}^d \mapsto \mathbb{R}$ at $m = o(N)$ many *lattice points* $\{l_i\}_{i \in [m]} \subseteq \mathbb{R}^d$. Akin to Eq. (10), we use a transformation $u_j = \sum_{i \in [m]: \mathcal{H}(i)=j} (\mathcal{P}x)_i \cdot \mathcal{G}(l_i)$.

The second difficulty is that the hashing is no longer perfect. For compressed sensing, a coordinate $i \in [N]$ contributes 100% to a target bin, and 0% to the other $(\mathcal{B} - 1)$ bins. For the discrete Fourier transform, however, besides the target bin (which still gets 100%), any other bin should get a $\delta > 0$ fraction of mass from a coordinate $i \in [N]$. This modification (a.k.a. “leakage” [IK14]) is to make the “hashing into \mathcal{B} bins” efficient. Because of the imperfect hashing, the current sample complexity must involve an extra $\log(1/\delta)$ factor.

To get a better sense, let us briefly review the techniques in [IK14]. In any dimension $d \geq 1$, the frequency domain $\{\xi_i\}_{i \in [n^d]} = [n]^d$ is “on-the-grid”. Partition the domain $[n]^d = \text{HEAD} \sqcup \text{TAIL}$ into the head and tail frequencies (i.e., $|\text{HEAD}| = k$ and $|\text{TAIL}| = n^d - k = N - k$) and denote the magnitudes by $\hat{x}[\xi_i] \in \mathbb{C}$. Roughly speaking, the permutation by [IK14] works as follows:

$$\widehat{\mathcal{P}x}[\Sigma\xi_i - b \pmod{n}] = \hat{x}[\xi_i] \cdot e^{-\frac{2\pi i}{n} \cdot \xi_i^\top a},$$

where the modulo operation is taken coordinate-wise, $\Sigma \in [n]^{d \times d}$ is a random matrix, and $b, a \in [n]^d$ are random vectors.

The matrix $\Sigma \in [n]^{d \times d}$ is sampled uniformly at random among all integer matrices with *odd* determinants. So the inverse $\Sigma^{-1} \pmod{n}$ exists, making the permutation *one-to-one*. The vector $b \sim \text{Unif}[n]^d$ is uniform random, i.e., the “anchor point” of the permuted frequency domain.

Also, Σ and b together determine the hashing \mathcal{H} . Since Σ is invertible, the linear transformation $\Sigma\xi_i - b \pmod{n}$ forms a bijection from the “grid” frequency domain $\{\xi_i\}_{i \in [n^d]} = [n]^d$ to itself. [IK14] partition the codomain $[n]^d$ into $\mathcal{B} = B^d$ isomorphic Cartesian sub-grid, each of which has $(\frac{n}{B})^d = \frac{N}{B}$ grid points. The sub-grids are exactly the desired bins. For a uniform random “anchor point” $b \sim \text{Unif}[n]^d$, a frequency $\xi_i \in [n]^d$ is equally likely to fall into one of the bins.

Another crucial observation is that, any two different frequencies $\xi_i \neq \xi_{i'} \in [n]^d$ fall into the same bin with probability $\leq 0.01 \cdot k^{-1}$ [IK14]. Thus, 90% head frequencies $\xi_i \in \text{HEAD}$ will not collide with other head frequencies, hence being *isolated*.

The above permutation samples a uniformly random vector $a \sim \text{Unif}[n]^d$, and thus rotates a magnitude $\hat{x}[\xi_i] \in \mathbb{C}$ by a certain angle $-(2\pi/n) \cdot \xi_i^\top a$, i.e., the rotated magnitude $\hat{x}[\xi_i] \cdot e^{-\frac{2\pi i}{n} \cdot \xi_i^\top a} \in \mathbb{C}$ has a random phase. This is crucial because, given any sufficiently large subset $S \subseteq \text{TAIL}$ of the tail magnitudes, a uniform random $a \sim \text{Unif}[n]^d$ makes the total rotated magnitude (over $\xi \in S$) much smaller than the sum of the individual magnitudes.

Let $S = \{\xi \in \text{TAIL} : \mathcal{H}(\xi) = j\}$ denote the tail frequencies hashed into a certain bin $j \in [B]^d$. Given the above discussions, the total tail magnitude $z_j := \sum_{\xi \in S} \hat{x}[\xi] \cdot e^{-\frac{2\pi i}{n} \cdot \xi^\top a} \in \mathbb{C}$ is small enough such that (i) $z_j \in \mathbb{C}$ will not be identified as a *spurious* head frequency, when no head frequency is hashed into the j -th bin; and (ii) $z_j \in \mathbb{C}$ will not falsify an isolated head frequency $\xi_i \in \text{HEAD}$ too much, when $\xi_i \in \text{HEAD}$ is the unique head frequency in the j -th bin.

Different from [IK14], some other works like [Kap16, Kap17] use more complicated hash schemes, to improve the sample complexity and/or the running time of the sparse FT algorithm.

2.4 Technical barriers against a better tone estimation duration

The claimed tone estimation guarantee (Theorem 1.1) requires that $T \gtrsim d^{4.5}/\eta \cdot \log(kd/\delta) \cdot \log d$. Here the $\text{poly}(d)$ term stems from several places.

- (i) We sample the time points from a large range $|\text{supp}(a)| \approx T^d$ (Section 1.3.1). Since the vector $a = (a_r)_{r \in [d]}$ is in d dimension, we need $|\text{supp}(a_r)| \geq T - \Theta(T/d)$ in any single dimension. The second term $\Theta(T/d)$ (rather than $\Theta(T)$) incurs a factor- d loss in the duration bound.
- (ii) The procedure HASHTOBINS (Algorithm 1) switches the ℓ_2 -norm to the ℓ_∞ -norm, and thus incurs another factor- \sqrt{d} loss.
- (iii) How we generate the random matrix $\Sigma \in \mathbb{R}^{d \times d}$ loses a \sqrt{d} factor, to ensure a small collision probability $\Pr[\hat{h}(f_i) = \hat{h}(f_{i'})] \leq 0.01 \cdot k^{-1}$ for any two frequencies $f_i \neq f_{i'} \in \text{supp}(\hat{x}^*)$.
- (iv) Our filter function \mathcal{G} (see Appendix 4) is modified from the one by [CKPS16], which incurs a factor- d loss in the duration bound T . Without the modification, the approximation factor of our algorithm would be $2^{\Theta(d)}$ rather than $\text{poly}(d)$.
- (v) To select the k recovered tones $\{(v_i, f_i)\}_{i \in [k]}$ from $k' = \Theta_d(k)$ candidate tones (Algorithm 9), we amplify the duration bound by a $(d^{1.5} \log d)$ factor. In particular, we first pay a $(d \log d)$ factor because there are $k' = 2^{\Theta(d \log d)} \cdot k$ candidate tones. Moreover, in the selection process, we cannot afford the running time to query points in ℓ_2 -space (i.e., the memberships regarding some ℓ_2 -regions) even with the best data structure. Instead, we will work in the ℓ_∞ -space and choose the gap $\eta' = \eta/\sqrt{d}$, which incurs another factor- \sqrt{d} loss.

To sum up, we need a duration $T \gtrsim \eta'^{-1} \cdot d^3 \cdot \log(k'd/\delta) = \eta^{-1} \cdot d^{4.5} \cdot \log(kd/\delta) \cdot \log d := C_{\text{tone}} \cdot \eta^{-1}$.

3 Probability tools

In this section, we present a number of classical probability tools to be used in this paper: the Chernoff bound (Lemma 3.1), the Hoeffding bound (Lemma 3.2) and the Bernstein bound (Lemma 3.3) measure the tail bounds of random scalar variables. Further, Lemma 3.4 is a concentration result about random matrices.

We state the classical Chernoff bound below, which is named after Herman Chernoff but is due to Herman Rubin. It gives exponentially decreasing bounds for the tail distributions of the sums of independent random variables.

Lemma 3.1 (Chernoff bound [Che52]). *Let $\{X_i\}_{i \in [n]}$ be $n \geq 1$ independent Bernoulli random variables, such that $X_i = 1$ with probability $p_i \in [0, 1]$ and $X_i = 0$ with probability $1 - p_i$. Then the following hold for the random sum $X := \sum_{i \in [n]} X_i$ and the expectation $\mu := \mathbf{E}[X] = \sum_{i \in [n]} p_i$.*

Part (a): $\Pr[X \geq (1 + \delta)\mu] \leq e^{\delta \cdot \mu} \cdot (1 + \delta)^{-(1+\delta) \cdot \mu}$ for any $\delta > 0$.

Part (b): $\Pr[X \leq (1 - \delta)\mu] \leq e^{-\delta \cdot \mu} \cdot (1 - \delta)^{-(1-\delta) \cdot \mu}$ for any $0 < \delta < 1$.

We state the Hoeffding bound below:

Lemma 3.2 (Hoeffding bound [Hoe63]). *Let $\{X_i\}_{i \in [n]}$ be $n \geq 1$ independent random variables bounded between $\text{supp}(X_i) \subseteq [a_i, b_i]$, for some $a_i \leq b_i \in \mathbb{R}$. Then the following holds for the random sum $X := \sum_{i \in [n]} X_i$ and any $t \geq 0$.*

$$\Pr[|X - \mathbf{E}[X]| \geq t] \leq 2 \cdot \exp\left(-\frac{2t^2}{\sum_{i \in [n]} (b_i - a_i)^2}\right).$$

We state the Bernstein inequality below:

Lemma 3.3 (Bernstein inequality [Ber24]). *Let $\{X_i\}_{i \in [n]}$ be $n \geq 1$ independent zero-mean random variables $\mathbf{E}[X_i] = 0$. Suppose that $|X_i| \leq M$ almost surely, for every $i \in [n]$ and some $M \geq 0$. Then the following holds for the random sum $X := \sum_{i \in [n]} X_i$ and any $t \geq 0$.*

$$\Pr[X > t] \leq \exp\left(-\frac{t^2/2}{\sum_{i \in [n]} \mathbf{E}[X_i^2] + Mt/3}\right).$$

Matrix concentration inequalities have various applications. Below, we state a matrix Bernstein inequality by [Tro15], which can be regarded as a matrix version of Lemma 3.3.

Lemma 3.4 (Matrix Bernstein [Tro15, Theorem 6.1.1]). *Let $\{X_i\}_{i \in [m]} \subseteq \mathbb{R}^{n_1 \times n_2}$ be a set of $m \geq 1$ i.i.d. matrices with the expectation $\mathbf{E}[X_i] = 0^{n_1 \times n_2}$. For some $M \geq 0$, assume*

$$\|X_i\| \leq M, \quad \forall i \in [m].$$

Let $X = \sum_{i \in [m]} X_i$ be the random sum. Let $\text{Var}[X]$ be the matrix variance statistic of the sum:

$$\text{Var}[X] := \max \left\{ \left\| \sum_{i \in [m]} \mathbf{E}[X_i X_i^\top] \right\|, \left\| \sum_{i \in [m]} \mathbf{E}[X_i^\top X_i] \right\| \right\}.$$

Then

$$\mathbf{E}[\|X\|] \leq \sqrt{2 \cdot \text{Var}[X] \cdot \log(n_1 + n_2)} + \frac{M}{3} \cdot \log(n_1 + n_2).$$

Furthermore, the following holds for any $t \geq 0$.

$$\Pr[\|X\| \geq t] \leq (n_1 + n_2) \cdot \exp\left(-\frac{t^2/2}{\text{Var}[X] + Mt/3}\right).$$

Lemma 3.5 (Sub-gaussian rows [Ver10, Theorem 5.39]). *Let A be an $N \times n$ matrix whose rows A_i for $i \in [N]$ are independent sub-gaussian isotropic random vectors in \mathbb{R}^n . Then for every $t \geq 0$, with probability at least $1 - 2\exp(-ct^2)$, we have*

$$\sqrt{N} - C\sqrt{n} - t \leq s_{\min}(A) \leq s_{\max}(A) \leq \sqrt{N} + C\sqrt{n} + t.$$

where $s_{\max}(A)$ (resp. $s_{\min}(A)$) represents the largest (resp. smallest) singular value of matrix A , and absolute constants $C = C_K$, $c = c_K$ depend only on the sub-gaussian norm $K = \max_{i \in [N]} \|A_i\|_{\psi_2}$ of the rows.

4 Filter, permutation and hashing in multiple dimensions

Different from the previous sections, in this section $t = (t_s)_{s \in [d]} \in \mathbb{R}^d$ and $f = (f_r)_{r \in [d]} \in \mathbb{R}^d$ will respectively denote the d -dimensional vectors in the time domain and in the frequency domain, and $i \in \mathbb{N}_{\geq 0}^d$ and $j \in \mathbb{N}_{\geq 0}^d$ will denote the vector indices.

4.1 Construction of filter $(\mathcal{G}(t), \widehat{\mathcal{G}}(f))$

Definition 4.1 (The multi-dimensional filter). Recall the parameters defined in Definition B.1:

- The number of bins in a single dimension $B = \Theta(d \cdot k^{1/d})$ is a certain multiple of $d \in \mathbb{N}_{\geq 1}$. Over all the $d \in \mathbb{N}_{\geq 1}$ dimensions, we have $\mathcal{B} = B^d = 2^{\Theta(d \log d)} \cdot k$ many bins.
- The noise level parameter $\delta \in (0, 1)$.
- $\alpha = \Theta(1/d)$ is chosen such that $\frac{1}{100 \cdot (d+1) \cdot \alpha} \in \mathbb{N}_{\geq 1}$ is an integer; clearly $\alpha \leq \frac{1}{100 \cdot (d+1)} \leq \frac{1}{200}$.
- $s_1 = \frac{2B}{\alpha}$ and $s_2 = \frac{1}{B+B/d}$.
- $\ell = \Theta(\log(kd/\delta))$ is an even integer. We safely assume $\ell \geq 1000$.

Further, the *width parameter* $W = \Omega(d \cdot \frac{F}{B\eta})$ is chosen to be a sufficiently large integer. Then for any $t = (t_s)_{s \in [d]} \in \mathbb{R}^d$ and any $f = (f_r)_{r \in [d]} \in \mathbb{R}^d$, the filter function $(\mathcal{G}(t), \widehat{\mathcal{G}}(f))$ is given by

$$\mathcal{G}(t) = \prod_{s \in [d]} \mathcal{G}(t_s) \quad \text{and} \quad \widehat{\mathcal{G}}(f) = \prod_{r \in [d]} \widehat{\mathcal{G}}(f_r),$$

where the single-dimensional filter $(\mathcal{G}(t_s), \widehat{\mathcal{G}}(f_r))$ is constructed according to Definition B.1, under the same parameters $B, \delta, \alpha, s_1, s_2, \ell$ and W .

Definition 4.2 (Hypercube grid). Define

$$\Lambda_W(z) := \{f \in \mathbb{R}^d : \|f - i\|_\infty \leq z \text{ for some vector index } i \in [-W : W]^d\}.$$

This denotes the union of all the hypercubes that (for the chosen i 's) have edge length $2z \geq 0$ and are centered at $i \in [-W : W]^d$. Notice that $\Lambda_W(z) \supseteq \Lambda_W(z')$ for any $z \geq z' \geq 0$.

4.2 Properties of filter $(\mathcal{G}(t), \widehat{\mathcal{G}}(f))$

Lemma 4.3 (The multi-dimensional filter). *The filter $(\mathcal{G}(t), \widehat{\mathcal{G}}(f))[B, \delta, \alpha, \ell, W]$ given in Definition 4.1 satisfies the following:*

Property I: $e^{-\frac{\delta}{\text{poly}(k, d)}} \leq \widehat{\mathcal{G}}(f) \leq 1$ for any $f \in \Lambda_W(\frac{1-\alpha}{2B})$.

Property II: $\widehat{\mathcal{G}}(f) \in [0, 1]$ for any $f \in \Lambda_W(\frac{1}{2B}) \setminus \Lambda_W(\frac{1-\alpha}{2B})$.

Property III: $0 \leq \widehat{\mathcal{G}}(f) \leq \frac{\delta}{\text{poly}(k, d)}$ for any $f \in \mathbb{R}^d \setminus \Lambda_W(\frac{1}{2B})$.

Property IV: $\text{supp}(\mathcal{G}) \subseteq [-\ell \cdot \frac{B}{\alpha}, \ell \cdot \frac{B}{\alpha}]^d$.

Property V: $\sum_{i \in \mathbb{Z}^d} \mathcal{G}(i)^2 \leq e^2 \cdot B^{-d} = e^2 \cdot \mathcal{B}^{-1}$.

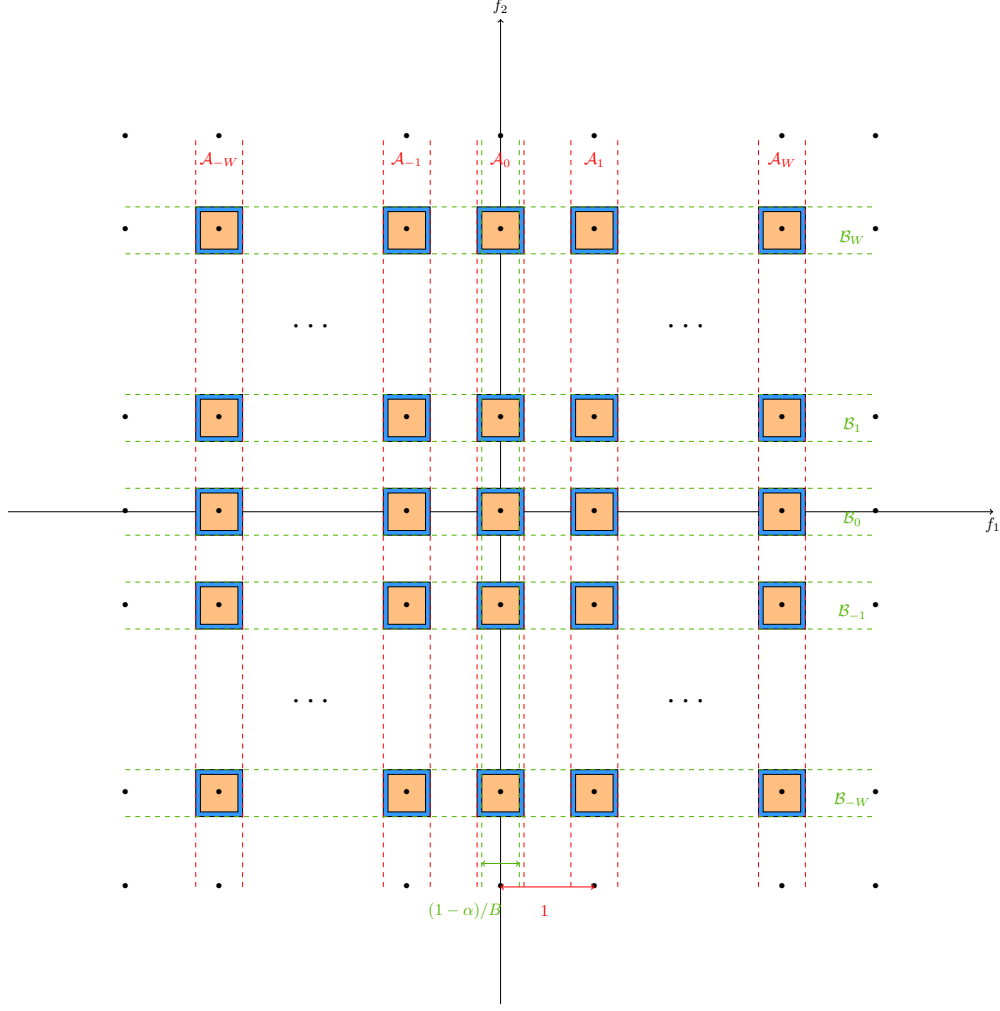


Figure 4: Demonstration for the filter $\widehat{\mathcal{G}}(f)$ in two dimension $d = 2$. “yellow” refers to Property I and $f \in \Lambda_W(\frac{1-\alpha}{2B})$, where $\widehat{\mathcal{G}}(f)$ ’s value is very close to 1; “blue” refers to Property II and $f \in \Lambda_W(\frac{1}{2B}) \setminus \Lambda_W(\frac{1-\alpha}{2B})$, where the value of $\widehat{\mathcal{G}}(f)$ drops sharply, and the other "white" region means Property III and $f \in \mathbb{R}^d \setminus \Lambda_W(\frac{1}{2B})$, where $\widehat{\mathcal{G}}(f)$ oscillates near 0. $\mathcal{A}_i = [i - 1/(2B), i + 1/(2B)] \times \mathbb{R}$ and $\mathcal{B}_i = \mathbb{R} \times [i - 1/(2B), i + 1/(2B)]$.

4.3 Proof of properties

Below we only present the proofs of Properties III and V, and the other properties directly follow from the corresponding properties of the single-dimensional filter $(\mathcal{G}(t_s), \widehat{\mathcal{G}}(f_r))$ that are given in Definition B.1 and Lemma B.2.

Claim 4.4 (Property III of Lemma 4.3). $0 \leq \widehat{\mathcal{G}}(f) \leq \frac{\delta}{\text{poly}(k, d)}$ for any $f \in \mathbb{R}^d \setminus \Lambda_W(\frac{1}{2B})$.

Proof. We let $r^* \in [d]$ denote (one of) the coordinate that maximizes, over all $r \in [d]$, the distance of f_r from the lattice $[-W : W]$. Because $f \in \mathbb{R}^d \setminus \Lambda_W(\frac{1}{2B})$, that maximum distance is at least $\frac{1}{2B}$. Then by construction (see Definition 4.1),

$$\widehat{\mathcal{G}}(f) = \prod_{r \in [d]} \widehat{\mathcal{G}}(f_r) \leq \widehat{\mathcal{G}}(f_{r^*}) \leq \frac{\delta}{\text{poly}(k, d)},$$

where the second step follows because $\widehat{\mathcal{G}}(f_r) \in [0, 1]$ for each coordinate $r \in [d] \setminus \{r^*\}$ (see Properties II to IV of Lemma B.2); and the last step follows from Property III of Lemma B.2.

This completes the proof of Claim 4.4. \square

Claim 4.5 (Property V of Lemma 4.3). $\sum_{i \in \mathbb{Z}^d} \mathcal{G}(i)^2 \leq e^2 \cdot B^{-d} = e^2 \cdot \mathcal{B}^{-1}$.

Proof. Due to Definition 4.1 that $\mathcal{G}(t) = \prod_{s \in [d]} \mathcal{G}(t_s)$ for any $t \in \mathbb{R}^d$, we have

$$\begin{aligned} \sum_{i \in \mathbb{Z}^d} \mathcal{G}(i)^2 &= \sum_{i \in \mathbb{Z}^d} \left(\prod_{s \in [d]} \mathcal{G}(i_s)^2 \right) \\ &= \prod_{s \in [d]} \left(\sum_{i_s \in \mathbb{Z}} \mathcal{G}(i_s)^2 \right) \\ &\leq \prod_{s \in [d]} \left(\left(1 + \frac{2}{d}\right) \cdot B^{-1} \right) \\ &= \left(1 + \frac{2}{d}\right)^d \cdot B^{-d} \\ &\leq e^2 \cdot B^{-d}, \end{aligned}$$

where the third step follows from Property VI of Lemma B.2 that $\sum_{i \in \mathbb{Z}} \mathcal{G}(i)^2 \leq (1 + \frac{2}{d}) \cdot B^{-1}$; and the last step follows because $(1 + \frac{1}{z})^z \leq e$ for any $z > 0$.

This completes the proof of Claim 4.5. \square

4.4 Construction and properties of standard window $(\mathcal{G}'(t), \widehat{\mathcal{G}}'(f))$

Now we associate our multi-dimensional filter $(\mathcal{G}(t), \widehat{\mathcal{G}}(f))$ given in Definition 4.1 with another standard window $(\mathcal{G}'(t), \widehat{\mathcal{G}}'(f))$ in a similar manner as Lemma B.11 and the counterpart results in [HIKP12a, HIKP12b], which is more convenient for our later use.

Lemma 4.6 (The multi-dimensional standard window). *Consider the filter function $(\mathcal{G}(t), \widehat{\mathcal{G}}(f))$ given in Definition 4.1, there is another function $(\mathcal{G}'(t), \widehat{\mathcal{G}}'(f))$ such that:*

Property I: $\widehat{\mathcal{G}}'(f) = 1$ for any $f \in \Lambda_W(\frac{1-\alpha}{2B})$.

Property II: $\widehat{\mathcal{G}}'(f) \in [0, 1]$ for any $f \in \Lambda_W(\frac{1}{2B}) \setminus \Lambda_W(\frac{1-\alpha}{2B})$.

Property III: $\widehat{\mathcal{G}}'(f) = 0$ for any $f \in \mathbb{R}^d \setminus \Lambda_W(\frac{1}{2B})$.

Property IV: $\|\widehat{\mathcal{G}}' - \widehat{\mathcal{G}}\|_\infty = \max_{f \in \mathbb{R}^d} |\widehat{\mathcal{G}}'(f) - \widehat{\mathcal{G}}(f)| \leq \frac{\delta}{\text{poly}(k, d)}$.

Proof. We define $\widehat{\mathcal{G}}'(f)$ as follows; noticeably, similar to $\widehat{\mathcal{G}}(f)$, this is also an even function in every coordinate $r \in [d]$ given that the other $(d-1)$ coordinates are fixed:

$$\widehat{\mathcal{G}}'(f) = \begin{cases} 1 & \forall f \in \Lambda_W(\frac{1-\alpha}{2B}) \\ \widehat{\mathcal{G}}(f) & \forall f \in \Lambda_W(\frac{1}{2B}) \setminus \Lambda_W(\frac{1-\alpha}{2B}) \\ 0 & \forall f \in \mathbb{R}^d \setminus \Lambda_W(\frac{1}{2B}) \end{cases}.$$

Then all the properties above can be inferred from Lemma 4.3.

This completes the proof of Lemma 4.6. \square

4.5 Permutation and hashing

We adopt the following notations for convenience:

- Let $\lfloor z \rfloor \in \mathbb{Z}$ denote the greatest integer that is less than or equal to a real number $z \in \mathbb{R}$. In the case that $z = (z_r)_{r=1}^d \in \mathbb{R}^d$ is a vector, we would abuse the notation $\lfloor z \rfloor = (\lfloor z_r \rfloor)_{r=1}^d \in \mathbb{Z}^d$.
- Let $\text{frac}(z) = z - \lfloor z \rfloor \in [0, 1)$ denote the fractional part of a real number $z \in \mathbb{R}$. In the case that $z = (z_r)_{r=1}^d \in \mathbb{R}^d$ is a vector, we would abuse the notation $\text{frac}(z) = (\text{frac}(z_r))_{r=1}^d \in [0, 1)^d$.
- Denote the set $[n] = \{0, 1, \dots, n-1\}$, for any positive integer $n \in \mathbb{N}_{\geq 1}$.
- Let $\bar{z} \in \mathbb{C}$ denote the conjugate of a complex number $z \in \mathbb{C}$. Notice that $|z|^2 = z\bar{z}$.

Definition 4.7 (Setup for permutation and hashing). We sample the random matrix $\Sigma \in \mathbb{R}^{d \times d}$ and the random vectors $a, b \in \mathbb{R}^d$, and define the parameter \mathcal{D} as follows:

- The d -to- d random matrix Σ is constructed in two steps. First, we sample an interim matrix $\Sigma' \sim \text{Unif}(\mathbf{SO}(d))$ uniformly at random from the *rotation group*, namely a rotation matrix of determinant $\det(\Sigma') = 1$. Then, we define $\Sigma := \beta \Sigma'$, where the scaling factor $\beta \sim \text{Unif}[\frac{2\sqrt{d}}{B\eta}, \frac{4\sqrt{d}}{B\eta}]$ is uniform random.
- The random vector $a \in \mathbb{R}^d$ will be specified later in Section 5.6. In this section, we only need the property of a given in Conditions 4.8 and 4.9, which also will be verified in Section 5.6.
- The random vector $b' = (b'_r)_{r \in [d]} \sim \text{Unif}[0, 1]^d$. Then, let $b := \Sigma^{-1}b'$.
- The parameter $D = \Theta(\ell/\alpha) = \Theta(d \cdot \log(kd/\delta))$ is a sufficiently large integer. Also, let $\mathcal{D} := D^d$.

Condition 4.8 (Duration requirement). *Given any $i \in [BD]^d$ and any choice of the random matrix $\Sigma \in \mathbb{R}^{d \times d}$ according to Definition 4.7, any choice of a ensures that $\Sigma^\top(i + a) \in [0, T]^d$ is within the duration.*

Condition 4.9 (Sampling requirement). *Given any $i \in [BD]^d$ and any choice of the random matrix $\Sigma \in \mathbb{R}^{d \times d}$ according to Definition 4.7, the following hold for the random vector a :*

$$\mathbf{E}_a \left[g(\Sigma^\top(i + a))^2 \right] \lesssim \frac{1}{T^d} \cdot \int_{t \in [0, T]^d} |g(t)|^2 \cdot dt,$$

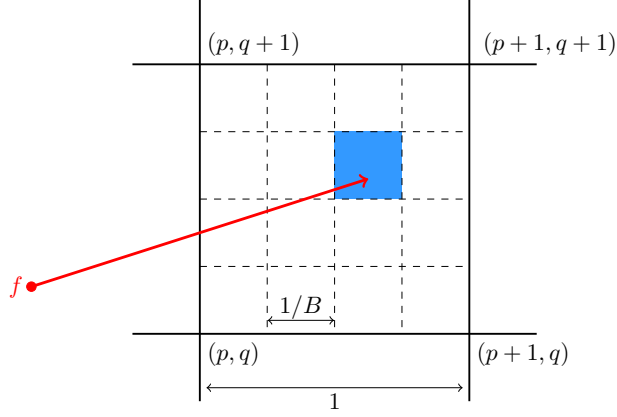


Figure 5: Demonstration for the hashing scheme (Definition 4.10) in two dimensions $d = 2$, where $p, q \in \mathbb{Z}$ are integers. The unit square $[p, p+1] \times [q, q+1]$ are divided into $\mathcal{B} = B^2$ subsquares, and the subsquare in which the frequency f is hashed into, is exactly the index $h_{\Sigma,b}(f) \in [B]^d$.

Definition 4.10 (Hashing). Define the vector-valued function

$$h_{\Sigma,b}(f) = \left\lfloor B \cdot \text{frac} \left(\frac{1}{2B} \cdot \mathbf{1} + \Sigma(f - b) \right) \right\rfloor \in [B]^d.$$

This function “hashes” any frequency $f \in [-F, F]^d$ into one of the $\mathcal{B} = B^d = 2^{\Theta(d \log d)} \cdot k \in \mathbb{N}_{\geq 1}$ bins. When $B = \Theta(d \cdot k^{1/d})$ is large enough, every bin $j \in [B]^d$ is likely to have at most one heavy hitter (namely one tone frequency $f \in \text{supp}(\widehat{x^*})$) and if so, we can recover the tone from the hitting bins via an 1-sparse algorithm. See Figure 5 for a demonstration.

Definition 4.11 (Offset). Define the vector-valued function

$$\sigma_{\Sigma,b}(f) = \text{frac} \left(\frac{1}{2B} \cdot \mathbf{1} + \Sigma(f - b) \right) - \frac{1}{B} \cdot h_{\Sigma,b}(f) - \frac{1}{2B} \cdot \mathbf{1} \in \left[-\frac{1}{2B}, \frac{1}{2B} \right]^d,$$

which measures the coordinate-wise distance from the center of the $h_{\Sigma,b}(f)$ -th bin to $f \in [-F, F]^d$.

Definition 4.12 (Collision). Consider a tone frequency $f \in \text{supp}(\widehat{x^*})$, the event $E_{\text{coll}}(f)$ occurs when $h_{\Sigma,b}(f') = h_{\Sigma,b}(f)$ for some other tone frequency $f' \in \text{supp}(\widehat{x^*}) \setminus \{f\}$, namely both $f \neq f' \in \text{supp}(\widehat{x^*})$ are hashed into the same bin. In this case, the algorithm cannot recover the two collided tone frequencies $f \neq f'$. See Figure 6(a) for a demonstration.

Definition 4.13 (Large offset). Consider a tone frequency $f \in \text{supp}(\widehat{x^*})$, the event $E_{\text{off}}(f)$ occurs when $\|\sigma_{\Sigma,b}(f)\|_{\infty} \geq \frac{1-\alpha}{2B}$, i.e. the frequency $f \in \text{supp}(\widehat{x^*})$ locates on the boundary of the $h_{\Sigma,b}(f)$ -th bin. In this case, the algorithm also cannot recover the tone frequency f . See Figure 6(b) for a demonstration.

Our multi-dimensional permutation scheme is a natural generalization of the single-dimensional one by [PS15, Definition A.5].

Definition 4.14 (Multi-dimensional permutation). Let $\mathcal{P}_{\Sigma,b,a}x(t) = x(\Sigma^{\top}(t + a)) \cdot e^{-2\pi i \cdot b^{\top} \Sigma^{\top} t}$ for any $t \in \mathbb{R}^d$.

Lemma 4.15 (Identities). The permutation given in Definition 4.14 satisfies that:

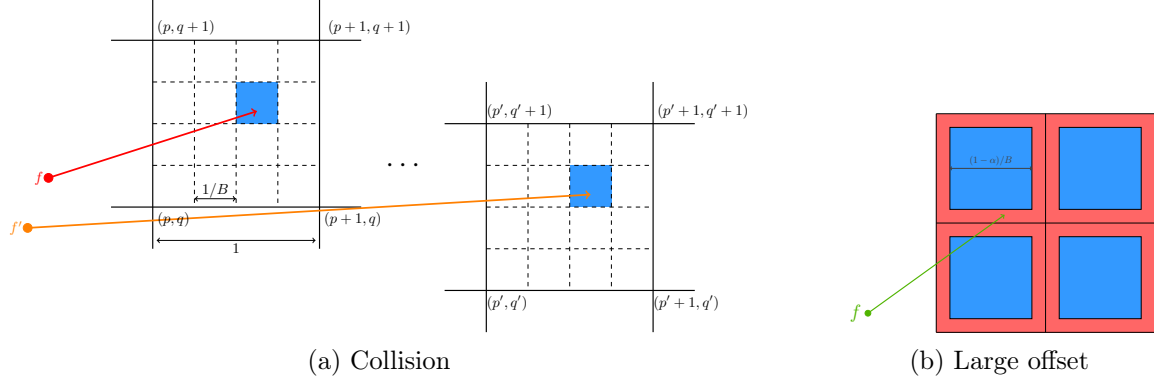


Figure 6: Demonstration for the bad events “collision” (Definition 4.12) and “large offset” (Definition 4.13) in two dimensions $d = 2$. In Figure 6(a), the two frequencies $f \neq f' \in \text{supp}(\widehat{x}^*)$ may be hashed into two different unit squares (i.e. possibly either $p \neq p'$ or $q \neq q'$ or both), but it is always the case that the two subsquares have the same index $\kappa_{\Sigma,b}(f) = \kappa_{\Sigma,b}(f') \in [B]^d$. In Figure 6(b), the red region (that gives a large offset) covers $1 - (1 - \alpha)^2$ fractions of the whole plane.

Property I: $\widehat{\mathcal{P}_{\Sigma,b,a}x}(\Sigma(f - b)) = \widehat{x}(f) \cdot \det(\Sigma)^{-1} \cdot e^{2\pi i \cdot a^\top \Sigma f}$ for any $f \in \mathbb{R}^d$.

Property II: $\widehat{\mathcal{P}_{\Sigma,b,a}x}(t) = \widehat{x}(\Sigma^{-1}t + b) \cdot \det(\Sigma)^{-1} \cdot e^{2\pi i \cdot a^\top (t + \Sigma^\top b)}$ for any $t \in \mathbb{R}^d$.

Proof. For Property I, by the definition of the CFT, the LHS equals

$$\begin{aligned}
\widehat{\mathcal{P}_{\Sigma,b,a}x}(\Sigma(f - b)) &= \int_{t \in \mathbb{R}^d} \mathcal{P}_{\Sigma,b,a}x(t) \cdot e^{-2\pi i \cdot (f^\top - b^\top) \Sigma^\top t} \cdot dt \\
&= \int_{t \in \mathbb{R}^d} x(\Sigma^\top(t + a)) \cdot e^{-2\pi i \cdot b^\top \Sigma^\top t} \cdot e^{-2\pi i \cdot (f^\top - b^\top) \Sigma^\top t} \cdot dt \\
&= e^{2\pi i \cdot f^\top \Sigma^\top a} \cdot \int_{t \in \mathbb{R}^d} x(\Sigma^\top(t + a)) \cdot e^{-2\pi i \cdot f^\top \Sigma^\top (t + a)} \cdot dt \\
&= e^{2\pi i \cdot f^\top \Sigma^\top a} \cdot \det(\Sigma)^{-1} \cdot \int_{\tau \in \mathbb{R}^d} x(\tau) \cdot e^{-2\pi i \cdot f^\top \tau} \cdot d\tau \\
&= e^{2\pi i \cdot f^\top \Sigma^\top a} \cdot \det(\Sigma)^{-1} \cdot \widehat{x}(f) \\
&= e^{2\pi i \cdot a^\top \Sigma f} \cdot \det(\Sigma)^{-1} \cdot \widehat{x}(f),
\end{aligned}$$

where the second step follows from Definition 4.14; the fourth step is by substitution; and the last step is by the definition of the CFT.

We directly infer Property II from Property I by substitution. Lemma 4.15 follows then. \square

4.6 HASHTOBINS: algorithm

Fact 4.16 (Identities under DFT/DTFT). *The following holds for each $j \in [B]^d$:*

$$\widehat{u}_j = \widehat{y}_{jD} = \widehat{\mathcal{G}} * \widehat{\mathcal{P}_{\Sigma,b,a}x}(B^{-1} \cdot j).$$

Proof. It is noteworthy that $y = (y_j)_{j \in [BD]^d}$ is a $(BD)^d$ -dimensional vector, and $u = (u_j)_{j \in [B]^d}$ is a B^d -dimensional vector. For the first equality $\widehat{u}_j = \widehat{y}_{jD}$, due to the definition of the DFT,

$$\widehat{u}_j = \sum_{i \in [B]^d} u_i \cdot e^{-\frac{2\pi i}{B} \cdot j^\top i}$$

Algorithm 1 HASHTOBINS in multiple dimensions

- 1: **procedure** HASHTOBINS(Σ, b, a, D)
 - 2: Define $(\mathcal{G}(t), \widehat{\mathcal{G}}(f))$ according to Definition B.1.
 - 3: Define $\mathcal{P}_{\Sigma, b, a}$ according to Definition 4.14.
 - 4: Define $y = (y_j)_{j \in [BD]^d}$, where $y_j = \mathcal{G}(j) \cdot \mathcal{P}_{\Sigma, b, a} x(j)$.
 - 5: Define $u = (u_j)_{j \in [B]^d}$, where $u_j = \sum_{i \in [D]^d} y_{Bj+i}$.
 - 6: **return** the DFT $\widehat{u} = (\widehat{u}_j)_{j \in [B]^d}$.
 - 7: **end procedure**
-

$$\begin{aligned}
 &= \sum_{i \in [B]^d} \sum_{l \in [D]^d} y_{Bl+i} \cdot e^{-\frac{2\pi i}{B} \cdot j^\top i} \\
 &= \sum_{i \in [B]^d} \sum_{l \in [D]^d} y_{Bl+i} \cdot e^{-\frac{2\pi i}{B} \cdot j^\top (Bl+i)} \\
 &= \sum_{i \in [BD]^d} y_i \cdot e^{-\frac{2\pi i}{B} \cdot j^\top i} \\
 &= \sum_{i \in [BD]^d} y_i \cdot e^{-\frac{2\pi i}{BD} \cdot (Dj)^\top i} \\
 &= \widehat{y}_{Dj},
 \end{aligned}$$

where the second step is by Line 5 of HASHTOBINS; the third step follows since both $j \in [B]^d$ and $l \in [D]^d$ are d -dimensional integer vectors and therefore, $e^{-\frac{2\pi i}{B} \cdot j^\top (Bl)} = e^{-2\pi i \cdot j^\top l} = 1$; the fourth step is by substitution; and the last step also applies the DFT.

For the second equality, again we know from the definition of the DFT that

$$\begin{aligned}
 \widehat{y}_{Dj} &= \sum_{i \in [BD]^d} y_i \cdot e^{-\frac{2\pi i}{BD} \cdot (Dj)^\top i} \\
 &= \sum_{i \in [BD]^d} \mathcal{G}(i) \cdot \mathcal{P}_{\Sigma, b, a} x(i) \cdot e^{-\frac{2\pi i}{B} \cdot j^\top i} \\
 &= \sum_{i \in \mathbb{Z}^d} \mathcal{G}(i) \cdot \mathcal{P}_{\Sigma, b, a} x(i) \cdot e^{-\frac{2\pi i}{B} \cdot j^\top i} \\
 &= \widehat{\mathcal{G} \cdot \widehat{\mathcal{P}_{\Sigma, b, a} x}}(B^{-1} \cdot j) \\
 &= \widehat{\mathcal{G}} * \widehat{\mathcal{P}_{\Sigma, b, a} x}(B^{-1} \cdot j),
 \end{aligned}$$

where the second step is by Line 4 of HASHTOBINS; the third step follows because $\mathcal{G}(t) = 0$ when $\|t\|_\infty \geq \ell \cdot B/\alpha$ (see Lemma 4.3), given a large enough $D = \Theta(d \cdot \log(kd/\delta))$ (see Definition 4.7); and the fourth step is by the definition of the DTFT.

This completes the proof of Fact 4.16. \square

Fact 4.17 (Sample complexity and time complexity). *The procedure HASHTOBINS takes $O(\mathcal{BD}) = 2^{O(d \cdot \log d)} \cdot k \cdot \log^d(k/\delta)$ samples and runs in $O(\mathcal{BD} + \mathcal{B} \log \mathcal{B}) = 2^{O(d \cdot \log d)} \cdot k \cdot \log^d(k/\delta)$ time.*

Proof. Recall that $\mathcal{B} = B^d = 2^{\Theta(d \log d)} \cdot k$ (Definition 4.1) and $\mathcal{D} = D^d = \log^d(kd/\delta)$. The sample complexity is easy to see, because we have exactly \mathcal{B} bins, and each bin $j \in [\mathcal{B}]$ requires exactly \mathcal{D} samples in Line 5 of HASHTOBINS.

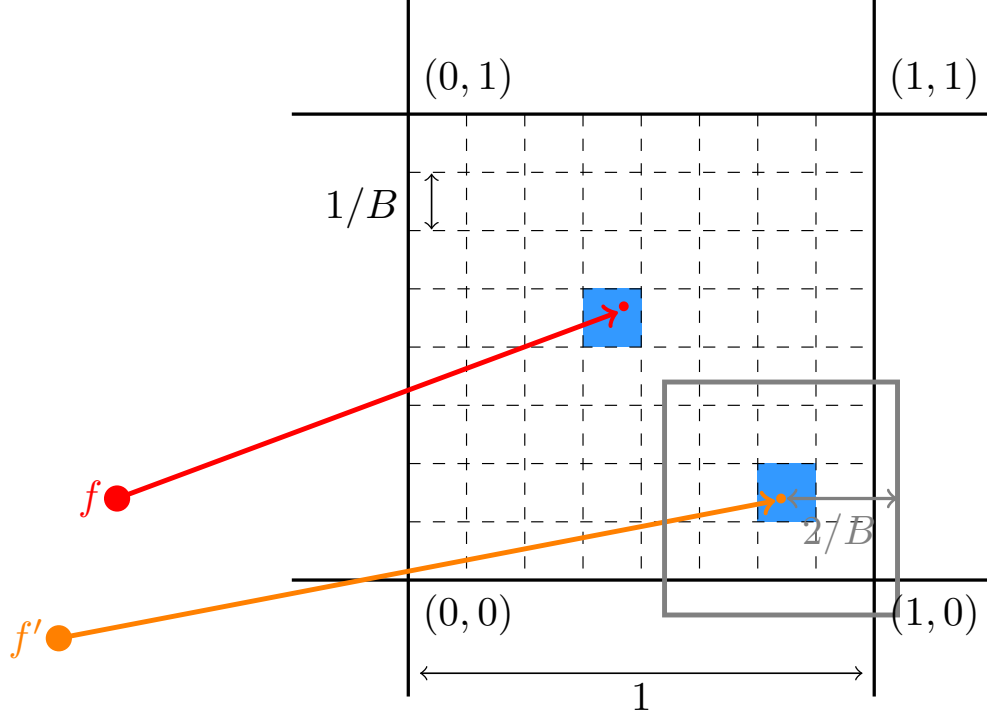


Figure 7: Demonstration for Lemma 4.18 in two dimensions $d = 2$, where the hashing of f and the hashing of f' are rounded into the same unit square. A sufficient condition to avoid the collision between $f \neq f' \in \text{supp}(\widehat{x^*})$ is that the ℓ_∞ distance $\|\hbar_{\Sigma,b}(f) - \hbar_{\Sigma,b}(f')\|_\infty \geq \frac{2}{B}$.

The \mathcal{BD} -dimensional vector $y = (y_j)_{j \in [BD]^d}$ can be computed $O(\mathcal{BD})$ time (assuming $O(1)$ -time query oracles to evaluating the filter $(\mathcal{G}(t), \widehat{\mathcal{G}}(f))$ and to sampling the signal $x(t)$; see Remark B.3). Furthermore, the \mathcal{B} -dimensional vector $u = (u_j)_{j \in [B]^d}$, where $u_j = \sum_{i \in [D]} y_{Bi+j}$, can be computed in $O(\mathcal{BD})$ time. Then we can derive its DFT $\widehat{u} = (\widehat{u}_j)_{j \in [B]^d}$ through any FFT algorithm in $O(\mathcal{B} \log \mathcal{B})$ time. The claimed time complexity follows as well. \square

4.7 HASHTOBINS: probabilities of bad events

Lemma 4.18 (Probability of collision). *Consider the random matrix $\Sigma \in \mathbb{R}^{d \times d}$ and the random vector $b \in \mathbb{R}^d$ given in Definition 4.7, for any pair of tone frequencies $f \neq f' \in \text{supp}(\widehat{x^*})$, the probability of collision*

$$\Pr_{\Sigma, b} [\hbar_{\Sigma, b}(f) = \hbar_{\Sigma, b}(f')] \leq 0.01 \cdot k^{-1}.$$

Proof. Recall that $\eta = \min\{\|f - f'\|_2 : f \neq f' \in \text{supp}(\widehat{x^*})\}$ is the minimum ℓ_2 -distance between any pair of tone frequencies. Due to Definition 4.10, two distinct tone frequencies $f \neq f' \in \text{supp}(\widehat{x^*})$ are hashed into bins

$$\begin{aligned} \hbar_{\Sigma, b}(f) &= \left\lfloor B \cdot \text{frac}\left(\frac{1}{2B} \cdot \mathbf{1} + \Sigma(f - b)\right) \right\rfloor \\ \hbar_{\Sigma, b}(f') &= \left\lfloor B \cdot \text{frac}\left(\frac{1}{2B} \cdot \mathbf{1} + \Sigma(f - b) + \Sigma(f' - f)\right) \right\rfloor. \end{aligned}$$

In order to hash f and f' into the same bin, a necessary condition (under any realization of the random vector b) is that $\|\Sigma(f' - f) - i\|_\infty \leq \frac{1}{B}$; in other words, $\Sigma(f' - f)$ locates in a hypercube that is centered at some integer vector $i = (i_r)_{r \in [d]} \in \mathbb{Z}^d$ and has edge length $\frac{2}{B}$.

Given the above two equations and as Figure 7 suggests, a necessary condition for the collision $h_{\Sigma,b}(f) = h_{\Sigma,b}(f')$ is that $\|\Sigma(f' - f) - i\|_\infty < \frac{2}{B}$ for some integer vector $i \in \mathbb{Z}^d$. Below we upper bound this probability based on case analysis.

Case (i): when $\eta \leq \|f - f'\|_2 \leq \frac{B-2}{4\sqrt{d}} \cdot \eta$. Recall Definition 4.7 that Σ is a *rotation matrix* scaled by a random factor $\beta \sim \text{Unif}[\frac{2\sqrt{d}}{B\eta}, \frac{4\sqrt{d}}{B\eta}]$. Given this, we have $\|\Sigma(f - f')\|_2 = \beta \cdot \|f - f'\|_2$. Further, since $\frac{2\sqrt{d}}{B\eta} \leq \beta \leq \frac{4\sqrt{d}}{B\eta}$ and $\eta \leq \|f - f'\|_2 \leq \frac{B-2}{4\sqrt{d}} \cdot \eta$, we have

$$\begin{aligned} \|\Sigma(f - f')\|_2 &\geq \frac{2\sqrt{d}}{B\eta} \cdot \eta = \frac{2\sqrt{d}}{B}, \\ \|\Sigma(f - f')\|_2 &\leq \frac{4\sqrt{d}}{B\eta} \cdot \frac{B-2}{4\sqrt{d}} \cdot \eta = 1 - \frac{2}{B}. \end{aligned}$$

Given these, we can easily see that $\|\Sigma(f' - f) - i\|_\infty \geq 2/B$ for any integer vector $i \in \mathbb{Z}^d$. Namely, the collision never occurs in this case.

Case (ii): when $\|f - f'\|_2 > \frac{B-2}{4\sqrt{d}} \cdot \eta$.

As for the case (ii), for simplicity, let $r \geq 1 - \frac{2\sqrt{d}}{B}$ denote $\|f - f'\|_2 \cdot (\frac{2\sqrt{d}}{B\eta})$, then we know $\|\Sigma(f - f')\|_2$ is distributed uniformly on $[r, 2r]$. Let $V_d(r) = \frac{\pi^{d/2}}{\Gamma(d/2+1)} \cdot (r)^d$ represent the volume of d -dimensional Euclidean ball of radius r and $S_d(r) = \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})} r^{d-1}$ represent the surface of area. The probability density function for $\Sigma(f - f') = x$ is $\text{PDF}(x) = \frac{1}{r \cdot S_d(\|x\|_2)}$ for $r \leq \|x\|_2 \leq 2r$. Then the probability for $\Sigma(f - f')$ falls into any region with volume V is at most

$$\frac{V}{r \cdot S_d(r)} \lesssim \frac{V \cdot \Gamma(d/2)}{(\sqrt{\pi}r)^d}.$$

Then we have

$$\begin{aligned} \Pr_{\Sigma,b} [h_{\Sigma,b}(f) = h_{\Sigma,b}(f')] &\leq (2 \cdot \lceil r + \frac{2\sqrt{d}}{B} \rceil)^d \cdot (2/B)^d \cdot \frac{\Gamma(d/2)}{(\sqrt{\pi}r)^d} \\ &\lesssim 2^{O(d)} / B^d \cdot \left(\frac{d}{2}\right)^{d/2} \\ &\leq 2^{O(d \log d)} \frac{1}{B}. \end{aligned}$$

where the first step is because that $\Sigma(f' - f)$ must locate in a hypercube that is centered at some integer vector and has edge length $2/B$, and there are at most $(2 \cdot \lceil r + \frac{2\sqrt{d}}{B} \rceil)^d$ different hypercubes in the sphere. The second step is by the Stirling's approximation.

This completes the proof of Lemma 4.18. \square

Lemma 4.19 (Probability of large offset). *Consider the random matrix $\Sigma \in \mathbb{R}^{d \times d}$ and the random vector $b \in \mathbb{R}^d$ given in Definition 4.7, for any tone frequency $f \in \text{supp}(\widehat{x^*})$, the probability of large offset*

$$\Pr_{\Sigma,b} [E_{\text{off}}(f)] = \Pr_{\Sigma,b} \left[\|\mathfrak{o}_{\Sigma,b}(f)\|_\infty \geq \frac{1-\alpha}{2B} \right] = 1 - (1-\alpha)^d \leq 0.01.$$

Proof. Recall (see Definition 4.10) that

$$h_{\Sigma,b}(f) = \lfloor B \cdot \text{frac}(\frac{1}{2B} \cdot \mathbf{1} + \Sigma(f - b)) \rfloor$$

and (see Definition 4.11) that

$$\sigma_{\Sigma,b}(f) = \text{frac}(\frac{1}{2B} \cdot \mathbf{1} + \Sigma(f - b)) - \frac{1}{B} \cdot h_{\Sigma,b}(f) - \frac{1}{2B} \cdot \mathbf{1} \in [-\frac{1}{2B}, \frac{1}{2B})^d.$$

It can be seen that a large offset $|\sigma_{\Sigma,b}(f)|_\infty \geq \frac{1-\alpha}{2B}$ occurs if and only if $\text{frac}(\frac{1}{2B} \cdot \mathbf{1} + \Sigma(f - b)) \notin S^d$, where (in each coordinate) the union of intervals

$$S = \bigcup_{j \in [B]} \left(\frac{j}{B} + \frac{\alpha/2}{B}, \frac{j}{B} + \frac{1-\alpha/2}{B} \right) \subseteq [0, 1).$$

Indeed, for any choice of Σ according to Definition 4.7, because $b \in \mathbb{R}^d$ is uniformly random, each i -th coordinate of the vector $(\frac{1}{2B} \cdot \mathbf{1} + \Sigma(f - b))$ is distributed uniformly on an interval of length $|\text{supp}(b_i)| = 1 \in \mathbb{N}_{\geq 1}$. Accordingly, each i -th coordinate of the fractional part $\text{frac}(\frac{1}{2B} \cdot \mathbf{1} + \Sigma(f - b))$ must be distributed independently and uniformly on $[0, 1)$.

To summarize, the conditional probability $\Pr_b[E_{\text{off}}(f) \mid \Sigma]$ always equals the probability that, an *coordinate-wise independent* uniform random vector $\tilde{b} \sim \text{Unif}[0, 1)^d$ locates outside the region $S^d \subseteq [0, 1)^d$. Thus, for any realized Σ given by Definition 4.7, we have

$$\begin{aligned} \Pr_b[E_{\text{off}}(f) \mid \Sigma] &= 1 - |S|^d \\ &= 1 - \left(\left(\frac{1-\alpha/2}{B} - \frac{\alpha/2}{B} \right) \cdot B \right)^d \\ &= 1 - (1 - \alpha)^d. \end{aligned}$$

Since $0 < \alpha \leq \frac{1}{100 \cdot (d+1)} < 1$ (see Definition 4.1), we also have $1 - (1 - \alpha)^d \leq d \cdot \alpha \leq \frac{d}{100 \cdot (d+1)} \leq 0.01$.

This completes the proof of Lemma 4.19. \square

Lemma 4.20 (Hashing into same bin). *For any tone frequency $f \in \text{supp}(\widehat{x^*})$, if the event $E_{\text{off}}(f)$ does not happen, then $h_{\Sigma,b}(f) = h_{\Sigma,b}(f')$ for any other frequency $f' \in \mathbb{R}^d$ that*

$$\|f - f'\|_2 < \frac{\alpha}{8\sqrt{d}} \cdot \eta = \Theta(d^{-1.5} \cdot \eta).$$

Proof. Since the event $E_{\text{off}}(f)$ does not happen, the offset $\|\sigma_{\Sigma,b}(f)\|_\infty < \frac{1-\alpha}{2B}$ (see Definitions 4.13 and 4.11). Given this, by construction (see Definition 4.10) a sufficient condition for the function $h_{\Sigma,b}$ to hash f and f' into the same bin is $\|\Sigma(f - f')\|_\infty < \frac{\alpha}{2B}$. We verify this condition as follows:

$$\begin{aligned} \|\Sigma(f - f')\|_\infty &\leq \|\Sigma(f - f')\|_2 \\ &\leq \frac{4\sqrt{d}}{B\eta} \cdot \|f - f'\|_2 \\ &< \frac{4\sqrt{d}}{B\eta} \cdot \frac{\alpha}{8\sqrt{d}} \cdot \eta \\ &= \frac{\alpha}{2B}, \end{aligned}$$

where the second step follows from Definition 4.7, i.e. Σ is a rotation matrix scaled by a random factor $\beta \sim \text{Unif}[\frac{2\sqrt{d}}{B\eta}, \frac{4\sqrt{d}}{B\eta}]$; and the third step follows from our premise $\|f - f'\|_2 < \frac{\alpha}{8\sqrt{d}} \cdot \eta$.

This completes the proof. \square

4.8 HASHTOBINS: error due to noise

Lemma 4.21 (The error due to noise). *Suppose that Condition 4.8 is true for the random vector $a \in \mathbb{R}^d$ and that $x^*(t) = 0$ for any $t \in [0, T]^d$, then the following holds for any random matrix $\Sigma \in \mathbb{R}^{d \times d}$ given in Definition 4.7:*

$$\mathbf{E}_{b,a} [\|\widehat{u}\|_2^2] \lesssim \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |g(\tau)|^2 \cdot d\tau.$$

Proof. We know from Parseval's theorem that $\|\widehat{u}\|_2^2 = \mathcal{B} \cdot \|u\|_2^2 = \mathcal{B} \cdot \sum_{j \in [B]^d} |u_j|^2$. To see the lemma, let us consider a specific coordinate $|u_j|^2$. By definition (see Line 5 of HASHTOBINS),

$$\begin{aligned} \mathbf{E}_{b,a} [|u_j|^2] &= \mathbf{E}_{b,a} \left[\left| \sum_{i \in [D]^d} y_{Bi+j} \right|^2 \right] \\ &= \mathbf{E}_{b,a} \left[\sum_{i \in [D]^d} y_{Bi+j} \cdot \sum_{i' \in [D]^d} \overline{y_{Bi'+j}} \right] \\ &= \sum_{i \in [D]^d} \mathbf{E}_{b,a} [y_{Bi+j} \cdot \overline{y_{Bi+j}}] + \sum_{i \neq i' \in [D]^d} \mathbf{E}_{b,a} [y_{Bi+j} \cdot \overline{y_{Bi'+j}}], \end{aligned} \quad (11)$$

where the second step follows as $|z|^2 = z\bar{z}$ for any complex number $z \in \mathbb{C}$; and the last step follows from the linearity of expectation.

As a premise of the current lemma, the signal $x(t) = x^*(t) + g(t) = g(t)$ for any $t \in [0, T]^d$. Due to Line 4 of HASHTOBINS, for any pair $i \in [D]^d$ and any $j \in [B]^d$ we have

$$\begin{aligned} y_{Bi+j} &= \mathcal{G}(Bi+j) \cdot \mathcal{P}_{\Sigma, b, a} g(Bi+j) \\ &= \mathcal{G}(Bi+j) \cdot g(\Sigma^\top (Bi+j+a)) \cdot e^{-2\pi i b^\top (Bi+j)} \\ &= S_{i,j} \cdot e^{-2\pi i b^\top (Bi+j)}, \end{aligned}$$

where the second step is by Definition 4.14; and in the last step we denote

$$S_{i,j} = \mathcal{G}(Bi+j) \cdot g(\Sigma^\top (Bi+j+a))$$

for ease of notation. Notice that $S_{i,j} \in \mathbb{R}$ is a real number and is determined by the random matrix $\Sigma \in \mathbb{R}^{d \times d}$ and the random vector $a \in \mathbb{R}^d$ (see Definition 4.7).

We can reformulate the first term in Equation (11) as follows: for each $i \in [D]^d$,

$$\mathbf{E}_{b,a} [y_{Bi+j} \cdot \overline{y_{Bi+j}}] = \mathbf{E}_a [S_{i,j} \cdot \overline{S_{i,j}}] = \mathbf{E}_a [S_{i,j}^2]$$

Indeed, the second term in Equation (11) equals zero. Particularly, for any $i \neq i' \in [D]^d$ we have

$$\begin{aligned} \mathbf{E}_{b,a} [y_{Bi+j} \cdot \overline{y_{Bi'+j}}] &= \mathbf{E}_{b,a} [S_{i,j} \cdot \overline{S_{i',j}} \cdot e^{-2\pi i b^\top (i-i')}] \\ &= \mathbf{E}_a [S_{i,j} \cdot \overline{S_{i',j}}] \cdot \mathbf{E}_{b,a} [e^{-2\pi i b^\top (i-i')}] = 0, \end{aligned} \quad (12)$$

where the second step follows since $b \in \mathbb{R}^d$ and $S_{i,j}$ are independent (see Definition 4.7).

Let us investigate the second term in Equation (12). We have $b^\top (i - i') = \sum_{r \in [d]} b_r \cdot (i_r - i'_r)$, in which at least one summand is non-zero (since $i \neq i' \in [D]^d$). Since both of B and $i_r - i'_r$ are

integers and each coordinate $b_r \sim \text{Unif}[0, 1]$ is independently, the fractional part $\text{frac}(B \cdot b^\top(i - i'))$ must follow the distribution $\text{Unif}[0, 1)$. Accordingly, the second term in Equation (12) is equal to

$$\mathbf{E}_b \left[e^{-2\pi i B \cdot b^\top(i - i')} \right] = 0,$$

Applying all of the above arguments to Equation (11) leads to $\mathbf{E}_{b,a}[|u_j|^2] = \sum_{i \in [D]^d} \mathbf{E}_a[S_{i,j}^2]$. Taking all vector indices $j \in [B]^d$ into account, we infer that

$$\begin{aligned} \mathbf{E}_{b,a} [\|\widehat{u}\|_2^2] &= \mathcal{B} \cdot \sum_{j \in [B]^d} \mathbf{E}_{b,a} [|u_j|^2] \\ &= \mathcal{B} \cdot \sum_{j \in [B]^d} \sum_{i \in [D]^d} \mathbf{E}_a [S_{i,j}^2] \\ &= \mathcal{B} \cdot \sum_{j \in [B]^d} \sum_{i \in [D]^d} \mathbf{E}_a \left[\mathcal{G}(Bi + j)^2 \cdot g(\Sigma^\top(Bi + j + a))^2 \right] \\ &= \mathcal{B} \cdot \sum_{i \in [BD]^d} \mathcal{G}(i)^2 \cdot \mathbf{E}_a \left[g(\Sigma^\top(i + a))^2 \right], \end{aligned} \quad (13)$$

where the third step is by the definition of $S_{i,j}$; and the last step is by substitution.

Due to Condition 4.8, given any $i \in [BD]^d$ and any choice of the random matrix $\Sigma \in \mathbb{R}^{d \times d}$ according to Definition 4.7, the random vector $a \in \mathbb{R}^d$ satisfies that

$$\mathbf{E}_a \left[g(\Sigma^\top(i + a))^2 \right] \lesssim \frac{1}{T^d} \cdot \int_{t \in [0, T]^d} |g(t)|^2 \cdot dt,$$

Plugging the above equation into Equation (13) gives

$$\begin{aligned} \mathbf{E}_{b,a} [\|\widehat{u}\|_2^2] &\lesssim \frac{1}{T^d} \cdot \int_{t \in [0, T]^d} |g(t)|^2 \cdot dt \cdot \left(\mathcal{B} \cdot \sum_{i \in [BD]^d} \mathcal{G}(i)^2 \right) \\ &\lesssim \frac{1}{T^d} \cdot \int_{t \in [0, T]^d} |g(t)|^2 \cdot dt, \end{aligned}$$

where the last step follows from Property V of Lemma 4.3 that $\sum_{i \in \mathbb{Z}^d} \mathcal{G}(i)^2 \leq e^2 \cdot \mathcal{B}^{-1}$.

This completes the proof of Lemma 4.21. \square

4.9 HASHTOBINS: error due to bad events

Lemma 4.22 (The error due to bad events). *Suppose $g(t) = 0$ for any $t \in [0, T]^d$. Given the hash function $h_{\Sigma,b}$ under any $\Sigma \in \mathbb{R}^{d \times d}$ and any $b \in \mathbb{R}^d$ (according to Definition 4.7), denote by*

$$H = \{f \in \text{supp}(\widehat{x^*}) : \text{neither } E_{\text{coll}}(f) \text{ nor } E_{\text{off}}(f) \text{ happens}\}$$

the set of “good” tone frequencies. Then the following hold:

$$\begin{aligned} \forall f \in H : \mathbf{E}_a \left[\left| \widehat{u}_{h_{\Sigma,b}(f)} - \widehat{x^*}[f] \cdot e^{2\pi i a^\top \Sigma f} \right|^2 \right] &\leq \frac{\delta}{\text{poly}(k, d)} \cdot \|\widehat{x^*}\|_1^2, \\ \sum_{f \in H} \mathbf{E}_a \left[\left| \widehat{u}_{h_{\Sigma,b}(f)} - \widehat{x^*}[f] \cdot e^{2\pi i a^\top \Sigma f} \right|^2 \right] &\leq \frac{\delta}{\text{poly}(k, d)} \cdot \|\widehat{x^*}\|_1^2, \end{aligned}$$

Proof. As promised by the lemma, the signal $x(t) = x^*(t) + g(t) = x^*(t)$ for any $t \in [0, T]^d$. For a specific frequency $f' \in H$, w.l.o.g. we assume that f' is hashed into the j -th bin, namely $h_{\Sigma, b}(f') = j$ (see Definition 4.10). According to Property II of Fact 4.16,

$$\begin{aligned}\widehat{u}_j &= \widehat{\mathcal{G}} * \widehat{\mathcal{P}_{\Sigma, b, a} x^*}(B^{-1} \cdot j) \\ &= \widehat{\mathcal{G}}' * \widehat{\mathcal{P}_{\Sigma, b, a} x^*}(B^{-1} \cdot j) + (\widehat{\mathcal{G}} - \widehat{\mathcal{G}}') * \widehat{\mathcal{P}_{\Sigma, b, a} x^*}(B^{-1} \cdot j).\end{aligned}\quad (14)$$

For the second summand in Equation (14), the corresponding function admits the ℓ_∞ norm of

$$\|(\widehat{\mathcal{G}} - \widehat{\mathcal{G}}') * \widehat{\mathcal{P}_{\Sigma, b, a} x^*}\|_\infty \leq \|(\widehat{\mathcal{G}} - \widehat{\mathcal{G}}')\|_\infty \cdot \|\widehat{\mathcal{P}_{\Sigma, b, a} x^*}\|_1 \leq \frac{\delta}{\text{poly}(k, d)} \cdot \|\widehat{\mathcal{P}_{\Sigma, b, a} x^*}\|_1$$

where the second step is due to Property IV of Lemma 4.6.

We then have

$$\begin{aligned}\|\widehat{\mathcal{P}_{\Sigma, b, a} x^*}\|_1 &= \int_{z \in \mathbb{R}^d} \left| \widehat{x^*}(\Sigma^{-1}z + b) \cdot \det(\Sigma)^{-1} \cdot e^{2\pi i \cdot a^\top (z + \Sigma b)} \right| \cdot dz \\ &= \int_{z \in \mathbb{R}^d} \left| \widehat{x^*}(\Sigma^{-1}z + b) \right| \cdot \det(\Sigma)^{-1} \cdot dz \\ &= \int_{z \in \mathbb{R}^d} |\widehat{x^*}(\xi)| \cdot d\xi \\ &= \|\widehat{x^*}\|_1,\end{aligned}$$

where the first step applies Property II of Lemma 4.15; the second step follows since $|e^{i\theta}| = 1$ for any $\theta \in \mathbb{R}$ and $\det(\Sigma) \neq 0$ (Definition 4.7); and the third step is by substitution.

Putting the equations together, we get

$$\|(\widehat{\mathcal{G}} - \widehat{\mathcal{G}}') * \widehat{\mathcal{P}_{\Sigma, b, a} x^*}\|_\infty \leq \frac{\delta}{\text{poly}(k, d)} \cdot \|\widehat{x^*}\|_1. \quad (15)$$

Moreover, the first summand in Equation (14) equals

$$\begin{aligned}\widehat{\mathcal{G}}' * \widehat{\mathcal{P}_{\Sigma, b, a} x^*}(B^{-1} \cdot j) &= \int_{\xi \in \mathbb{R}^d} \widehat{\mathcal{G}}'(B^{-1} \cdot j - \xi) \cdot \widehat{\mathcal{P}_{\Sigma, b, a} x^*}(\xi) \cdot d\xi \\ &= \int_{\xi \in \mathbb{R}^d} \widehat{\mathcal{G}}'(B^{-1} \cdot j - \xi) \cdot \widehat{x^*}(\Sigma^{-1}\xi + b) \cdot \det(\Sigma)^{-1} \cdot e^{2\pi i \cdot a^\top (\xi + \Sigma b)} \cdot d\xi \\ &= \int_{\xi \in \mathbb{R}^d} \widehat{\mathcal{G}}'(B^{-1} \cdot j - \Sigma(\xi - b)) \cdot \widehat{x^*}(\xi) \cdot e^{2\pi i \cdot a^\top \Sigma \xi} \cdot d\xi,\end{aligned}\quad (16)$$

where the first step applies the convolution operation; the second step follows from Property II of Lemma 4.15; and the third step is by substitution.

Notably (see Properties I to III of Lemma 4.6 and Definition 4.2), the standard window $\widehat{\mathcal{G}}'(\xi)$ is supported within the hypercube grid

$$\Lambda_W\left(\frac{1}{2B}\right) = \left\{ \xi \in \mathbb{R}^d : \|\xi - i\|_\infty \leq \frac{1}{2B} \text{ for some vector index } i \in [-W : W]^d \right\}.$$

Recall that Σ is a random rotation matrix scaled by a random factor $\beta \sim \text{Unif}[\frac{2\sqrt{d}}{B\eta}, \frac{4\sqrt{d}}{B\eta}]$ (see Definition 4.7). Further, the Fourier spectrum $\text{supp}(\widehat{x^*}) \subseteq [-F, F]^d$ is bounded. Under any choice

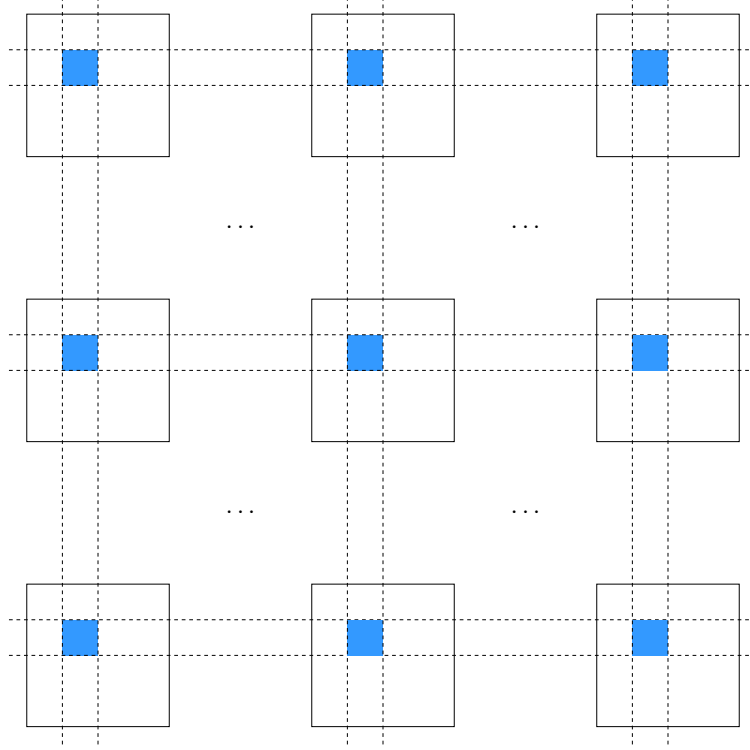


Figure 8: Demonstration for Lemma 4.22 in two dimensions $d = 2$, where the blue subsquares together denote the region $\{z \in \mathbb{R}^2 : |B^{-1} \cdot j - z - i|_\infty \leq \frac{1}{2B} \text{ for some } i \in \mathbb{Z}^2\}$.

of the random matrix Σ of the random vector $b \in \mathbb{R}^d$ (see Definition 4.7), for any $j \in [B]^d$ and any $\xi \in [-F, F]^d$ we have

$$\begin{aligned}
\|B^{-1} \cdot j - \Sigma(\xi - b)\|_\infty &\lesssim \|\Sigma(F, F, \dots, F)^\top\|_\infty \\
&\leq \|\Sigma(F, F, \dots, F)^\top\|_2 \\
&\leq \frac{4\sqrt{d}}{B\eta} \cdot \|(F, F, \dots, F)^\top\|_2 \\
&\lesssim d \cdot \frac{F}{B\eta}.
\end{aligned}$$

Thus, a sufficiently large width parameter $W = \Theta(d \cdot \frac{F}{B\eta})$ (see Definition 4.1) guarantees that

$$\left\{ \frac{1}{B} \cdot j - \Sigma(\xi - b) : \xi \in [-F, F]^d \right\} \subseteq [-W, W]^d,$$

for any choice of Σ and b (according to Definition 4.7) and any $j \in [B]^d$.

Given the above arguments and as Figure 8 suggests, Equation (16) suffices to integrate the tone frequencies hashed into the j -th bin, namely

$$\begin{aligned}
\Phi_j &= \left\{ \xi \in \text{supp}(\widehat{x^*}) : |B^{-1} \cdot j - \Sigma(\xi - b) - i|_\infty \leq \frac{1}{2B} \text{ for some } i \in \mathbb{Z}^d \right\} \\
&= \left\{ \xi \in \text{supp}(\widehat{x^*}) : -\frac{1}{2B} \cdot \mathbf{1} \preceq B^{-1} \cdot j - \Sigma(\xi - b) - i \preceq \frac{1}{2B} \cdot \mathbf{1} \text{ for some } i \in \mathbb{Z}^d \right\}
\end{aligned}$$

$$\begin{aligned}
&= \left\{ \xi \in \text{supp}(\widehat{x^*}) : \frac{1}{B} \cdot j \preceq \frac{1}{2B} \cdot \mathbf{1} + \Sigma(\xi - b) + i \preceq \frac{1}{B} \cdot (j + \mathbf{1}) \text{ for some } i \in \mathbb{Z}^d \right\} \\
&= \left\{ \xi \in \text{supp}(\widehat{x^*}) : j \preceq B \cdot \left(\frac{1}{2B} \cdot \mathbf{1} + \Sigma(\xi - b) + i \right) \preceq j + \mathbf{1} \text{ for some } i \in \mathbb{Z}^d \right\}.
\end{aligned}$$

In the above condition, $\frac{1}{2B} \cdot \mathbf{1} + \Sigma(\xi - b) + i$ must be bounded within $[0, 1]^d$, as the concerning bin $j \in [B]^d = \{0, 1, \dots, B-1\}^d$. In particular, the case $\|\frac{1}{2B} \cdot \mathbf{1} + \Sigma(\xi - b) + i\|_\infty = 1$ occurs with zero probability, since $b \in \mathbb{R}^d$ follows a *continuous* uniform distribution (Definition 4.7); we safely ignore this case. Given the hash function $h_{\Sigma, b}(\xi) \in [B]^d$ in Definition 4.10, we conclude that

$$\Phi_j = \left\{ \xi \in \text{supp}(\widehat{x^*}) : h_{\Sigma, b}(\xi) = j \right\}.$$

The concerning tone frequency $f' \in H$ ensures that neither $E_{\text{coll}}(f)$ nor $E_{\text{off}}(f)$ happens:

- $E_{\text{coll}}(f')$ does not happen. No other tone frequencies $f \in \text{supp}(\widehat{x^*}) \setminus \{f'\}$ collide with f' after the hashing. That is, the j -th bin contains f' as the only tone frequency, namely $\Phi_j = \{f'\}$.
- $E_{\text{off}}(f')$ does not happen. The offset $\|\sigma_{\Sigma, b}(f')\|_\infty < \frac{1-\alpha}{2B}$ is small enough, namely the frequency f' lies within the hypercube grid $\Lambda_W(\frac{1-\alpha}{2B})$. We know from Property I of Lemma B.11 that

$$\widehat{\mathcal{G}}'(B^{-1} \cdot j - \Sigma(f' - b)) = 1$$

Also, recall Observation 2.1 that $\widehat{x^*}(\xi)$ is the combination of k many scaled d -dimensional Dirac delta functions (at the tone frequencies $\xi \in \text{supp}(\widehat{x^*})$). In precise, for any frequency $\xi \in [-F, F]^d$,

$$\widehat{x^*}(\xi) = \sum_{f \in \text{supp}(\widehat{x^*})} \widehat{x^*}[f] \cdot \text{Delta}_{=f}(\xi).$$

Applying all of the above arguments to Equation (16) results in

$$\begin{aligned}
\widehat{\mathcal{G}}' * \widehat{\mathcal{P}_{\Sigma, b, a} x^*}(j/B) &= \int_{\xi \in \mathbb{R}^d} \widehat{x^*}[f'] \cdot \text{Delta}_{=f'}(\xi) \cdot e^{2\pi i \cdot a^\top \Sigma \xi} \cdot d\xi \\
&= \widehat{x^*}[f'] \cdot e^{2\pi i \cdot a^\top \Sigma f'}.
\end{aligned}$$

This equation, together with Equation (14) and Equation (15), implies that

$$\left| \widehat{u}_j - \widehat{x^*}[f'] \cdot e^{2\pi i \cdot a^\top \Sigma f'} \right| \leq \frac{\delta}{\text{poly}(k, d)} \cdot \|\widehat{x^*}\|_1$$

Taking square on the both sides:

$$\left| \widehat{u}_j - \widehat{x^*}[f'] \cdot e^{2\pi i \cdot a^\top \Sigma f'} \right|^2 \leq \frac{\delta^2}{\text{poly}(k, d)} \cdot \|\widehat{x^*}\|_1^2 \leq \frac{\delta}{\text{poly}(k, d)} \cdot \|\widehat{x^*}\|_1^2,$$

where the last step is by $0 < \delta < 1$; note that $j = h_{\Sigma, b}(f') \in [B]^d$.

Finally, we note that $|H| \leq k$, since $H \subseteq \text{supp}(\widehat{x^*})$ and there are just k many tone frequencies $f' \in \text{supp}(\widehat{x^*})$. Apply the last inequality over all $f' \in H$ and take the expectation over the random vector $a \in \mathbb{R}^d$ (see Definition 4.7), then Lemma 4.22 follows. \square

4.10 Performance guarantees

Lemma 4.23 (Performance guarantee for HASHTOBINS). *Recall Theorem 1.1 for the ℓ_2 -norm noise level*

$$\mathcal{N}^2 := \underbrace{\frac{1}{T^d} \cdot \int_{t \in [0, T]^d} |g(t)|^2 \cdot dt}_{\mathcal{N}_g^2} + \delta \cdot \underbrace{\sum_{i \in [k]} |\widehat{x^*}[f_i]|^2}_{\mathcal{N}_v^2}.$$

Sample the matrix $\Sigma \in \mathbb{R}^{d \times d}$ and the vectors $b \in \mathbb{R}^d$ according to Definition 4.7, and suppose that Conditions 4.8 and 4.9 hold for the random vector $a \in \mathbb{R}^d$. Consider the “good” frequencies

$$H := \{f \in \text{supp}(\widehat{x^*}) : \text{neither } E_{\text{coll}}(\xi) \text{ nor } E_{\text{off}}(\xi) \text{ happens}\},$$

and the bins $I := [\mathcal{B}] \setminus \mathcal{H}_{\Sigma, b}(\text{supp}(\widehat{x^*}))$ with no frequency $\{f_i\}_{i \in [k]} = \text{supp}(\widehat{x^*})$ hashed into. Then given any $\Sigma \in \mathbb{R}^{d \times d}$, the following holds for each good frequency $f \in H$:

$$\mathbf{E}_{b, a} \left[\left| \widehat{u}_{\mathcal{H}_{\Sigma, b}(f)} - \widehat{x}[f] \cdot e^{2\pi i a^\top \Sigma f} \right|^2 \right] \lesssim \mathcal{B}^{-1} \cdot \mathcal{N}_g^2 + k^{-1} \cdot \frac{\delta}{\text{poly}(k, d)} \cdot \mathcal{N}_v^2$$

And take all good frequencies $f \in H$ and all bins $j \in I$ into account:

$$\mathbf{E}_{b, a} \left[\sum_{f \in H} \left| \widehat{u}_{\mathcal{H}_{\Sigma, b}(f)} - \widehat{x^*}[f] \cdot e^{2\pi i a^\top \Sigma f} \right|^2 + \sum_{j \in I} |\widehat{u}_j|^2 \right] \lesssim \mathcal{N}^2$$

Proof. This can be easily seen by combining Lemmas 4.21 and 4.22. □

5 Locate inner

Statement	Section	Algorithm	Comment
Definitions 5.1 and 5.2	Section 5.1	Algorithm 2	Definitions
Lemma 5.4	Section 5.2	Algorithm 2	Sample complexity and running time
Lemma 5.5	Section 5.3	Algorithm 2	Voting process
Lemma 5.10	Section 5.4	Algorithm 2	Election process
Lemma 5.13	Section 5.5	Algorithm 2	Guarantees
Lemmas 5.14 and 5.15	Section 5.6	Algorithm 3	Sampling scheme
Lemma 5.16	Section 5.7	Algorithm 4	Stronger guarantees

Table 2: List of Lemmas/Algorithms in locate inner section.

5.1 Definitions and algorithm

Definition 5.1 (Setup for LOCATEINNER). We adopt the following notations:

- The guessed approximation ratio $\mathcal{C} \in [120, \rho]$.
- Let $M = 4 \cdot \lceil 4\sqrt{d} \cdot \mathcal{C}^{2/3} \rceil \in \mathbb{N}_{\geq 1}$.
- Let $\varpi = \mathcal{C}^{-2/3}$; this parameter will be used in the voting scheme (see Definition 5.3).

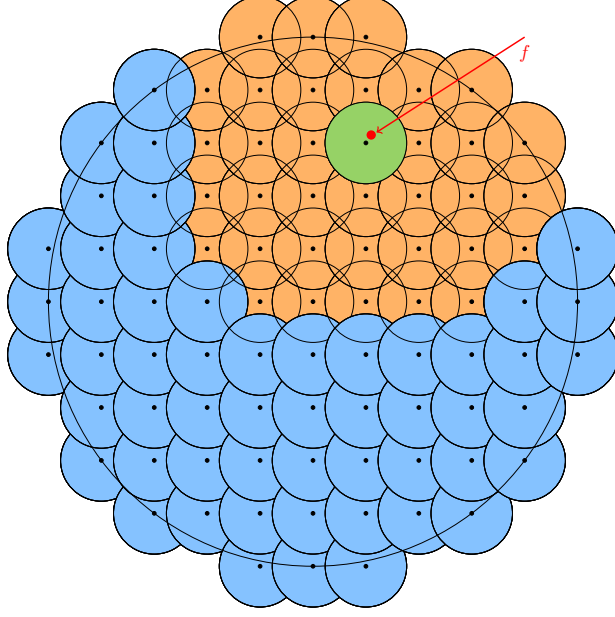


Figure 9: Demonstration for Definition 5.2 when $d = 2$. The “red” point means the true frequency f . The “lime” region means the *true* sub-hyperball $\mathbf{HB}(f_{q^*}^{\text{grid}}[j], \frac{1}{M} \cdot L^{\text{dia}}) \ni f$. The “blue” regions mean the *wrong* sub-hyperballs, and the remaining “yellow” regions (with overlapping parts) represent the *intermediate* sub-hyperballs.

- $\varphi_j = \arg(\hat{u}_j) - \arg(\hat{u}'_j)$ denotes the phase difference between \hat{u}_j and \hat{u}'_j ;
- The number of iterations $\mathcal{R}_{\text{vote}} = \Theta(d \cdot \log(\mathcal{C} \cdot d) + \log \log(F/\eta))$ is sufficiently large.

Definition 5.2 (Hyperball and sub-hyperballs). For any frequency $\text{List}[j] \in \mathbb{R}^d$ and any $L^{\text{dia}} \geq 0$, $\mathbf{HB}(\text{List}[j], L^{\text{dia}})$ denotes the ℓ_2 -norm hyperball with center $\text{List}[j]$ and diameter L^{dia} :

$$\mathbf{HB}(\text{List}[j], L^{\text{dia}}) := \left\{ \xi \in \mathbb{R}^d : \|\xi - \text{List}[j]\|_2 \leq L^{\text{dia}}/2 \right\}.$$

Let $\bigcup_{q \in Q} \mathbf{HB}(f_q^{\text{grid}}[j], \frac{1}{M} \cdot L^{\text{dia}})$ denote a cover of $\mathbf{HB}(\text{List}[j], L^{\text{dia}})$, by using a minimum amount of sub-hyperballs that have the diameter $\frac{1}{M} \cdot L^{\text{dia}}$ each. At most $\mathcal{M} := |Q| = (4M \cdot \sqrt{d})^d = 2^{\Theta(d \cdot \log(\mathcal{C} \cdot d))}$ many sub-hyperballs can be used, namely the *external covering number* [SSBD14, Page 337].

Given that the hyperball $\mathbf{HB}(\text{List}[j], L^{\text{dia}})$ contains the targeted tone frequency $f \in [-F, F]^d$, all the sub-hyperballs can be classified into three groups (as Figure 9 shows):

- The *true* sub-hyperball $\mathbf{HB}(f_{q^*}^{\text{grid}}[j], \frac{1}{M} \cdot L^{\text{dia}}) \ni f$, for some index $q^* \in Q$. For convenience, assume that the true sub-hyperball is unique, namely the targeted tone frequency is not on the boundary of two or more sub-hyperballs.¹³
- The *wrong* sub-hyperballs $q \in Q \setminus \{q^*\}$ have the ℓ_2 -distances

$$\left\| f_{q^*}^{\text{grid}}[j] - f_q^{\text{grid}}[j] \right\|_2 \geq \frac{1}{M} \cdot L^{\text{dia}} \cdot \lceil 4\sqrt{d}/\varpi \rceil = \frac{1}{M} \cdot L^{\text{dia}} \cdot \lceil 4\sqrt{d} \cdot \mathcal{C}^{2/3} \rceil.$$

- The remaining sub-hyperballs are called the *intermediate* sub-hyperballs.

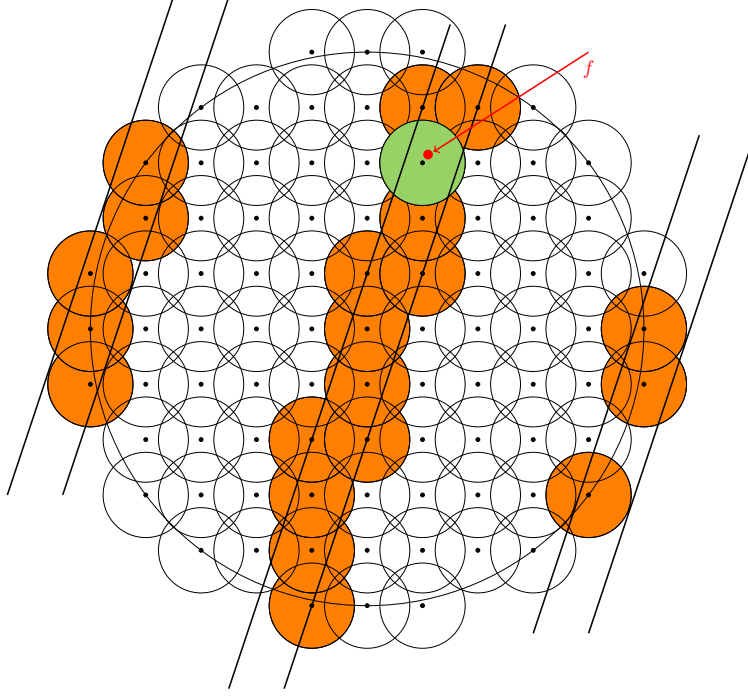


Figure 10: Demonstration for the voting scheme (Definition 5.3)

Definition 5.3 (Voting scheme). Let $\|\theta\|_{\bigcirc} \in [0, \pi]$ denote the “phase distance” from $e^{i0} = 1$ to any $\theta \in \mathbb{R}$. As Figure 10 suggests, given any $\varphi_j \in \mathbb{R}$ and any $\Delta_a := \Delta_a^r$, let $\mathcal{V}ote_j[q] \leftarrow \mathcal{V}ote_j[q] + 1$ (namely adding a vote to any sub-hyperball $q \in \mathcal{Q}$) for which

$$\left\| \varphi_j - 2\pi \cdot \Delta_a^\top f_q^{grid[j]} \right\|_{\bigcirc} \leq \pi \cdot \varpi.$$

5.2 Sample complexity and running time

The goal of this section is to prove Lemma 5.4.

Lemma 5.4 (Sample complexity and running time of LOCATEINNER). *The procedure LOCATEINNER (Algorithm 2) has the following performance guarantees:*

- The sample complexity is

$$\Theta(\mathcal{R}_{\text{vote}} \cdot \mathcal{BD}) = 2^{\Theta(d \cdot \log d)} \cdot (\log \mathcal{C} + \log \log(F/\eta)) \cdot k \cdot \mathcal{D}.$$

- The running time is

$$\Theta(\mathcal{R}_{\text{vote}} \cdot (\mathcal{BD} + \mathcal{B} \log \mathcal{B} + \mathcal{BM})) = 2^{\Theta(d \cdot (\log d + \log \mathcal{C}))} \cdot \log \log(F/\eta) \cdot k \cdot (\mathcal{D} + \log k)$$

Proof. Throughout the procedure LOCATEINNER, the subroutine HASHTOBINS (Algorithm 1) is invoked $2 \cdot \mathcal{R}_{\text{vote}}$ times. Recall Definition 5.1 that $\mathcal{R}_{\text{vote}} = \Theta(d \cdot \log(\mathcal{C} \cdot d) + \log \log(F/\eta))$.

¹³We make this assumption just to specify the true sub-hyperball; our proof does not rely on the assumption.

Algorithm 2 LOCATEINNER, Lemmas 5.4, 5.5, 5.10, 5.13

```

1: procedure LOCATEINNER( $\Sigma, b, D, \mathcal{L}ist, L^{dia}, \mathcal{C}, T$ )
2:   Define  $M \in \mathbb{N}_{\geq 1}$  according to Definition 5.1. ▷ Definition 5.2
3:   for  $j \in [B]^d$  that  $\mathcal{L}ist[j] \neq \text{NIL}$  do
4:     Cover  $\mathbf{HB}(\mathcal{L}ist[j], L^{dia})$  via sub-hyperballs  $\bigcup_{q \in Q} \mathbf{HB}(f_q^{grid[j]}, \frac{1}{M} \cdot L^{dia})$ .
5:   end for

6:   Initialize  $\mathcal{V}ote_j[q] = 0$  for all  $q \in Q$  and each  $j \in [B]^d$ . ▷ Voting process. Lemma 5.5
7:   Define  $\mathcal{R}_{\text{vote}} \in \mathbb{N}_{\geq 1}$  according to Definition 5.1.
8:   for  $r = 1, 2, \dots, \mathcal{R}_{\text{vote}}$  do
9:      $(a_r, \Delta_a^r) \leftarrow \text{SAMPLETIMEPOINT}(M, L^{dia}, \mathcal{C}, T)$ . ▷ Algorithm 3
10:     $\hat{u} \leftarrow \text{HASHTOBINS}(x, \Sigma, b, a_r, D)$ .
11:     $\hat{u}' \leftarrow \text{HASHTOBINS}(x, \Sigma, b, a_r + \Delta_a^r, D)$ .
12:    Let  $\varphi = (\varphi_j)_{j=1}^B$ , where  $\varphi_j = \arg(\hat{u}_j) - \arg(\hat{u}'_j)$ .
13:    for  $j \in [B]^d$  that  $\mathcal{L}ist[j] \neq \text{NIL}$  do
14:      Update  $\mathcal{V}ote_j[q]$  according to Definition 5.3.
15:    end for
16:  end for

17:  Initialize  $\mathcal{L}ist_{\text{new}}[j] = \text{NIL}$  for each  $j \in [B]^d$ . ▷ Election process. Lemma 5.10
18:  for  $j \in [B]^d$  that  $\mathcal{L}ist[j] \neq \text{NIL}$  do
19:     $\mathcal{W}inner[j] \leftarrow \bigcup_{q \in Q: \mathcal{V}ote_j[q] \geq \frac{1}{2} \cdot \mathcal{R}_{\text{vote}}} \mathbf{HB}(f_q^{grid[j]}, \frac{1}{M} \cdot L^{dia})$ .
20:    if  $\mathcal{W}inner[j]$  is non-empty then
21:      Let  $\mathcal{L}ist_{\text{new}}[j]$  be any frequency so that  $\mathbf{HB}(\mathcal{L}ist_{\text{new}}[j], \frac{1}{2} \cdot L^{dia}) \supseteq \mathcal{W}inner[j]$ .
22:    end if
23:  end for
24:  return the frequencies  $\mathcal{L}ist_{\text{new}}$ .
25: end procedure

```

Sample complexity. The procedure LOCATEINNER takes samples only by invoking the subroutine HASHTOBINS. Due to Fact 4.17, HASHTOBINS has the sample complexity $O(\mathcal{B}\mathcal{D})$. Recall Definition 4.1 that $\mathcal{B} = 2^{\Theta(d \cdot \log d)} \cdot k$. Thus, LOCATEINNER has the sample complexity

$$\begin{aligned}
\#\text{sample}(\text{LOCATEINNER}) &= \Theta(\mathcal{R}_{\text{vote}} \cdot \mathcal{B} \cdot \mathcal{D}) \\
&= \underbrace{\Theta(d \cdot \log(\mathcal{C} \cdot d) + \log \log(F/\eta))}_{\mathcal{R}_{\text{vote}}} \cdot \underbrace{2^{\Theta(d \cdot \log d)} \cdot k}_{\mathcal{B}} \cdot \mathcal{D} \\
&= 2^{\Theta(d \cdot \log d)} \cdot (\log \mathcal{C} + \log \log(F/\eta)) \cdot k \cdot \mathcal{D}.
\end{aligned}$$

Running time. The running time of LOCATEINNER is dominated by the $\mathcal{R}_{\text{vote}}$ many loops for the voting process (namely the second for loop). Such a loop invokes the subroutine HASHTOBINS twice, and then update $\mathcal{V}ote_j[q]$ for all $q \in Q$ and all $j \in [B]^d$. (The subroutine SAMPLETIMEPOINT runs in $O_d(1)$ time; see Algorithm 3.)

Due to Fact 4.17, HASHTOBINS has the running time $O(\mathcal{B}\mathcal{D} + \mathcal{B} \log \mathcal{B})$ time. Thus, the time that LOCATEINNER spends on hashing is

$$\#\text{time}(\text{hashing}) = \Theta(\mathcal{R}_{\text{vote}}) \cdot (\mathcal{B}\mathcal{D} + \mathcal{B} \log \mathcal{B})$$

$$\begin{aligned}
&= \underbrace{\Theta(d \cdot \log(\mathcal{C} \cdot d) + \log \log(F/\eta))}_{\mathcal{R}_{\text{vote}}} \cdot \underbrace{2^{\Theta(d \cdot \log d)} \cdot k \cdot (\mathcal{D} + \log k)}_{\mathcal{B}\mathcal{D} + \mathcal{B} \log \mathcal{B}} \\
&= 2^{\Theta(d \cdot \log d)} \cdot (\log \mathcal{C} + \log \log(F/\eta)) \cdot k \cdot (\mathcal{D} + \log k)
\end{aligned}$$

Further, the time that LOCATEINNER spends on voting is

$$\begin{aligned}
\#\text{time}(\text{voting}) &= \Theta(\mathcal{R}_{\text{vote}} \cdot \mathcal{B} \cdot \mathcal{M}) \\
&= \underbrace{\Theta(d \cdot \log(\mathcal{C} \cdot d) + \log \log(F/\eta))}_{\mathcal{R}_{\text{vote}}} \cdot \underbrace{2^{\Theta(d \cdot \log d)} \cdot k}_{\mathcal{B}} \cdot \underbrace{2^{\Theta(d \cdot \log(\mathcal{C} \cdot d))}}_{\mathcal{M}} \\
&= 2^{\Theta(d \cdot \log(\mathcal{C} \cdot d))} \cdot k \cdot \log \log(F/\eta).
\end{aligned}$$

In total, the procedure LOCATEINNER has the running time

$$\begin{aligned}
\#\text{time}(\text{LOCATEINNER}) &= \#\text{time}(\text{hashing}) + \#\text{time}(\text{voting}) \\
&= 2^{\Theta(d \cdot \log(\mathcal{C} \cdot d))} \cdot k \cdot (\mathcal{D} + \log k) \cdot \log \log(F/\eta).
\end{aligned}$$

This completes the proof. \square

5.3 Voting process

The goal of this section is to prove Lemma 5.5.

Lemma 5.5 (The voting process of LOCATEINNER). *Given any realized matrix Σ and any realized vector b , assume three premises for a particular good tone frequency $f \in H = \{\xi \in \text{supp}(\widehat{x}^*) : \text{neither } E_{\text{coll}}(\xi) \text{ nor } E_{\text{off}}(\xi) \text{ happens}\}$:*

- *W.l.o.g. the tone frequency $f \in H$ is hashed into the bin $h_{\Sigma, b}(f) = j \in [B]^d$ (Definition 4.10).*
- *The tone frequency $f \in H$ locates within the hyperball $\mathbf{HB}(\text{List}[j], L^{\text{dia}})$.*
- *Given the guessed approximation ratio $\mathcal{C} \in [120, \rho]$, the following holds for both $a = a_r$ and $a = a_r + \Delta_a^r$, in every single iteration $r \in [\mathcal{R}_{\text{vote}}]$ of the procedure LOCATEINNER:*

$$\mathbf{E}_a \left[\left| \widehat{u}_j - \widehat{x}[f] \cdot e^{2\pi i \cdot a^\top \Sigma f} \right|^2 \right] \leq \mathcal{C}^{-2} \cdot |\widehat{x}[f]|^2.$$

Then the following hold in every single iteration of procedure LOCATEINNER (Algorithm 2):

Property I: *The (unique) true sub-hyperball gets a vote with probability at least*

$$1 - \frac{4}{(\mathcal{C} \cdot \varpi)^2} = 1 - \frac{4}{\mathcal{C}^{2/3}} > \frac{1}{2}.$$

Property II: *Any wrong sub-hyperball gets a vote with probability at most*

$$8\varpi + \frac{4}{(\mathcal{C} \cdot \varpi)^2} = \frac{12}{\mathcal{C}^{2/3}} < \frac{1}{2}.$$

Claim 5.6 (Property I of Lemma 5.5). *The (unique) true sub-hyperball gets a vote with probability at least*

$$1 - \frac{4}{(\mathcal{C} \cdot \varpi)^2} = 1 - \frac{4}{\mathcal{C}^{2/3}} > \frac{1}{2}.$$

Proof. For brevity, we rewrite a_r and Δ_a^r respectively as a and Δ_a in this proof. Note that all of the probabilities and the expectations given below are taken over the random vectors a and Δ_a .

Combining the second premise of the lemma and Chebyshev's inequality together, we know that the following holds with probability at least $1 - \frac{2}{(\mathcal{C} \cdot \varpi)^2}$:

$$\left| \hat{u}_j - \hat{x}[f] \cdot e^{2\pi i \cdot a^\top \Sigma f} \right| \leq (\varpi/\sqrt{2}) \cdot |\hat{x}[f]|,$$

which is equivalent to

$$\left| \hat{u}_j / \hat{x}[f] \cdot e^{-2\pi i \cdot a^\top \Sigma f} - 1 \right| \leq \varpi/\sqrt{2}.$$

I.e., the complex number $\hat{u}_j / \hat{x}[f] \cdot e^{-2\pi i \cdot a^\top \Sigma f}$ lies in the circle $\{z \in \mathbb{C} : |z - 1| \leq \varpi/\sqrt{2}\}$. Clearly, any complex number in this circle has the phase $\leq \sin^{-1}(\varpi/\sqrt{2})$. In particular,

$$\left\| \underbrace{\arg(\hat{u}_j) - \arg(\hat{x}[f]) - 2\pi \cdot a^\top \Sigma f}_{A_1} \right\|_{\bigcirc} \leq \sin^{-1}(\varpi/\sqrt{2}),$$

where $\|\theta\|_{\bigcirc} \in [-\pi, \pi)$ denotes the “phase distance” from $e^{i0} = 1$ to any $\theta \in \mathbb{R}$.

Similarly, when a is replaced with $(a + \Delta_a)$, with probability $1 - \frac{2}{(\mathcal{C} \cdot \varpi)^2}$ we also have

$$\left\| \underbrace{\arg(\hat{u}'_j) - \arg(\hat{x}[f]) - 2\pi \cdot (a + \Delta_a)^\top \Sigma f}_{A_2} \right\|_{\bigcirc} \leq \sin^{-1}(\varpi/\sqrt{2}),$$

Put the above two inequalities together, (by the union bound) the following holds for the phase difference $\varphi_j = \arg(\hat{u}_j) - \arg(\hat{u}'_j)$ with probability $1 - \frac{4}{(\mathcal{C} \cdot \varpi)^2}$:

$$\begin{aligned} \left\| \varphi_j - 2\pi \cdot \Delta_a^\top \Sigma f \right\|_{\bigcirc} &= \|A_1 - A_2\|_{\bigcirc} \\ &\leq \|A_1\|_{\bigcirc} + \|A_2\|_{\bigcirc} \\ &\leq 2 \sin^{-1}(\varpi/\sqrt{2}) \\ &\leq (\pi/2) \cdot \varpi, \end{aligned} \tag{17}$$

where the second step applies the triangle inequality; and last step follows since for any $z \in (0, 1)$, we have $\sin^{-1}(z/\sqrt{2}) \leq (\pi/4) \cdot z$.

Let $q^* \in Q$ be the index of the true sub-hyperball $\mathbf{HB}(f_{q^*}^{grid[j]}, \frac{1}{M} \cdot L^{dia}) \ni f$. Compared with the center frequency $f_{q^*}^{grid[j]}$ of this sub-hyperball, the tone frequency f differs by has the ℓ_2 -distance

$$\left\| f - f_{q^*}^{grid[j]} \right\|_2 \leq \frac{1}{2M} \cdot L^{dia}.$$

Equation (17) suggests that the phase $\varphi_j = \arg(\hat{u}_j) - \arg(\hat{u}'_j) \in \mathbb{R}$ is likely to be a good approximation to $2\pi \cdot \Delta_a^\top \Sigma f$. Indeed, that inequality (if true) guarantees a vote for the true sub-hyperball $\mathbf{HB}(f_{q^*}^{grid[j]}, \frac{1}{M} \cdot L^{dia})$:

$$\begin{aligned} \left| 2\pi \cdot \Delta_a^\top \Sigma (f - f_{q^*}^{grid[j]}) \right| &\leq 2\pi \cdot \|\Sigma^\top \Delta_a\|_2 \cdot \left\| f - f_{q^*}^{grid[j]} \right\|_2 \\ &\leq 2\pi \cdot \|\Sigma^\top \Delta_a\|_2 \cdot \left(\frac{1}{2M} \cdot L^{dia} \right) \end{aligned}$$

$$\begin{aligned}
&\leq 2\pi \cdot \left(\frac{\varpi \cdot M}{2L^{dia}} \right) \cdot \left(\frac{1}{2M} \cdot L^{dia} \right) \\
&= (\pi/2) \cdot \varpi,
\end{aligned} \tag{18}$$

where the second step uses the ℓ_2 -distance derived above; and the third step follows because the vector Δ_a is sampled such that $\|\Sigma^\top \Delta_a\|_2 \sim \text{Unif}[\frac{\varpi \cdot M}{4L^{dia}}, \frac{\varpi \cdot M}{2L^{dia}}]$ (see Algorithm 3).

Combining everything together, with probability at least $1 - \frac{4}{(\mathcal{C} \cdot \varpi)^2}$ we have

$$\begin{aligned}
\left\| \varphi_j - 2\pi \cdot \Delta_a^\top \Sigma f_{q^*}^{grid[j]} \right\|_{\bigcirc} &= \left\| \varphi_j - 2\pi \cdot \Delta_a^\top \Sigma f \right\|_{\bigcirc} + \left| 2\pi \cdot \Delta_a^\top \Sigma (f - f_{q^*}^{grid[j]}) \right| \\
&\leq (\pi/2) \cdot \varpi + (\pi/2) \cdot \varpi \\
&= \pi \cdot \varpi,
\end{aligned}$$

where the first step follows from the triangle inequality; and the second step follows by applying inequalities (17) and (18).

Recall Definitions 5.1 and 5.3 that $\varpi = \mathcal{C}^{-2/3}$ and $\mathcal{C} \geq 120$. Via elementary calculation, it can be seen that

$$1 - \frac{4}{(\mathcal{C} \cdot \varpi)^2} = 1 - \frac{4}{\mathcal{C}^{2/3}} \geq 1 - \frac{4}{120^{2/3}} \approx 0.8356 > \frac{1}{2}.$$

This completes the proof of Claim 5.6. \square

Claim 5.7 (Property II of Lemma 5.5). *Any wrong sub-hyperball gets a vote with probability at most*

$$8\varpi + \frac{4}{(\mathcal{C} \cdot \varpi)^2} = \frac{12}{\mathcal{C}^{2/3}} < \frac{1}{2}.$$

Proof. Once again, we rewrite a_r and Δ_a^r respectively as a and Δ_a for simplicity, and all of the probabilities and the expectations in this proof are taken over the random vectors a and Δ_a .

Let $q^* \in Q$ be the index of the true sub-hyperball $\mathbf{HB}(f_{q^*}^{grid[j]}, \frac{1}{M} \cdot L^{dia}) \ni f$. For a specific wrong sub-hyperball $\mathbf{HB}(f_q^{grid[j]}, \frac{1}{M} \cdot L^{dia})$, where $q \in Q \setminus \{q^*\}$, the next inequality turns out to hold with probability at least $1 - 8\varpi$:

$$\left\| 2\pi \cdot \Delta_a^\top \Sigma (f_{q^*}^{grid[j]} - f_q^{grid[j]}) \right\|_{\bigcirc} \geq 2\pi \cdot \varpi. \tag{19}$$

We assume this fact for a while, and will justify this fact in the last part of this proof.

As shown in the proof of Claim 5.6, the following holds with probability at least $1 - 4 \cdot (\mathcal{C} \varpi)^{-2}$:

$$\left\| \varphi_j - 2\pi \cdot \Delta_a^\top \Sigma f_{q^*}^{grid[j]} \right\|_{\bigcirc} \leq \pi \cdot \varpi. \tag{20}$$

Conditioned on both Inequalities (19) and (20), we must have

$$\begin{aligned}
\left\| \varphi_j - 2\pi \cdot \Delta_a^\top \Sigma f_q^{grid[j]} \right\|_{\bigcirc} &= \left\| 2\pi \cdot \Delta_a^\top \Sigma (f_{q^*}^{grid[j]} - f_q^{grid[j]}) + (\varphi_j - 2\pi \cdot \Delta_a^\top \Sigma f_{q^*}^{grid[j]}) \right\|_{\bigcirc} \\
&\geq \pi \cdot \varpi.
\end{aligned}$$

Given this, we know from Definition 5.3 that the q -th (wrong) sub-hyperball is guaranteed to lose a vote. And based on the union bound, we derive Claim 5.7 as desired:

$$\Pr[q\text{-th sub-hyperball gets a vote}] \leq \Pr[\text{Equation (19) does not hold}]$$

$$\begin{aligned}
& + \Pr [\text{Equation (20) does not hold}] \\
& = 8\varpi + \frac{4}{(\mathcal{C}\varpi)^2} \\
& = \frac{12}{\mathcal{C}^{2/3}} \\
& \leq \frac{12}{120^{2/3}} \\
& \leq 0.4933 \\
& \leq \frac{1}{2}
\end{aligned}$$

where the third step follows because $\varpi = \mathcal{C}^{-2/3}$ (see Definition 5.3); and the fourth step follows because $\mathcal{C} \geq 120$ (see Definition 5.1).

To establish the claim, we are left to justify that Equation (19) holds with probability at least $1 - 8\varpi$. Indeed, an equivalent condition of Equation (19) is that

" $\Delta_a^\top \Sigma(f_{q^}^{grid[j]} - f_q^{grid[j]})$ differs from its closest integer $\lfloor \Delta_a^\top \Sigma(f_{q^*}^{grid[j]} - f_q^{grid[j]}) + \frac{1}{2} \rfloor$ by at least ϖ ",*

because $\|z\|_\circ \in [-\pi, \pi)$ denotes the "phase distance" from $0 = \arg(e^{i0})$ to any $z \in \mathbb{R}$.

We then observe that the vector Δ_a is sampled (see Algorithm 3) such that $\Sigma^\top \Delta_a$ has a uniform random direction, and the ℓ_2 -norm follows the uniform distribution

$$\|\Sigma^\top \Delta_a\|_2 \sim \text{Unif} \left[\frac{\varpi \cdot M}{4L^{dia}}, \frac{\varpi \cdot M}{2L^{dia}} \right].$$

Given these, we infer (e.g. from [CFJ13]) that $\Delta_a^\top \Sigma(f_{q^*}^{grid[j]} - f_q^{grid[j]})$ has the same distribution as the random variable $(\tilde{w} \cdot \cos \tilde{\theta})$, where

- $\tilde{\theta} \in [0, \pi]$ is the angle between the vectors $\Sigma^\top \Delta_a$ and $(f_{q^*}^{grid[j]} - f_q^{grid[j]})$. It is known that $\tilde{\theta}$ has the following probability density function: for all $\tilde{\theta} \in [0, \pi]$,

$$\text{PDF}(\tilde{\theta}) = \frac{\sin^{d-2}(\tilde{\theta})}{\int_0^\pi \sin^{d-2}(z) \cdot dz}.$$

- $\tilde{w} \sim \text{Unif}[w, 2w]$ with the parameter

$$\begin{aligned}
w &= \frac{\varpi \cdot M}{4L^{dia}} \cdot \left\| f_{q^*}^{grid[j]} - f_q^{grid[j]} \right\|_2 \\
&\geq \frac{\varpi \cdot M}{4L^{dia}} \cdot \frac{1}{M} \cdot L^{dia} \cdot \left\lceil \frac{4\sqrt{d}}{\varpi} \right\rceil \\
&\geq \sqrt{d},
\end{aligned}$$

where the second step follows from the definition of a *wrong* sub-hyperball (see Definition 5.2).

We conclude from the above that

$$\Pr_{\Delta_a} [\text{Equation (19) does not hold}] = \Pr_{\tilde{w}, \tilde{\theta}} \left[\left| \tilde{w} \cdot \cos \tilde{\theta} - \left\lfloor \tilde{w} \cdot \cos \tilde{\theta} + 1/2 \right\rfloor \right| \leq \varpi \right] \quad (21)$$

It turns out that $\varpi \leq 1/5$ and that $w \geq 1$. Concretely, we know from Definitions 5.1 and 5.3 that

$$\varpi = C^{-2/3} \leq 120^{-2/3} \approx 0.0411 < \frac{1}{5}.$$

Further, we have shown that the parameter $w \geq \sqrt{d}$. Given these, Claim 5.8 (presented below) is applicable to the RHS of Equation (21). By doing so, we accomplish Claim 5.7.

This completes the proof. \square

Claim 5.8 (Technical result for Claim 5.7). *Given any $u \in (d/2^d, 1/5]$ and any $w \geq \sqrt{d}$, it follows*

$$\Pr_{\tilde{w}, \tilde{\theta}} \left[\left| \tilde{w} \cdot \cos(\tilde{\theta}) - \left\lfloor \tilde{w} \cdot \cos(\tilde{\theta}) + 1/2 \right\rfloor \right| \leq u \right] \lesssim u,$$

where $\tilde{w} \sim \text{Unif}[w, 2w]$, and the random phase $\tilde{\theta} \in [0, \pi]$ has the probability density function

$$\text{PDF}(\tilde{\theta}) = \frac{\sin^{d-2}(\tilde{\theta})}{\int_0^\pi \sin^{d-2}(z) \cdot dz}, \quad \forall \tilde{\theta} \in [0, \pi].$$

Proof. Fix $\tilde{\theta}$ first. We know that

$$\begin{aligned} \int_0^\pi \sin^{d-2}(z) \cdot dz &= \pi \cdot \frac{(2(d-2)-1)!!}{(2(d-2))!!} \\ &\approx 1/d, \end{aligned}$$

where the first step is by induction and the second step follows from the Wallis formula.

We need some asymptotic evaluations:

- $|\sin(z)| \approx 1 - \frac{\cos^2 z}{2}$ when $|\cos(z)| \ll 1$.
- $\sin x \approx x$ when $|x| \ll 1$.

In the next a few paragraphs, we discuss the three cases for $|\cos(\tilde{\theta})|$.

- Case 1. $|\cos(\tilde{\theta})| \leq c/\sqrt{d}$
- Case 2. $c/\sqrt{d} \leq |\cos(\tilde{\theta})| \leq 1/2$
- Case 3. $|\cos(\tilde{\theta})| > 1/2$

Case 1. If $|\cos(\tilde{\theta})| \leq c/\sqrt{d}$.

First consider the range that $|\cos(\tilde{\theta})| \leq c/\sqrt{d}$ for some small constant $c > 1$. Then in this range we know that $|\sin z| \approx 1 - \frac{\cos^2(\tilde{\theta})}{2} \geq 1 - \frac{c^2}{d}$, which implies that $|\sin(\tilde{\theta})|^{d-2} = \Omega(1)$. In other word, we can treat $\tilde{\theta}$ as nearly uniform distributed in this range. By similar arguments in Claim 5.9, we can prove in this range

$$\Pr_{\tilde{\theta}} \left[\left| \tilde{w} \cdot \cos(\tilde{\theta}) - \left\lfloor \tilde{w} \cdot \cos(\tilde{\theta}) + 1/2 \right\rfloor \right| \leq u, |\cos(\tilde{\theta})| \leq c/\sqrt{d} \right] \lesssim u$$

Case 2. If $c/\sqrt{d} \leq |\cos(\tilde{\theta})| \leq 1/2$.

We fix an integer $i > c$.

We have that $|\sin(z)| \geq 1/2$. We can use a straight line to simulate \cos function. For any integer $i \geq c$, we have that

$$\cos^{-1}\left(\frac{i-u}{\tilde{w}}\right) - \cos^{-1}\left(\frac{i+u}{\tilde{w}}\right) \lesssim u \cdot \left(\cos^{-1}\left(\frac{i-1+u}{\tilde{w}}\right) - \cos^{-1}\left(\frac{i-u}{\tilde{w}}\right)\right).$$

We know that $\text{PDF}(\tilde{\theta})$ is increasing, then we have that

$$u \cdot \Pr_{\tilde{\theta}}[\tilde{\theta} \in [\cos^{-1}(\frac{i-u}{\tilde{w}}), \cos^{-1}(\frac{i-1+u}{\tilde{w}})]] \gtrsim \Pr_{\tilde{\theta}}[\tilde{\theta} \in [\cos^{-1}(\frac{i+u}{\tilde{w}}), \cos^{-1}(\frac{i-u}{\tilde{w}})]].$$

We have that

$$\Pr_{\tilde{\theta}} \left[\left| \tilde{w} \cdot \cos(\tilde{\theta}) - i \right| \leq u, \frac{1}{2} \geq |\cos(\tilde{\theta})| \geq c/\sqrt{d} \right] \lesssim u \cdot \Pr_{\tilde{\theta}} \left[\left| \tilde{w} \cdot \cos(\tilde{\theta}) - i \right| \leq 1, \frac{1}{2} \geq |\cos(\tilde{\theta})| \geq c/\sqrt{d} \right].$$

Combine this together, we know that

$$\Pr_{\tilde{\theta}} \left[\left| \tilde{w} \cdot \cos(\tilde{\theta}) - \left\lfloor \tilde{w} \cdot \cos(\tilde{\theta}) + 1/2 \right\rfloor \right| \leq u, 1/2 \geq |\cos(\tilde{\theta})| \geq c/\sqrt{d} \right] \lesssim u.$$

Case 3. If $|\cos(\tilde{\theta})| > 1/2$.

Then we have

$$\begin{aligned} \Pr_{\tilde{\theta}} \left[|\cos(\tilde{\theta})| \geq 1/2 \right] &\leq \frac{\pi \cdot (1/2)^{d-2}}{\int_0^\pi \sin^{d-2}(z) \cdot dz} \\ &\lesssim \frac{d}{2^d} \\ &\leq u. \end{aligned}$$

The first step is because $|\sin(\tilde{\theta})| < 1/2$.

Combine three cases. Then combine these three cases together, we have

$$\begin{aligned} &\Pr_{\tilde{\theta}} \left[\left| \tilde{w} \cdot \cos(\tilde{\theta}) - \left\lfloor \tilde{w} \cdot \cos(\tilde{\theta}) + 1/2 \right\rfloor \right| \leq u \right] \\ &\leq \Pr_{\tilde{\theta}} \left[\left| \tilde{w} \cdot \cos(\tilde{\theta}) - \left\lfloor \tilde{w} \cdot \cos(\tilde{\theta}) + 1/2 \right\rfloor \right| \leq u, 1/2 \geq |\cos(\tilde{\theta})| \geq c/\sqrt{d} \right] \\ &+ \Pr_{\tilde{\theta}}[|\cos(\tilde{\theta})| > 1/2] + \Pr_{\tilde{\theta}} \left[\left| \tilde{w} \cdot \cos(\tilde{\theta}) - \left\lfloor \tilde{w} \cdot \cos(\tilde{\theta}) + 1/2 \right\rfloor \right| \leq u, |\cos(\tilde{\theta})| \leq c/\sqrt{d} \right] \\ &\lesssim u. \end{aligned}$$

This completes the proof. □

Claim 5.9 (Technical result for Claim 5.7). *Given any $u \in (0, 1/5]$ and any $w \geq 1$, the following holds for the random variables $\tilde{w} \sim \text{Unif}[w, 2w]$ and $\tilde{\theta} \sim \text{Unif}[-\pi, \pi)$:*

$$\Pr_{\tilde{w}, \tilde{\theta}} \left[\left| \tilde{w} \cdot \sin(\tilde{\theta}) - \left\lfloor \tilde{w} \cdot \sin(\tilde{\theta}) + 1/2 \right\rfloor \right| \leq u \right] \leq 8u.$$

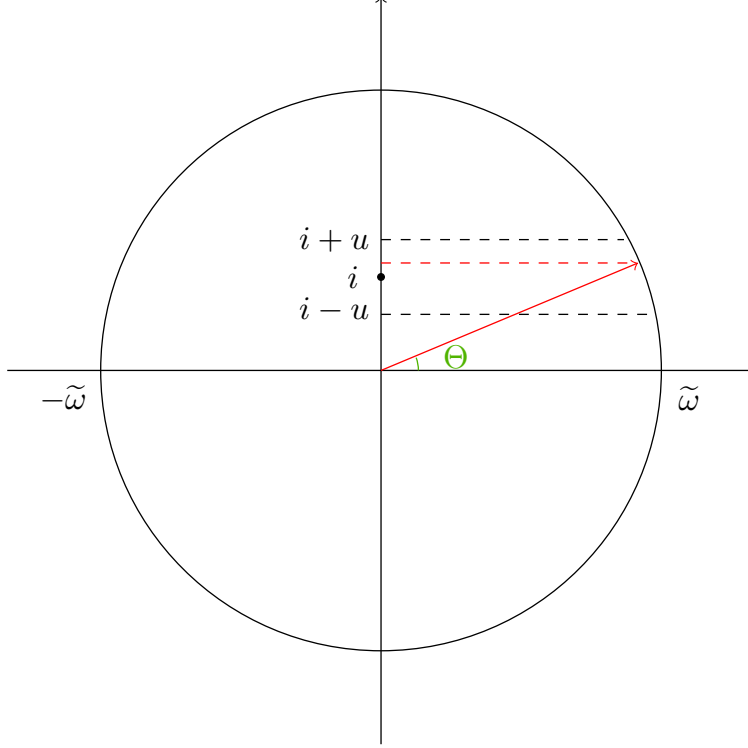


Figure 11: Demonstration for Claim 5.9, where $i \in \mathbb{Z}$ is an integer.

Proof. To improve the readability, we provide Figure 11 for demonstration. We denote

$$\tilde{d}_\theta = |\tilde{w} \cdot \sin(\tilde{\theta}) - \lfloor \tilde{w} \cdot \sin(\tilde{\theta}) + 1/2 \rfloor|$$

for ease of notation. Notice that \tilde{d}_θ represents the distance between $(\tilde{w} \cdot \sin(\tilde{\theta}))$ and its closest integer. By symmetry, the following random distance \tilde{d}_ψ has the same distribution as \tilde{d}_θ :

$$\tilde{d}_\psi = \left| \tilde{w} \cdot \sin(\tilde{\psi}) - \left\lfloor \tilde{w} \cdot \sin(\tilde{\psi}) + 1/2 \right\rfloor \right|,$$

where the new random phase $\tilde{\psi}$ is distributed uniformly on $[0, \frac{\pi}{2}]$ rather than on $[-\pi, \pi)$.

Let us investigate the new random distance \tilde{d}_ψ via case analysis.

Case (i): when $(1 - u)/(2w) \leq \sin(\tilde{\psi}) \leq 1$.

Notice that this case is non-empty, since $u \in (0, 1/5]$ and $w \geq 1$. Suppose the random phase $\tilde{\psi}$ is fixed. Because the random variable $\tilde{w} \sim \text{Unif}[w, 2w]$, the random closest integer $\lfloor \tilde{w} \cdot \sin(\tilde{\psi}) + 1/2 \rfloor$ admits the following lower and upper bounds:

$$\begin{aligned} \left\lfloor \tilde{w} \cdot \sin(\tilde{\psi}) + 1/2 \right\rfloor &\geq \tilde{w} \cdot \sin(\tilde{\psi}) - 1/2 \geq w \cdot \sin(\tilde{\psi}) - 1/2 \\ \left\lfloor \tilde{w} \cdot \sin(\tilde{\psi}) + 1/2 \right\rfloor &\leq \tilde{w} \cdot \sin(\tilde{\psi}) + 1/2 \leq 2w \cdot \sin(\tilde{\psi}) + 1/2 \end{aligned}$$

Namely, the random closest integer $\lfloor \tilde{w} \cdot \sin(\tilde{\psi}) + 1/2 \rfloor$ has at most $(w \cdot \sin(\tilde{\psi}) + 1)$ many possibilities. Consider the set \tilde{A}_ψ of all possible $(\tilde{w} \cdot \sin(\tilde{\psi}))$ such that the random distance $\tilde{d}_\psi \leq u$:

$$\tilde{A}_\psi = \left\{ \tilde{w} \cdot \sin \tilde{\psi} : \tilde{w} \in [w, 2w] \text{ and } \tilde{d}_\psi \leq u \right\}.$$

Since the closed integer $\lfloor \tilde{w} \cdot \sin(\tilde{\psi}) + 1/2 \rfloor$ has at most $(w \cdot \sin(\tilde{\psi}) + 1)$ many possibilities, the total length of this set $\widetilde{A_\psi}$ is at most

$$|\widetilde{A_\psi}| \leq 2u \cdot (w \cdot \sin(\tilde{\psi}) + 1). \quad (22)$$

Hence, under any choice of the random phase $\tilde{\psi}$, the conditional probability (over the uniform random variable $\tilde{w} \sim \text{Unif}[w, 2w]$) below is at most

$$\begin{aligned} \Pr_{\tilde{w}} \left[\widetilde{d_\psi} \leq u \mid \text{case (i)} \right] &= \frac{|\widetilde{A_\psi}|}{(2w - w) \cdot \sin(\tilde{\psi})} \\ &\leq \frac{2u \cdot (w \cdot \sin(\tilde{\psi}) + 1)}{(2w - w) \cdot \sin(\tilde{\psi})} \\ &= 2u \cdot \left(1 + (w \cdot \sin(\tilde{\psi}))^{-1} \right) \\ &\leq 2u \cdot \left(1 + \frac{2}{1 - u} \right) \\ &\leq 7u, \end{aligned}$$

where the second step applies Equation (22); the fourth step follows because (in this case) we assume that $\sin(\tilde{\psi}) \geq \frac{1-u}{2w}$; and the last step is because $u \in (0, 1/5]$.

Case (ii): when $u/w < \sin(\tilde{\psi}) < (1 - u)/(2w)$.

Notice that this case is non-empty, since $u < \frac{1-u}{2}$ for any $u \in (0, 1/5]$. Of course, any realized random variable $\tilde{w} \sim \text{Unif}[w, 2w]$ satisfies $\tilde{w} \geq w$ and $\tilde{w} \leq 2w$. On the lower-bound part:

$$\tilde{w} \cdot \sin(\tilde{\psi}) \geq w \cdot \sin(\tilde{\psi}) > w \cdot \frac{u}{w} = u.$$

Further, on the upper-bound part:

$$\tilde{w} \cdot \sin(\tilde{\psi}) \leq 2w \cdot \sin(\tilde{\psi}) < 2w \cdot \frac{1 - u}{2w} = 1 - u.$$

Combining both inequalities together, regardless of the realized $\tilde{w} \sim \text{Unif}[w, 2w]$, the random variable $(\tilde{w} \cdot \sin(\tilde{\psi}))$ locates between $(u, 1 - u)$ and differs from its closest integer by at least u .

From the above arguments, we conclude that the next conditional probability equals zero.

$$\Pr_{\tilde{w}, \tilde{\psi}} \left[\widetilde{d_\psi} \leq u \mid \text{case (ii)} \right] = 0.$$

Case (iii): when $0 \leq \sin(\tilde{\psi}) \leq u/w$.

Of course, the following conditional probability is at most one:

$$\Pr_{\tilde{w}, \tilde{\psi}} \left[\widetilde{d_\psi} \leq u \mid \text{case (iii)} \right] \leq 1.$$

Observe that $u/w \leq u \leq 1/5$, because $u \in (0, 1/5]$ and $w \geq 1$. Then, since the random phase $\tilde{\psi}$ is distributed uniformly on $[0, \frac{\pi}{2}]$, this case happens with probability

$$\Pr_{\tilde{w}, \tilde{\psi}} [\text{case (iii)}] = \frac{2}{\pi} \cdot \sin^{-1}(u/w)$$

$$\begin{aligned}
&\leq \frac{2}{\pi} \cdot \sin^{-1}(u) \\
&\leq \frac{2}{3} \cdot u,
\end{aligned}$$

where the last step follows since $z \leq \sin(\frac{\pi}{3} \cdot z)$ for any $z \in [0, 1/2]$ and we have $u \in (0, 1/5] \subseteq [0, 1/2]$.

Putting all the three cases together, we conclude that

$$\begin{aligned}
\Pr_{\tilde{w}, \tilde{\psi}} \left[\widetilde{d_{\psi}} \leq u \right] &\leq 7u \cdot \Pr_{\tilde{w}, \tilde{\psi}} [\text{case (i)}] + 0 \cdot \Pr_{\tilde{w}, \tilde{\psi}} [\text{case (ii)}] + 1 \cdot \Pr_{\tilde{w}, \tilde{\psi}} [\text{case (iii)}] \\
&= 7u \cdot \Pr_{\tilde{w}, \tilde{\psi}} [\text{case (i)}] + 1 \cdot \Pr_{\tilde{w}, \tilde{\psi}} [\text{case (iii)}] \\
&\leq 7u \cdot 1 + 1 \cdot \frac{2}{3} \cdot u \\
&\leq 8u,
\end{aligned}$$

where the first step applies the bounds on the conditional probabilities derived before.

This completes the proof of Claim 5.9. \square

5.4 Election process

The goal of this section is to prove Lemma 5.10.

Lemma 5.10 (The election process of LOCATEINNER). *For any matrix $\Sigma \in \mathbb{R}^{d \times d}$ and any vector $b \in \mathbb{R}^d$, assume three premises for a particular good tone frequency $f \in H = \{\xi \in \text{supp}(\widehat{x}^*) : \text{neither } E_{\text{coll}}(\xi) \text{ nor } E_{\text{off}}(\xi) \text{ happens}\}$:*

- The tone frequency $f \in H$ is hashed into the bin $h_{\Sigma, b}(f) = j \in [B]^d$ (Definition 4.10).
- The tone frequency $f \in H$ locates within the hyperball $\mathbf{HB}(\text{List}[j], L^{\text{dia}})$.
- Given the guessed approximation ratio $\mathcal{C} \in [120, \rho]$, the following holds for both $a = a_r$ and $a = a_r + \Delta_a^r$, in every single iteration $r \in [\mathcal{R}_{\text{vote}}]$ of the procedure LOCATEINNER (Algorithm 2):

$$\mathbf{E}_a \left[\left| \widehat{u}_j - \widehat{x}[f] \cdot e^{2\pi i \cdot a^\top \Sigma f} \right|^2 \right] \leq \mathcal{C}^{-2} \cdot |\widehat{x}[f]|^2.$$

Then with probability at least $1 - \mathcal{M} \cdot 2^{-\Omega(\mathcal{R}_{\text{vote}})}$, the following hold for the algorithm LOCATEINNER:

Property I: The tone frequency $f \in H$ locates in one of the winning sub-hyperballs

$$\text{Winner}[j] = \bigcup_{q \in Q: \text{Vote}_j[q] \geq \frac{1}{2} \cdot \mathcal{R}_{\text{vote}}} \mathbf{HB}(f_q^{\text{grid}[j]}, M^{-1} \cdot L^{\text{dia}}).$$

Property II: All the winning frequencies $\text{Winner}[j]$ can be included within the smaller hyperball

$$\text{Winner}[j] \subseteq \mathbf{HB}(f_{q^*}^{\text{grid}[j]}, L_{\text{new}}^{\text{dia}}),$$

which is centered at $f_{q^*}^{\text{grid}[j]}$ and has the new diameter $L_{\text{new}}^{\text{dia}} := \frac{1}{2} \cdot L^{\text{dia}}$.

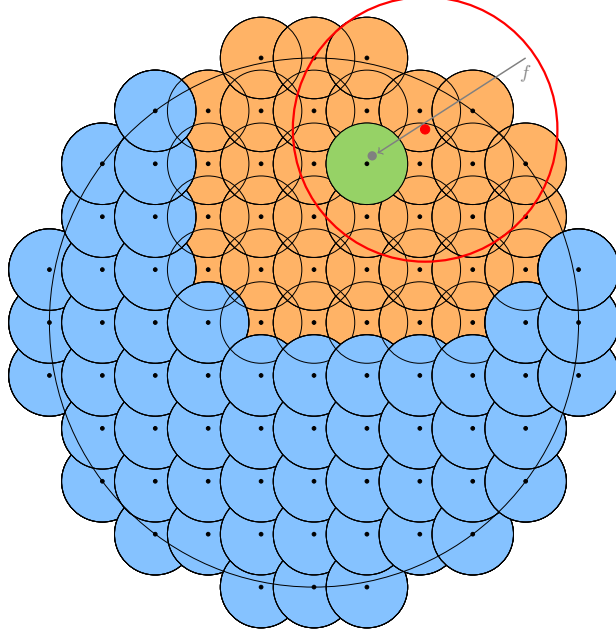


Figure 12: Demonstration for Property III of Lemma 5.10 in two dimensions $d = 2$. The given (black) circle $\mathbf{HB}(\mathcal{L}ist[j], L^{dia})$ narrows down into a smaller (red) circle $\mathbf{HB}(\mathcal{L}ist_{\text{new}}[j], L_{\text{new}}^{dia})$.

Property III: *The output frequency $\mathcal{L}ist_{\text{new}}[j] \in \mathbb{R}^d$ makes the tone frequency $f \in H$ locate in a new hyperball that is centered at $\mathcal{L}ist_{\text{new}}[j]$ and has a new diameter $L_{\text{new}}^{dia} = \frac{1}{2} \cdot L^{dia}$:*

$$f \in \mathbf{HB}(\mathcal{L}ist_{\text{new}}[j], L_{\text{new}}^{dia}).$$

To make Lemma 5.10 meaningful, later we will choose a large enough $\mathcal{R}_{\text{vote}} \in \mathbb{N}_{\geq 1}$ such that the failure probability $\mathcal{M} \cdot 2^{-\Omega(\mathcal{R}_{\text{vote}})} \ll 1$. Furthermore, we observe that Property III of Lemma 5.10 (see Figure 12 for demonstration) is a direct follow-up to Properties I and II. Below, we would show that Property I holds with probability $1 - 2^{-\Omega(\mathcal{R}_{\text{vote}})}$, and that Property II holds with probability $1 - (\mathcal{M} - 1) \cdot 2^{-\Omega(\mathcal{R}_{\text{vote}})}$. Then, all the properties can be inferred via the union bound.

Claim 5.11 (Property I of Lemma 5.10). *With probability at least $1 - 2^{-\Omega(\mathcal{R}_{\text{vote}})}$, the tone frequency $f \in H$ locates in one of the winning sub-hyperballs*

$$\text{Winner}[j] = \bigcup_{q \in Q: \text{Vote}_j[q] \geq \frac{1}{2} \cdot \mathcal{R}_{\text{vote}}} \mathbf{HB}(f_q^{\text{grid}[j]}, M^{-1} \cdot L^{dia}).$$

Proof. Given the second premise of Lemma 5.10 that, the tone frequency $f \in H$ locates within the hyperball $\mathbf{HB}(\mathcal{L}ist[j], L^{dia})$, there is a unique true sub-hyperball $\mathbf{HB}(f_{q^*}^{\text{grid}[j]}, \frac{1}{M} \cdot L^{dia})$ containing $f \in H$ (see Definition 5.2), for some vector index $q^* \in Q$. Apparently, a necessary condition for $f \in H$ to locate in none of the winning sub-hyperballs is the event

$$\overline{E_{q^*}} = \left\{ \text{the } q^*\text{-th sub-hyperball in total gets less than } \frac{1}{2} \cdot \mathcal{R}_{\text{vote}} \text{ votes} \right\}.$$

Based on Property I of Lemma 5.5, in every iteration $r \in [\mathcal{R}_{\text{vote}}]$, the q^* -th sub-hyperball independently loses a vote with probability at most $4 \cdot \mathcal{C}^{-2/3} < 1/2$. Combining a simple coupling

argument together with the Chernoff bound (see Part (a) of Lemma 3.1), the event $\overline{E_{q^*}}$ happens with probability at most

$$\begin{aligned} \Pr [\overline{E_{q^*}}] &\leq \exp \left(-\frac{\mathcal{R}_{\text{vote}}}{2} \cdot \left(\ln \left(\frac{\mathcal{C}^{2/3}}{8} \right) + \frac{8}{\mathcal{C}^{2/3}} - 1 \right) \right) \\ &\leq \exp \left(-\frac{\mathcal{R}_{\text{vote}}}{2} \cdot \left(\ln \left(\frac{120^{2/3}}{8} \right) + \frac{8}{120^{2/3}} - 1 \right) \right) \\ &= \exp (-\Omega(\mathcal{R}_{\text{vote}})) \end{aligned}$$

where the second step follows because the formula $\ln z + \frac{1}{z}$ is increasing in $z \in \mathbb{R}_{>0}$ (and $\mathcal{C} \geq 120$; see Definition 5.1); and the last step follows as $\ln(\frac{120^{2/3}}{8}) + \frac{8}{120^{2/3}} - 1 \approx 0.1174 = \Omega(1)$.

This completes the proof of Claim 5.11. \square

Claim 5.12 (Property II of Lemma 5.10). *All the winning frequencies $\text{Winner}[j]$ can be included within the smaller hyperball*

$$\text{Winner}[j] \subseteq \mathbf{HB}(f_{q^*}^{\text{grid}[j]}, L_{\text{new}}^{\text{dia}}),$$

which is centered at $f_{q^*}^{\text{grid}[j]}$ and has the new diameter $L_{\text{new}}^{\text{dia}} := \frac{1}{2} \cdot L^{\text{dia}}$.

Proof. Recall Definition 5.2 that we cover the hyperball $\mathbf{HB}(\text{List}[j], L^{\text{dia}})$ by using $\mathcal{M} = 2^{\Theta(d \cdot \log(\mathcal{C} \cdot d))}$ many sub-hyperballs. Given the second premise of Lemma 5.10, one particular sub-hyperball $q^* \in Q$ is the true sub-hyperball, and there are at most $(\mathcal{M} - 1)$ many wrong sub-hyperballs.

We first demonstrate that, a specific wrong sub-hyperball $q \in Q$ “wins” with probability at most $2^{-\Omega(\mathcal{R}_{\text{vote}})}$. By definition (see Line 19 of LOCATEINNER), this wrong sub-hyperball “wins” if and only if the following event happens:

$$E_q = \left\{ \text{the } q\text{-th sub-hyperball in total gets at least } \frac{1}{2} \cdot \mathcal{R}_{\text{vote}} \text{ votes} \right\}.$$

According to Property II of Lemma 5.5, in each iteration $r \in [\mathcal{R}_{\text{vote}}]$, the q -th sub-hyperball independently gets a vote with probability at most $12 \cdot \mathcal{C}^{-2/3} < 1/2$. Combining a simple coupling argument together with the Chernoff bound (see Part (a) of Lemma 3.1), the event E_q happens with probability at most

$$\begin{aligned} \Pr [E_q] &\leq \exp \left(-\frac{\mathcal{R}_{\text{vote}}}{2} \cdot \left(\ln \left(\frac{\mathcal{C}^{2/3}}{24} \right) + \frac{24}{\mathcal{C}^{2/3}} - 1 \right) \right) \\ &\leq \exp \left(-\frac{\mathcal{R}_{\text{vote}}}{2} \cdot \left(\ln \left(\frac{120^{2/3}}{24} \right) + \frac{24}{120^{2/3}} - 1 \right) \right) \\ &= \exp (-\Omega(\mathcal{R}_{\text{vote}})) \end{aligned}$$

where the second step follows because the formula $\ln z + \frac{1}{z}$ is increasing in $z \in \mathbb{R}_{>0}$ (and $\mathcal{C} \geq 120$; see Definition 5.1); and the last step follows as $\ln(\frac{120^{2/3}}{24}) + \frac{24}{120^{2/3}} - 1 \approx 9.2161 \times 10^{-5} = \Omega(1)$.

Since there are at most $(\mathcal{M} - 1)$ many wrong sub-hyperballs, we can apply the union bound for all of them. Hence, with probability at least $1 - (\mathcal{M} - 1) \cdot 2^{-\Omega(\mathcal{R}_{\text{vote}})}$, none of the wrong sub-hyperballs “win” in the election process.

According to Definition 5.2, the diameter of a sub-hyperball is $\frac{1}{M} \cdot L^{dia}$, and any *intermediate* sub-hyperball $q \in Q$ (or the true sub-hyperball $q^* \in Q$ itself) satisfies that

$$\|f_{q^*}^{grid[j]} - f_q^{grid[j]}\|_2 \geq \frac{1}{M} \cdot L^{dia} \cdot \lceil 4\sqrt{d} \cdot \mathcal{C}^{2/3} \rceil.$$

For these reasons, the ℓ_2 -distance between any $f \in \mathcal{W}inner[j]$ and the center frequency $f_{q^*}^{grid[j]}$ of the true sub-hyperball is at most

$$\begin{aligned} \|f - f_{q^*}^{grid[j]}\|_2 &\leq \left(\lceil 4\sqrt{d} \cdot \mathcal{C}^{2/3} \rceil - 1 + \frac{1}{2} \right) \cdot \frac{1}{M} \cdot L^{dia} \\ &\leq \frac{1}{4} \cdot L^{dia}, \end{aligned}$$

where the last step is because $M = 4 \cdot \lceil 4\sqrt{d} \cdot \mathcal{C}^{2/3} \rceil$ (see Definition 5.1).

Thus, any frequency $f \in \mathcal{W}inner[j]$ can be included in a smaller hyperball $\mathbf{HB}(f_{q^*}^{grid[j]}, L_{\text{new}}^{dia})$ that is centered at $f_{q^*}^{grid[j]}$ and has the new diameter $L_{\text{new}}^{dia} := \frac{1}{2} \cdot L^{dia}$.

This accomplishes the proof of Claim 5.12. \square

5.5 Performance guarantees

The goal of this section is to prove Corollary 5.13.

Corollary 5.13 (The guarantee of LOCATEINNER). *Given Σ and b (according to Definition 4.7), let $H \subseteq \text{supp}(\widehat{x^*})$ be a subset of “good” tone frequencies:*

$$H = \{\xi \in \text{supp}(\widehat{x^*}) : \text{neither } E_{\text{off}}(\xi) \text{ nor } E_{\text{coll}}(\xi) \text{ happens}\}$$

Let $j := h_{\Sigma, b}(f) \in [B]^d$ where a good frequency $f \in H$ is hashed into. Suppose $f \in \mathbf{HB}(\text{List}[j], L^{dia})$ at the beginning, then with failure probability at most $\mathcal{M} \cdot 2^{-\Omega(\mathcal{R}_{\text{vote}})}$, the procedure LOCATEINNER outputs a new frequency $\text{List}_{\text{new}}[j] \in [-F, F]^d$ so that

$$f \in \mathbf{HB}(\text{List}_{\text{new}}[j], L_{\text{new}}^{dia}),$$

where the new diameter $L_{\text{new}}^{dia} = L^{dia}/2$.

Proof. This follows immediately from Property III of Lemma 5.10. \square

5.6 Sampling time points

The procedure SAMPLETIMEPOINT is given in Algorithm 3, which is illustrated in Figure 13.

5.6.1 Duration requirement

The goal of this part is to prove Lemma 5.14, and thus to obtain the duration bound required by Condition 4.8.

Lemma 5.14 (Duration of LOCATEINNER). *To satisfy Condition 4.8, the sampling duration requirement of the procedure LOCATEINNER (Algorithm 2) is*

$$T = \Omega(d^3 \cdot \eta^{-1} \cdot \log(kd/\delta)).$$

Algorithm 3 SAMPLETIMEPOINT, Lemmas 5.14 and 5.15

- 1: **procedure** SAMPLETIMEPOINT($M, L^{dia}, \mathcal{C}, T$)
 - 2: Define $\varpi \in (0, 1)$ according to Definition 5.1.
 - 3: Sample $\Delta_a \in \mathbb{R}^d$ such that $\Sigma^\top \Delta_a \sim \text{Unif}\{z \in \mathbb{R}^d : \|z\|_2 = 1\}$. $\triangleright |\Sigma| \neq 0$; Definition 4.7
 - 4: Scale Δ_a by a random factor $\beta \sim \text{Unif}[\frac{\varpi \cdot M}{4L^{dia}}, \frac{\varpi \cdot M}{2L^{dia}}]$.
 - 5: Let $A := \{z \in \mathbb{R}^d : \{z, z + \Sigma^\top \Delta_a\} \subseteq [\frac{0.01}{d} \cdot T, (1 - \frac{0.01}{d}) \cdot T]^d\}$.
 - 6: Sample $a \in \mathbb{R}^d$ such that $\Sigma^\top a \sim \text{Unif}(A)$.
 - 7: **return** a and Δ_a .
 - 8: **end procedure**
-

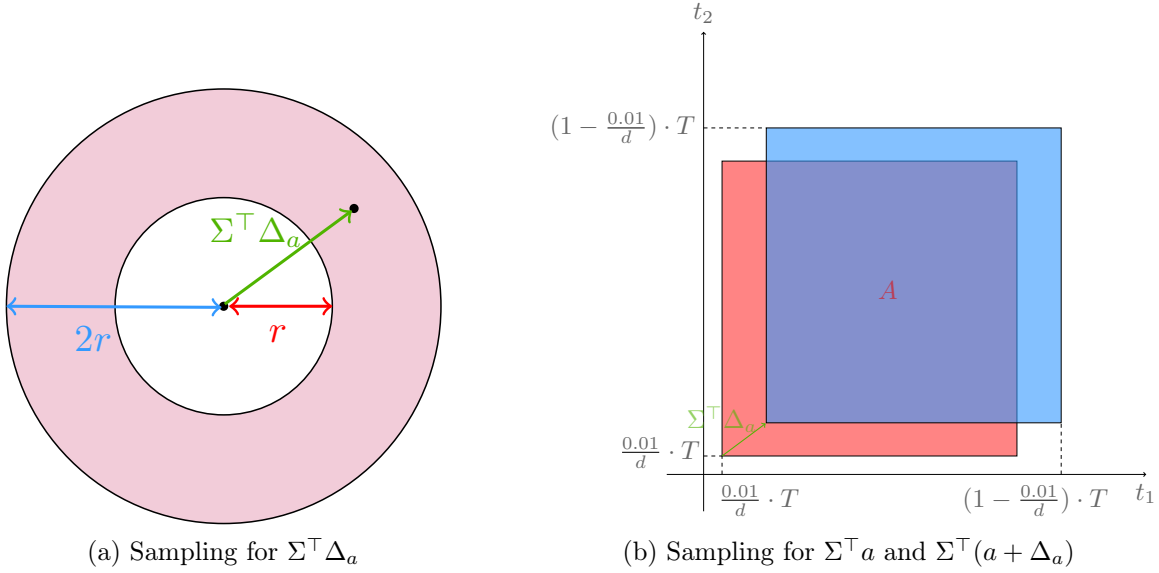


Figure 13: Demonstration of the sampling for Δ_a , and then for a and $(a + \Delta_a)$, where in Figure 13(a) the parameter $r := \frac{\varpi \cdot M}{4L^{dia}}$.

Proof. The procedure LOCATEINNER uses the samples in the time domain by invoking the subroutine HASHTOBINS (Algorithm 1) with a number of pairs $a' \in \{a, a + \Delta_a\}$ output by SAMPLETIMEPOINT (Algorithm 3). In particular (see Line 4 of HASHTOBINS), we take the following sample for all $i \in [BD]^d$ and both $a' \in \{a, a + \Delta_a\}$:

$$\mathcal{P}_{\Sigma, b, a'} x(i) = x(\Sigma^\top(i + a')) \cdot e^{-2\pi i \cdot b^\top i},$$

where the equation follows from Definition 4.14.

To meet Condition 4.8, we shall have

$$\Sigma^\top(i + a') \in [0, T]^d, \tag{23}$$

for all $i \in [BD]^d$ and both $a' \in \{a, a + \Delta_a\}$, under any choice of the random matrix $\Sigma \in \mathbb{R}^{d \times d}$ (according to Definition 4.7).

We know from Line 5 that both $a' \in \{a, a + \Delta_a\}$ satisfy that

$$\Sigma^\top a' \in \left[\frac{0.01}{d} \cdot T, \left(1 - \frac{0.01}{d}\right) \cdot T\right]^d.$$

Given this, a sufficient condition for Equation (23) is that

$$\|\Sigma^\top i\|_\infty \leq \frac{0.01}{d} \cdot T,$$

for all $i \in [BD]^d$, under any choice of the random matrix $\Sigma \in \mathbb{R}^{d \times d}$.

For the above equation, we deduce that

$$\begin{aligned} \|\Sigma^\top i\|_\infty &\leq \|\Sigma^\top i\|_2 \\ &\leq \frac{4\sqrt{d}}{B\eta} \cdot \|i\|_2 \\ &\leq \frac{4\sqrt{d}}{B\eta} \cdot \sqrt{d} \cdot BD \\ &\lesssim d^2 \cdot \eta^{-1} \cdot \log(kd/\delta), \end{aligned}$$

where the second step follows because $\Sigma \in \mathbb{R}^{d \times d}$ is a *rotation matrix* scaled by a random factor $\beta \sim \text{Unif}[\frac{2\sqrt{d}}{B\eta}, \frac{4\sqrt{d}}{B\eta}]$ (Definition 4.7); the third step follows since $i \in [BD]^d = \{0, 1, \dots, BD-1\}^d$; and the last step follows because $D = \Theta(d \cdot \log(kd/\delta))$ (see Definition 4.7).

Putting the above arguments together, we know that Condition 4.8 holds for any sufficiently large $T = \Omega(d^3 \cdot \eta^{-1} \cdot \log(kd/\delta))$.

This completes the proof. \square

5.6.2 Performance guarantees

The goal of this part is to prove Lemma 5.15, and thus to verify Condition 4.9.

Lemma 5.15 (Performance guarantees). *Suppose that Condition 4.8 is true and that $L^{dia} \geq \frac{20d}{T}$ (which will be ensured by Definition 6.1), then Condition 4.9 holds for both $a' \in \{a, a + \Delta_a\}$ derived from the procedure SAMPLETIMEPOINT (Algorithm 3):*

$$\mathbf{E}_{a'} \left[g(\Sigma^\top (i + a'))^2 \right] \lesssim \frac{1}{T^d} \cdot \int_{t \in [0, T]^d} |g(t)|^2 \cdot dt,$$

for all $i \in [BD]^d$, under any choice of the random matrix $\Sigma \in \mathbb{R}^{d \times d}$ (according to Definition 4.7).

Proof. Since both time points $a' = a$ and $a' = a + \Delta_a$ are constructed in a symmetric fashion (see Line 5 of SAMPLETIMEPOINT), we only need to reason about the time point $a \sim \text{Unif}(A)$ given in SAMPLETIMEPOINT. Denote $T' := (1 - 0.02/d) \cdot T \geq 0.98 \cdot T$. By construction (see Line 3),

$$\begin{aligned} \|\Sigma^\top \Delta_a\|_2 &\leq \frac{\varpi \cdot M}{2L^{dia}} \\ &\leq \frac{17\sqrt{d}}{2L^{dia}} \\ &\leq \frac{17}{40\sqrt{d}} \cdot T \\ &\leq \frac{1}{2\sqrt{d}} \cdot T', \end{aligned}$$

where the second step follows since $\mathcal{C} \geq 120$ and $\varpi = \mathcal{C}^{-2/3}$ and $M = 4 \cdot \lceil 4\sqrt{d} \cdot \mathcal{C}^{2/3} \rceil \leq 17\sqrt{d} \cdot \mathcal{C}^{2/3}$ (see Definition 5.1); the third step follows from the premise that $L^{dia} \geq \frac{20d}{T}$; and the last step holds because $T' \geq 0.98T$.

We have

$$\|\Sigma^\top \Delta_a\|_\infty \leq \|\Sigma^\top \Delta_a\|_1 \leq \sqrt{d} \cdot \|\Sigma^\top \Delta_a\|_2 \leq \frac{1}{2} \cdot T'.$$

Let $(\Sigma^\top \Delta_a)_r$ denote the r -th coordinate of $\Sigma^\top \Delta_a \in \mathbb{R}^d$. For any choice of Δ_a by `SAMPLETIMEPOINT`, the volume of the sampling range $\Sigma^\top a \sim \text{Unif}(A)$ is

$$\begin{aligned} \text{vol}(A) &= \prod_{r \in [d]} (T' - |(\Sigma^\top \Delta_a)_r|) \\ &= T'^d \cdot \prod_{r \in [d]} (1 - |(\Sigma^\top \Delta_a)_r| \cdot T'^{-1}) \\ &\geq T'^d \cdot \prod_{r \in [d]} \exp(-2 \cdot |(\Sigma^\top \Delta_a)_r| \cdot T'^{-1}) \\ &= T'^d \cdot \exp(-2 \cdot \|\Sigma^\top \Delta_a\|_1 \cdot T'^{-1}) \\ &\geq T'^d \cdot e^{-1} \\ &\geq T^d \cdot 0.98 \cdot e^{-1}, \end{aligned}$$

where the first step is by Line 5 of `SAMPLETIMEPOINT`; the third step follows since $|(\Sigma^\top \Delta_a)_r| \leq \|\Sigma^\top \Delta_a\|_\infty \leq \frac{1}{2} \cdot T'$ and $1 - z \geq e^{-2z}$ when $z \in [0, \frac{1}{2}]$; the fifth step follows since $\|\Sigma^\top \Delta_a\|_1 \leq \frac{1}{2} \cdot T'$; and the last step is because $T'^d = (1 - 0.02/d)^d \cdot T^d \geq 0.98 \cdot T^d$.

We conclude from the above that, for any choice of Δ_a by `SAMPLETIMEPOINT`, the time point $\Sigma^\top a \sim \text{Unif}(A)$ is sampled uniformly from a *constant* proportion of the duration $t \in [0, T]^d$. And because $\Sigma^\top(i + a)$ is guaranteed to be within the duration $t \in [0, T]^d$, for any choice of $\Sigma \in \mathbb{R}^{d \times d}$ and any $i \in [BD]^d$, we have

$$\mathbf{E}_a \left[g(\Sigma^\top(i + a))^2 \right] \lesssim \frac{1}{T^d} \cdot \int_{t \in [0, T]^d} |g(t)|^2 \cdot dt.$$

This completes the proof. □

5.7 Stronger Guarantee

Lemma 5.16 (Stronger guarantees). *Let*

$$\rho^2 = |\hat{x}[f]|^2 / \mathbf{E}_a[|\hat{u}_j - \hat{x}[f] \cdot e^{2\pi i \cdot a^\top \Sigma f}|^2]$$

Let $C_* = d^2$. Let $\mathcal{R}_{\text{reg}} = C \cdot d$ for some constant C . Let $L^{\text{dia}} = 20d/T$. There is an algorithm (procedure `LOCATEINNER*` in Algorithm 4) that output a list of frequencies such that there is a mapping $\pi : [k] \rightarrow [m]$,

$$\|f'_{\pi(i)} - f_i\|_2 \lesssim \frac{C_*}{\rho T}, \quad \forall i \in [k].$$

Proof. We provide Figure 14 for demonstration. Let $d' = \mathcal{R}_{\text{reg}} = Cd$ for simplicity. By Markov inequality, we know that the following holds with probability at least $1 - \frac{1}{10d'}$:

$$\left| \hat{u}_j - \hat{x}[f] \cdot e^{2\pi i \cdot a^\top \Sigma f} \right| \leq |\hat{x}[f]| \cdot 10\sqrt{d'}/\rho,$$

Algorithm 4 A stronger version of LOCATEINNER when search range is small

```

1: procedure LOCATEINNER*( $\Sigma, b, D, \mathcal{L}ist, L^{dia}, \mathcal{C}, T$ ) ▷ Lemma 5.16
2:   for  $r = 1, 2, \dots, \mathcal{R}_{reg}$  do
3:      $(a_r, \Delta_a^r) \leftarrow \text{SAMPLETIMEPOINT}(M, L^{dia}, \mathcal{C}, T)$ . ▷ Algorithm 3
4:      $\hat{u} \leftarrow \text{HASHTOBINS}(x, \Sigma, b, a_r, D)$ .
5:      $\hat{u}' \leftarrow \text{HASHTOBINS}(x, \Sigma, b, a_r + \Delta_a^r, D)$ .
6:      $\varphi_{j,r} = \arg(\hat{u}_j) - \arg(\hat{u}'_j)$ .
7:   end for
8:    $\mathcal{L}ist_{new} \leftarrow \emptyset$ 
9:   Form matrix  $\Delta^\top := [\Sigma^\top \Delta_a^1, \dots, \Sigma^\top \Delta_a^{\mathcal{R}_{reg}}] \in \mathbb{R}^{d \times \mathcal{R}_{reg}}$ .
10:  for  $j \in [B]$  do
11:    Form vector  $\varphi_j \in \mathbb{R}^{\mathcal{R}_{reg}}$ .
12:     $\mathcal{L}ist_{new} \leftarrow \mathcal{L}ist_{new} \cup \{\frac{1}{2\pi} \cdot \Delta^\top \varphi_j\}$ .
13:  end for
14:  return  $\mathcal{L}ist_{new}$ .
15: end procedure

```

which is equivalent to

$$\left| \hat{u}_j / \hat{x}[f] \cdot e^{-2\pi i a^\top \Sigma f} - 1 \right| \leq 10\sqrt{d'}/\rho.$$

Namely, the complex number $\hat{u}_j / \hat{x}[f] \cdot e^{-2\pi i a^\top \Sigma f}$ lies in the circle $\{z \in \mathbb{C} : |z - 1| \leq \varpi/\sqrt{2}\}$. Clearly, any complex number in this circle has the phase less than $\sin^{-1}(10\sqrt{d'}/\rho)$. In particular,

$$\left\| \underbrace{\arg(\hat{u}_j) - \arg(\hat{x}[f]) - 2\pi \cdot a^\top \Sigma f}_{A_1} \right\|_{\bigcirc} \leq \sin^{-1}(10\sqrt{d'}/\rho),$$

where $\|\theta\|_{\bigcirc} \in [-\pi, \pi)$ denotes the “phase distance” $\min_{z \in \mathbb{Z}} |\theta - 2\pi \cdot z|$.

Similarly, when a is replaced with $(a + \Delta_a)$, with probability $1 - \frac{1}{10d}$ we also have

$$\left\| \underbrace{\arg(\hat{u}'_j) - \arg(\hat{x}[f]) - 2\pi \cdot (a + \Delta_a)^\top \Sigma f}_{A_2} \right\|_{\bigcirc} \leq \sin^{-1}(10\sqrt{d'}/\rho),$$

Put the above two inequalities together, (by the union bound) the following holds for the phase difference $\varphi_{j,r} = \arg(\hat{u}_j) - \arg(\hat{u}'_j)$ with probability $1 - \frac{2}{10d}$:

$$\begin{aligned}
\|\varphi_{j,r} - 2\pi \cdot \Delta_a^r{}^\top \Sigma f\|_{\bigcirc} &= \|A_1 - A_2\|_{\bigcirc} \\
&\leq \|A_1\|_{\bigcirc} + \|A_2\|_{\bigcirc} \\
&\leq 2\sin^{-1}(10\sqrt{d'}/\rho) \\
&\leq 10\sqrt{d'}/\rho.
\end{aligned}$$

where the second step applies the triangle inequality; and last step follows since for any $z \in (0, 1)$, we have $\sin^{-1}(z/\sqrt{2}) \leq (\pi/4) \cdot z$.

Then with probability at least 0.8, we have that

$$\|\varphi - 2\pi \cdot \Delta \cdot f\|_{\infty} \leq 10\sqrt{d'}/\rho,$$

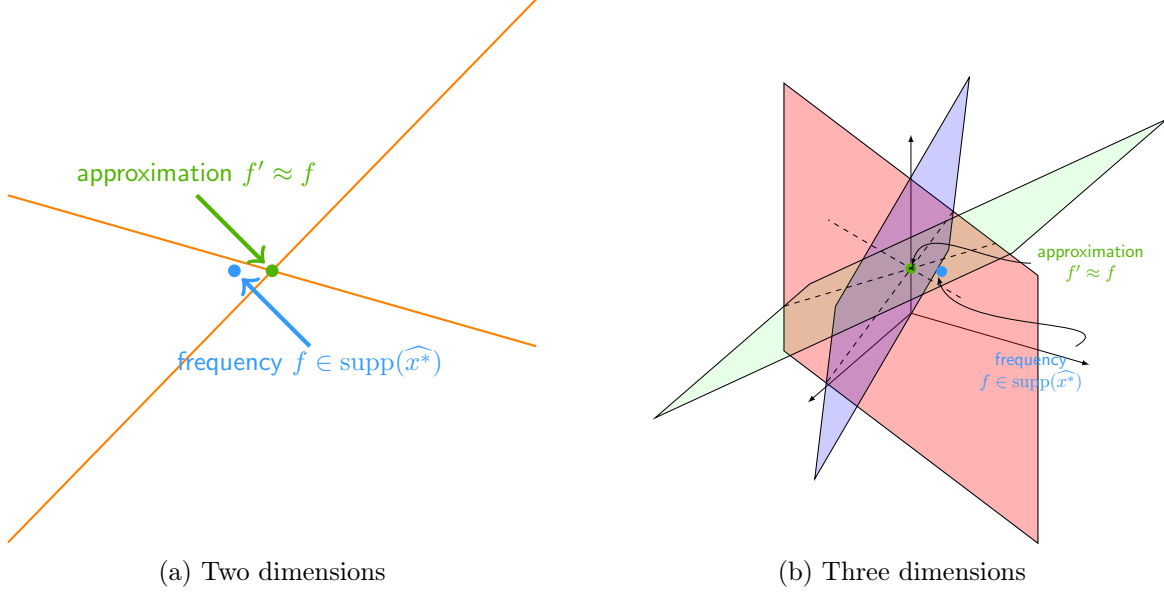


Figure 14: Demonstration for Algorithm 4 in two dimensions $d = 2$ and three dimensions $d = 3$.

where

$$\varphi := \begin{bmatrix} \varphi_{j,1} \\ \varphi_{j,2} \\ \vdots \\ \varphi_{j,d'} \end{bmatrix} \in \mathbb{R}^{d'} \quad \text{and} \quad \Delta := \begin{bmatrix} \Delta_a^{1\top} \Sigma \\ \Delta_a^{2\top} \Sigma \\ \vdots \\ \Delta_a^{d'\top} \Sigma \end{bmatrix} \in \mathbb{R}^{d' \times d}.$$

We deduce from the above that

$$\|\varphi - 2\pi \cdot \Delta f\|_2 \leq \sqrt{d'} \cdot 10\sqrt{d'}/\rho = 10d'/\rho.$$

$\{\Sigma^\top \Delta_a^r\}_{r \in [d']}$ are uniformly distributed on a sphere. Consider any $r \in [d]$, by Theorem 3.4.6 in [Ver18], we know that $\Sigma^\top \Delta_a^r$ is sub-gaussian. Besides, the value of each coordinate of $\Sigma^\top \Delta_a^r$ follows a *Beta*-distribution, and $\frac{d}{\|\Sigma^\top \Delta_a^r\|_2^2} \cdot \mathbf{E}[\Delta_a^{r\top} \Sigma \Sigma^\top \Delta_a^r] = I$. Thus we know that $\frac{\sqrt{d}}{\|\Sigma^\top \Delta_a^r\|_2} \Sigma^\top \Delta_a^r$ is a sub-gaussian isotropic random vector. By selecting $\|\Sigma^\top \Delta_a^r\|_2 \approx T/d$ and the constant $C = d'/d$ large enough, then by Lemma 3.5, we can show that with probability at least $1 - 1/\text{poly}(d)$, we have $s_{\min}(\Delta) \geq \sqrt{d} \cdot \frac{T}{d^{1.5}} = \frac{T}{d}$. Let Δ^\dagger represent the Generalized inverse of Δ , let $f_{LS} = \Delta^\dagger \varphi$ represent the least squares solution, then we have

$$2\pi \|\Delta(f - f_{LS})\|_2 \leq 2\|\varphi - 2\pi \cdot \Delta \cdot f\|_2 \lesssim d/\rho.$$

Then we have

$$\begin{aligned} \|f - f_{LS}\|_2 &\leq \|\Delta(f - f_{LS})\|_2 / s_{\min}(\Delta) \\ &\lesssim \frac{d}{\rho} \cdot \frac{d}{T} \\ &= C_* \cdot \frac{1}{T\rho}. \end{aligned}$$

where the first step is by $\|\Delta x\|_2 / \|x\|_2 \geq s_{\min}(\Delta)$; the second step follows from $s_{\min}(\Delta) \geq T/d$ and the last step follows because we define $C_* := d^2$.

This completes the proof. □

6 Locate signal

Statement	Section	Algorithm	Comment
Definition 6.1	Section 6.1	Algorithm 5	Definitions
Lemma 6.2	Section 6.2	Algorithm 5	Sample complexity and running time
Lemma 6.3	Section 6.3	Algorithm 5	Duration
Lemma 6.4	Section 6.4	Algorithm 5	Guarantees, without Alg. 4
Lemma 6.5	Section 6.5	Algorithm 5	Stronger guarantees, with Alg. 4

Table 3: List of Lemmas/Algorithms in locate signal section.

6.1 Algorithm

Denote $H := \{\xi \in \text{supp}(\widehat{x^*}) : \text{neither } E_{\text{coll}}(\xi) \text{ nor } E_{\text{off}}(\xi) \text{ happens}\}$. Recall the performance guarantees given in Corollary 5.13:

Assume that a specific “good” tone frequency $f \in H$ locates in a hyperball $\mathbf{HB}(\mathcal{L}ist[j], L^{dia})$ that is centered at some frequency $\mathcal{L}ist[j] \in \mathbb{R}^d$ and has the diameter $L^{dia} > 0$.

Then with probability at least $1 - \mathcal{M} \cdot 2^{-\Omega(\mathcal{R}_{\text{vote}})}$, then procedure LOCATEINNER (Algorithm 2) outputs $\mathcal{L}ist_{\text{new}}[j]$ for which

$$f \in \mathbf{HB}(\mathcal{L}ist_{\text{new}}[j], L_{\text{new}}^{dia}).$$

where the new diameter $L_{\text{new}}^{dia} := \frac{1}{2} \cdot L^{dia}$.

Given this, we would estimate the good frequencies by invoking the procedure LOCATEINNER repeatedly. This idea is implemented as the procedure LOCATESIGNAL (Algorithm 5).

Definition 6.1 (Setup for LOCATESIGNAL). The procedure LOCATESIGNAL keeps track of a number of $\mathcal{B} = 2^{\Theta(d \cdot \log d)} \cdot k$ hyperballs. These hyperballs have

- The same initial diameter $L^{dia} := 2\sqrt{d} \cdot F$.
- The final diameter $L^{dia} \in (\frac{20d}{T}, \frac{40d}{T}]$ is chosen so that $\log_2(\frac{\text{initial } L^{dia}}{\text{final } L^{dia}})$ is an integer; clearly, this final L^{dia} is well defined and is unique.
- The number of iteration $\mathcal{R}_{\text{search}} := \log_2(\frac{\text{initial } L^{dia}}{\text{final } L^{dia}}) = O(\log(T \cdot F))$.

That is, each hyperball is initialized to be $\mathbf{HB}(\mathbf{0}, 2\sqrt{d} \cdot F) \supseteq [-F, F]^d$. Clearly, such a hyperball contains all the “good” tone frequencies $f \in H$ at the beginning. Then, the subroutine LOCATEINNER is invoked $\mathcal{R}_{\text{search}}$ times, until the diameter shrinks to the final $L^{dia} \in (\frac{20d}{T}, \frac{40d}{T}]$.

Algorithm 5 LOCATESIGNAL, Lemmas 6.2, 6.3, 6.4

```

1: procedure LOCATESIGNAL( $\Sigma, b, D, \mathcal{C}, T$ )
2:    $\mathcal{List}[j] \leftarrow \mathbf{0} \in \mathbb{R}^d$  for each  $j \in [B]^d$ . ▷ Initialize the center frequency
3:    $L^{dia} = 2\sqrt{d} \cdot F$ . ▷ Initialize the diameter
4:   for  $r = 1, 2, \dots, \mathcal{R}_{\text{search}}$  do
5:      $\mathcal{List}_{\text{new}} \leftarrow \text{LOCATEINNER}(\Sigma, b, D, \mathcal{List}, L^{dia}, \mathcal{C}, T)$ . ▷ Algorithm 2
6:      $\mathcal{List} \leftarrow \mathcal{List}_{\text{new}}$ .
7:      $L^{dia} \leftarrow \frac{1}{2} \cdot L^{dia}$ .
8:   end for
9:   ▷  $L^{dia} = \Theta(d/T)$ 
10:   $\mathcal{List} \leftarrow \text{LOCATEINNER}^*(\Sigma, b, D, \mathcal{List}, L^{dia}, \mathcal{C}, T)$  ▷ Algorithm 4
11:   $\mathcal{List}^* \leftarrow \mathcal{List}$  (after removing the NIL's) ▷ the frequencies
12:  return  $\mathcal{List}^*$ .
13: end procedure

```

6.2 Sample complexity and running time

The goal of this section is to prove Lemma 6.2.

Lemma 6.2 (Sample complexity and running time of LOCATESIGNAL). *The procedure LOCATESIGNAL (Algorithm 5) has the following performance guarantees:*

- The sample complexity is $2^{\Theta(d \cdot \log d)} \cdot k \cdot \mathcal{D} \cdot (\log \mathcal{C} + \log \log(F/\eta)) \cdot \log(T \cdot F)$.
- The running time is $2^{\Theta(d \cdot \log(\mathcal{C} \cdot d))} \cdot k \cdot (\mathcal{D} + \log k) \cdot \log \log(F/\eta) \cdot \log(T \cdot F)$.
- The output \mathcal{List}^* contains at most $O(\mathcal{B}) = O(B^d) = 2^{O(d \cdot \log d)} \cdot k$ many candidate frequencies.

Proof. How many frequencies the output \mathcal{List}^* contains is easy to see, since \mathcal{List}^* is indexed by the bins $j \in [B]^d$. Below we quantify the sample complexity and the running time.

Sample complexity. The procedure LOCATESIGNAL invokes the subroutine LOCATEINNER $\mathcal{R}_{\text{search}}$ times. Due to Lemma 5.4, the subroutine LOCATEINNER has the sample complexity

$$\#\text{sample}(\text{LOCATEINNER}) = 2^{\Theta(d \cdot \log d)} \cdot k \cdot \mathcal{D} \cdot (\log \mathcal{C} + \log \log(F/\eta)).$$

Thus, LOCATESIGNAL has the sample complexity

$$\begin{aligned} \#\text{sample}(\text{LOCATESIGNAL}) &= \#\text{sample}(\text{LOCATEINNER}) \cdot \mathcal{R}_{\text{search}} \\ &= 2^{\Theta(d \cdot \log d)} \cdot k \cdot \mathcal{D} \cdot (\log \mathcal{C} + \log \log(F/\eta)) \cdot \log(T \cdot F). \end{aligned}$$

Running time. Due to Lemma 5.4, the subroutine LOCATEINNER has the running time

$$\#\text{time}(\text{LOCATEINNER}) = 2^{\Theta(d \cdot \log(\mathcal{C} \cdot d))} \cdot k \cdot (\mathcal{D} + \log k) \cdot \log \log(F/\eta).$$

Thus, LOCATESIGNAL has the running

$$\begin{aligned} \#\text{time}(\text{LOCATESIGNAL}) &= \#\text{time}(\text{LOCATEINNER}) \cdot \mathcal{R}_{\text{search}} \\ &= 2^{\Theta(d \cdot \log(\mathcal{C} \cdot d))} \cdot k \cdot (\mathcal{D} + \log k) \cdot \log \log(F/\eta) \cdot \log(T \cdot F). \end{aligned}$$

This completes the proof of Lemma 6.2. □

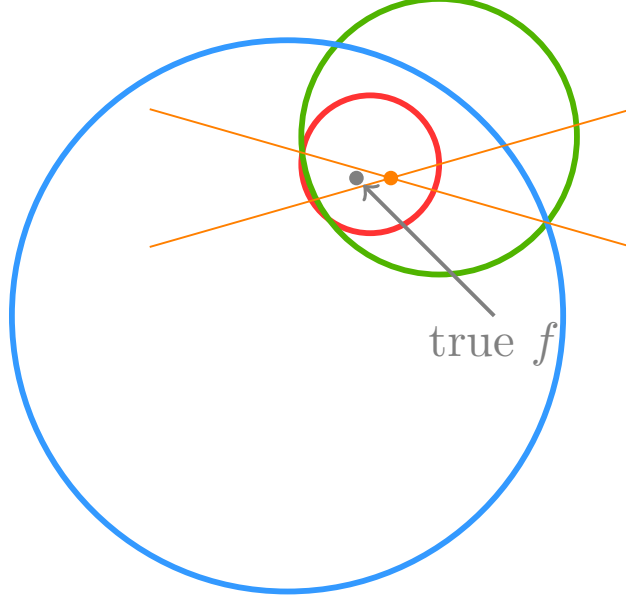


Figure 15: Demonstration of Algorithm 5 for a single frequency $f \in \text{supp}(\widehat{x^*})$ in two dimensions $d = 2$. The blue/green/red circles refer to the coarse-grained location, and the “orange” lines refer to the fine-grained location.

6.3 Duration requirement

The goal of this section is to prove Lemma 6.3.

Lemma 6.3 (Duration of LOCATESIGNAL). *The sampling duration requirement of the procedure LOCATESIGNAL (Algorithm 5) is*

$$T = \Omega(d^3 \cdot \eta^{-1} \cdot \log(kd/\delta)).$$

Proof. This follows immediately from Lemma 5.14. □

6.4 Performance guarantees

The goal of this section is to prove Lemma 6.4.

Lemma 6.4 (Guarantees of LOCATESIGNAL). *Given $\Sigma \in \mathbb{R}^{d \times d}$ and $b \in \mathbb{R}^d$, the output list List^* of procedure LOCATESIGNAL (Algorithm 5) contains at most $\mathcal{B} = 2^{O(d \cdot \log d)} \cdot k$ many frequencies with minimum separation $\Omega(\eta)$. Let $H \subseteq \text{supp}(\widehat{x^*})$ be a subset of “good” tone frequencies:*

$$H = \{\xi \in \text{supp}(\widehat{x^*}) : \text{neither } E_{\text{off}}(\xi) \text{ nor } E_{\text{coll}}(\xi) \text{ happens}\}$$

For any good frequency $f \in H$, suppose that its signal-to-noise ratio $\rho(i) \geq \mathcal{C}$ (see Definition 7.1), then with probability at least 99%, there exists an output frequency $f' \in \text{List}^$ such that*

$$\|f - f'\|_2 \lesssim \frac{d}{T}.$$

Proof. The concerning frequency $f \in H$ w.l.o.g. is hashed into the bin $j := h_{\Sigma,b}(f) \in [B]^d$. Recall Definition 6.1 that the procedure LOCATESIGNAL keeps track of a number of $\mathcal{B} = 2^{\Theta(d \cdot \log d)} \cdot k$

hyperballs. The j -th hyperball is initialized to be $\mathbf{HB}(\mathbf{0}, 2\sqrt{d} \cdot F) \supseteq [-F, F]^d$, and thus contains the frequency $f \in H$. Then, the procedure LOCATESIGNAL invokes the subroutine LOCATEINNER $\mathcal{R}_{\text{search}}$ times, each of which shrinks the diameter of the j -th hyperball by half, until the diameter drops down to the final $L^{\text{dia}} \in (\frac{20d}{T}, \frac{40d}{T}]$.

Failure probability. For the concerning frequency $f \in H$, we know from Corollary 5.13 that each invocation of LOCATEINNER fails with probability at most $\mathcal{M} \cdot 2^{-\Omega(\mathcal{R}_{\text{vote}})}$. By the union bound, the failure probability of LOCATESIGNAL is at most

$$\begin{aligned} \mathcal{R}_{\text{search}} \cdot \mathcal{M} \cdot 2^{-\Omega(\mathcal{R}_{\text{vote}})} &= \underbrace{O(\log(F \cdot T))}_{\mathcal{R}_{\text{search}}} \cdot \underbrace{2^{\Theta(d \cdot \log(\mathcal{C} \cdot d))}}_{\mathcal{M}} \cdot 2^{-\Omega(\mathcal{R}_{\text{vote}})} \\ &\leq 1\%, \end{aligned}$$

where the first step follows because the parameters $\mathcal{R}_{\text{search}} = O(\log(T \cdot F))$ and $\mathcal{M} = 2^{\Theta(d \cdot \log(\mathcal{C} \cdot d))}$ (see Definitions 6.1 and 5.1); and the last step holds since we choose in Definition 5.1 a sufficiently large $\mathcal{R}_{\text{vote}} = \Theta(d \cdot \log(\mathcal{C} \cdot d) + \log \log(F/\eta))$.

Performance guarantee. At the beginning, the initial j -th hyperball $\mathbf{HB}(\mathbf{0}, 2F) = [-F, F]^d$ contains the concerning frequency $f \in H$. If the procedure LOCATESIGNAL succeeds in all of the first $r \in [\mathcal{R}_{\text{search}}]$ iterations, then (Corollary 5.13) we locate $f \in H$ within a hyperball that is centered at some frequency $\text{List}_{\text{new}}[j] \in [-F, F]^d$ and has the diameter $L_{\text{new}}^{\text{dia}} = 2\sqrt{d} \cdot F \cdot 2^{-r}$. Formally, we have

$$f \in \mathbf{HB}(\text{List}_{\text{new}}[j], L_{\text{new}}^{\text{dia}}).$$

In particular, if all of the $\mathcal{R}_{\text{search}}$ iterations succeed, the diameter drops down to the final $L^{\text{dia}} \in (\frac{20d}{T}, \frac{40d}{T}]$. As a consequence, we have

$$f \in \mathbf{HB}(\text{List}^*[j], 40d/T).$$

That is, the ℓ_2 -distance between the concerning tone frequency $f \in H$ and the output frequency $\text{List}^*[j]$ is at most $\frac{1}{2} \cdot \frac{40d}{T} = \frac{20d}{T}$.

This completes the proof of Lemma 6.4. \square

6.5 Stronger guarantees

The goal of this section is to improve Lemma 6.4.

Lemma 6.5 (Stronger guarantees, compared to Lemma 6.4). *Given $\Sigma \in \mathbb{R}^{d \times d}$ and $b \in \mathbb{R}^d$ (according to Definition 4.7), the output list List^* of procedure LOCATESIGNAL (Algorithm 5) contains at most $\mathcal{B} = 2^{O(d \cdot \log d)} \cdot k$ many frequencies with minimum separation $\Omega(\eta)$. Let $H \subseteq \text{supp}(\widehat{x^*})$ be a subset of “good” tone frequencies:*

$$H = \{\xi \in \text{supp}(\widehat{x^*}) : \text{neither } E_{\text{off}}(\xi) \text{ nor } E_{\text{coll}}(\xi) \text{ happens}\}$$

For any good frequency $f \in H$, suppose that its signal-to-noise ratio $\rho(i) \geq \mathcal{C}$ (see Definition 7.1), then with probability at least 99%, there exists an output frequency $f' \in \text{List}^$ such that*

$$\|f - f'\|_2 \lesssim C_* \cdot \frac{1}{\rho T}.$$

Proof. The proof follows from Lemma 5.16. \square

Statement	Section	Algorithm	Comment
Definition 7.1	Section 7.1	None	Definitions and facts
Lemma 7.3	Section 7.2	Algorithm 6	Estimate signal
Lemma 7.4	Section 7.3	Algorithm 7	One stage, sample complexity and running time
Lemma 7.5	Section 7.4	Algorithm 7	One stage, guarantees
Lemma 7.9	Section 7.5	Algorithm 8	Multi sage
Lemma 7.11	Section 7.6	Algorithm 9	Merged stage, running time
Lemma 7.14	Section 7.7	Algorithm 9	Merged stage, guarantees
Lemma 7.16	Section 7.8	Algorithm 10	Running merged stage twice
Theorem 7.18	Section 7.9	Algorithm 10	Recovery stage

Table 4: List of Lemmas/Algorithms in sparse recovery section

7 Sparse recovery

7.1 Definitions and facts

Definition 7.1 (Signal-to-noise ratio). For the i -th tone (v_i, f_i) , define the signal-to-noise ratio $\rho_i := |v_i|/\mu_i \geq 0$, where the noise $\mu_i \geq 0$ is given by

$$\mu_i^2 = \mathbf{E}_{\Sigma, b, a} [|\hat{u}'_j \cdot e^{-2\pi i \cdot a^\top f_i} - v_i|^2],$$

where $j = h_{\Sigma, b}(f_i) \in [B]^d$ is the bin that the tone frequency $f_i \in [-F, F]^d$ is hashed into according to Definition 4.10, and $\hat{u}'_j = \hat{u}_j \cdot e^{-(\pi i/B) \cdot \|j\|_1}$.

Definition 7.2 (Hypercube). For any frequency $f \in \mathbb{R}^d$ and any $L^{edge} \geq 0$, we denote by $\mathbf{HC}(f, L^{edge})$ the ℓ_∞ -norm hypercube with center $f \in \mathbb{R}^d$ and the edge length L^{edge} :

$$\mathbf{HC}(f, L^{edge}) := \left\{ \xi \in \mathbb{R}^d : \|\xi - f\|_\infty \leq L^{edge}/2 \right\}.$$

7.2 ESTIMATESIGNAL

The goal of this section is to prove Lemma 7.3.

Algorithm 6 ESTIMATESIGNAL

- 1: **procedure** ESTIMATESIGNAL($\Sigma, b, a, D, T, List$)
 - 2: Sample $a \in \mathbb{R}^d$ according to Definition 5.1.
 - 3: Let $\hat{u} \leftarrow \text{HASHTOBINS}(\Sigma, b, a, D)$. ▷ Algorithm 1
 - 4: Let $v'(\xi) = \hat{u}_{h_{\Sigma, b}(\xi)} \cdot e^{-(\pi i/B) \cdot \|h_{\Sigma, b}(\xi)\|_1} \cdot e^{-2\pi i \cdot a^\top \xi}$ for $\xi \in List$.
 - 5: **return** $\{v'(\xi)\}_{\xi \in List}$.
 - 6: **end procedure**
-

Lemma 7.3 (ESTIMATESIGNAL). *The procedure ESTIMATESIGNAL (Algorithm 6) satisfies that:*

- *The sample complexity is upper bounded by the sample complexity of the procedure LOCATESIGNAL (Algorithm 5).*
- *The running time is upper bounded by the running time of the procedure LOCATESIGNAL (Algorithm 5).*

Denote by $H \subseteq [k]$ the indices of a subset of true tones $\{(v_i, f_i)\}_{i \in [k]}$ for which neither $E_{\text{coll}}(f_i)$ nor $E_{\text{off}}(f_i)$ happens. There is a subset $S \subseteq H$ and an injection $\pi : S \mapsto [k]$ such that

Property I: For the tones in set S , the (partial) tone estimation error

$$\sum_{i \in S} \mathbf{E}_{\Sigma, b} \left[\frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |v'_i \cdot e^{2\pi i \cdot f'_i \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)} \tau}|^2 \cdot d\tau \right] \lesssim (\mathcal{C}^2 + dC_*^2) \cdot \mathcal{N}^2.$$

Property III: For each tone $i \in S$, the (single) tone estimation error

$$|v'_i - v_{\pi(i)}| \lesssim (\mathcal{C} + \sqrt{d}C_*) \cdot \mathcal{N}.$$

Proof. The bounds on the sample complexity and the running time are direct follow-ups to the previous lemmas. Also, Properties I and II will be proved soon after in Lemma 7.5; particularly, we will specify the subset $S \subseteq H$ therein. \square

7.3 ONESTAGE: algorithm, sample complexity and running time

The goal of this section is to prove Lemma 7.4.

Algorithm 7 ONESTAGE, Lemma 7.4, 7.5

```

1: procedure ONESTAGE( $x, \Sigma, b, D, \mathcal{C}, T$ )
2:    $\text{List} \leftarrow \text{LOCATESIGNAL}(\Sigma, b, D, M, \mathcal{C}, T)$ . ▷ Algorithm 5
3:   for  $\xi \in \text{List}$  do
4:     if either  $E_{\text{coll}}(\xi)$  or  $E_{\text{off}}(\xi)$  or both happen then
5:       Remove  $\xi$  from  $\text{List}$ .
6:     end if
7:   end for
8:    $\{v'(\xi)\}_{\xi \in \text{List}} \leftarrow \text{ESTIMATESIGNAL}(\Sigma, b, a, D, T, \text{List})$  ▷ Algorithm 6
9:   Add an supplementary list  $\text{List}_{\text{sup}} = \{(0, \xi_i)\}_{i=1}^k$ , for which  $\min_{\xi, \xi' \in \text{List}_{\text{sup}}} \|\xi - \xi'\|_2 \geq \eta$ , and
   ▷ Used in Claim 7.7
    $\min_{\xi \in \text{List}, \xi' \in \text{List}_{\text{sup}}} \|\xi - \xi'\|_2 \geq \eta$ .
10:  return  $\{(v'(\xi), \xi)\}_{\xi \in \text{List}} \cup \{(0, \xi)\}_{\xi \in \text{List}_{\text{sup}}}$ .
11: end procedure

```

Lemma 7.4 (Sample complexity and running time of ONESTAGE). *The procedure ONESTAGE (Algorithm 7) has the following performance guarantees:*

- The sample complexity is $2^{\Theta(d \cdot \log d)} \cdot (\log \mathcal{C} + \log \log(F/\eta)) \cdot k \cdot \log(F \cdot T) \cdot \mathcal{D}$.
- The running time is $O(2^{\Theta(d \cdot (\log d + \log \mathcal{C}))} \cdot \log(F \cdot T) \cdot \log \log(F/\eta) \cdot k \cdot (\mathcal{D} + \log k))$.
- The output $\{(v'(\xi), \xi)\}_{\xi \in \text{List}}$ contains at most $O(\mathcal{B}) = 2^{O(d \cdot \log d)} \cdot k$ many candidate tones.

Proof. It follows directly from previous Lemma. \square

7.4 ONESTAGE: performance guarantees

The goal of this section is to prove Lemma 7.5.

Lemma 7.5 (Guarantees of ONESTAGE). *The procedure ONESTAGE (Algorithm 7) has the following performance guarantees. For each true tone (v_i, f_i) , it “succeeds” in LOCATESIGNAL (Algorithm 5) with probability at least 0.99. More specifically, let $S \subseteq H$ denote the set of successful tones in LOCATESIGNAL (Algorithm 5). There exists an injection $\pi : S \mapsto [k]$ such that*

Property I: *Each true tone (v_i, f_i) is estimated well with probability $\Pr[i \in S] \geq 0.9$ and if so, those tones whose signal-to-noise ratio $\rho(i) \geq \mathcal{C}$ (see Definition 7.1) has the estimation error*

$$\|f'_i - f_{\pi(i)}\|_2 \lesssim C_* \frac{1}{\rho_{\pi(i)} \cdot T}.$$

Property II: *For all the successfully recovered tones S , the (partial) tone estimation error*

$$\sum_{i \in S} \mathbf{E}_{\Sigma, b} \left[\frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |v'_i \cdot e^{2\pi i \cdot f'_i{}^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}{}^\top \tau}|^2 \cdot d\tau \right] \lesssim (\mathcal{C}^2 + dC_*^2) \cdot \mathcal{N}^2.$$

Property III: *For each successfully recovered tone $i \in S$, if its signal-to-noise ratio $\rho(i) \geq \mathcal{C}$, the (single) tone estimation error*

$$|v'_i - v_{\pi(i)}| \lesssim (\mathcal{C} + \sqrt{d}C_*) \cdot \mathcal{N}.$$

The performance guarantees on the sample complexity, the duration, the success probability, and the running time are controlled by the counterpart performance guarantees of the subroutine LOCATESIGNAL. For ease of presentation, here we omit the formal proofs of these performance guarantees.

Claim 7.6 (Property I of Lemma 7.5). *For each successfully recovered tone $i \in S$ with large enough signal-to-noise ratio, the frequency estimation error*

$$\|f'_i - f_{\pi(i)}\|_2 \lesssim C_* \frac{1}{\rho_{\pi(i)} \cdot T}.$$

Proof. This follows directly from Lemma 6.5. □

Claim 7.7 (Property II of Lemma 7.5). *For all the successfully recovered tones S , the (partial) tone estimation error*

$$\sum_{i \in S} \mathbf{E}_{\Sigma, b} \left[\frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |v'_i \cdot e^{2\pi i \cdot f'_i{}^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}{}^\top \tau}|^2 \cdot d\tau \right] \lesssim (\mathcal{C}^2 + dC_*^2) \cdot \mathcal{N}^2.$$

Proof. Consider a specific true tone $(v_{\pi(i)}, f_{\pi(i)})$ that $i \in S \subseteq H$, for which neither $E_{\text{coll}}(f_{\pi(i)})$ nor $E_{\text{off}}(f_{\pi(i)})$ happens, and this tone “succeeds” in LOCATESIGNAL (Algorithm 5). Assume w.l.o.g. that the tone frequency is hashed into the bin $j = h_{\Sigma, b}(f_{\pi(i)}) \in [B]^d$ (according to Definition 4.10). For simplicity, we adopt the following notations in this proof:

- $j' = h_{\Sigma, b}(f'_i) \in [B]^d$ is the bin where the estimation frequency f'_i hashed into;

- $v'_i = \hat{u}'_{j'} \cdot e^{-2\pi i \cdot a^\top f'_i} \in \mathbb{C}$ is the estimation magnitude returned by the procedure ESTIMATESIGNAL (Algorithm 6);
- $v''_i = \hat{u}'_{j'} \cdot e^{-2\pi i \cdot a^\top f_i} \in \mathbb{C}$; and
- $\mu_{\pi(i)}^2 = \mathbf{E}_{\Sigma, b, a} [|\hat{u}'_j \cdot e^{-2\pi i \cdot a^\top f_{\pi(i)}} - v_{\pi(i)}|^2] \geq 0$ according to Definition 7.1.

We discuss two cases for the signal-to-noise ratio $\rho_{\pi(i)} = |v_i|/\mu_{\pi(i)}$.

Case (i): when the signal-to-noise ratio $\rho_{\pi(i)} = |v_i|/\mu_{\pi(i)} \geq \mathcal{C}$, i.e. when the premise for Lemma 5.5 holds.

According to Markov inequality, the equation below holds with probability at least $1 - \mathcal{C}^{-2} \geq 0.9999$ (given that $\mathcal{C} \geq 120$; see Definition 5.1). In what follows, We assume that this equation holds.

$$\left| \hat{u}'_j \cdot e^{-2\pi i \cdot a^\top f_{\pi(i)}} - v_{\pi(i)} \right|^2 \leq \mathcal{C}^2 \cdot \mu_i^2. \quad (24)$$

The procedure LOCATESIGNAL and the follow-up procedures have the desired performance guarantees. In particular, we derive a “good” frequency estimation $f'_i \in [-F, F]^d$ from the procedure LOCATESIGNAL. Also, it follows from Lemma 6.5 that

$$\|f'_i - f_{\pi(i)}\|_2 \lesssim C_* \frac{1}{\rho_{\pi(i)} \cdot T}. \quad (25)$$

Since we choose a large enough duration $T = \Omega(\frac{d^{4.5} \log(dk/\delta) \log d}{\eta})$ finally according to Theorem 1.1 and $\rho_{\pi(i)} \geq \mathcal{C} \geq 120$, the ℓ_2 -norm frequency error $\|f'_i - f_{\pi(i)}\|_2$ is sufficiently small and thus Lemma 4.20 is applicable. That is, the tone frequency $f_{\pi(i)}$ and the estimation frequency f'_i are hashed into the same bin $j = j' \in [B]^d$ (see Lemma 4.20). Combining the above arguments together, we have

$$\begin{aligned} & \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |v'_i \cdot e^{2\pi i \cdot f'_i{}^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}{}^\top \tau}|^2 \cdot d\tau \\ &= \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |\hat{u}'_{j'} \cdot e^{-2\pi i \cdot a^\top f'_i} \cdot e^{2\pi i \cdot f'_i{}^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}{}^\top \tau}|^2 \cdot d\tau \\ &= \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |\hat{u}'_j \cdot e^{-2\pi i \cdot a^\top f'_i} \cdot e^{2\pi i \cdot f'_i{}^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}{}^\top \tau}|^2 \cdot d\tau \\ &\leq \underbrace{\frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |\hat{u}'_j \cdot e^{-2\pi i \cdot a^\top f'_i} \cdot e^{2\pi i \cdot f'_i{}^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}{}^\top \tau}|^2 \cdot d\tau}_{A_1(\tau)} \\ &\quad + \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |v_{\pi(i)} \cdot e^{2\pi i \cdot f'_i{}^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}{}^\top \tau}|^2 \cdot d\tau \\ &= A_1(\tau) + 2 \cdot |v_{\pi(i)}|^2 \cdot (1 - \text{sinc}_T(f'_i - f_{\pi(i)})) \\ &\leq A_1(\tau) + 2 \cdot |v_{\pi(i)}|^2 \cdot \left(\frac{\pi^2}{6} \cdot T^2 \cdot \|f'_i - f_{\pi(i)}\|_2^2\right) \\ &\lesssim A_1(\tau) + C_*^2 \cdot |v_{\pi(i)}|^2 / \rho_{\pi(i)}^2 \end{aligned} \quad (26)$$

where the first step is by the definition of v'_i (see Algorithm 6); the second step follows because the tone frequency $f_{\pi(i)}$ and the estimation frequency f'_i are hashed into the same bin $j = j' \in [B]^d$;

the third step applies the triangle inequality; the forth step applies Property II of Lemma 8.4 to the second summand; the fifth step applies Part (e) of Fact 8.2; the last step applies Equation (25).

Then we consider $A_1(\tau)$. We have

$$\begin{aligned}
& A_1(\tau) \\
&= \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |\widehat{u}_j' \cdot e^{-2\pi i \cdot a^\top f_i'} \cdot e^{2\pi i \cdot f_i'^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_i'^\top \tau}|^2 \cdot d\tau \\
&\leq \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |\widehat{u}_j' \cdot e^{-2\pi i \cdot a^\top f_i'} \cdot e^{2\pi i \cdot f_i'^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_i'^\top \tau} \cdot e^{-2\pi i \cdot a^\top (f_i' - f_{\pi(i)})}|^2 \cdot d\tau \\
&\quad + \frac{1}{T^d} \int_{\tau \in [0, T]^d} |v_{\pi(i)} \cdot e^{2\pi i \cdot f_i'^\top \tau} \cdot e^{-2\pi i \cdot a^\top (f_i' - f_{\pi(i)})} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_i'^\top \tau}|^2 \cdot d\tau \\
&= |\widehat{u}_j \cdot e^{-2\pi i \cdot a^\top f_{\pi(i)}} - v_{\pi(i)}|^2 \cdot \frac{1}{T^d} \int_{\tau \in [0, T]^d} |e^{2\pi i \cdot f_i'^\top \tau}|^2 \cdot d\tau \\
&\quad + \frac{1}{T^d} \int_{\tau \in [0, T]^d} |v_{\pi(i)} \cdot e^{2\pi i \cdot f_i'^\top \tau} \cdot e^{-2\pi i \cdot a^\top (f_i' - f_{\pi(i)})} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_i'^\top \tau}|^2 \cdot d\tau \\
&\leq \mathcal{C}^2 \cdot \mu_i^2 + \frac{1}{T^d} \int_{\tau \in [0, T]^d} |v_{\pi(i)} \cdot e^{2\pi i \cdot f_i'^\top \tau} \cdot e^{-2\pi i \cdot a^\top (f_i' - f_{\pi(i)})} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_i'^\top \tau}|^2 \cdot d\tau \\
&= \mathcal{C}^2 \cdot \mu_i^2 + |v_{\pi(i)}|^2 \cdot |e^{-2\pi i \cdot a^\top (f_i' - f_{\pi(i)})} - 1|^2 \\
&\lesssim \mathcal{C}^2 \cdot \mu_i^2 + |v_{\pi(i)}|^2 \cdot |a^\top (f_i' - f_{\pi(i)})|^2 \\
&\lesssim \mathcal{C}^2 \cdot \mu_i^2 + |v_{\pi(i)}|^2 \cdot \|a\|_2 \cdot \|f_i' - f_{\pi(i)}\|_2 \\
&\lesssim \mathcal{C}^2 \cdot \mu_i^2 + dC_*^2 |v_{\pi(i)}|^2 / \rho_{\pi(i)}^2
\end{aligned} \tag{27}$$

where the second step is by triangle inequality; the fourth step is by the integral equals to 1 and follows from Equation (24); the sixth step is because $|e^{2\pi i x} - 1| \leq |x|$; the last step is because $\|a\|_2 \leq \sqrt{dT}$ (see Algorithm 3) and $\|f_i' - f_{\pi(i)}\|_2 \lesssim C_*/(T\rho_{\pi(i)})$, which implies $|a^\top (f_i' - f_{\pi(i)})| \lesssim \sqrt{d}C_*/\rho_{\pi(i)}$.

Combine Equation (26) and Equation (27) together, we prove that

$$\begin{aligned}
\frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |v_i' \cdot e^{2\pi i \cdot f_i'^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}^\top \tau}|^2 \cdot d\tau &\lesssim \mathcal{C}^2 \cdot \mu_i^2 + dC_*^2 |v_{\pi(i)}|^2 / \rho_{\pi(i)}^2 \\
&\lesssim (\mathcal{C}^2 + dC_*^2) \cdot \mu_i^2
\end{aligned}$$

The last step follows because the signal-to-noise ratio $\rho_{\pi(i)} = |v_{\pi(i)}|/\mu_{\pi(i)}$ (see Definition 7.1). This accomplishes Case (i).

Case (ii): when the signal-to-noise ratio $\rho_{\pi(i)} = |v_{\pi(i)}|/\mu_{\pi(i)} \leq \mathcal{C}$, i.e. when the premise for Lemma 5.5 does not hold. Under this case, we can use $(0, f_i')$ to recover the true tone (v_i, f_i) , as

$$\frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |v_i' \cdot e^{2\pi i \cdot f_i'^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}^\top \tau}|^2 \cdot d\tau = |v_{\pi(i)}|^2 \leq \mathcal{C}^2 \cdot \mu_i^2.$$

Recall that we add a supplementary list $List_{sup}$ in the Line 9 of ONESTAGE, and there are enough candidates with zero magnitude and minimum separation in the list. We can let S include some tones $(0, \xi)$ from $List_{sup}$ when needed.

This completes the proof. \square

Claim 7.8 (Property III of Lemma 7.5). *For each successfully recovered tone $i \in S$, the (single) tone estimation error*

$$|v_i' - v_{\pi(i)}| \lesssim (\mathcal{C} + \sqrt{d}C_*) \cdot \mathcal{N}.$$

Proof. This can be directly inferred from Claim 7.7. \square

7.5 MULTISTAGE

The goal of this section is to prove Lemma 7.9.

Lemma 7.9 (MULTISTAGE). *The procedure MULTISTAGE (Algorithm 8) satisfies the following:*

- *The sample complexity is $\mathcal{R}_{\text{merge}}$ times the sample complexity of ONESTAGE (Algorithm 7).*
- *The running time is $\mathcal{R}_{\text{merge}}$ times the running time of ONESTAGE (Algorithm 7).*

Proof. All these properties can be easily inferred from Lemma 7.4. \square

Algorithm 8 MULTISTAGE

```

1: procedure MULTISTAGE( $x, D, \mathcal{C}, T, \mathcal{R}_{\text{merge}}$ ) ▷ Lemma 7.9
2:   Let  $\text{List}^* \leftarrow \emptyset$ .
3:   for  $r = 1, 2, \dots, \mathcal{R}_{\text{merge}}$  do
4:     Sample  $\Sigma \in \mathbb{R}^{d \times d}$  and  $b \in \mathbb{R}^d$  according to Definition 4.7.
5:      $\text{List}_{\text{new}} \leftarrow \text{ONESTAGE}(x, \Sigma, b, D, \mathcal{C}, T)$ . ▷ Algorithm 7
6:      $\text{List}^* \leftarrow \text{List}^* \cup \text{List}_{\text{new}}$ .
7:   end for
8:   return the tones  $\text{List}^*$ .
9: end procedure

```

Lemma 7.10 (Guarantees of MULTISTAGE). *The procedure MULTISTAGE (Algorithm 8) repeat the procedure ONESTAGE (Algorithm 7) for $\mathcal{R}_{\text{merge}} = \Theta(d \cdot \log d \cdot \log k)$ times, and returns a set $\text{List} = \{(v'_i, f'_i)\}_{i \in [m]}$ of $m = |\text{List}| = 2^{O(d \cdot \log d)} \cdot k \cdot \log k \in \mathbb{N}_{\geq 1}$ many candidate tones. With probability at least $1 - 1/\text{poly}(k)$, there are at least $\mathcal{R}_{\text{merge}}$ different disjoint subsets $\{S_r\}$ of List , where for each r we have that $S_r \subset \text{List}$ and $|S_r| \leq k$, and for $r \neq r'$ we have that $S_r \cap S_{r'} = \emptyset$. For each S_r , there is a injective projection $\pi_r : S_r \rightarrow [k]$ and has the following properties:*

Property I: *For each true tone (v_i, f_i) , $\Pr[i \in S_r] \geq 0.9$ and if the signal-to-noise ratio is large enough, the frequency estimation error*

$$\|f'_i - f_{\pi_r(i)}\|_2 \lesssim C_* \frac{1}{\rho_{\pi_r(i)} \cdot T}.$$

Property II: *For all the successfully recovered tones S_r , the (partial) tone estimation error*

$$\sum_{i \in S_r} \mathbf{E}_{\Sigma, b} \left[\frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |v'_i \cdot e^{2\pi i \cdot f'_i{}^\top \tau} - v_{\pi_r(i)} \cdot e^{2\pi i \cdot f_{\pi_r(i)}{}^\top \tau}|^2 \cdot d\tau \right] \lesssim (C^2 + dC_*^2) \cdot \mathcal{N}^2.$$

Property III: *For each successfully recovered tone $i \in S_r$, if its signal-to-noise ratio is large enough, then we can bound the (single) tone estimation error*

$$|v'_i - v_{\pi_r(i)}| \lesssim (C + \sqrt{d}C_*) \cdot \mathcal{N}.$$

Proof. This lemma can be proved directly by Lemma 7.5. \square

7.6 MERGEDSTAGE: algorithm and running time

The goal of this section is to prove Lemma 7.11. The concerning Algorithm 9 (MERGEDSTAGE) is demonstrated in Figure 16.

Lemma 7.11 (MERGEDSTAGE, Input size and Running time). *The procedure MERGEDSTAGE (Algorithm 9) has the following properties:*

- The input $List = \{(v_j, f_j)\}_{j \in [m]}$ is a multi-set of $m = |List| = 2^{O(d \cdot \log d)} \cdot k \cdot \mathcal{R}_{\text{merge}}$ many candidate tones, where $\mathcal{R}_{\text{merge}} = \Theta(d \cdot \log d \cdot \log k)$ is sufficiently large.
- The running time is $2^{O(d \cdot \log d)} \cdot k \cdot \mathcal{R}_{\text{merge}} \cdot \log^d(k \cdot \mathcal{R}_{\text{merge}}) = 2^{O(d \cdot \log d)} \cdot k \cdot \log^{O(d)} k$.

Proof. The first property about the input $List$ is guaranteed by Lemma 7.9.

The second property is proved by using a well-known data-structure. We use a textbook d -dimensional range tree data-structure (see section 5 in [KSBO00]).

Theorem 7.12 (Theorem 5.11 in [KSBO00]). *Let P be a set of n points in d -dimensional space, with $d \geq 2$. A layered range tree for P uses $O(n \log^{d-1} n)$ storage and it can be constructed in $O(n \log^{d-1} n)$ time. With this range tree one can report the points in P that lie in a rectangular query range in $O(\log^{d-1} n + q)$ time, where q is the number of reported points.*

The above theorem works for ℓ_∞ -norm. By choosing

$$n = m = |List|$$

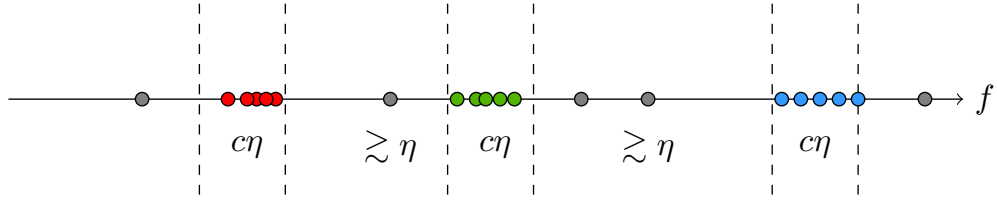
we complete the proof of running time. □

Algorithm 9 MERGEDSTAGE, Lemmas 7.11 and 7.14

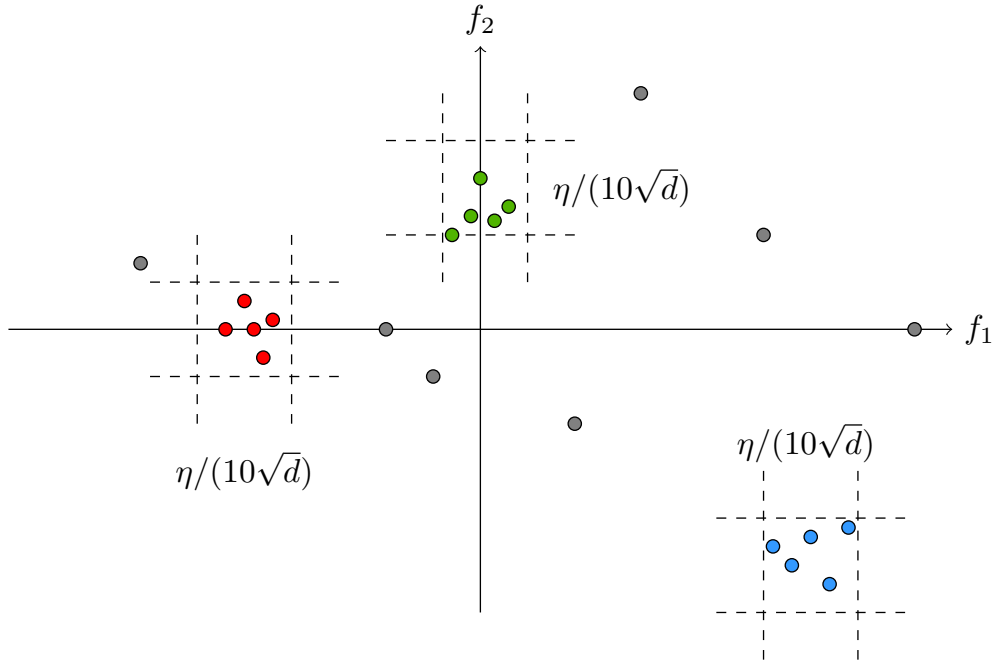
```

1: procedure MERGEDSTAGE( $List, \mathcal{R}_{\text{merge}}$ )
2:   Denote  $List = \{(v_j, f_j)\}_{j \in [m]}$  for  $m = |List|$ .
3:   Build a  $d$ -dimensional segment tree TREE on the frequencies  $\{f_j\}_{j \in [m]} \subseteq [-F, F]^d$ .
4:   All these frequencies  $\{f_j\}_{j \in [m]}$  are unmarked.
5:    $List^* \leftarrow \emptyset$ .
6:   while TREE has at least one unmarked frequency do
7:     Choose an arbitrary unmarked frequency  $\xi_i$  from TREE.
8:     if TREE.COUNT( $\mathbf{HC}(\xi_i, \eta/d^3)$ )  $\geq 8/10 \cdot \mathcal{R}_{\text{merge}}$  then ▷ Theorem 7.12
9:        $f^* \leftarrow \xi_i$ .
10:       $v^* \leftarrow \text{median}\{v_j : j \in [m] \text{ and } f_j \in \mathbf{HC}(\xi_i, \eta/d^3)\}$ .
11:       $List^* \leftarrow List^* \cup (v^*, f^*)$ .
12:      Delete  $\{f_j : j \in [m] \text{ and } f_j \in \mathbf{HC}(\xi_i, \eta/(10\sqrt{d}))\}$  from TREE.
13:      Delete  $\{(v_i, f_j) : j \in [m] \text{ and } f_j \in \mathbf{HC}(\xi_i, \eta/(10\sqrt{d}))\}$  from  $List$ .
14:     else
15:       Mark the chosen frequency  $\xi_i$  in TREE.
16:     end if
17:   end while
18:   return the tones  $List^*$ .
19: end procedure

```



(a) MERGEDSTAGE in one dimension



(b) MERGEDSTAGE in two dimensions

Figure 16: Demonstration for Algorithm 9 in one dimension ($d = 1$) and two dimensions ($d = 2$).

7.7 MERGEDSTAGE: performance guarantees

The goal of this section is to prove Lemma 7.14.

Claim 7.13 (Approximate formula for tone-wise error). *The following holds for any pair of tones $(v, f) \in \mathbb{C} \times \mathbb{R}^d$ and $(v^*, f^*) \in \mathbb{C} \times \mathbb{R}^d$:*

$$\begin{aligned} \text{err} &:= \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |v \cdot e^{2\pi i \cdot f^\top \tau} - v^* \cdot e^{2\pi i \cdot f^{*\top} \tau}|^2 \cdot d\tau \\ &\approx |v - v^*|^2 + |v^*|^2 \cdot (1 - \text{sinc}_T(f - f^*)) \\ &\approx |v - v^*|^2 + |v^*|^2 \cdot \min\{1, T^2 \cdot \|f - f^*\|_2^2\}. \end{aligned}$$

Proof. We first prove the second part of the claim, which is equivalent to

$$1 - \text{sinc}_T(f - f^*) \approx \min\{1, T^2 \cdot \|f - f^*\|_2^2\}. \quad (28)$$

Indeed, when $T^2 \cdot \|f - f^*\|_2^2 \geq (\frac{2.05}{\pi})^2 \approx 1$, we know from Part (d) of Fact 8.2 that

$$1 - \text{sinc}_T(f - f^*) = 1 \pm \frac{1}{2} \approx 1.$$

And when $T^2 \cdot \|f - f^*\|_2^2 < (\frac{2.05}{\pi})^2 \approx 1$, we know from Part (c) of Fact 8.2 that

$$\begin{aligned} 1 - \text{sinc}_T(f - f^*) &\geq 1 - \exp\left(-(\pi^2/6) \cdot T^2 \cdot \|f - f^*\|_2^2\right) \\ &\gtrsim T^2 \cdot \|f - f^*\|_2^2, \end{aligned}$$

and that

$$\begin{aligned} 1 - \text{sinc}_T(f - f^*) &\leq 1 - \exp\left(-(\pi^2/5) \cdot T^2 \cdot \|f - f^*\|_2^2\right) \\ &\lesssim T^2 \cdot \|f - f^*\|_2^2, \end{aligned}$$

Combining the above arguments together implies Equation (28).

In what follows, we prove the first part of the claim that

$$\text{err} \approx |v - v^*|^2 + |v^*|^2 \cdot (1 - \text{sinc}_T(f - f^*)). \quad (29)$$

We know Property II of Lemma 8.4 that

$$\text{err} = |v|^2 + |v^*|^2 - (v \cdot \overline{v^*} + \overline{v} \cdot v^*) \cdot \text{sinc}_T(f - f^*).$$

For brevity, we denote $w_1 \cdot e^{i\theta} = v/v^*$ for some norm $w_1 \geq 0$ and some phase $\theta \in [0, 2\pi)$, and denote $w_2 = \text{sinc}_T(f - f^*) \in [-\frac{1}{4}, 1]$ (see Part (e) of Fact 8.2). We notice that the formula

$$|v - v^*|^2 + |v^*|^2 \cdot (1 - \text{sinc}_T(f - f^*))$$

is non-negative. As a consequence, to verify Equation (29), it suffices to show that the following function $L(w_1, w_2, \theta) \approx 1$, for any $w_1 \geq 0$, any $w_2 \in [-\frac{1}{4}, 1]$ and any $\theta \in [0, 2\pi)$:

$$\begin{aligned} L(w_1, w_2, \theta) &:= \frac{\text{err}}{\text{RHS of (29)}} \\ &= \frac{|v^* \cdot w_1 \cdot e^{i\theta}|^2 + |v^*|^2 - 2 \cdot |v^* \cdot w_1 \cdot e^{i\theta}| \cdot |v^*| \cdot \cos(\theta) \cdot w_2}{|v^* \cdot w_1 \cdot e^{i\theta} - v^*|^2 + |v^*|^2 \cdot (1 - w_2)} \end{aligned}$$

$$\begin{aligned}
&= \frac{w_1^2 + 1 - 2 \cdot w_1 \cdot w_2 \cdot \cos(\theta)}{|w_1 \cdot e^{i \cdot \theta} - 1|^2 + 1 - w_2} \\
&= \frac{w_1^2 + 1 - 2 \cdot w_1 \cdot w_2 \cdot \cos(\theta)}{w_1^2 - w_2 + 2 - 2 \cdot w_1 \cos(\theta)},
\end{aligned}$$

where the second step is by the definition of w_1 , w_2 and θ ; the third step divides both the numerator and the denominator by $|v'_i|^2$; and the last step can be seen via elementary calculation.

Let us investigate the partial derivative $\frac{\partial L}{\partial \theta}$ in $\theta \in [0, 2\pi)$:

$$\begin{aligned}
\frac{\partial L}{\partial \theta} &= \frac{2 \cdot w_1 \cdot w_2 \cdot \sin(\theta)}{w_1^2 - w_2 + 2 - 2 \cdot w_1 \cos(\theta)} - \frac{w_1^2 + 1 - 2 \cdot w_1 \cdot w_2 \cdot \cos(\theta)}{(w_1^2 - w_2 + 2 - 2 \cdot w_1 \cos(\theta))^2} \cdot 2 \cdot w_1 \cdot \sin(\theta) \\
&= -\sin(\theta) \cdot 2 \cdot w_1 \cdot \underbrace{\frac{w_1^2 \cdot (1 - w_2) + (1 - w_2)^2}{(w_1^2 - w_2 + 2 - 2 \cdot w_1 \cos(\theta))^2}}_{A_4},
\end{aligned}$$

where the second step can be seen via elementary calculation.

Because $w_1 \geq 0$ and $w_2 \in [-\frac{1}{4}, 1]$, we must have $A_4 \geq 0$. As a result, for any fixed w_1 and w_2 , the function $L(w_1, w_2, \theta)$ is non-increasing when $\theta \in [0, \pi]$, and is non-decreasing when $\theta \in [\pi, 2\pi)$. The functions $L_{\min}(w_1, w_2) := \min_{\theta \in [0, 2\pi)} L(w_1, w_2, \theta)$ and $L_{\max}(w_1, w_2) := \max_{\theta \in [0, 2\pi)} L(w_1, w_2, \theta)$ for any $w_1 \geq 0$ and any $w_2 \in [-\frac{1}{4}, 1]$ are given by

$$\begin{aligned}
L_{\min}(w_1, w_2) &= L(w_1, w_2, \pi) = \frac{A_5(w_1, w_2)}{A_6(w_1, w_2)}, \\
A_5(w_1, w_2) &:= w_1^2 + 1 + 2 \cdot w_1 \cdot w_2, \\
A_6(w_1, w_2) &:= w_1^2 + 2 \cdot w_1 - w_2 + 2,
\end{aligned}$$

and

$$\begin{aligned}
L_{\max}(w_1, w_2) &= L(w_1, w_2, 0) = \frac{A_7(w_1, w_2)}{A_8(w_1, w_2)}, \\
A_7(w_1, w_2) &:= w_1^2 + 1 - 2 \cdot w_1 \cdot w_2, \\
A_8(w_1, w_2) &:= w_1^2 - 2 \cdot w_1 - w_2 + 2.
\end{aligned}$$

We now justify the lower-bound part of Equation (29) by exploring the function $L_{\min}(w_1, w_2)$. For any fixed $w_1 \geq 1$, the numerator $A_5(w_1, w_2)$ is a non-decreasing function in $w_2 \in [-\frac{1}{4}, 1]$, while the denominator $A_6(w_1, w_2)$ is a non-increasing *non-negative* function in $w_2 \in [-\frac{1}{4}, 1]$. Given these, we can infer the lower-bound part of Equation (29) as follows:

$$\begin{aligned}
L(w_1, w_2, \theta) &\geq \min_{w_1 \in [0, 1]} \min_{w_2 \in [-\frac{1}{4}, 1]} L_{\min}(w_1, w_2) \\
&= \min_{w_1 \in [0, 1]} L_{\min}(w_1, -1/4) \\
&= \min_{w_1 \in [0, 1]} \frac{w_1^2 + 1 - (1/2) \cdot w_1}{w_1^2 + 2 \cdot w_1 - (1/4) + 2} \\
&\approx 0.3107,
\end{aligned}$$

where the last step can be seen via numeric calculation.

We next show the upper-bound part of Equation (29) by exploring the function $L_{\max}(w_1, w_2)$. For any fixed $w_1 \geq 1$, both of the numerator $A_7(w_1, w_2)$ and the denominator $A_8(w_1, w_2)$ are *linear*

functions in $w_2 \in [-\frac{1}{4}, 1]$. Accordingly, $L_{\max}(w_1, w_2)$ itself is a monotone function in $w_2 \in [-\frac{1}{4}, 1]$. We can infer the upper-bound part of Equation (29) as follows:

$$\begin{aligned} \max_{w_1 \in [0,1]} \max_{w_2 \in [-\frac{1}{4}, 1]} L_{\max}(w_1, w_2) &= \max_{w_1 \in [0,1]} \max \{L_{\max}(w_1, -1/4), L_{\max}(w_1, 1)\} \\ &= \max_{w_1 \in [0,1]} \max \left\{ \frac{w_1^2 + 1 + (1/2) \cdot w_1}{w_1^2 - 2 \cdot w_1 + (9/4)}, 1 \right\} \\ &\approx 2.7247, \end{aligned}$$

where the last step can be seen via numeric calculation.

This completes the proof. \square

Lemma 7.14 (Guarantees for MERGEDSTAGE). *The procedures MERGEDSTAGE (Algorithm 9) returns a set $\mathcal{List} = \{(v'_i, f'_i)\}_{i \in [m]}$ of $m = |\mathcal{List}| = 2^{O(d \cdot \log d)} \cdot k \in \mathbb{N}_{\geq 1}$ many candidate tones. With probability at least $1 - 1/\text{poly}(k)$, the outputs $\mathcal{List} = \{(v'_i, f'_i)\}_{i \in [m]}$ satisfies the following:*

Property I: *The set size $m = 2^{O(d \cdot \log d)} \cdot k$, and the frequency separation*

$$\min_{i, j \in [m]} \|f'_i - f'_j\|_2 \gtrsim \eta/\sqrt{d}.$$

Property II: *For the true tones $\{(v_i, f_i)\}_{i \in [k]}$, there is an injection $\pi : [k] \mapsto [m]$ such that*

$$\sum_{i \in [k]} \frac{1}{Td} \cdot \int_{\tau \in [0, T]^d} \left| v'_{\pi(i)} \cdot e^{2\pi i \cdot f'_{\pi(i)} \tau} - v_i \cdot e^{2\pi i \cdot f_i \tau} \right|^2 \cdot d\tau \leq (C^2 + dC_*^2) \cdot \mathcal{N}^2.$$

Proof. The size of output is straightforward from Algorithm 9. If we add one candidate tone into the output, we will delete at least $8/10 \cdot \mathcal{R}_{\text{merge}}$ tones.

Property I: The set size can be induced from proof of the Property II. As for the frequency separation, it comes from that if we choose to take the median of $\mathbf{HC}(\xi_i, \eta/d^3)$ for ξ_i , we will clear a larger region $\mathbf{HC}(\xi_i, \eta/(10\sqrt{d}))$.

It is safe to clear the larger region, as we have an assumption that $\min_{i \neq j} \|f_i - f_j\|_2 \geq \eta$, which implies that $\min_{i \neq j} \|f_i - f_j\|_\infty \geq \eta/\sqrt{d}$ for true tones $\{(v_i, f_i)\}$. Suppose ξ_i is a successful recovery of true tone f_i . Then if we find a cluster of successful recovered tones $\mathbf{HC}(\xi_i, \eta/d^3)$, for all other successful recovered tones ξ_j where $j \neq i$, we have that

$$\begin{aligned} \|\xi_i - \xi_j\|_\infty &= \|\xi_i - f_i + f_i - f_j + f_j - \xi_j\|_\infty \\ &\geq \|f_i - f_j\|_\infty - \|\xi_i - f_i\|_\infty - \|f_j - \xi_j\|_\infty \\ &\geq \|f_i - f_j\|_2/\sqrt{d} - \|\xi_i - f_i\|_2 - \|f_j - \xi_j\|_2 \\ &\gtrsim \eta/\sqrt{d} - 2C_*/(\rho T) \\ &\gtrsim \eta/\sqrt{d} - 2C_*/T \\ &\gtrsim \eta/\sqrt{d} \end{aligned}$$

where the second step follows from triangle inequality, the third step follows from $\|\cdot\|_2/\sqrt{d} \leq \|\cdot\|_\infty \leq \|\cdot\|_2$, the last step follows from $T \geq C_*\sqrt{d}/\eta$.

This means that $\xi_j \notin \mathbf{HC}(\xi_i, \eta/(10\sqrt{d}))$ and proves the safety of the operation.

Property II:

For each true tone (v_i, f_i) , by Lemma 7.10, with probability at least $1 - 1/\text{poly}(k)$, there are at least $0.8\mathcal{R}_{\text{merge}}$ a successful recovery $\{(v'_i, f'_i)\}$ of it, where $\|f'_i - f_i\|_2 \lesssim C_*/(\rho T)$. By the choice of duration $T = \Omega(d^3 \cdot \eta^{-1} \cdot \log(kd/\delta))$ by Lemma 5.14, we know that $\|f'_i - f_i\|_\infty \leq \|f'_i - f_i\|_2 \ll \eta/d^3$. And let $\mu^2(f_i)$ denote the expected error of successful recovery (v'_i, f'_i) :

$$\mu^2(f_i) = \mathbf{E}_{\Sigma, b, v'_i, f'_i} \left[\frac{1}{T^d} \int_{t \in [0, T]^d} |v'_i \cdot e^{2\pi i \cdot f'_i{}^\top t} - v_i \cdot e^{2\pi i \cdot f_i{}^\top t}|^2 \cdot dt \right]$$

Then by Markov Inequality, we know that

$$\Pr \left[\int_{t \in [0, T]^d} |v'_i \cdot e^{2\pi i \cdot f'_i{}^\top t} - v_i \cdot e^{2\pi i \cdot f_i{}^\top t}|^2 \cdot dt \geq 10\mu^2(f_i) \right] \leq 1/10. \quad (30)$$

By Lemma 7.5, we can bound the summation of expected errors of successful recovery:

$$\sum_{i \in [k]} \mu^2(f_i) \lesssim (\mathcal{C}^2 + dC_*^2) \cdot \mathcal{N}^2. \quad (31)$$

As a summary, for each true tone (v_i, f_i) , we have shown that there are at least $0.8\mathcal{R}_{\text{merge}}$ successful recovery $\{(v'_i, f'_i)\}$ of it, ie. $\text{TREE.COUNT}(\mathbf{HC}(\xi_i, \eta/d^3)) \geq 8/10 \cdot \mathcal{R}_{\text{merge}}$. Then we will take the any frequency f_i^* in $\mathbf{HC}(f_i, \eta/d^3)$ in Line 9 and coordinate-wise median of magnitude v_i^* of successful recovery in $\mathbf{HC}(f_i, \eta/d^3)$ in Line 10.

Among the successful recovery $\{(v'_i, \xi_i) | f'_i \in \mathbf{HC}(\xi_i, \eta/d^3)\}$ of (v_i, f_i) , with probability $1 - 1/\text{poly}(k)$, at least half of them will have error less than $10\mu^2(f_i)$. Note that $f^* = \xi_i$.¹⁴ To be more specific, with probability at least $1 - 1/\text{poly}(k)$,

$$\frac{1}{T^d} \cdot \int_{t \in [0, T]^d} |v_i^* \cdot e^{2\pi i \cdot f_i^{*\top} t} - v_i \cdot e^{2\pi i \cdot f_i{}^\top t}|^2 dt \lesssim \mu^2(f_i),$$

Then we have

$$\sum_{i \in [k]} \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |v'_i \cdot e^{2\pi i \cdot f'_i{}^\top \tau} - v_i \cdot e^{2\pi i \cdot f_i{}^\top \tau}|^2 \cdot d\tau \lesssim \sum_{i \in [k]} \mu^2(f_i) \leq (\mathcal{C}^2 + dC_*^2) \cdot \mathcal{N}^2.$$

This completes the proof. \square

7.8 Running MERGEDSTAGE twice

The goal of this section is to prove Lemma 7.16.

Definition 7.15 (Setup for RECOVERYSTAGE). Given two sets

$$\mathcal{L}ist_1^* = \{(v'_i, f'_i)\}_{i \in [k']} \quad \text{and} \quad \mathcal{L}ist_2^* = \{(v''_i, f''_i)\}_{i \in [k'']}$$

of sizes $k', k'' = 2^{O(d \cdot \log d)} \cdot k \in \mathbb{N}_{\geq 1}$, output each pair (v''_i, f''_i) in the second set (for $i \in [k'']$) that has a small frequency distance $\|f''_i - f'_j\|_2 \leq c/T$, against some frequency f'_j in the first set (for $j \in [k']$). Denote the resulting set by $\{(v''_i, f''_i)\}_{i \in S} \subseteq \{(v''_i, f''_i)\}_{i \in [k'']}$ of size $|S| = k^* \leq k'' = 2^{O(d \cdot \log d)} \cdot k$.

¹⁴Note that we only need to take coordinate wise median for v , for frequency f , using ξ_i is good enough. Since ξ_i is close to the true f .

Algorithm 10 RECOVERYSTAGE

```

1: procedure RECOVERYSTAGE( $x, D, \mathcal{C}, T$ ) ▷ Theorem 7.18
2:    $\mathcal{R}_{\text{merge}} \leftarrow \Theta(d \cdot \log d \cdot \log k)$ .
3:    $\text{List}' \leftarrow \text{MULTISTAGE}(x, D, \mathcal{C}, T, \mathcal{R}_{\text{merge}})$ . ▷ Algorithm 8
4:    $\text{List}_1^* \leftarrow \text{MERGEDSTAGE}(\text{List}', \mathcal{R}_{\text{merge}})$ . ▷ Algorithm 9
5:    $\text{List}'' \leftarrow \text{MULTISTAGE}(x, D, \mathcal{C}, T, \mathcal{R}_{\text{merge}})$ . ▷ Algorithm 8
6:    $\text{List}_2^* \leftarrow \text{MERGEDSTAGE}(\text{List}'', \mathcal{R}_{\text{merge}})$ . ▷ Algorithm 9
7:   Derive  $\text{List}^*$  from  $\text{List}_1^*$  and  $\text{List}_2^*$  according to Definition 7.15. ▷ Lemma 7.16
8:   Sort  $\text{List}^* = \{(v_i^*, f_i^*)\}_{i=1}^{|\text{List}^*|}$  in decreasing order of magnitudes  $|v_i^*|$ .
9:    $\text{List}_{[k]}^* \leftarrow$  the top- $k$  tones  $\{(v_i^*, f_i^*)\}_{i=1}^k$  in  $\text{List}^*$ 
10:  return  $\text{List}_{[k]}^*$ .
11: end procedure

```

Lemma 7.16 (Running MERGEDSTAGE twice). *Given two sets $\{(v'_i, f'_i)\}_{i \in [k']}$ and $\{(v''_i, f''_i)\}_{i \in [k']}$ of sizes $k', k'' = 2^{O(d \log d)} \cdot k \in \mathbb{N}_{\geq 1}$, assume w.l.o.g. that Definition 7.15 selects $k^* \leq k''$ pairs $\{(v''_i, f''_i)\}_{i \in [k^*]}$ of the second set, then these $k^* = 2^{O(d \log d)} \cdot k$ pairs can be reindexed such that*

$$\sum_{i \in [k]} \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |v_i \cdot e^{2\pi i \cdot f_i^\top \tau} - v'_i \cdot e^{2\pi i \cdot f'_i{}^\top \tau}|^2 \cdot d\tau + \sum_{i \in [k^*] \setminus [k]} |v'_i|^2 \lesssim \mathcal{C}^2 \cdot \mathcal{N}^2.$$

Proof. By Claim 7.13, the following holds for any pair of tones $(v, f) \in \mathbb{C} \times \mathbb{R}^d$ and $(v^*, f^*) \in \mathbb{C} \times \mathbb{R}^d$:

$$\begin{aligned} \text{err}((v, f), (v^*, f^*)) &= \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} |v \cdot e^{2\pi i \cdot f^\top \tau} - v^* \cdot e^{2\pi i \cdot f^{*\top} \tau}|^2 \cdot d\tau \\ &\approx |v - v^*|^2 + (|v^*|^2 + |v|^2) \cdot \min\{1, T^2 \cdot \|f - f^*\|_2^2\} \end{aligned}$$

Then by Lemma 7.14, with probability at least $1 - 1/\text{poly}(k)$, there is a permutation of the output of the first run $\{(v'_i, f'_i)\}_{i \in [k']}$ and an injective projection $\pi : [k] \rightarrow [k]$, subject to

$$\sum_{i=1}^k ((|v'_i|^2 + |v_{\pi(i)}|^2) \cdot \min\{1, T^2 \cdot \|f'_i - f_{\pi(i)}\|_2^2\} + |v'_i - v_{\pi(i)}|^2) \lesssim (\mathcal{C}^2 + dC_*^2) \mathcal{N}^2$$

If $\|f'_i - f_{\pi(i)}\|_2 > 1/T$, then $\text{err}((0, f'_i), (v_{\pi(i)}, f_{\pi(i)})) \leq \text{err}((v'_i, f'_i), (v_{\pi(i)}, f_{\pi(i)}))$. Let $S = \{i \in [k] : \|f'_i - f_{\pi(i)}\|_2 \leq c/T\}$ for any $c = O(1)$. We can rewrite the result:

$$\begin{aligned} &\sum_{i \in S} ((|v'_i|^2 + |v_{\pi(i)}|^2) \cdot \min\{1, T^2 \cdot \|f'_i - f_{\pi(i)}\|_2^2\} + |v'_i - v_{\pi(i)}|^2) + \sum_{i \in [k] \setminus S} (|v_{\pi(i)}|^2 + |v'_i|^2) \\ &\lesssim (\mathcal{C}^2 + dC_*^2) \mathcal{N}^2. \end{aligned}$$

If we can know the set S and the right permutation of the output of the first run, we can output a set of tones that meet this lemma easily. But the problem is that we do not have the information. This is why we run the MERGEDSTAGE twice. Recall that the signal x^* we want to recover is defined by $\{v_i, f_i\}_{i=1}^k$, then it is equivalent to define x^* by $\{v_i, f_i\}_{i=1}^k \cup \{0, f'_i\}_{i=1}^{k'}$, where f'_i is the output of the first run of MERGEDSTAGE. Then the number of frequencies is $2^{O(d \log d)} k$ and the separation gap is $\Omega(\eta/\sqrt{d})$, then Lemma 7.14 applies again.

Define

$$S' = \{i \in [k''] : \exists j \in [k'], \|f'_j - f''_i\|_2 \leq 1/T\},$$

and we can reindex $\{v_i'', f_i''\}$ such that:

$$\begin{aligned}
(\mathcal{C}^2 + dC_*^2)\mathcal{N}^2 &\gtrsim \sum_{S' \cap [k]} ((|v_i''|^2 + |v_i|^2) \min\{1, T^2 \cdot \|f_i'' - f_i\|_2^2\} + (|v_i''|^2 + |v_{\pi(i)}|^2)) \\
&\quad + \sum_{i \in S' \setminus [k]} ((0^2 + |v_i''|^2) \cdot \min\{1, T^2 \|f_i'' - f_i'\|_2^2\} + |v_i'' - 0|^2) \\
&\quad + \sum_{i \in [k] \setminus S'} (|v_i|^2 + |v_i''|^2) + \sum_{i \in [k'] \setminus (S' \cup [k])} |v_i''|^2 \\
&\geq \sum_{S' \cap [k]} ((|v_i''|^2 + |v_i|^2) \min\{1, T^2 \cdot \|f_i'' - f_i\|_2^2\} + (|v_i''|^2 + |v_{\pi(i)}|^2)) \\
&\quad + \sum_{i \in S' \setminus [k]} |v_i''|^2 + \sum_{i \in [k] \setminus S'} |v_i|^2
\end{aligned}$$

This is exactly the summation of error of $\{(v_i'', f_i'')\}_{i \in S'}$, and $|S'| = k^* \leq k'' = 2^{O(d \log d)} k$ which complete the proof. \square

7.9 RECOVERYSTAGE

The goal of this section is to prove Theorem 7.18. Before the proof of the main result, we need the following lemma:

Lemma 7.17. *The following holds for any three tones $(v_{\pi(i)}, f_{\pi(i)}) \in \mathbb{C} \times \mathbb{R}^d$ and $(v_i^*, f_i^*) \in \mathbb{C} \times \mathbb{R}^d$ and $(v_i', f_i') \in \mathbb{C} \times \mathbb{R}^d$ that $|v_i'| \gtrsim |v_i^*| \approx |v_{\pi(i)}|$:*

$$\begin{aligned}
&\underbrace{\frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} \left| v_i' \cdot e^{2\pi i \cdot f_i'^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}^\top \tau} \right|^2 \cdot d\tau}_{A_1} \\
&\lesssim \underbrace{\frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} \left| v_i^* \cdot e^{2\pi i \cdot f_i^{*\top} \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}^\top \tau} \right|^2 \cdot d\tau}_{A_2} + |v_i'|^2.
\end{aligned}$$

Proof. We will show in Property II of Lemma 8.4 (see Section 8.3) that

$$\begin{aligned}
A_1 &= |v_i'|^2 + |v_{\pi(i)}|^2 - (v_i' \cdot \overline{v_{\pi(i)}} + \overline{v_i'} \cdot v_{\pi(i)}) \cdot \text{sinc}_T(f_i' - f_{\pi(i)}), \\
A_2 &= |v_i^*|^2 + |v_{\pi(i)}|^2 - (v_i^* \cdot \overline{v_{\pi(i)}} + \overline{v_i^*} \cdot v_{\pi(i)}) \cdot \text{sinc}_T(f_i^* - f_{\pi(i)}).
\end{aligned}$$

Because $|v_i'| \gtrsim |v_i^*| \approx |v_{\pi(i)}|$, we can easily verify the lemma by elementary calculation (notice that $|\text{sinc}_T(f_i' - f_{\pi(i)})| \leq 1$ and $|\text{sinc}_T(f_i^* - f_{\pi(i)})| \leq 1$ for any $f_{\pi(i)}, f_i^*, f_i' \in \mathbb{R}^d$). \square

Theorem 7.18 (RECOVERYSTAGE, formal of Theorem 1.1). *Let*

$$T \geq \frac{d^{4.5} \log(kd/\delta) \log d}{\eta}.$$

Let \mathcal{C} be some universal constant and $C_ = d^2$. The procedure RECOVERYSTAGE (Algorithm 10) takes*

$$2^{\Theta(d \log d)} \cdot k \cdot \log^{d+1}(k/\delta) \cdot \log(F/\eta) \cdot \log \log(F/\eta)$$

samples over $[0, T]$, runs in

$$2^{O(d \cdot \log d)} \cdot k \cdot \log^{O(d)}(k/\delta) \cdot \log(F/\eta) \cdot \log \log(F/\eta).$$

time and outputs a set $\{(v'_i, f'_i)\}_{i \in [k]} \subset \mathbb{C} \times \mathbb{R}^d$ of size $k \in \mathbb{N}_{\geq 1}$ such that the following hold with probability $1 - 1/\text{poly}(k)$

Property I *Magnitude estimation*

$$|v_i - v'_i| \leq (\mathcal{C} + \sqrt{d}C_*) \cdot \mathcal{N}, \forall i \in [k].$$

Property II *Frequency estimation*

$$\|f_i - f'_i\|_2 \leq C_* \frac{1}{\rho \cdot T}, \forall i \in [k].$$

Property III *Tone estimation (Total)*

$$\sum_{i \in [k]} \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} \left| v_i \cdot e^{2\pi i \cdot f_i^\top \tau} - v'_i \cdot e^{2\pi i \cdot f'_i{}^\top \tau} \right|^2 \cdot d\tau \lesssim (\mathcal{C}^2 + dC_*^2) \cdot \mathcal{N}^2.$$

Property IV *The frequency separation of output frequencies*

$$\min_{i \neq j \in [k]} \|f'_i - f'_j\|_2 \geq \eta/2.$$

Recall that $C_* = d^2$ is stated in the statement in Lemma 5.16.

Claim 7.19 (Sample complexity, running time and duration of Theorem 7.18).

Proof. **Sample complexity.**

$$\begin{aligned} & \mathcal{R}_{\text{merge}} \cdot 2^{\Theta(d \cdot \log d)} \cdot (\log \mathcal{C} + \log \log(F/\eta)) \cdot k \cdot \log(F \cdot T) \cdot \mathcal{D} \\ &= 2^{\Theta(d \log d)} \cdot k \cdot \log^{d+1}(k/\delta) \cdot \log \log(F/\eta) \cdot \log(F/\eta) \end{aligned}$$

Running time.

$$\begin{aligned} & \mathcal{R}_{\text{merge}} (2^{O(d \cdot \log d)} \cdot k \cdot \log^d(k \cdot \mathcal{R}_{\text{merge}}) + 2^{\Theta(d \cdot (\log d + \log \mathcal{C}))} \cdot \log(F \cdot T) \cdot \log \log(F/\eta) \cdot k \cdot (\mathcal{D} + \log k)) \\ &= 2^{O(d \cdot \log d)} \cdot k \cdot \log^{O(d)} k + 2^{\Theta(d \log d)} \cdot \log(F/\eta) \cdot k \cdot \log \log(F/\eta) \cdot \log^{d+1}(k/\delta) \\ &= 2^{O(d \cdot \log d)} \cdot k \cdot \log^{O(d)}(k/\delta) \cdot \log(F/\eta) \cdot \log \log(F/\eta). \end{aligned}$$

Duration. As we run MERGEDSTAGE twice, the first run for k -sparsity signal and the second run for $k' = 2^{O(d \log d)} k$ -sparsity signal, $\eta' = \frac{\eta}{\sqrt{d}}$, by Lemma 6.3, the duration is

$$T = \Omega\left(\frac{d^3 \log(dk'/\delta)}{\eta'}\right) = \Omega\left(\frac{d^{4.5} \log(dk/\delta) \log d}{\eta}\right).$$

□

Claim 7.20 (Property I of Lemma 7.18).

$$|v_i - v'_i| \leq (\mathcal{C} + \sqrt{d}C_*) \cdot \mathcal{N}, \forall i \in [k]$$

Proof. This proof is a direct application of previous Lemma. \square

Claim 7.21 (Property II of Lemma 7.18).

$$\|f_i - f'_i\|_2 \leq C_* \frac{1}{\rho \cdot T}, \forall i \in [k]$$

Proof. The proof is a direct application of previous Lemma. \square

Claim 7.22 (Property III of Lemma 7.18).

$$\sum_{i \in [k]} \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} \left| v_i \cdot e^{2\pi i \cdot f_i^\top \tau} - v'_i \cdot e^{2\pi i \cdot f'_i{}^\top \tau} \right|^2 \cdot d\tau \lesssim (\mathcal{C}^2 + dC_*^2) \cdot \mathcal{N}^2.$$

Proof. We denote by $\{(v_i^*, f_i^*)\}_{i \in [k'']}$ the set of tones derived according to Definition 7.15, where $k'' = 2^{O(d \cdot \log d)} \cdot k$ (and we safely assume $k'' \geq k$ in view of Lemma 7.14). We assume w.l.o.g. that each $(v_i^*, f_i^*) \in \mathbb{C} \times \mathbb{R}^d$ of the top- k largest-magnitude tones (for each $i \in [k]$) is mapped to a *true* tone $(v_{\pi(i)}, f_{\pi(i)}) \in \mathbb{C} \times \mathbb{R}^d$ according to Lemma 7.16.

Let $\{(v'_i, f'_i)\}_{i \in [k]}$ be a subset of the recovered tones $\{(v_i^*, f_i^*)\}_{i \in [k'']}$ that have the top- k largest magnitudes; these $k \in \mathbb{N}_{\geq 1}$ tones together form the output of the procedure RECOVERYSTAGE (Algorithm 10). Upon reindexing, we safely assume that $|v'_i| \geq |v_i^*|$ for each $i \in [k]$. Also, we know from Lemma 7.16 that $\min_{i \neq j \in [k]} \|f'_i - f'_j\| \gtrsim \eta$.

For each $i \in [k]$, let us consider these three tones $(v_{\pi(i)}, f_{\pi(i)})$ and (v_i^*, f_i^*) and (v'_i, f'_i) . In the case $i \in S$ for which $(v_i^*, f_i^*) \neq (v'_i, f'_i)$, it follows from Lemma 7.17 that

$$\begin{aligned} & \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} \left| v'_i \cdot e^{2\pi i \cdot f'_i{}^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}^\top \tau} \right|^2 \cdot d\tau \\ & \lesssim \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} \left| v_i^* \cdot e^{2\pi i \cdot f_i^{*\top} \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}^\top \tau} \right|^2 \cdot d\tau + |v'_i|^2. \end{aligned}$$

And in the other case $i \in S \subseteq [k]$ for which $(v_i^*, f_i^*) = (v'_i, f'_i)$, of course we have

$$\begin{aligned} & \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} \left| v'_i \cdot e^{2\pi i \cdot f'_i{}^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}^\top \tau} \right|^2 \cdot d\tau \\ & = \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} \left| v_i^* \cdot e^{2\pi i \cdot f_i^{*\top} \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}^\top \tau} \right|^2 \cdot d\tau. \end{aligned}$$

Taking all the indices $i \in [k]$ into account, we know from the above two equations that

$$\begin{aligned} & \sum_{i \in [k]} \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} \left| v'_i \cdot e^{2\pi i \cdot f'_i{}^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}^\top \tau} \right|^2 \cdot d\tau \\ & \lesssim \sum_{i \in [k]} \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} \left| v_i^* \cdot e^{2\pi i \cdot f_i^{*\top} \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}^\top \tau} \right|^2 \cdot d\tau + \sum_{i \in S} |v'_i|^2. \end{aligned}$$

Because $\{(v'_i, f'_i)\}_{i \in [k]}$ are chosen to be the top- k largest-magnitude recovered tones among $\{(v_i^*, f_i^*)\}_{i \in [k'']}$, the set S involved in the summation $\sum_{i \in S} |v'_i|^2$ only includes those small-magnitude recovered tones. We thus conclude that

$$\sum_{i \in [k]} \frac{1}{T^d} \cdot \int_{\tau \in [0, T]^d} \left| v'_i \cdot e^{2\pi i \cdot f'_i{}^\top \tau} - v_{\pi(i)} \cdot e^{2\pi i \cdot f_{\pi(i)}^\top \tau} \right|^2 \cdot d\tau \lesssim (\mathcal{C}^2 + dC_*^2) \cdot \mathcal{N}^2.$$

This completes the proof of Property III of Theorem 7.18. \square

Claim 7.23 (Property IV of Lemma 7.18).

$$\min_{i \neq j \in [k]} \|f'_i - f'_j\|_2 \geq \eta/2.$$

Proof. Since $\min_{i \neq j \in [k]} \|f_i - f_j\|_2 \geq \eta$ and for all $i \in [k]$, $\|f_i - f'_i\|_2 \leq \min(C_* \frac{1}{\rho T}, d/T) \leq \eta/10$ (given Lemma 6.4, Claim 7.21 and the duration

$$T = \Omega\left(\frac{d^{4.5} \log(kd/\delta) \log d}{\eta}\right),$$

we can infer the current claim. \square

8 Converting tone estimation into signal estimation

This section is structured in the following way:

- Section 8.1 briefly discusses the high-level idea of the proof.
- Section 8.2 provides some basic definitions and mathematical facts.
- Section 8.3 splits the signal estimation error into the tone-wise errors (which we call the diagonal terms) and the cross-tone errors (which we call the off-diagonal terms).
- Section 8.4 provides an upper bound for the cross-tone errors (i.e. the off-diagonal terms) via some advanced analytic tools.
- Section 8.5 combines everything together, converting the tone estimation error into the signal estimation error as desired.
- Section 8.6 states several geometry properties.
- Section 8.7 presents our main result.

For ease of presentation, throughout this section we would shift the sampling time domain from $t \in [0, T]^d$ to $t \in [-T/2, T/2]^d$.

8.1 Improvement of signal estimation duration

The claimed tone estimation guarantee holds when the duration $T \gtrsim \eta^{-1} \cdot C_{\text{tone}}$ (Theorem 1.1; see Section 2.4 for more discussions), for $C_{\text{tone}} := d^{4.5} \cdot \log(kd/\delta) \cdot \log(d)$. To further get the signal estimation guarantee (Theorem 1.2), we adopt the proof framework of [PS15] but provide a better analysis. Particularly, we will show that the signal estimation holds when

$$T \gtrsim \eta^{-1} \cdot (C_{\text{tone}} + d^{1.5} \cdot k^{1-1/d} \cdot \log k).$$

In one dimension $d = 1$, this bound is $\eta^{-1} \cdot \log(k/\delta)$, which improves the $\eta^{-1} \cdot \log^2(k/\delta)$ bound by [PS15] and answers an open question in the thesis [Son19].

Following [PS15], we rewrite the signal estimation error as LHS of Eq. (2) = $\sum_{i,j \in [k]} \text{err}_{i,j}$, where

$$\text{err}_{i,j} := \frac{1}{T^d} \cdot \int_{t \in [0,T]^d} (x'_i(t) - x_i^*(t)) \cdot \overline{(x'_j(t) - x_j^*(t))} \cdot dt.$$

To get the signal estimation, [PS15] proves that the *cross-tone* errors $|\text{err}_{i,j}|$ for $i \neq j \in [k]$ converge to zero at the rate

$$|\text{err}_{i,j}| \lesssim \sqrt{\text{err}_{i,i} \cdot \text{err}_{j,j}} \cdot \sqrt{d} \cdot \frac{\log(1+\|f_i-f_j\|_{2,T})}{\|f_i-f_j\|_{2,T}} = O(\sqrt{d} \cdot T^{-1} \cdot \log T). \quad (32)$$

Based on a new application of Parseval's theorem and the convolution theorem, we will prove that the $O(\log T)$ term in Eq. (32) can be removed. More concretely, Parseval's theorem gives the analytic formulas of the errors $|\text{err}_{i,i}|$ and $|\text{err}_{i,j}|$, in the case of an *infinite* duration $t \in \mathbb{R}^d$. Inspired by this, we access the proof details of Parseval's theorem. Following the involved arguments and the convolution theorem, we get the counterpart formulas (Lemma 8.4) in the case of a *finite* duration $t \in [0, T]^d$. These analytic formulas and other arguments together give the faster convergence rate $|\text{err}_{i,j}| = O(\sqrt{d} \cdot T^{-1})$. In contrast, [PS15] just uses the approximate formulas of the errors $|\text{err}_{i,i}|$ and $|\text{err}_{i,j}|$, which incurs the factor- $\log(k/\delta)$ loss in their signal estimation duration.

Indeed, we have a concrete example for which $|\text{err}_{i,j}| = \Omega(\sqrt{d} \cdot T^{-1})$, matching our new convergence rate in the duration T and the dimension $d \geq 1$. We believe that this tight convergence rate, as well as the analytic formulas in Lemma 8.4, can find their applications in the future.

8.2 Preliminaries and mathematical facts

In this part, we introduce some useful notations and mathematical facts. Recall Definition A.1 for the functions $\text{rect}_{s_1}(\xi)$ and $\text{sin}_{s_1}(\tau)$ in the single-dimensional setting (namely $\xi, \tau \in \mathbb{R}$). Below, we define in Definition 8.1 two counterpart functions (redenoted by $\text{rect}_{s_1}(\xi)$ and $\text{sin}_{s_1}(\tau)$ for convenience) when $\xi, \tau \in \mathbb{R}^d$ are d -dimensional vectors, and then show in Fact 8.2 several properties of these functions (which can be easily inferred from Fact A.2 or the previous literature like [CKPS16]).

Definition 8.1 (Two basic functions). Given any $s_1 > 0$, for all $\xi, \tau \in \mathbb{R}^d$, the $\text{rect}_{s_1}(\xi)$ function and the $\text{sin}_{s_1}(\tau)$ function are defined as follows:

- $\text{rect}_{s_1}(\xi) = \prod_{r \in [d]} \text{rect}_{s_1}(\xi_r)$ for any $\xi \in \mathbb{R}^d$. When $s_1 = 1$, we shorthand it as $\text{rect}(\xi)$.
- $\text{sin}_{s_1}(\tau) = \prod_{r \in [d]} \text{sin}_{s_1}(\tau_r)$ for any $\tau \in \mathbb{R}^d$. When $s_1 = 1$, we shorthand it as $\text{sin}(\tau)$.

Fact 8.2 (Facts about basic functions). *Given any $s_1 > 0$, the following hold for the functions $\text{sin}_{s_1}(\tau)$ and $\text{rect}_{s_1}(\xi)$ in the d -dimensional setting, as Figure 17 suggests:*

Part (a): $|\text{sin}_{s_1}(\tau)| \leq \prod_{r \in [d]} \min\{1, \frac{1}{\pi \cdot s_1 \cdot |\tau_r|}\}$ for any $\tau \in \mathbb{R}^d$.

Part (b): $\text{sin}_{s_1}(\tau) = \widehat{\text{rect}_{s_1}}(\tau)$ for any $\tau \in \mathbb{R}^d$, and $\text{rect}_{s_1}(\xi) = \widehat{\text{sin}_{s_1}}(\xi)$ for any $\xi \in \mathbb{R}^d$.

Part (c): $\exp(-\frac{\pi^2 \cdot s_1^2}{5} \cdot \|\tau\|_2^2) \leq \text{sin}_{s_1}(\tau) \leq \exp(-\frac{\pi^2 \cdot s_1^2}{6} \cdot \|\tau\|_2^2)$ for any $\tau \in \mathbb{R}^d$ that $\|\tau\|_2 \leq \frac{2.05}{\pi s_1}$.

Part (d): $|\text{sin}_{s_1}(\tau)| \leq \exp(-\frac{2.05^2}{6}) < \frac{1}{2}$ for any $\tau \in \mathbb{R}^d$ that $\|\tau\|_2 \geq \frac{2.05}{\pi s_1}$.

Part (e): $-\frac{1}{4} \leq \text{sin}_{s_1}(\tau) \leq 1$ and $\text{sin}_{s_1}(\tau) \geq 1 - \frac{\pi^2}{6} \cdot s_1^2 \cdot \|\tau\|_2^2$ for any $\tau \in \mathbb{R}^d$.

Also, in the single-dimensional setting:

Part (f): $|\frac{d}{d\tau} \text{sin}_{s_1}(\tau)| = |\frac{\cos(\pi \cdot s_1 \cdot \tau)}{\tau} - \frac{\text{sin}_{s_1}(\tau)}{\tau}| \leq \frac{7}{5} \cdot \min\{s_1, \frac{1}{|\tau|}\}$ for any $\tau \in \mathbb{R}$.

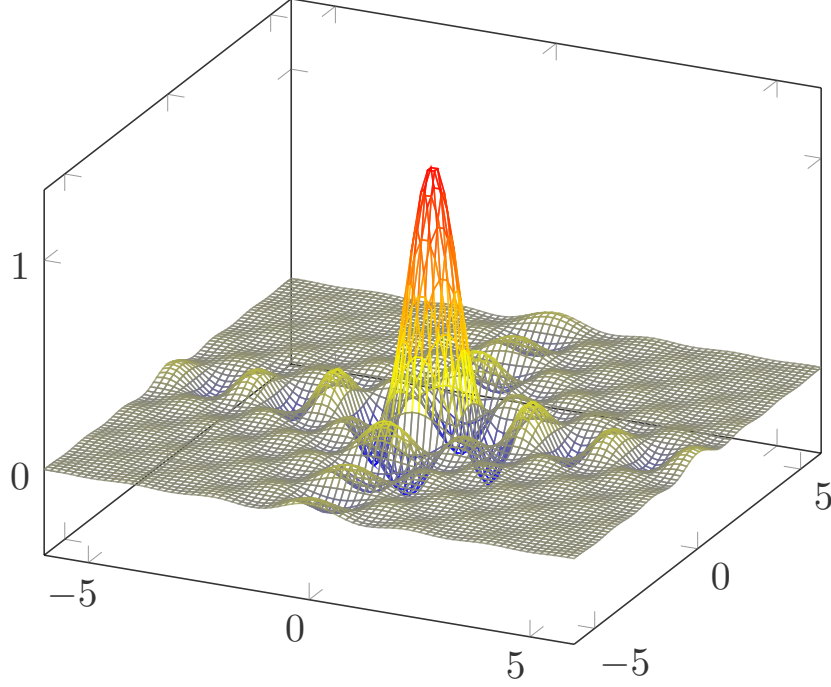


Figure 17: Demonstration for the two-dimensional sinc function.

8.3 Tone-wise errors and cross-tone errors

The goal of this section is to prove Lemma 8.4. We first start with the following definitions.

Definition 8.3 (Tone-wise errors in time domain). Given any pair of tones $(v_i, f_i) \in \mathbb{C} \times \mathbb{R}^d$ and $(v'_i, f'_i) \in \mathbb{C} \times \mathbb{R}^d$, where $i \in [k]$, the error is given by the complex-valued function $a_i(\tau) \in \mathbb{C}$:

- Define $a_i(\tau) = v_i \cdot e^{2\pi i \cdot f_i^\top \tau} - v'_i \cdot e^{2\pi i \cdot f'_i{}^\top \tau}$ for all $\tau \in [-T/2, T/2]^d$ for notational brevity.
- The CFT is given by $\widehat{a_i(\xi)} = v_i \cdot \text{Delta}_{=f_i(\xi)} - v'_i \cdot \text{Delta}_{=f'_i(\xi)}$ for all $\xi \in \mathbb{R}^d$.
- Define the error $\|a_i\|_T = \sqrt{\mathbf{E}_\tau[|a_i(\tau)|^2]} = \sqrt{\mathbf{E}_\tau[a_i(\tau) \cdot \overline{a_i(\tau)}]}$, where $\tau \sim \text{Unif}[-T/2, T/2]^d$ is uniformly random, or equivalently,

$$\begin{aligned} \|a_i\|_T &= \left(\frac{1}{T^d} \cdot \int_{\tau \in [-T/2, T/2]^d} |a_i(\tau)|^2 \cdot d\tau \right)^{1/2} \\ &= \left(\frac{1}{T^d} \cdot \int_{\tau \in [-T/2, T/2]^d} a_i(\tau) \cdot \overline{a_i(\tau)} \cdot d\tau \right)^{1/2} \end{aligned}$$

Lemma 8.4 (Tone-wise and cross-tone errors in time domain). *Respecting a pair of error functions $a_i(\tau) \in \mathbb{C}$ and $a_j(\tau) \in \mathbb{C}$ given in Definition 8.3, where $i, j \in [k]$, the following hold:*

Property I: *When $\tau \sim \text{Unif}[-T/2, T/2]^d$ is uniformly random,*

$$\begin{aligned} \mathbf{E}_\tau \left[a_i(\tau) \cdot \overline{a_j(\tau)} \right] &= v_i \cdot \overline{v_j} \cdot \text{sinc}_T(f_i - f_j) - v_i \cdot \overline{v'_j} \cdot \text{sinc}_T(f_i - f'_j) \\ &\quad - v'_i \cdot \overline{v_j} \cdot \text{sinc}_T(f'_i - f_j) + v'_i \cdot \overline{v'_j} \cdot \text{sinc}_T(f'_i - f'_j). \end{aligned}$$

Property II: In the special case that $i = j$,

$$\|a_i\|_T^2 = \mathbf{E}_\tau \left[a_i(\tau) \cdot \overline{a_i(\tau)} \right] = |v_i|^2 + |v'_i|^2 - (v_i \cdot \overline{v'_i} + \overline{v_i} \cdot v'_i) \cdot \text{sinc}_T(f'_i - f_i).$$

Proof. Assume Property I to be true, then we can infer Property II by elementary calculation.

Before proving Property I, let us consider the following function $y_i(\tau)$ for all $\tau \in \mathbb{R}^d$:

$$\begin{aligned} y_i(\tau) &= a_i(\tau) \cdot \mathbb{I}\{\tau \in [-T/2, T/2]^d\} \\ &= T^d \cdot a_i(\tau) \cdot \text{rect}_T(\tau), \end{aligned}$$

as well as its CFT $\widehat{y}_i(\xi)$ for all $\xi \in \mathbb{R}^d$:

$$\begin{aligned} \widehat{y}_i(\xi) &= T^d \cdot \widehat{a_i * \text{rect}_T}(\xi) \\ &= T^d \cdot \widehat{a_i} * \widehat{\text{rect}_T}(\xi) \\ &= T^d \cdot \widehat{a_i} * \text{sinc}_T(\xi) \\ &= T^d \cdot v_i \cdot \text{sinc}_T(f_i - \xi) - T^d \cdot v'_i \cdot \text{sinc}_T(f'_i - \xi), \end{aligned} \tag{33}$$

where the second step applies the convolution theorem; the third step is due to Part (d) of Fact A.2; and the last step follows from Definition 8.3 that $\widehat{a_i}(\xi) = v_i \cdot \text{Delta}_{=f_i}(\xi) - v'_i \cdot \text{Delta}_{=f'_i}(\xi)$ for $\xi \in \mathbb{R}^d$.

Similar to Equation (33), we also have

$$\overline{\widehat{y}_j(\xi)} = T^d \cdot \overline{v_j} \cdot \text{sinc}_T(f_j - \xi) - T^d \cdot \overline{v'_j} \cdot \text{sinc}_T(f'_j - \xi). \tag{34}$$

Based on the above arguments, we deduce that when $\tau \sim \text{Unif}[-T/2, T/2]^d$ is uniformly random,

$$\begin{aligned} \mathbf{E}_\tau \left[a_i(\tau) \cdot \overline{a_j(\tau)} \right] &= \frac{1}{T^d} \cdot \int_{\tau \in [-T/2, T/2]^d} a_i(\tau) \cdot \overline{a_j(\tau)} \cdot d\tau \\ &= \frac{1}{T^d} \cdot \int_{\tau \in \mathbb{R}^d} \Re(y_i(\tau) \cdot \overline{y_j(\tau)}) \cdot d\tau \\ &= \frac{1}{T^d} \cdot \int_{\xi \in \mathbb{R}^d} \widehat{y}_i(\xi) \cdot \overline{\widehat{y}_j(\xi)} \cdot d\xi \\ &= T^d \cdot v_i \cdot \overline{v_j} \cdot \underbrace{\int_{\xi \in \mathbb{R}^d} \text{sinc}_T(f_i - \xi) \cdot \text{sinc}_T(f_j - \xi) \cdot d\xi}_{A_1} \\ &\quad - T^d \cdot v_i \cdot \overline{v'_j} \cdot \underbrace{\int_{\xi \in \mathbb{R}^d} \text{sinc}_T(f_i - \xi) \cdot \text{sinc}_T(f'_j - \xi) \cdot d\xi}_{A_2} \\ &\quad - T^d \cdot v'_i \cdot \overline{v_j} \cdot \underbrace{\int_{\xi \in \mathbb{R}^d} \text{sinc}_T(f'_i - \xi) \cdot \text{sinc}_T(f_j - \xi) \cdot d\xi}_{A_3} \\ &\quad + T^d \cdot v'_i \cdot \overline{v'_j} \cdot \underbrace{\int_{\xi \in \mathbb{R}^d} \text{sinc}_T(f'_i - \xi) \cdot \text{sinc}_T(f'_j - \xi) \cdot d\xi}_{A_4}, \end{aligned}$$

where the second step follows because $y_i(\tau) = T^d \cdot a_i(\tau) \cdot \text{rect}_T(\tau) = 0$ for any $\tau \notin [-T/2, T/2]^d$; the third step applies Parseval's theorem; and the last step employs Equations (33) and (34).

We next give in Equation (35) an explicit formula for A_1 , and similar formulas respectively for A_2 and A_3 and A_4 can be obtained in the same way. Concretely, we have

$$\begin{aligned}
A_1 &= \int_{\xi \in \mathbb{R}^d} \text{sinc}_T(f_i - \xi) \cdot \text{sinc}_T(f_j - \xi) \cdot d\xi \\
&= \frac{1}{T^d} \cdot \int_{\xi \in \mathbb{R}^d} \text{sinc}(Tf_i - \xi) \cdot \text{sinc}(Tf_j - \xi) \cdot d\xi \\
&= \frac{1}{T^d} \cdot \int_{\xi \in \mathbb{R}^d} \text{sinc}(\xi + Tf_i - Tf_j) \cdot \text{sinc}(\xi) \cdot d\xi \\
&= \frac{1}{T^d} \cdot \text{sinc}(Tf_i - Tf_j) \\
&= \frac{1}{T^d} \cdot \text{sinc}_T(f_i - f_j),
\end{aligned} \tag{35}$$

where the second step follows by substitution; the third step also follows by substitution; the fourth step follows from Part (b) of Fact 8.5; and the last step follows by substitution.

Applying Equation (35) and the counterpart formulas for A_2 and A_3 and A_4 , we conclude that when $\tau \sim \text{Unif}[-T/2, T/2]^d$ is uniformly random,

$$\begin{aligned}
\mathbf{E}_\tau \left[a_i(\tau) \cdot \overline{a_j(\tau)} \right] &= v_i \cdot \overline{v_j} \cdot \text{sinc}_T(f_i - f_j) - v_i \cdot \overline{v'_j} \cdot \text{sinc}_T(f_i - f'_j) \\
&\quad - v'_i \cdot \overline{v_j} \cdot \text{sinc}_T(f'_i - f_j) + v'_i \cdot \overline{v'_j} \cdot \text{sinc}_T(f'_i - f'_j).
\end{aligned}$$

This completes the proof. \square

8.4 Upper bounding cross-tone errors

Fact 8.5. *The following hold for the single-/multi-dimensional sinc function:*

Part (a): *Single dimension.* $\text{sinc}(\Delta_r) = \int_{\xi_r \in \mathbb{R}} \text{sinc}(\xi_r + \Delta_r) \cdot \text{sinc}(\xi_r) \cdot d\xi_r$ for any $\Delta_r \in \mathbb{R}$.

Part (b): *Multi dimension.* $\text{sinc}(\Delta) = \int_{\xi \in \mathbb{R}^d} \text{sinc}(\xi + \Delta) \cdot \text{sinc}(\xi) \cdot d\xi$ for any $\Delta \in \mathbb{R}^d$.

Proof. Part (b) can be easily inferred from Part (a), since $\text{sinc}(\Delta) = \prod_{r \in [d]} \text{sinc}(\Delta_r)$ is a product and we deal with all the coordinates $r \in [d]$ separately.

We deduce Part (a) as follows:

$$\begin{aligned}
\int_{\xi_r \in \mathbb{R}} \text{sinc}(\xi_r + \Delta_r) \cdot \text{sinc}(\xi_r) \cdot d\xi_r &= \int_{\xi_r \in \mathbb{R}} \text{sinc}(\Delta_r - \xi_r) \cdot \text{sinc}(\xi_r) \cdot d\xi_r \\
&= \text{sinc} * \text{sinc}(\Delta_r) \\
&= \int_{\tau \in \mathbb{R}} \widehat{\text{sinc} * \text{sinc}}(\tau) \cdot e^{-2\pi i \cdot \Delta_r \cdot \tau} \cdot d\tau \\
&= \int_{\tau \in \mathbb{R}} \text{rect}^2(\tau) \cdot e^{-2\pi i \cdot \Delta_r \cdot \tau} \cdot d\tau \\
&= \int_{-1/2}^{1/2} e^{-2\pi i \cdot \Delta_r \cdot \tau} \cdot d\tau \\
&= \text{sinc}(\Delta_r),
\end{aligned}$$

where the first step follows by substitution and the fact that $\text{sinc}(\xi_r)$ is an even function; the second step follows from Definition 2.2; the third step follows the definition of the CFT; the fourth step

applies the convolution Theorem as $\widehat{\text{sinc} * \text{sinc}}(\tau) = \text{rect}(\tau) \cdot \text{rect}(\tau) = \text{rect}^2(\tau)$; the fifth follows because $\text{rect}(\tau) = \mathbb{I}\{|\tau| \leq 1/2\}$; and the last step can be seen via elementary calculation.

This completes the proof. \square

Lemma 8.6 (Upper bounds on the cross-tone errors). *For any pair of indices $i < j \in [k]$. Assume the following for both $(v, f, v', f') = (v_i, f_i, v'_i, f'_i)$ and $(v, f, v', f') = (v_j, f_j, v'_j, f'_j)$:*

- $\|f - f'\|_2 \leq \Delta f_{i,j}$, where the distance $\Delta f_{i,j} \geq 0$ is given by

$$\Delta f_{i,j} := \min \left\{ \|f''_i - f''_j\|_2 : f''_i \in \{f_i, f'_i\} \text{ and } f''_j \in \{f_j, f'_j\} \right\}.$$

Then for the functions $a_i(\tau) : \mathbb{R}^d \rightarrow \mathbb{C}$ and $a_j(\tau) : \mathbb{R}^d \rightarrow \mathbb{C}$ given in Definition 8.3, the cross-tone error satisfies the following when $T = \Omega(\frac{d}{\Delta f_{i,j}})$ is large enough:

$$|\mathbf{E}_\tau[a_i(\tau) \cdot \overline{a_j(\tau)} + \overline{a_i(\tau)} \cdot a_j(\tau)]| \lesssim \frac{\sqrt{d}}{\Delta f_{i,j} \cdot T} \cdot \|a_i\|_T \cdot \|a_j\|_T,$$

where $\tau \sim \text{Unif}[-T/2, T/2]^d$ is uniformly random.

Proof. In this proof, we use $f_{i,s} \in \mathbb{R}$ to denote the s -th coordinate of the i -th frequency $f_i \in \mathbb{R}^d$.

For simplicity of notation, we define

$$\text{err}_{i,j} := \mathbf{E}_\tau[a_i(\tau) \cdot \overline{a_j(\tau)} + \overline{a_i(\tau)} \cdot a_j(\tau)].$$

According to Property II of Lemma 8.4, we can rewrite $\text{err}_{i,j}$ as follows:

$$\begin{aligned} \text{err}_{i,j} &= \mathbf{E}_\tau \left[a_i(\tau) \cdot \overline{a_j(\tau)} + \overline{a_i(\tau)} \cdot a_j(\tau) \right] \\ &= (v_i \cdot \overline{v_j} + \overline{v_i} \cdot v_j) \cdot \text{sinc}_T(f_i - f_j) - (v_i \cdot \overline{v'_j} + \overline{v_i} \cdot v'_j) \cdot \text{sinc}_T(f_i - f'_j) \\ &\quad - (v'_i \cdot \overline{v_j} + \overline{v'_i} \cdot v_j) \cdot \text{sinc}_T(f'_i - f_j) + (v'_i \cdot \overline{v'_j} + \overline{v'_i} \cdot v'_j) \cdot \text{sinc}_T(f'_i - f'_j). \end{aligned} \quad (36)$$

Below, we would prove the lemme based on case analysis.

In total there are four cases:

- $T \cdot \|f'_i - f_i\|_2 \geq \frac{2.05}{\pi}$ and $T \cdot \|f'_j - f_j\|_2 \geq \frac{2.05}{\pi}$ (see Claim 8.7).
- $T \cdot \|f'_i - f_i\|_2 < \frac{2.05}{\pi}$ and $T \cdot \|f'_j - f_j\|_2 < \frac{2.05}{\pi}$ (see Claim 8.8).
- $T \cdot \|f'_i - f_i\|_2 \geq \frac{2.05}{\pi}$ and $T \cdot \|f'_j - f_j\|_2 < \frac{2.05}{\pi}$ (see Claim 8.9).
- $T \cdot \|f'_i - f_i\|_2 < \frac{2.05}{\pi}$ and $T \cdot \|f'_j - f_j\|_2 \geq \frac{2.05}{\pi}$ (see Claim 8.9).

Combining all the four cases completes the proof. \square

8.4.1 Both pairs are far

Claim 8.7 (Case (i) for Lemma 8.6). *If $T \cdot \|f'_i - f_i\|_2 \geq \frac{2.05}{\pi}$ and $T \cdot \|f'_j - f_j\|_2 \geq \frac{2.05}{\pi}$, then we have*

$$|\text{err}_{i,j}| \lesssim \frac{1}{\Delta f_{i,j} T} \cdot \|a_i\|_T \cdot \|a_j\|_T.$$

Proof. Let us first bound the tone-wise errors $\|a_i\|_T^2$ and $\|a_j\|_T^2$ from below. Respecting the i -th pair of tones (v_i, f_i) and (v'_i, f'_i) , we know from Part (d) of Fact 8.2 that

$$|\text{sinc}_T(f'_i - f_i)| < \frac{1}{2}.$$

Then, the tone-wise error $\|a_i\|_T^2$ between (v_i, f_i) and (v'_i, f'_i) admits the lower bound

$$\begin{aligned} \|a_i\|_T^2 &= |v_i|^2 + |v'_i|^2 - (v_i \cdot \overline{v'_i} + \overline{v_i} \cdot v'_i) \cdot \text{sinc}_T(f'_i - f_i) \\ &= |v_i|^2 + |v'_i|^2 - 2 \cdot |v_i| \cdot |v'_i| \cdot \cos(\arg(v'_i/v_i)) \cdot \text{sinc}_T(f'_i - f_i) \\ &\geq |v_i|^2 + |v'_i|^2 - 2 \cdot |v_i| \cdot |v'_i| \cdot \left| \cos(\arg(v'_i/v_i)) \right| \cdot \left| \text{sinc}_T(f'_i - f_i) \right| \\ &\geq |v_i|^2 + |v'_i|^2 - 2 \cdot |v_i| \cdot |v'_i| \cdot 1 \cdot \frac{1}{2} \\ &\geq \frac{1}{4} \cdot (|v_i| + |v'_i|)^2, \end{aligned} \tag{37}$$

where the first step is by Property II of Lemma 8.4; the fourth step follows since $|\cos(\arg(v'_i/v_i))| \leq 1$ and $|\text{sinc}_T(f'_i - f_i)| < \frac{1}{2}$; and the last step applies the AM-GM inequality.

Applying the same arguments to the j -th tone-wise error $\|a_j\|_T^2$, we also have

$$\|a_j\|_T^2 \geq \frac{1}{4} \cdot (|v_j| + |v'_j|)^2. \tag{38}$$

We next establish an upper bound on the cross-tone error $|\mathbf{E}_\tau[a_i(\tau) \cdot \overline{a_j(\tau)} + \overline{a_i(\tau)} \cdot a_j(\tau)]|$, where $\tau \sim \text{Unif}[-\frac{T}{2}, \frac{T}{2}]^2$ is uniformly random. For simplicity, we denote

$$\text{sinc}_{T,i,j}^{\max} = \max \left\{ |\text{sinc}_T(f''_i - f''_j)| : f''_i \in \{f_i, f'_i\} \text{ and } f''_j \in \{f_j, f'_j\} \right\} \geq 0. \tag{39}$$

Following Equation (36), we deduce that

$$\begin{aligned} |\text{err}_{i,j}| &= \left| (v_i \cdot \overline{v_j} + \overline{v_i} \cdot v_j) \cdot \text{sinc}_T(f_i - f_j) - (v_i \cdot \overline{v'_j} + \overline{v_i} \cdot v'_j) \cdot \text{sinc}_T(f_i - f'_j) \right. \\ &\quad \left. - (v'_i \cdot \overline{v_j} + \overline{v'_i} \cdot v_j) \cdot \text{sinc}_T(f'_i - f_j) + (v'_i \cdot \overline{v'_j} + \overline{v'_i} \cdot v'_j) \cdot \text{sinc}_T(f'_i - f'_j) \right| \\ &\leq 2 \cdot |v_i| \cdot |v_j| \cdot |\text{sinc}_T(f_i - f_j)| + 2 \cdot |v_i| \cdot |v'_j| \cdot |\text{sinc}_T(f_i - f'_j)| \\ &\quad + 2 \cdot |v'_i| \cdot |v_j| \cdot |\text{sinc}_T(f'_i - f_j)| + 2 \cdot |v'_i| \cdot |v'_j| \cdot |\text{sinc}_T(f'_i - f'_j)| \\ &\leq 2 \cdot |v_i| \cdot |v_j| \cdot \text{sinc}_{T,i,j}^{\max} + 2 \cdot |v_i| \cdot |v'_j| \cdot \text{sinc}_{T,i,j}^{\max} \\ &\quad + 2 \cdot |v'_i| \cdot |v_j| \cdot \text{sinc}_{T,i,j}^{\max} + 2 \cdot |v'_i| \cdot |v'_j| \cdot \text{sinc}_{T,i,j}^{\max} \\ &= 2 \cdot (|v_i| + |v'_i|) \cdot (|v_j| + |v'_j|) \cdot \text{sinc}_{T,i,j}^{\max} \\ &\leq 2 \cdot (2 \cdot \|a_i\|_T) \cdot (2 \cdot \|a_j\|_T) \cdot \text{sinc}_{T,i,j}^{\max} \\ &= 8 \cdot \|a_i\|_T \cdot \|a_j\|_T \cdot \text{sinc}_{T,i,j}^{\max}, \end{aligned} \tag{40}$$

where the second step uses the triangle inequality; the third step follows from the definition of $\text{sinc}_{T,i,j}^{\max}$ (see Equation (39)); the fifth step follows from Equations (37) and (38); and the last step follows from the AM-GM inequality.

To accomplish Case (i), given Equation (40), we are left to justify that $\text{sinc}_{T,i,j}^{\max} \geq 0$ diminishes to zero when $T > 0$ goes to the infinity (at the claimed rate). We safely assume $T = \Omega(\frac{d}{\Delta f_{i,j}})$ to be large enough, and consider a specific pair of frequencies $f''_i \in \{f_i, f'_i\}$ and $f''_j \in \{f_j, f'_j\}$. For

simplicity, we denote $\delta_r = \max(0, \pi \cdot T \cdot |f''_{i,r} - f''_{j,r}| - 1) \geq 0$ for each coordinate $r \in [d]$. Given these, one can easily see that

$$\begin{aligned}
\sum_{r \in [d]} \delta_r &\geq \sum_{r \in [d]} (\pi \cdot T \cdot |f''_{i,r} - f''_{j,r}| - 1) \\
&= \pi \cdot T \cdot \|f''_i - f''_j\|_1 - d \\
&= \pi \cdot T \cdot \|f''_i - f''_j\|_2 - d \\
&\geq \pi \cdot T \cdot \Delta f_{i,j} - d \geq 0.
\end{aligned} \tag{41}$$

In addition, we have

$$\begin{aligned}
|\text{sinc}_T(f''_i - f''_j)| &\leq \prod_{r \in [d]} \min \left\{ 1, \frac{1}{\pi \cdot T \cdot |f''_{i,r} - f''_{j,r}|} \right\} \\
&= \prod_{r \in [d]} \frac{1}{1 + \delta_r} \\
&\leq \frac{1}{1 + \sum_{r \in [d]} \delta_r} \\
&\leq \frac{1}{\pi \cdot T \cdot \Delta f_{i,j} - (d - 1)} \\
&\lesssim \frac{1}{T \cdot \Delta f_{i,j}},
\end{aligned} \tag{42}$$

where the first step applies Part (a) of Fact 8.2; the second step is due to the definition of δ_r 's; the third step follows because $\delta_r \geq 0$ for each $r \in [d]$; the fourth step follows from Equation (41); and the last step holds whenever $T = \Omega(\frac{d}{\Delta f_{i,j}})$ is large enough.

We observe that Equation (42) holds for any pair of frequencies $f''_i \in \{f_i, f'_i\}$ and $f''_j \in \{f_j, f'_j\}$. In other words,

$$\text{sinc}_{T,i,j}^{\max} \lesssim \frac{1}{T \cdot \Delta f_{i,j}}$$

Combining the above equation and Equation (40) together completes the proof. \square

8.4.2 Both pairs are close

Claim 8.8 (Case (ii) for Lemma 8.6). *If $T \cdot \|f'_i - f_i\|_2 < \frac{2.05}{\pi}$ and $T \cdot \|f'_j - f_j\|_2 < \frac{2.05}{\pi}$, then we have*

$$\text{err}_{i,j} \lesssim \frac{\sqrt{d}}{\Delta f_{i,j} T} \cdot \|a_i\|_T \cdot \|a_j\|_T.$$

Proof. Let us first bound the tone-wise errors $\|a_i\|_T^2$ and $\|a_j\|_T^2$ from below. Respecting the i -th pair of tones (v_i, f_i) and (v'_i, f'_i) , we know from Part (d) of Fact 8.2 that

$$\exp \left(-\frac{\pi^2}{5} \cdot T^2 \cdot \|f'_i - f_i\|_2^2 \right) \leq \text{sinc}_T(f'_i - f_i) \leq \exp \left(-\frac{\pi^2}{6} \cdot T^2 \cdot \|f'_i - f_i\|_2^2 \right).$$

Then, the tone-wise error $\|a_i\|_T^2$ between (v_i, f_i) and (v'_i, f'_i) admits the lower bound

$$\|a_i\|_T^2 = |v_i|^2 + |v'_i|^2 - (v_i \cdot \overline{v'_i} + \overline{v_i} \cdot v'_i) \cdot \text{sinc}_T(f'_i - f_i)$$

$$= |v_i|^2 + |v'_i|^2 - 2 \cdot |v_i| \cdot |v'_i| \cdot \cos(\arg(v'_i/v_i)) \cdot \text{sinc}_T(f'_i - f_i), \quad (43)$$

where the first step applies Property II of Lemma 8.4.

Given Equation (43), we would prove that $\|a_i\|_T^2$ is lower bounded by

$$\|a_i\|_T^2 \geq \frac{3}{13} \cdot \left(|v_i - v'_i|^2 + (|v_i|^2 + |v'_i|^2) \cdot (1 - \text{sinc}_T(f'_i - f_i)) \right). \quad (44)$$

To see so, we denote $w_1 \cdot e^{i\theta} = v_i/v'_i$ for some norm $w_1 \geq 0$ and some phase $\theta \in [0, 2\pi)$, and denote $w_2 = \text{sinc}_T(f'_i - f_i) \in [-\frac{1}{4}, 1]$ (see Part (e) of Fact 8.2). We notice that the RHS of Equation (44) is non-negative. Thus, it suffices to show that the following function $L(w_1, w_2, \theta) \geq \frac{3}{11}$, for any $w_1 \geq 0$, any $w_2 \in [-\frac{1}{4}, 1]$ and any $\theta \in [0, 2\pi)$:

$$\begin{aligned} L(w_1, w_2, \theta) &:= \frac{\text{RHS of (43)}}{\text{RHS of (44)}} \\ &= \frac{|v'_i \cdot w_1 \cdot e^{i\theta}|^2 + |v'_i|^2 - 2 \cdot |v'_i \cdot w_1 \cdot e^{i\theta}| \cdot |v'_i| \cdot \cos(\theta) \cdot w_2}{|v'_i \cdot w_1 \cdot e^{i\theta} - v'_i|^2 + (|v'_i \cdot w_1 \cdot e^{i\theta}|^2 + |v'_i|^2) \cdot (1 - w_2)} \\ &= \frac{w_1^2 + 1 - 2 \cdot w_1 \cdot w_2 \cdot \cos(\theta)}{|w_1 \cdot e^{i\theta} - 1|^2 + (w_1^2 + 1) \cdot (1 - w_2)} \\ &= \frac{w_1^2 + 1 - 2 \cdot w_1 \cdot w_2 \cdot \cos(\theta)}{(w_1^2 + 1) \cdot (2 - w_2) - 2 \cdot w_1 \cdot \cos(\theta)} \end{aligned}$$

where the second step is by the definition of w_1 , w_2 and θ ; the third step divides both the numerator and the denominator by $|v'_i|^2$; and the last step can be seen via elementary calculation.

Let us investigate the partial derivative $\frac{\partial L}{\partial \theta}$ in $\theta \in [0, 2\pi)$:

$$\begin{aligned} \frac{\partial L}{\partial \theta} &= \frac{2 \cdot w_1 \cdot w_2 \cdot \sin(\theta)}{(w_1^2 + 1) \cdot (2 - w_2) - 2 \cdot w_1 \cdot \cos(\theta)} - \frac{(w_1^2 + 1 - 2 \cdot w_1 \cdot w_2 \cdot \cos(\theta)) \cdot (2 \cdot w_1 \cdot \sin(\theta))}{((w_1^2 + 1) \cdot (2 - w_2) - \cos(\theta))^2} \\ &= -\sin(\theta) \cdot \underbrace{\frac{2 \cdot w_1 \cdot (w_1^2 + 1) \cdot (w_2 + 1)^2}{((w_1^2 + 1) \cdot (2 - w_2) - 2 \cdot w_1 \cdot \cos(\theta))^2}}_{A_5}, \end{aligned}$$

where the second step can be seen via elementary calculation.

Because $w_1 \geq 0$ and $w_2 \in [-\frac{1}{4}, 1]$, we must have $A_5 \geq 0$. Hence, for any fixed w_1 and w_2 , the function $L(w_1, w_2, \theta)$ is non-increasing when $\theta \in [0, \pi]$, and is non-decreasing when $\theta \in [\pi, 2\pi)$. Then we conclude that the function $L_1(w_1, w_2) := \min_{\theta \in [0, 2\pi)} L(w_1, w_2, \theta)$ for any $w_1 \geq 0$ and any $w_2 \in [-\frac{1}{4}, 1]$ is given by

$$\begin{aligned} L_1(w_1, w_2) &= L(w_1, w_2, \pi) = \frac{A_6(w_1, w_2)}{A_7(w_1, w_2)}, \\ A_6(w_1, w_2) &:= w_1^2 + 1 + 2 \cdot w_1 \cdot w_2, \\ A_7(w_1, w_2) &:= (w_1^2 + 1) \cdot (2 - w_2) + 2 \cdot w_1. \end{aligned}$$

Clearly, for any fixed $w_1 \geq 1$, the numerator $A_6(w_1, w_2)$ is a non-decreasing function in $w_2 \in [-\frac{1}{4}, 1]$, while the denominator $A_7(w_1, w_2)$ is a non-increasing *non-negative* function in $w_2 \in [-\frac{1}{4}, 1]$. Given these, we deduce that

$$\min_{w_1 \in [0, 1]} \min_{w_2 \in [-\frac{1}{4}, 1]} L_1(w_1, w_2) = \min_{w_1 \in [0, 1]} L_1(w_1, -1/4)$$

$$\begin{aligned}
&= \min_{w_1 \in [0,1]} \frac{w_1^2 + 1 - (1/2) \cdot w_1}{(9/4) \cdot (w_1^2 + 1) + 2 \cdot w_1} \\
&= \min_{w_1 \in [0,1]} \left(\frac{4}{9} - \frac{50/81}{w_1 + (1/w_1) + 8/9} \right) \\
&= \frac{4}{9} - \frac{50/81}{1 + 1 + 8/9} \\
&= \frac{3}{13},
\end{aligned}$$

which implies Equation (44) immediately.

Following Equation (44), we further have

$$\begin{aligned}
\|a_i\|_T^2 &\gtrsim |v_i - v'_i|^2 + (|v_i|^2 + |v'_i|^2) \cdot (1 - \text{sinc}_T(f'_i - f_i)) \\
&\geq |v_i - v'_i|^2 + (|v_i|^2 + |v'_i|^2) \cdot \left(1 - \exp\left(-\frac{\pi^2}{6} \cdot T^2 \cdot \|f'_i - f_i\|_2^2\right)\right) \\
&\geq |v_i - v'_i|^2 + (|v_i|^2 + |v'_i|^2) \cdot T^2 \cdot \|f'_i - f_i\|_2^2 \\
&\geq \frac{1}{2} \cdot \left(|v_i - v'_i| + \sqrt{|v_i|^2 + |v'_i|^2} \cdot T \cdot \|f'_i - f_i\|_2\right)^2 \\
&\geq \frac{1}{2} \cdot \left(|v_i - v'_i| + \frac{\sqrt{2}}{2} \cdot (|v_i| + |v'_i|) \cdot T \cdot \|f'_i - f_i\|_2\right)^2 \\
&\geq \frac{1}{4} \cdot \left(|v_i - v'_i| + (|v_i| + |v'_i|) \cdot T \cdot \|f'_i - f_i\|_2\right)^2,
\end{aligned} \tag{45}$$

where the first step applies Equation (44); the second step applies Part (d) of Fact 8.2; the third step follows from the premise that $T^2 \cdot \|f'_i - f_i\|_2^2 < (\frac{2.05}{\pi})^2$, together with the fact that, for any $0 \leq z < (\frac{2.05}{\pi})^2 \approx 0.4258$, we have $\exp(-\frac{\pi^2}{6} \cdot z) \leq 1 - z$; and both of the fourth step and the fifth step apply the AM-GM inequality.

Applying the same arguments to the j -th tone-wise error $\|a_j\|_T^2$, we also have

$$\|a_j\|_T^2 \geq \frac{1}{9} \cdot \left(|v_j - v'_j| + (|v_j| + |v'_j|) \cdot T \cdot \|f'_j - f_j\|_2\right)^2. \tag{46}$$

Following Equation (36), we deduce that

$$\begin{aligned}
|\text{err}_{i,j}| &= \left| (v_i \cdot \overline{v_j} + \overline{v_i} \cdot v_j) \cdot \text{sinc}_T(f_i - f_j) - (v_i \cdot \overline{v'_j} + \overline{v_i} \cdot v'_j) \cdot \text{sinc}_T(f_i - f'_j) \right. \\
&\quad \left. - (v'_i \cdot \overline{v_j} + \overline{v'_i} \cdot v_j) \cdot \text{sinc}_T(f'_i - f_j) + (v'_i \cdot \overline{v'_j} + \overline{v'_i} \cdot v'_j) \cdot \text{sinc}_T(f'_i - f'_j) \right| \\
&= \left| ((v_i - v'_i) \cdot \overline{(v_j - v'_j)} + \overline{(v_i - v'_i)} \cdot (v_j - v'_j)) \cdot \text{sinc}_T(f_i - f_j) \right. \\
&\quad + (v_i \cdot \overline{v'_j} + \overline{v_i} \cdot v'_j) \cdot (\text{sinc}_T(f_i - f_j) - \text{sinc}_T(f_i - f'_j)) \\
&\quad + (v'_i \cdot \overline{v_j} + \overline{v'_i} \cdot v_j) \cdot (\text{sinc}_T(f_i - f_j) - \text{sinc}_T(f'_i - f_j)) \\
&\quad \left. - (v'_i \cdot \overline{v'_j} + \overline{v'_i} \cdot v'_j) \cdot (\text{sinc}_T(f_i - f_j) - \text{sinc}_T(f'_i - f'_j)) \right| \\
&\leq \left| ((v_i - v'_i) \cdot \overline{(v_j - v'_j)} + \overline{(v_i - v'_i)} \cdot (v_j - v'_j)) \cdot \text{sinc}_T(f_i - f_j) \right| \\
&\quad + \left| (v_i \cdot \overline{v'_j} + \overline{v_i} \cdot v'_j) \cdot (\text{sinc}_T(f_i - f_j) - \text{sinc}_T(f_i - f'_j)) \right| \\
&\quad + \left| (v'_i \cdot \overline{v_j} + \overline{v'_i} \cdot v_j) \cdot (\text{sinc}_T(f_i - f_j) - \text{sinc}_T(f'_i - f_j)) \right|
\end{aligned}$$

$$\begin{aligned}
& + \left| (v'_i \cdot \overline{v'_j} + \overline{v'_i} \cdot v'_j) \cdot (\text{sinc}_T(f_i - f_j) - \text{sinc}_T(f'_i - f'_j)) \right| \\
\leq & 2 \cdot |v_i - v'_i| \cdot |v_j - v'_j| \cdot \underbrace{\left| \text{sinc}_T(f_i - f_j) \right|}_{A_8} \\
& + 2 \cdot |v_i| \cdot |v'_j| \cdot \underbrace{\left| \text{sinc}_T(f_i - f_j) - \text{sinc}_T(f_i - f'_j) \right|}_{A_9} \\
& + 2 \cdot |v'_i| \cdot |v_j| \cdot \underbrace{\left| \text{sinc}_T(f_i - f_j) - \text{sinc}_T(f'_i - f_j) \right|}_{A_{10}} \\
& + 2 \cdot |v'_i| \cdot |v'_j| \cdot \underbrace{\left| \text{sinc}_T(f_i - f_j) - \text{sinc}_T(f'_i - f'_j) \right|}_{A_{11}}, \tag{47}
\end{aligned}$$

where the second step follows by elementary calculation; and the third step follows from the triangle inequality.

In what follows, we safely assume $T = \Omega(\frac{d}{\Delta f_{i,j}})$ to be large enough, and upper bound the terms $|A_8|$ and $|A_9|$ and $|A_{10}|$ and $|A_{11}|$ one by one.

Bound on $|A_8|$. Recall the quantity $\text{sinc}_{T,i,j}^{\max}$ defined in Equation (39). We have

$$\begin{aligned}
|A_8| & \leq \text{sinc}_{T,i,j}^{\max} \\
& \lesssim \frac{1}{T \cdot \Delta f_{i,j}} \tag{48}
\end{aligned}$$

where the last step is by Equation (42), and holds whenever $T = \Omega(\frac{d}{\Delta f_{i,j}})$ is large enough.

Bound on $|A_9|$. Under the premises $\|f'_j - f_j\|_2 \leq \Delta_{i,j}$ and $\|f_i - f_j\|_2 \geq \Delta_{i,j}$ and $\|f_i - f'_j\|_2 \geq \Delta_{i,j}$ (see the statement of Lemma 8.6), via a standard geometric argument, we know that the next equation holds for any $\lambda \in [0, 1]$:

$$\|f_i - f''_j(\lambda)\|_2 \geq \frac{\sqrt{3}}{2} \cdot \Delta_{i,j}, \tag{49}$$

where $f''_j(\lambda) = \lambda \cdot f_j + (1 - \lambda) \cdot f'_j$.

Due to the mean value theorem, there exists a particular $\lambda \in [0, 1]$ such that

$$A_9 = \text{sinc}_T(f_i - f_j) - \text{sinc}_T(f_i - f'_j) = (\nabla(f_i - f''_j))^\top (f'_j - f_j), \tag{50}$$

where the gradient $\nabla(f_i - f''_j) \in \mathbb{R}^d$ is given by $\nabla_l(f_i - f''_j) = (\frac{\partial}{\partial \xi_l} \text{sinc}_T(\xi))|_{\xi=f_i-f''_j}$ for each coordinate $l \in [d]$.

Consider a specific coordinate $l \in [d]$. The corresponding partial derivative is

$$\begin{aligned}
\left| \nabla_l(f_i - f''_j) \right| & = \left| \left(\frac{d}{d\xi_l} \text{sinc}_T(\xi_l) \right) \Big|_{\xi_l=f_{i,l}-f''_{j,l}} \right| \cdot \prod_{r \in [d] \setminus \{l\}} \left| \text{sinc}_T(f_{i,r} - f''_{j,r}) \right| \\
& \leq \frac{7}{5} \cdot T \cdot \min \left\{ 1, \frac{1}{T \cdot |f_{i,l} - f''_{j,l}|} \right\} \cdot \prod_{r \in [d] \setminus \{l\}} \left| \text{sinc}_T(f_{i,r} - f''_{j,r}) \right| \\
& \leq \frac{7}{5} \cdot T \cdot \min \left\{ 1, \frac{1}{T \cdot |f_{i,l} - f''_{j,l}|} \right\} \cdot \prod_{r \in [d] \setminus \{l\}} \min \left\{ 1, \frac{1}{\pi \cdot T \cdot |f_{i,r} - f''_{j,r}|} \right\} \\
& \lesssim T \cdot \min \left\{ 1, \frac{1}{T \cdot |f_{i,l} - f''_{j,l}|} \right\} \cdot \prod_{r \in [d] \setminus \{l\}} \min \left\{ 1, \frac{1}{\pi \cdot T \cdot |f_{i,r} - f''_{j,r}|} \right\}
\end{aligned}$$

$$\begin{aligned}
&\lesssim T \cdot \frac{1}{T \cdot \|f_i - f_j''\|_2} \\
&\lesssim \frac{1}{\Delta f_{i,j}},
\end{aligned} \tag{51}$$

where the second step uses Part (f) of Fact 8.2; the third step uses Part (a) of Fact 8.2; the fifth step holds whenever $T = \Omega(\frac{d}{\Delta f_{i,j}})$ is large enough, and can be seen by reusing the arguments for Equation (42); and the last step follows from Equation (49).

We emphasize that Equation (51) holds for any coordinate $l \in [d]$, and therefore holds for the ℓ_∞ -norm $\|\nabla(f_i - f_j'')\|_\infty$ as well. Putting everything together,

$$\begin{aligned}
|A_9| &= |(\nabla(f_i - f_j''))^\top (f_j' - f_j)| \\
&\leq \|\nabla(f_i - f_j'')\|_\infty \cdot \|f_j' - f_j\|_1 \\
&\lesssim \frac{1}{\Delta f_{i,j}} \cdot \|f_j' - f_j\|_1 \\
&\lesssim \frac{\sqrt{d}}{\Delta f_{i,j}} \cdot \|f_j' - f_j\|_2,
\end{aligned} \tag{52}$$

where the first step is by Equation (50); the third step is by Equation (51); and the last step follows because $\sqrt{d} \cdot \|f_j' - f_j\|_2 \geq \|f_j' - f_j\|_1$.

Reapplying the above arguments for $|A_9|$, we also have

$$|A_{10}| \lesssim \frac{\sqrt{d}}{\Delta f_{i,j}} \cdot \|f_i' - f_i\|_2, \tag{53}$$

$$|A_{11}| \lesssim \frac{\sqrt{d}}{\Delta f_{i,j}} \cdot (\|f_i' - f_i\|_2 + \|f_j' - f_j\|_2). \tag{54}$$

Plugging Equations (48) and (52) and (53) and (54) into Equation (47) results in

$$\begin{aligned}
|\text{err}_{i,j}| &\lesssim |v_i - v_i'| \cdot |v_j - v_j'| \cdot |A_8| \\
&\quad + |v_i| \cdot |v_j'| \cdot |A_9| + |v_i'| \cdot |v_j| \cdot |A_{10}| + |v_i'| \cdot |v_j'| \cdot |A_{11}| \\
&\lesssim |v_i - v_i'| \cdot |v_j - v_j'| \cdot \frac{1}{T \cdot \Delta f_{i,j}} \\
&\quad + (|v_i| + |v_i'|) \cdot (|v_j| + |v_j'|) \cdot \frac{\sqrt{d}}{\Delta f_{i,j}} \cdot (\|f_i' - f_i\|_2 + \|f_j' - f_j\|_2) \\
&\lesssim \|a_i\|_T \cdot \|a_j\|_T \cdot \frac{\sqrt{d}}{T \cdot \Delta f_{i,j}},
\end{aligned}$$

where the second step uses Equations (48) and (52) and (53) and (54); and the last step uses Equations (45) and (46).

This completes the proof. \square

8.4.3 One pair is far and one pair is close

Claim 8.9 (Case (iii) for Lemma 8.6). *If $T \cdot \|f_i' - f_i\|_2 \geq \frac{2.05}{\pi}$ and $T \cdot \|f_j' - f_j\|_2 < \frac{2.05}{\pi}$, then we have*

$$|\text{err}_{i,j}| \lesssim \frac{\sqrt{d}}{\Delta f_{i,j} T} \cdot \|a_i\|_T \cdot \|a_j\|_T.$$

Proof. We have shown in Equation (37) that

$$\|a_i\|_T \gtrsim |v_i| + |v'_i|, \quad (55)$$

and have shown in Equation (46) that

$$\|a_j\|_T \gtrsim |v_j - v'_j| + (|v_j| + |v'_j|) \cdot T \cdot \|f'_j - f_j\|_2. \quad (56)$$

Following Equation (47), we deduce that

$$\begin{aligned} |\text{err}_{i,j}| &\lesssim |v_i - v'_i| \cdot |v_j - v'_j| \cdot \left| \text{sinc}_T(f_i - f_j) \right| \\ &\quad + |v_i| \cdot |v'_j| \cdot \left| \text{sinc}_T(f_i - f_j) - \text{sinc}_T(f_i - f'_j) \right| \\ &\quad + |v'_i| \cdot |v_j| \cdot \left| \text{sinc}_T(f_i - f_j) - \text{sinc}_T(f'_i - f_j) \right| \\ &\quad + |v'_i| \cdot |v'_j| \cdot \left| \text{sinc}_T(f_i - f_j) - \text{sinc}_T(f'_i - f'_j) \right| \\ &\lesssim |v_i - v'_i| \cdot |v_j - v'_j| \cdot \underbrace{\left| \text{sinc}_T(f_i - f_j) \right|}_{A_{12}} \\ &\quad + |v_i| \cdot |v'_j| \cdot \underbrace{\left| \text{sinc}_T(f_i - f_j) - \text{sinc}_T(f_i - f'_j) \right|}_{A_{13}} \\ &\quad + |v'_i| \cdot |v_j| \cdot \underbrace{\left(\left| \text{sinc}_T(f_i - f_j) \right| + \left| \text{sinc}_T(f'_i - f_j) \right| \right)}_{A_{14}} \\ &\quad + |v'_i| \cdot |v'_j| \cdot \underbrace{\left(\left| \text{sinc}_T(f_i - f_j) \right| + \left| \text{sinc}_T(f'_i - f'_j) \right| \right)}_{A_{15}} \\ &\lesssim |v_i - v'_i| \cdot |v_j - v'_j| \cdot \frac{1}{T \cdot \Delta f_{i,j}} \\ &\quad + |v_i| \cdot |v'_j| \cdot \frac{\sqrt{d}}{\Delta f_{i,j}} \cdot \|f'_j - f_j\|_2 \\ &\quad + |v'_i| \cdot |v_j| \cdot \frac{1}{T \cdot \Delta f_{i,j}} \\ &\quad + |v'_i| \cdot |v'_j| \cdot \frac{1}{T \cdot \Delta f_{i,j}}, \end{aligned} \quad (57)$$

where the first step applies the triangle inequality; the second step applies Equation (42) to A_{12} , applies Equation (52) to A_{13} , applies Equation (42) to A_{14} , and applies Equation (42) to A_{15} .

Combining Equations (55) and (56) and (57) together, it can be easily seen that

$$|\text{err}_{i,j}| \lesssim \frac{\sqrt{d}}{T \cdot \Delta f_{i,j}} \cdot \|a_i\|_T \cdot \|a_j\|_T.$$

This completes the proof. \square

8.5 Combining tone-wise errors and cross-tone errors

Let $\Re(z) \in \mathbb{R}$ denote the real part of a complex number $z \in \mathbb{C}$.

Lemma 8.10. Let $\{(v_i, f_i)\}_{i \in [k]}$ and $\{(v'_i, f'_i)\}_{i \in [k]}$ be two sets of $k \in \mathbb{N}_{\geq 1}$ tones, for which

$$\min_{i \neq j} \|f_i - f_j\|_1 \geq \eta \quad \text{and} \quad \min_{i \neq j} \|f'_i - f'_j\|_1 \geq \eta \quad \text{and} \quad \min_{i \in [k]} \|f_i - f'_i\|_1 \leq \eta/100$$

Then these two sets can be reindexed such that

$$\frac{1}{T^d} \cdot \int_{\tau \in [-T/2, T/2]^d} \left| \sum_{i \in [k]} a_i(\tau) \right|^2 \cdot d\tau \leq (1 + \alpha) \cdot \sum_{i \in [k]} \frac{1}{T^d} \int_{\tau \in [-T/2, T/2]^d} |a_i(\tau)|^2 d\tau. \quad (58)$$

where

$$\alpha := O(\eta^{-1} \cdot T^{-1}) \cdot \sqrt{d} \cdot \min \left\{ k, \sum_{j=1}^{k-1} \sqrt{d} \cdot j^{-1/d} \right\},$$

which further implies

$$\alpha = \begin{cases} O(\eta^{-1} \cdot T^{-1}) \cdot \log k, & \text{if } d = 1; \\ O(\eta^{-1} \cdot T^{-1}) \cdot \sqrt{d} \cdot \min\{k, \sqrt{d} \cdot k^{1-1/d}\}, & \text{if } d \geq 2. \end{cases}$$

Proof. It follows that

$$\begin{aligned} \text{LHS of (58)} &= \frac{1}{T^d} \cdot \int_{\tau \in [-T/2, T/2]^d} \sum_{i \in [k]} a_i(\tau) \cdot \sum_{i \in [k]} \overline{a_i(\tau)} \cdot d\tau \\ &= \text{diagonal terms} + \text{off-diagonal terms} \end{aligned}$$

where

$$\begin{aligned} \text{diagonal terms} &= \sum_{i \in [k]} \frac{1}{T^d} \cdot \int_{\tau \in [-T/2, T/2]^d} |a_i(\tau)|^2 \cdot d\tau = \sum_{i \in [k]} \|a_i\|_T^2, \\ \text{off-diagonal terms} &= \sum_{i < j} \frac{1}{T^d} \cdot \int_{\tau \in [-T/2, T/2]^d} (a_i(\tau) \cdot \overline{a_j(\tau)} + \overline{a_i(\tau)} \cdot a_j(\tau)) \cdot d\tau \\ &= \sum_{i < j} \mathbf{E}_{\tau \sim \text{Unif}[-T/2, T/2]^d} [a_i(\tau) \cdot \overline{a_j(\tau)} + \overline{a_i(\tau)} \cdot a_j(\tau)] \end{aligned}$$

First we can simplify the off-diagonal terms in the following sense:

$$\begin{aligned} |\text{off-diagonal terms}| &= \left| \sum_{i < j} \mathbf{E}_{\tau \sim \text{Unif}[-T/2, T/2]^d} [a_i(\tau) \cdot \overline{a_j(\tau)} + \overline{a_i(\tau)} \cdot a_j(\tau)] \right| \\ &\leq \sum_{i < j} \left| \mathbf{E}_{\tau \sim \text{Unif}[-T/2, T/2]^d} [a_i(\tau) \cdot \overline{a_j(\tau)} + \overline{a_i(\tau)} \cdot a_j(\tau)] \right| \\ &\lesssim \sum_{i < j} \frac{\sqrt{d}}{T \cdot \Delta f_{i,j}} \cdot \|a_i\|_T \cdot \|a_j\|_T \\ &\leq \sum_{i < j} \frac{\sqrt{d}}{T \cdot \Delta f_{i,j}} \cdot (\|a_i\|_T^2 + \|a_j\|_T^2) \end{aligned}$$

where the second step uses the triangle inequality, and the third step applies Lemma 8.6.

We consider two cases. Case 1. $d = 1$. Case $d \geq 2$. The reason we consider $d = 1$ separately because, for $d = 1$ we can get a much better bound than general d .

Case 1. $d = 1$.

We have

$$\begin{aligned} |\text{off-diagonal terms}| &\lesssim \frac{1}{T\eta} \cdot \sum_{i < j} \frac{1}{|i - j|} \cdot (\|a_i\|_T^2 + \|a_j\|_T^2) \\ &\leq \frac{1}{T\eta} \cdot \sum_{i=1}^k \|a_i\|_T^2 \sum_{j=1}^k \frac{1}{j} \\ &\leq \frac{1}{T\eta} \cdot \log k \cdot \sum_{i=1}^k \|a_i\|_T^2. \end{aligned}$$

Case 2. $d \geq 2$. We give two bounds which are not comparative.

Case 2a.

We have

$$\begin{aligned} |\text{off-diagonal terms}| &\lesssim \frac{\sqrt{d}}{T\eta} \cdot \sum_{i < j} (\|a_i\|_T^2 + \|a_j\|_T^2) \\ &\leq \frac{\sqrt{d}}{T\eta} \cdot k \cdot \sum_{i=1}^k \|a_i\|_T^2. \end{aligned}$$

Case 2b.

We have

$$\begin{aligned} |\text{off-diagonal terms}| &\lesssim \frac{\sqrt{d}}{T\eta} \cdot \sum_{i < j} \frac{\sqrt{d}}{|i - j|^{1/d}} \cdot (\|a_i\|_T^2 + \|a_j\|_T^2) \\ &\leq \frac{\sqrt{d}}{T\eta} \cdot \sqrt{d} \cdot \sum_{i=1}^k \|a_i\|_T^2 \sum_{j=1}^k \frac{1}{j^{1/d}} \\ &\leq \frac{d}{T\eta} \cdot k^{1-1/d} \cdot \sum_{i=1}^k \|a_i\|_T^2. \end{aligned}$$

where the first step follows from Lemma 8.11.

This completes the proof. □

8.6 Geometric property

Lemma 8.11 (Geometric property). *Given a set $\{f_j\}_{j \in [k]} \subseteq \mathbb{R}^d$ of $k \in \mathbb{N}_{\geq 1}$ many d -dimensional frequencies with the minimum ℓ_2 -norm separation $\eta := \min_{i \neq j \in [k]} \|f_i - f_j\|_2 > 0$. Consider any particular frequency f in the set, then these frequencies can be reindexed such that $f_1 = f$ and*

$$\|f_1 - f_j\|_2 \gtrsim j^{1/d} \cdot \eta / \sqrt{d}, \quad \forall j \in [k]$$

which further implies

$$\sum_{j \in [2:k]} \frac{1}{\|f_1 - f_j\|_2} \lesssim k^{1-1/d} \cdot \sqrt{d} / \eta.$$

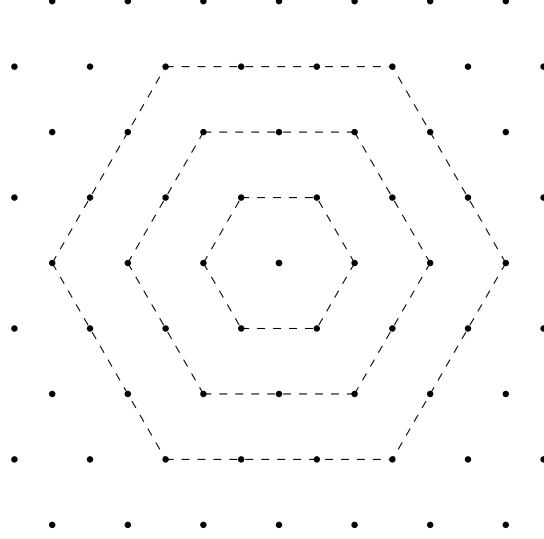


Figure 18: Demonstration for the proof of Lemma 8.11.

Proof. Fix an arbitrary $f_1 = f$ from the set, and w.l.o.g. reindex the frequencies $\{f_j\}_{j \in [k]}$ such that

$$\|f_2 - f_1\|_2 \leq \dots \leq \|f_j - f_1\|_2 \leq \dots \leq \|f_1 - f_k\|_2.$$

We would prove that $\|f_j - f_1\|_2 \gtrsim (|j - 1|^{1/d} / \sqrt{d}) \cdot \eta$ for each $j \in [2 : k]$.

Consider the ℓ_2 -balls that are centered at the frequencies $\{f_j\}_{j \in [k]}$ and have the radius $\eta/2$ each. Those ℓ_2 -balls are disjoint, because the distance of every two frequencies is at least $\eta > 0$.

For a specific $j \in [2 : k]$, let us denote $L := \|f_j - f_1\|_2 \geq \eta$. Clearly (as Figure 18 suggests), all the radius- $(\eta/2)$ balls centered at f_1, \dots, f_j are contained within the bigger ℓ_2 -ball that is centered at f_1 and has the radius $(L + \eta/2)$. Now consider another geometric question:

How many disjoint radius- $(\eta/2)$ balls we can pack within a bigger radius- $(L + \eta/2)$ ball?

Let $m \geq 1$ denote this number, and it is easy to see $m \geq j$.

Indeed, the $m \geq 1$ is call the *packing number*. According to [SSBD14, Page 337], we have

$$m \leq (2\sqrt{d} \cdot (L + \eta/2) / (\eta/2))^d,$$

which after being rearranged gives $L \gtrsim m^{1/d} \cdot \eta / \sqrt{d}$ and further

$$\|f_j - f_1\|_2 \gtrsim j^{1/d} \cdot \eta / \sqrt{d}, \quad \forall j \in [2 : k].$$

Then it is easy to get

$$\sum_{j \in [2:k]} \frac{1}{\|f_1 - f_j\|_2} \lesssim \sum_{j \in [2:k]} \frac{1}{j^{1/d} \cdot \eta / \sqrt{d}} \lesssim k^{1-1/d} \cdot \sqrt{d} / \eta.$$

This finishes the proof. □

8.7 Main result

Theorem 8.12 (Signal reconstruction. Formal version of Theorem 1.2). *Let \mathcal{C} be some universal constant and $C_* = d^2$. When RECOVERYSTAGE observes the signal $x(t)$ over a duration*

$T \gtrsim \eta^{-1} \cdot (d^{4.5} \cdot \log(d) + d^{1.5} \cdot k^{1-1/d}) \cdot \log(kd/\delta)$, the signal estimation error of the k -Fourier-sparse recovered signal $x'(t) := \sum_{i \in [k]} x'_i(t)$ against the observed signal $x(t) = x^*(t) + g(t)$ is bounded as follows:

$$\frac{1}{T^d} \int_{[0, T]^d} |x'(t) - x(t)|^2 \cdot dt \leq (1 + \alpha)(\mathcal{C}^2 + dC_*^2)\mathcal{N}^2, \quad (59)$$

where

$$\alpha = \begin{cases} O(\eta^{-1} \cdot T^{-1}) \cdot \log k, & \text{if } d = 1; \\ O(\eta^{-1} \cdot T^{-1}) \cdot \sqrt{d} \cdot \min\{k, \sqrt{d} \cdot k^{1-1/d}\}, & \text{if } d \geq 2. \end{cases}$$

Proof. This result follows directly by Property III of Theorem 7.18 and Lemma 8.10. \square

9 Conclusion, future directions, other related work

In this paper, we designed a randomized non-adaptive algorithm for the multi-dimensional continuous sparse Fourier transform problem, which achieves a constant approximation under the ℓ_2/ℓ_2 guarantee and, in any constant dimension, takes sublinear samples and running time. Many attractive directions deserve exploring in the future, for which we give a short discussion below.

9.1 Future directions

Approximation ratio. First, whether we can improve the approximation guarantees to $O(\mathcal{N}^2)$ (i.e., making it independent of the dimension $d \geq 1$) or even $(1 + \varepsilon) \cdot \mathcal{N}^2$? In the discrete settings, what enables the $(1 + \varepsilon)$ approximation algorithms is that the noise spectrum $\widehat{g} = \widehat{x} - \widehat{x}^*$ is on the grid (i.e., the whole spectrum except the top- k frequencies) and the noise $g(t)$ is the sum of the sine/cosine functions with given frequencies. The past works like [IK14, Kap16] use $\Theta_d(k/\varepsilon)$ bins, so the average noise in most of bins is $\Theta_d(\varepsilon\mathcal{N}^2/k)$ each.

But in the continuous model, achieving an $(1 + \varepsilon)$ approximation seems difficult, and the hurdles come from the current hashing and sampling methods. First, because the noise $g(t)$ is arbitrary, we cannot hope the noise energy to distribute almost uniformly among the bins after the hashing. Second, as mentioned in Section 1.3.1, the sampling range of the time points $a \in [0, T]^d$ cannot be too large. Namely, we can only hope $|\text{supp}(a)| = c \cdot T^d$ for some $0 < c < 1$, limiting the precision to which the frequencies $f'_i \approx f_i$ can be recovered. These are the two main reasons why, even in the one-dimensional case $d = 1$, the past work [PS15] can only get a constant approximation rather than an $(1 + \varepsilon)$ approximation. (See [PS15, Lemma 3.3] for more details.)

Deterministic algorithm. Actually, no deterministic sublinear-sample algorithm can achieve the ℓ_∞/ℓ_2 -guarantee or the ℓ_2/ℓ_2 -guarantee [DBIPW10]. But under the (weaker) ℓ_∞/ℓ_1 -guarantee, the past works [MZIC19, LN20] design an $\widetilde{O}(k^2)$ -sample deterministic algorithm for the discrete Fourier transform; both works rely on the tools from functional analysis. It would be interesting to see a deterministic algorithm for the continuous Fourier transform (even in one dimension $d = 1$).

$\widetilde{O}(N)$ -time algorithms. For the discrete model (i.e., recover top- k frequencies out of $N = n^d$ ones), several past works improve the sample complexity or other performance guarantees by allowing an $\widetilde{O}(N)$ -time Fourier transform (instead of a sublinear-time one). In the multi-dimensional case, the past works [IK14, NSW19] implement the “point-query” idea (which originates from the sparse recovery/heavy hitter literature) in a clever way, and thus optimize the sample complexity. Can we obtain such results in the continuous model? The main difficulty is that, different from the discrete

cases where the “on-the-grid” frequencies can be checked coordinate by coordinate, the “continuous” frequencies have infinitely many possibilities.

Sample complexity. As mentioned in the introduction, another potential direction is to reduce the sample complexity. Up to the iterated logarithmic factors, our algorithm RECOVERYSTAGE takes $k \cdot (\log k)^{d+O(1)} \cdot \log(F/\eta) \cdot 2^{O(d \log d)}$ samples/running time. Here the term $\log^d k$ is a consequence of our “precise” filter function. As quoted:

[Kap16, Kap17] *“in the discrete settings ... the price to pay for the precision of the filter, however, is that each hashing becomes a $\log^d k$ factor more costly in terms of sample complexity and running time than in the idealized case ...”*

To shave the $\log^d k$ term in the discrete model, the past works [IK14, Kap16] randomize the noise by using the “crude” filters. However, randomizing the noise does not work in the continuous model, since two noise frequencies $f, f' \in \text{TAIL}$ can be arbitrarily close and, no matter how we randomized the noise, the errors can accumulate in the estimation. The exponential dependence on dimension seems to be intrinsic to the current sampling methods, and avoiding it seems need completely different methods.

Set query. A problem in the “sparse recovery” paradigm has two primary tasks: (i) to recover the heavy locations; and (ii) to pin down the masses/densities in those locations. Price [Pri11] pulls the second task out from the sparse recovery literature and defines the “set query” problem. Kapralov [Kap17] introduces and studies the Fourier set query problem in the discrete settings. It would be interesting to explore such problems in the continuous settings.

9.2 Further related works

Over the last two decades, the Sparse FT problem has been investigated and extended in various directions. By now we can even say that it constitutes a “subarea” within sublinear algorithms. These former works can be classified into two lines: (i) those in the one-/multi-dimensional discrete settings [HIKP12a, HIKP12b, IKP14, IK14, Kap16, Kap17, NSW19, KVZ19, BKM⁺21] and follow-ups. (ii) those in the one-dimensional continuous setting [BCG⁺12, Moi15, PS15, CKPS16] and follow-ups.

Compressed sensing is initiated by [CT06, Don06]. Since then, there is a long line of works exploring and extending it in various directions [GLPS10, GLPS10, IP11, IPW11, IR13, PW13, Pri13, AZGR16, LNNT16, BIP⁺16, KP19, NS19]. Compressed sensing allows us to design the sensing matrices, which is the main difference between it and the Sparse FT problem.

Apart from the one-/multi-dimensional discrete/continuous Sparse FT problems that we have considered thus far, where the sampling is carried out in an arbitrary yet *non-adaptive* way, there are other meaningful adjustments to the model.

For example, there is (i) a line of works studying the model where the sampling is conducted in a (more restricted) *uniform* way [RV08, BD08, CGV13, Bou14, HR16, BLL⁺19, and the references therein]; and (ii) another line of works studying the model that allows an algorithm to *adaptively* take the samples and recover the Fourier spectrum [PW13, KCSZ17, and the references therein].

Within theoretical computer science (TCS), Fourier Transform also finds an abundance of applications: integer multiplication [Für09], SUBSET SUM and 3SUM [CLRS09], linear programming [LSZ19, BLSS20, JSWZ21], learning mixture of regressions [CLS20], and fast Johnson-Lindenstrauss transform [LDFU13] etc.

A Building-block function $(G(t), \widehat{G}(f))$ in a single dimension

This appendix presents the construction of a basic function $(G(t), \widehat{G}(f))$ as well as its properties, which serves as the building block of our single-dimensional filter function (see Appendix B) and multi-dimensional filter function (see Appendix 4).

A.1 Construction of function $(G(t), \widehat{G}(f))$

To introduce the building-block function $(G(t), \widehat{G}(f))$, we will employ the rectangular function $\text{rect}_{s_1}(f)$ and the sinc function $\text{sinc}_{s_1}(t)$. Both functions are widely used in the previous literature, and we shall be familiar with their properties given in Fact A.2 (e.g. see [CKPS16]).

Definition A.1 (Two basic functions). Given any $s_1 > 0$, the $\text{rect}_{s_1}(f)$ function and the $\text{sinc}_{s_1}(t)$ function are defined as follows:

- $\text{rect}_{s_1}(f) = 1/s_1 \cdot \mathbb{I}\{|f| \leq s_1/2\}$ for any $f \in \mathbb{R}$. When $s_1 = 1$, we shorthand it as $\text{rect}(f)$.
- $\text{sinc}_{s_1}(t) = \frac{\sin(\pi s_1 t)}{\pi s_1 t}$ for any $t \neq 0$ and $\text{sinc}_{s_1}(0) = 1$. When $s_1 = 1$, we shorthand it as $\text{sinc}(t)$.

Fact A.2 (Facts about basic functions [CKPS16, Appendix C]). *Given any $s_1 > 0$, the following hold for the functions $\text{sinc}_{s_1}(t)$ and $\text{rect}_{s_1}(f)$:*

Part (a): $1 - \frac{\pi^2}{6} \cdot (s_1 t)^2 \leq |\text{sinc}_{s_1}(t)| \leq 1$ for any $t \in \mathbb{R}$.

Part (b): $|\text{sinc}_{s_1}(t)| \leq 1 - \frac{\pi^2}{8} \cdot (s_1 t)^2$ for any $|t| \leq \frac{2.3}{\pi s_1}$.

Part (c): $|\text{sinc}_{s_1}(t)| \leq \min(1, \frac{1}{\pi \cdot |s_1 t|})$ for any $t \in \mathbb{R}$.

Part (d): $\text{sinc}_{s_1}(t) = \widehat{\text{rect}_{s_1}}(t)$ for any $t \in \mathbb{R}$, and $\text{rect}_{s_1}(f) = \widehat{\text{sinc}_{s_1}}(f)$ for any $f \in \mathbb{R}$.

Our building-block function $(G(t), \widehat{G}(f))$ is constructed in the following Definition A.3. This construction is similar to [CKPS16, Definition C.11], and we carefully modify the involved parameters for our later use. We present several important properties of $(G(t), \widehat{G}(f))$ in Section A.2, and then prove these properties in Section A.3.

Definition A.3 (Building-block function in a single dimension). We set the parameters as follows:

- The number of bins in a single dimension $B = \Theta(d \cdot k^{1/d})$ is a certain multiple of $d \in \mathbb{N}_{\geq 1}$.
- The noise level parameter $\delta \in (0, 1)$.
- $\alpha = \Theta(1/d)$ is chosen such that $\frac{1}{100 \cdot (d+1) \cdot \alpha} \in \mathbb{N}_{\geq 1}$ is an integer; clearly $\alpha \leq \frac{1}{100 \cdot (d+1)} \leq \frac{1}{200}$.
- $s_1 = \frac{2B}{\alpha}$ and $s_2 = \frac{1}{B+B/d}$.
- $\ell = \Theta(\log(kd/\delta))$ is an even integer. We safely assume $\ell \geq 1000$.

Then for any $t, f \in \mathbb{R}$ the building-block function $(G(t), \widehat{G}(f))$ is given by

$$\begin{aligned} G(t) &= s_0 \cdot \text{rect}_{s_1}^{*\ell}(t) \cdot \text{sinc}_{s_2}(t) \\ &= s_0 \cdot \text{rect}_{2B/\alpha}^{*\ell}(t) \cdot \text{sinc}_{1/(B+B/d)}(t), \\ \widehat{G}(f) &= s_0 \cdot (\text{sinc}_{s_1}(f))^\ell * \text{rect}_{s_2}(f) \\ &= s_0 \cdot (\text{sinc}_{2B/\alpha}(f))^\ell * \text{rect}_{1/(B+B/d)}(f), \end{aligned}$$

where the scalar $s_0 > 0$ achieves the normalization $\widehat{G}(0) = 1$. Notice that both $G(t)$ and $\widehat{G}(f)$ take real values, and are even functions.

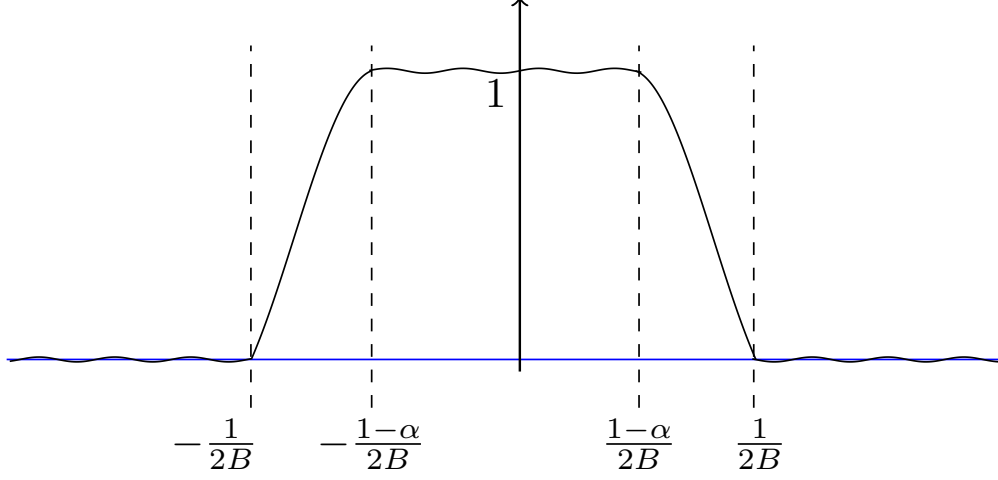


Figure 19: Demonstration for the function $\widehat{G}(f)$ in Lemma A.4.

A.2 Properties of function $(G(t), \widehat{G}(f))$

Lemma A.4 (Building-block function in a single dimension). *The function $(G(t), \widehat{G}(f))[B, \delta, \alpha, \ell]$ given in Definition A.3 satisfies the following (as Figure 19 illustrates):*

- Property I:** *The scalar $s_0 \approx s_1 s_2 \sqrt{\ell} \approx \sqrt{\ell}/\alpha$.*
- Property II:** $1 - \frac{\delta}{\text{poly}(k,d)} \leq \widehat{G}(f) \leq 1$ *when $|f| \leq \frac{1-\alpha}{2B}$.*
- Property III:** $\widehat{G}(f) \in [0, 1]$ *when $\frac{1-\alpha}{2B} \leq |f| \leq \frac{1}{2B}$.*
- Property IV:** $0 \leq \widehat{G}(f) \leq (\pi B f)^{-\ell} \leq \frac{\delta}{\text{poly}(k,d)}$ *when $|f| \geq \frac{1}{2B}$.*
- Property V:** $\text{supp}(G) \subseteq [-\ell \cdot \frac{B}{\alpha}, \ell \cdot \frac{B}{\alpha}]$.
- Property VI:** $\max_{t \in \mathbb{R}} |G(t)| = G(0) \in [\frac{1-\alpha-\delta/(4kd)}{B}, \frac{1+\delta/(4kd)}{B}]$.
- Property VII:** $\sum_{i \in \mathbb{Z}} G(i + 1/2)^2 \leq (1 + \frac{\delta}{4kd})^2 \cdot (1 + \frac{1}{d}) \cdot B^{-1} \lesssim B^{-1}$.

A.3 Proof of properties

Claim A.5 (Property I of Lemma A.4). *The scalar $s_0 \approx s_1 s_2 \sqrt{\ell} \approx \sqrt{\ell}/\alpha$.*

Proof. Recall that the scalar $s_0 > 0$ achieves the normalization $\widehat{G}(0) = 1$. By definition,

$$\begin{aligned}
 \widehat{G}(0) &= s_0 \cdot \int_{-\infty}^{+\infty} (\text{sinc}_{s_1}(\xi))^\ell \cdot \text{rect}_{s_2}(0 - \xi) \cdot d\xi \\
 &= \frac{s_0}{s_2} \cdot \int_{-s_2/2}^{+s_2/2} (\text{sinc}_{s_1}(\xi))^\ell \cdot d\xi \\
 &= \frac{2s_0}{s_2} \cdot \int_0^{s_2/2} (\text{sinc}_{s_1}(\xi))^\ell \cdot d\xi \\
 &= \frac{2s_0}{s_2} \cdot \int_0^{s_2/2} |\text{sinc}_{s_1}(\xi)|^\ell \cdot d\xi,
 \end{aligned}$$

where the second step follows because $\text{rect}_{s_2}(\xi) = \frac{1}{s_2} \cdot \mathbb{I}\{|\xi| \leq \frac{s_2}{2}\}$ for any $\xi \in \mathbb{R}$ (see Definition A.1); the third step follows because $\text{sinc}_{s_1}(\xi)$ is an even function in $\xi \in \mathbb{R}$; and the last step is because $\ell \in \mathbb{N}_{\geq 1}$ is an even integer (see Definition A.3).

Given that $s_1 = \frac{2B}{\alpha}$ and $s_2 = \frac{1}{B+B/d}$ and $0 < \alpha \leq \frac{1}{100 \cdot (d+1)}$ (see Definition A.3), one can easily check that $\frac{2}{\pi s_1} \leq \frac{s_2}{2}$. Accordingly, we know from the additivity of integration that

$$\begin{aligned} \widehat{G}(0) &= \frac{2s_0}{s_2} \cdot \int_0^{2/(\pi s_1)} |\text{sinc}_{s_1}(\xi)|^\ell \cdot d\xi + \frac{2s_0}{s_2} \cdot \int_{2/(\pi s_1)}^{s_2/2} |\text{sinc}_{s_1}(\xi)|^\ell \cdot d\xi \\ &= \frac{2s_0}{s_2} \cdot \underbrace{\int_0^{2/(\pi s_1)} |\text{sinc}_{s_1}(\xi)|^\ell \cdot d\xi}_{A_1} \pm \frac{2s_0}{s_2} \cdot \underbrace{\int_{2/(\pi s_1)}^{+\infty} |\text{sinc}_{s_1}(\xi)|^\ell \cdot d\xi}_{A_2}, \end{aligned}$$

where the second step follows because $(\text{sinc}_{s_1}(\xi))^\ell \geq 0$ for any $\xi \in \mathbb{R}$ (note that ℓ is an even integer; see Definition A.3), and thus $\widehat{G}(0) \cdot \frac{s_2}{2s_0}$ is bounded between $(A_1 - A_2)$ and $(A_1 + A_2)$.

We will verify respectively in Claims A.13 and A.14 (see Section A.5 for the proofs of both claims) that $A_1 \approx \frac{1}{s_1} \cdot \ell^{-1/2}$ and $A_2 = O(\frac{1}{s_1} \cdot 2^{-\ell})$. Under our choice of $\ell = \Theta(\log(kd/\delta))$, it follows that $A_1 \gg A_2$ and thus, that

$$1 = \widehat{G}(0) \approx \frac{2s_0}{s_2} \cdot A_1 \approx \frac{s_0}{s_1 s_2} \cdot \ell^{-1/2},$$

which implies $s_0 \approx s_1 s_2 \sqrt{\ell} \approx \sqrt{\ell}/\alpha$ (since $s_1 = \frac{2B}{\alpha}$ and $s_2 = \frac{1}{B+B/d}$).

This completes the proof of Claim A.5. \square

Claim A.6 (Property II of Lemma A.4). $1 - \frac{\delta}{\text{poly}(k,d)} \leq \widehat{G}(f) \leq 1$ when $|f| \leq \frac{1-\alpha}{2B}$.

Proof. We first prove the upper-bound part that $\widehat{G}(f) \leq \widehat{G}(0) = 1$ for any $f \in \mathbb{R}$. Since $\widehat{G}(f)$ is an even function (see Definition A.3), it suffices to deal with the case that $f \geq 0$. By definition,

$$\begin{aligned} \widehat{G}(f) - \widehat{G}(0) &= s_0 \cdot \int_{-\infty}^{+\infty} (\text{sinc}_{s_1}(\xi))^\ell \cdot (\text{rect}_{s_2}(f - \xi) - \text{rect}_{s_2}(0 - \xi)) \cdot d\xi \\ &= \frac{s_0}{s_2} \cdot \int_{-s_2/2+f}^{s_2/2+f} (\text{sinc}_{s_1}(\xi))^\ell \cdot d\xi - \frac{s_0}{s_2} \cdot \int_{-s_2/2}^{s_2/2} (\text{sinc}_{s_1}(\xi))^\ell \cdot d\xi \\ &= \frac{s_0}{s_1 s_2} \cdot \int_{-s_1 s_2/2+s_1 f}^{s_1 s_2/2+s_1 f} (\text{sinc}(\xi))^\ell \cdot d\xi - \frac{s_0}{s_1 s_2} \cdot \int_{-s_1 s_2/2}^{s_1 s_2/2} (\text{sinc}(\xi))^\ell \cdot d\xi \\ &= \frac{s_0}{s_1 s_2} \cdot \left(\int_{s_1 s_2/2}^{s_1 s_2/2+s_1 f} - \int_{-s_1 s_2/2}^{-s_1 s_2/2+s_1 f} \right) (\text{sinc}(\xi))^\ell \cdot d\xi \\ &= \frac{s_0}{s_1 s_2} \cdot \left(\int_{s_1 s_2/2}^{s_1 s_2/2+s_1 f} - \int_{s_1 s_2/2-s_1 f}^{s_1 s_2/2} \right) (\text{sinc}(\xi))^\ell \cdot d\xi \\ &= \frac{s_0}{s_1 s_2} \cdot \int_0^{s_1 f} \underbrace{\left((\text{sinc}(\frac{s_1 s_2}{2} + \xi))^\ell - (\text{sinc}(\frac{s_1 s_2}{2} - \xi))^\ell \right)}_{A_3} \cdot d\xi, \end{aligned} \tag{60}$$

where the second step follows because $\text{rect}_{s_2}(\xi) = \frac{1}{s_2} \cdot \mathbb{I}\{|\xi| \leq \frac{s_2}{2}\}$ for any $\xi \in \mathbb{R}$ (see Definition A.1); the third step is by substitution; the fourth step applies the additivity of integration; the fifth step follows because $\text{sinc}(\xi)$ is an even function in $\xi \in \mathbb{R}$; and the last step is by substitution.

Given Equation (60), it suffices to show that $A_3 \leq 0$ when $\xi \in [0, s_1 f]$. Recall that $\ell \in \mathbb{N}_{\geq 1}$ is an even integer, and $\frac{s_1 s_2}{2} = (\frac{2B}{\alpha}) \cdot (\frac{1}{B+B/d}) \cdot \frac{1}{2} = \frac{1}{100 \cdot (d+1) \cdot \alpha} \cdot (100 \cdot d) \in \mathbb{N}_{\geq 1}$ is an integer (see Definition A.3). Therefore, for any $\xi \in [0, s_1 f]$ we have

$$\begin{aligned}
A_3 &= \left| \operatorname{sinc}\left(\frac{s_1 s_2}{2} + \xi\right) \right|^\ell - \left| \operatorname{sinc}\left(\frac{s_1 s_2}{2} - \xi\right) \right|^\ell \\
&= \frac{|\sin(\pi \cdot \frac{s_1 s_2}{2} + \pi \cdot \xi)|^\ell}{|\pi \cdot \frac{s_1 s_2}{2} + \pi \cdot \xi|^\ell} - \frac{|\sin(\pi \cdot \frac{s_1 s_2}{2} - \pi \cdot \xi)|^\ell}{|\pi \cdot \frac{s_1 s_2}{2} - \pi \cdot \xi|^\ell} \\
&= \frac{|\sin(\pi \cdot \xi)|^\ell}{|\pi \cdot \frac{s_1 s_2}{2} + \pi \cdot \xi|^\ell} - \frac{|\sin(-\pi \cdot \xi)|^\ell}{|\pi \cdot \frac{s_1 s_2}{2} - \pi \cdot \xi|^\ell} \\
&= |\sin(\pi \cdot \xi)|^\ell \cdot \left(\left| \pi \cdot \frac{s_1 s_2}{2} + \pi \cdot \xi \right|^{-\ell} - \left| \pi \cdot \frac{s_1 s_2}{2} - \pi \cdot \xi \right|^{-\ell} \right) \\
&\leq 0,
\end{aligned}$$

where the first step follows because $\ell \in \mathbb{N}_{\geq 1}$ is an even integer; the third step follows because $|\sin(\pi \cdot \xi)|$ is a periodic function in $\xi \in \mathbb{R}$ and its basic period is 1 (notice that $\frac{s_1 s_2}{2}$ is an integer); and the fourth step follows because $|\sin(\pi \cdot \xi)|$ is an even function.

To see the lower-bound part, due to the normalization $\widehat{G}(0) = 1$, we have

$$\begin{aligned}
1 - \widehat{G}(f) &= \widehat{G}(0) - \widehat{G}(f) \\
&= \frac{s_0}{s_1 s_2} \cdot \int_0^{s_1 f} \left(\left(\operatorname{sinc}\left(\frac{s_1 s_2}{2} + \xi\right) \right)^\ell - \left(\operatorname{sinc}\left(\frac{s_1 s_2}{2} - \xi\right) \right)^\ell \right) \cdot d\xi \\
&\leq \frac{s_0}{s_1 s_2} \cdot \int_0^{s_1 f} \left| \operatorname{sinc}\left(\frac{s_1 s_2}{2} + \xi\right) \right|^\ell \cdot d\xi \\
&\leq \frac{s_0}{s_1 s_2} \cdot \int_0^{s_1 f} \frac{1}{\pi^\ell \cdot |s_1 s_2 / 2 + \xi|^\ell} \cdot d\xi \\
&\leq \frac{s_0}{s_1 s_2} \cdot s_1 f \cdot \frac{1}{\pi^\ell \cdot (s_1 s_2 / 2)^\ell} \\
&= s_0 \cdot (B + B/d) \cdot f \cdot \left(\frac{(d+1) \cdot \alpha}{\pi \cdot d} \right)^\ell \\
&\leq s_0 \cdot 2B \cdot f \cdot \left(\frac{(d+1) \cdot \alpha}{\pi \cdot d} \right)^\ell,
\end{aligned}$$

where the second step follows from Equation (60); the third step follows because ℓ is an even integer (see Definition A.3), namely $(\operatorname{sinc}(\xi))^\ell \geq 0$ for any $\xi \in \mathbb{R}$; the fourth step is by Part (c) of Fact A.2; and the sixth step is by $s_1 = \frac{2B}{\alpha}$ and $s_2 = \frac{1}{B+B/d}$ (see Definition A.3).

According to Claim A.5, for some universal constant $C_0 > 0$, we have $s_0 \leq C_0 \cdot \sqrt{\ell}/\alpha$. Also, as promised by the concerning claim, $|f| \leq \frac{1-\alpha}{2B} \leq \frac{1}{2B}$. Plugging these into the above inequality:

$$\begin{aligned}
1 - \widehat{G}(f) &\leq \left(C_0 \cdot \sqrt{\ell}/\alpha \right) \cdot 2B \cdot \frac{1}{2B} \cdot \left(\frac{(d+1) \cdot \alpha}{\pi \cdot d} \right)^\ell \\
&= C_0 \cdot \sqrt{\ell} \cdot \frac{d+1}{\pi \cdot d} \cdot \left(\frac{(d+1) \cdot \alpha}{\pi \cdot d} \right)^{\ell-1} \\
&\leq C_0 \cdot \sqrt{\ell} \cdot \left(\frac{(d+1) \cdot \alpha}{\pi \cdot d} \right)^{\ell-1}
\end{aligned}$$

$$\begin{aligned}
&\leq C_0 \cdot \sqrt{\ell} \cdot \left(\frac{1}{100\pi \cdot d} \right)^{\ell-1} \\
&\leq \frac{\delta}{\text{poly}(k, d)},
\end{aligned}$$

where the third step follows because $\frac{d+1}{\pi \cdot d} \leq \frac{2}{\pi} \leq 1$; the fourth step follows because $0 < \alpha \leq \frac{1}{100 \cdot (d+1)}$ (see Definition A.3); and the last step holds for any large enough $\ell = \Theta(\log(kd/\delta))$.

This completes the proof of Claim A.6. \square

Claim A.7 (Property III of Lemma A.4). $\widehat{G}(f) \in [0, 1]$ when $\frac{1-\alpha}{2B} \leq |f| \leq \frac{1}{2B}$.

Proof. The upper-bound part has been shown in the proof of Claim A.6, namely $\widehat{G}(f) \leq \widehat{G}(0) = 1$ for any $f \in \mathbb{R}$. The lower-bound part is trivial, since both functions $(\text{sinc}_{s_1}(f))^\ell$ and $\text{rect}_{s_2}(f)$ are nonnegative (note that $\ell \in \mathbb{N}_{\geq 1}$ is an even integer; see Definition A.3).

This completes the proof of Claim A.7. \square

Claim A.8 (Property IV of Lemma A.4). $0 \leq \widehat{G}(f) \leq (\pi B f)^{-\ell} \leq \frac{\delta}{\text{poly}(k, d)}$ when $|f| \geq \frac{1}{2B}$.

Proof. The lower-bound part has been shown in the proof of Claim A.7, namely $\ell \in \mathbb{N}_{\geq 1}$ is an even integer (see Definition A.3) and thus both functions $(\text{sinc}_{s_1}(f))^\ell$ and $\text{rect}_{s_2}(f)$ are nonnegative.

For the upper-bound part, since $\widehat{G}(f)$ is an even function, it suffices to handle the case $f \geq \frac{1}{2B}$. By definition,

$$\begin{aligned}
\widehat{G}(f) &= s_0 \cdot \int_{-\infty}^{+\infty} (\text{sinc}_{s_1}(\xi))^\ell \cdot \text{rect}_{s_2}(f - \xi) \cdot d\xi \\
&= \frac{s_0}{s_2} \cdot \int_{f-s_2/2}^{f+s_2/2} (\text{sinc}_{s_1}(\xi))^\ell \cdot d\xi \\
&= \frac{s_0}{s_2} \cdot \int_{f-s_2/2}^{f+s_2/2} |\text{sinc}_{s_1}(\xi)|^\ell \cdot d\xi \\
&\leq \frac{s_0}{s_2} \cdot \int_{f-s_2/2}^{f+s_2/2} \frac{1}{\pi^\ell \cdot |s_1 \xi|^\ell} \cdot d\xi,
\end{aligned}$$

where the second step follows because $\text{rect}_{s_2}(\xi) = \frac{1}{s_2} \cdot \mathbb{I}\{|\xi| \leq \frac{s_2}{2}\}$ for any $\xi \in \mathbb{R}$ (see Definition A.1); the third step follows because $\ell \in \mathbb{N}_{\geq 1}$ is an even integer (see Definition A.3); and the last step is by Part (c) of Fact A.2.

Recall Definition A.3 that $s_2 = \frac{1}{B+B/d}$. Given this and since we assume $f \geq \frac{1}{2B}$, one can easily check that the above interval of integral is lower bounded by $f - s_2/2 \geq f/(d+1)$. Hence,

$$\begin{aligned}
\widehat{G}(f) &\leq \frac{s_0}{s_2} \cdot s_2 \cdot \frac{1}{\pi^\ell \cdot |s_1 \xi|^\ell} \Big|_{\xi=f/(d+1)} \\
&= s_0 \cdot \left(\frac{(d+1) \cdot \alpha}{2\pi B f} \right)^\ell,
\end{aligned}$$

where the second step follows because $s_1 = \frac{2B}{\alpha}$ and $s_2 = \frac{1}{B+B/d}$. According to Claim A.5, for some universal constant $C_0 > 0$, we have $s_0 \leq C_0 \sqrt{\ell}/\alpha$. As a consequence,

$$\widehat{G}(f) \leq C_0 \cdot \frac{\sqrt{\ell}}{\alpha} \cdot \left(\frac{(d+1) \cdot \alpha}{2\pi B f} \right)^\ell$$

$$\begin{aligned}
&\leq C_0 \cdot 100 \cdot (d+1) \cdot \sqrt{\ell} \cdot (200\pi Bf)^{-\ell} \\
&\leq (\pi Bf)^{-\ell},
\end{aligned}$$

where the second step follows because, given that $\ell \geq 1000$, the concerning formula $C_0 \cdot \frac{\sqrt{\ell}}{\alpha} \cdot (\frac{(d+1) \cdot \alpha}{2\pi Bf})^\ell$ is an increasing function when $0 < \alpha \leq \frac{1}{100 \cdot (d+1)}$ (see Definition A.3); and the last step, which is equivalent to $\frac{C_0 \cdot 100 \cdot (d+1) \cdot \sqrt{\ell}}{200^\ell} \leq 1$, holds for any large enough $\ell = \Theta(\log(kd/\delta))$.

Following the above calculation, for any $f \geq \frac{1}{2B}$ we have

$$\widehat{G}(f) \leq (\pi Bf)^{-\ell} \leq (\pi/2)^{-\ell} \leq \frac{\delta}{\text{poly}(k, d)},$$

where the last step holds for any large enough $\ell = \Theta(\log(kd/\delta))$.

This completes the proof of Claim A.8. \square

Claim A.9 (Property V of Lemma A.4). $\text{supp}(G) \subseteq [-\ell \cdot \frac{B}{\alpha}, \ell \cdot \frac{B}{\alpha}]$.

Proof. Recall that $s_1 = \frac{2B}{\alpha}$. By definition, the function $\text{rect}_{s_1}(t) = \frac{1}{s_1} \cdot \mathbb{I}\{|t| \leq \frac{s_1}{2}\}$ is supported on the interval $t \in [-\frac{s_1}{2}, \frac{s_1}{2}]$, and thus $\text{rect}_{s_1}^{\ast \ell}(t)$ is supported on $t \in [-\ell \cdot \frac{s_1}{2}, \ell \cdot \frac{s_1}{2}] = [-\ell \cdot \frac{B}{\alpha}, \ell \cdot \frac{B}{\alpha}]$. Clearly, the later interval contains the support of the function $G(t) = s_0 \cdot \text{rect}_{s_1}^{\ast \ell}(t) \cdot \text{sinc}_{s_2}(t)$.

This completes the proof of Claim A.9. \square

Claim A.10 (Property VI of Lemma A.4). $\max_{t \in \mathbb{R}} |G(t)| = G(0) \in [\frac{1-\alpha-\delta/(4kd)}{B}, \frac{1+\delta/(4kd)}{B}]$.

Proof. Observe that $\widehat{G}(\xi) = s_0 \cdot (\text{sinc}_{s_1}(f))^\ell \ast \text{rect}_{s_2}(f)$ is an even function in $\xi \in \mathbb{R}$, since both $\text{sinc}_{s_1}(\xi)$ and $\text{rect}_{s_2}(\xi)$ are even functions.

We first prove that $\max_{t \in \mathbb{R}} |G(t)| = G(0)$. By the definition of the inverse CFT,

$$\int_{-\infty}^{+\infty} \widehat{G}(\xi) \cdot d\xi = G(0) \leq \max_{t \in \mathbb{R}} G(t) \leq \max_{t \in \mathbb{R}} |G(t)|.$$

Also, for any $t \in \mathbb{R}$ we can derive $G(t)$ from $\widehat{G}(f)$ via the inverse CFT:

$$\begin{aligned}
|G(t)| &= \left| \int_{-\infty}^{+\infty} \widehat{G}(\xi) \cdot e^{2\pi i t \cdot \xi} \cdot d\xi \right| \\
&= \left| \int_{-\infty}^{+\infty} \widehat{G}(\xi) \cdot (\cos(2\pi t \cdot \xi) + \mathbf{i} \cdot \sin(2\pi t \cdot \xi)) \cdot d\xi \right| \\
&= \left| \int_{-\infty}^{+\infty} \widehat{G}(\xi) \cdot \cos(2\pi t \cdot \xi) \cdot d\xi \right| \\
&\leq \int_{-\infty}^{+\infty} |\widehat{G}(\xi)| \cdot |\cos(2\pi t \cdot \xi)| \cdot d\xi \\
&\leq \int_{-\infty}^{+\infty} |\widehat{G}(\xi)| \cdot d\xi \\
&= \int_{-\infty}^{+\infty} \widehat{G}(\xi) \cdot d\xi
\end{aligned}$$

where the third step follows because $\widehat{G}(\xi)$ is an even function in $\xi \in \mathbb{R}$ (see Definition A.3), whereas $\sin(2\pi t \cdot \xi)$ is an odd function; the fifth step is because $|\cos(2\pi t \cdot \xi)| \leq 1$ for any $\xi \in \mathbb{R}$; and the last step follows as $\widehat{G}(\xi) \geq 0$ for any $\xi \in \mathbb{R}$ (see Claims A.6, A.7 and A.8).

We conclude from the above that

$$\max_{t \in \mathbb{R}} |G(t)| = \int_{-\infty}^{+\infty} \widehat{G}(\xi) \cdot d\xi = 2 \cdot \int_0^{+\infty} \widehat{G}(\xi) \cdot d\xi = 2 \cdot (A_4 + A_5 + A_6), \quad (61)$$

where the second step follows as $\widehat{G}(\xi)$ is an even function in $\xi \in \mathbb{R}$; and for the third step we denote the terms A_4 and A_5 and A_6 as follows:

$$\begin{aligned} A_4 &= \int_0^{(1-\alpha)/(2B)} \widehat{G}(\xi) \cdot d\xi, \\ A_5 &= \int_{(1-\alpha)/(2B)}^{1/(2B)} \widehat{G}(\xi) \cdot d\xi, \\ A_6 &= \int_{1/(2B)}^{+\infty} \widehat{G}(\xi) \cdot d\xi. \end{aligned}$$

Let us quantify the three terms A_4 and A_5 and A_6 respectively:

- $A_4 \in [\frac{1-\alpha-\delta/(4kd)}{2B}, \frac{1-\alpha}{2B}]$. This is because $1 - \frac{\delta}{4kd} \leq \widehat{G}(\xi) \leq 1$ for any $\xi \in [0, \frac{1-\alpha}{2B}]$ (Claim A.6); we shall notice that $0 < \alpha \leq \frac{1}{100 \cdot (d+1)} < 1$ and that $B > 1$ (see Definition A.3).
- $A_5 \in [0, \frac{\alpha}{2B}]$. This is because $\widehat{G}(\xi) \in [0, 1]$ for any $\xi \in [\frac{1-\alpha}{2B}, \frac{1}{2B}]$ (see Claim A.7).
- $A_6 \in [0, \frac{\delta/(4kd)}{2B}]$. Based on Claim A.8, we have $0 \leq \widehat{G}(\xi) \leq (\pi B \xi)^{-\ell}$ for any $\xi \geq \frac{1}{2B}$. Then the lower-bound part $A_6 \geq 0$ follows immediately. For the upper-bound part, we have

$$\begin{aligned} A_6 &\leq \int_{1/(2B)}^{+\infty} (\pi B \xi)^{-\ell} \cdot d\xi \\ &= \frac{1}{\pi B} \cdot \int_{\pi/2}^{+\infty} \xi^{-\ell} \cdot d\xi \\ &= \frac{1}{2B} \cdot \frac{1}{\ell-1} \cdot (\pi/2)^{-\ell} \\ &\leq \frac{\delta/(4kd)}{2B}, \end{aligned}$$

where the second step is by substitution; the third step is by elementary calculation; and the last step, given Definition A.3, holds for any large enough $\ell = \Theta(\log(kd/\delta))$.

Applying the above bounds to Equation (61) completes the proof of Claim A.10. \square

Claim A.11 (Property VII of Lemma A.4). $\sum_{i \in \mathbb{Z}} G(i + 1/2)^2 \leq (1 + \frac{\delta}{4kd})^2 \cdot (1 + \frac{1}{d}) \cdot B^{-1} \lesssim B^{-1}$.

Proof. We first prove by induction that $\text{rect}_{s_1}^{*\ell}(t)$ is an even function and is non-increasing for any $t \geq 0$. Obviously, $\text{rect}_{s_1}(t)$ itself meets the both properties. Given any $\ell' < \ell$, w.l.o.g. we assume $\text{rect}_{s_1}^{*\ell'}(t)$ to satisfy the two properties as well. Then for any $t \in \mathbb{R}$, it follows that

$$\begin{aligned} \text{rect}_{s_1}^{*\ell'+1}(-t) &= \int_{-\infty}^{+\infty} \text{rect}_{s_1}(-t - \tau) \cdot \text{rect}_{s_1}^{*\ell'}(\tau) \cdot d\tau \\ &= \int_{-\infty}^{+\infty} \text{rect}_{s_1}(t + \tau) \cdot \text{rect}_{s_1}^{*\ell'}(-\tau) \cdot d\tau \\ &= \text{rect}_{s_1}^{*\ell'+1}(t), \end{aligned}$$

namely $\text{rect}_{s_1}^{*\ell'+1}(t)$ is also an even function. In addition, for any $t' \geq t \geq 0$ we have

$$\begin{aligned}
\text{rect}_{s_1}^{*\ell'+1}(t') - \text{rect}_{s_1}^{*\ell'+1}(t) &= \int_{-\infty}^{+\infty} \text{rect}_{s_1}(t' - \tau) \cdot \text{rect}_{s_1}^{*\ell'}(\tau) \cdot d\tau \\
&\quad - \int_{-\infty}^{+\infty} \text{rect}_{s_1}(t - \tau) \cdot \text{rect}_{s_1}^{*\ell'}(\tau) \cdot d\tau \\
&= 1/s_1 \cdot \int_{t'-s_2/2}^{t'+s_2/2} \text{rect}_{s_1}^{*\ell'}(\tau) \cdot d\tau - 1/s_1 \cdot \int_{t-s_2/2}^{t+s_2/2} \text{rect}_{s_1}^{*\ell'}(\tau) \cdot d\tau \\
&= 1/s_1 \cdot \int_{t+s_2/2}^{t'+s_2/2} \text{rect}_{s_1}^{*\ell'}(\tau) \cdot d\tau - 1/s_1 \cdot \int_{t-s_2/2}^{t'-s_2/2} \text{rect}_{s_1}^{*\ell'}(\tau) \cdot d\tau \\
&= 1/s_1 \cdot \int_t^{t'} \left(\text{rect}_{s_1}^{*\ell'}(\tau + s_2/2) - \text{rect}_{s_1}^{*\ell'}(\tau - s_2/2) \right) \cdot d\tau \\
&\leq 0,
\end{aligned}$$

where the second step follows since $\text{rect}_{s_1}(\xi) = \frac{1}{s_1} \cdot \mathbb{I}\{|\xi| \leq \frac{s_1}{2}\}$ for any $\xi \in \mathbb{R}$ (see Definition A.1); the third step is by the additivity of integration; the fourth step is by substitution; and the last step uses our induction hypotheses that $\text{rect}_{s_1}^{*\ell'}(t)$ is an even function and is non-increasing when $t \geq 0$. Thus, $\text{rect}_{s_1}^{*\ell'+1}(t)$ also meets the properties, and our claim follows by induction.

Further, it is easy to see that $\text{rect}_{s_1}^{*\ell}(t)$ is a non-negative function. Put everything together:

$$\begin{aligned}
\max_{\tau \in \mathbb{R}} \{s_0^2 \cdot \text{rect}_{s_1}^{*\ell}(\tau)^2\} &= s_0^2 \cdot \text{rect}_{s_1}^{*\ell}(0)^2 \\
&= G(0)^2 / \text{sinc}_{s_2}(0)^2 \\
&= G(0)^2 \\
&\leq (1 + \delta/(4kd))^2 \cdot B^{-2},
\end{aligned}$$

where the third step follows because $\text{sinc}_{s_1}(0) = 1$ (see Definition A.1); and the fourth step follows from Claim A.10.

For any $t \in \mathbb{R}$, we infer from the above that

$$\begin{aligned}
G(t)^2 &= s_0^2 \cdot \text{rect}_{s_1}^{*\ell}(t)^2 \cdot \text{sinc}_{s_2}(t)^2 \\
&\leq \max_{\tau \in \mathbb{R}} \{s_0^2 \cdot \text{rect}_{s_1}^{*\ell}(\tau)^2\} \cdot \text{sinc}_{s_2}(t)^2 \\
&\leq (1 + \delta/(4kd))^2 \cdot B^{-2} \cdot \text{sinc}_{s_2}(t)^2
\end{aligned}$$

Further, given that $s_2 = \frac{1}{B+B/d} < 1$ (see Definition A.3), we have

$$\begin{aligned}
\sum_{i \in \mathbb{Z}} G(i + 1/2)^2 &\leq (1 + \delta/(4kd))^2 \cdot B^{-2} \cdot \sum_{i \in \mathbb{Z}} \text{sinc}_{1/(B+B/d)}(i + 1/2)^2 \\
&= (1 + \delta/(4kd))^2 \cdot B^{-2} \cdot (B + B/d) \\
&= (1 + \delta/(4kd))^2 \cdot (1 + 1/d) \cdot B^{-1},
\end{aligned}$$

where the second step, which is equivalent to $\sum_{i=0}^{+\infty} \text{sinc}_{1/(B+B/d)}(i + 1/2)^2 = B + B/d$, can be directly inferred from [BJP73, Equation (1)].

This completes the proof of Claim A.11. \square

A.4 Construction and properties of standard window function $(G'(t), \widehat{G}'(f))$

We associate the building-block function with the standard window function $(G'(t), \widehat{G}'(f))[B, \delta, \alpha, \ell]$ (similar to the ones used in [HIKP12a, HIKP12b]), which is more convenient for our later use.

Lemma A.12 (Standard window function in a single dimension). *Consider the building-block function $(G(t), \widehat{G}(f))[B, \delta, \alpha, \ell]$ given in Definition A.3, there exists another function $(G'(t), \widehat{G}'(f))$ such that:*

Property I: $\widehat{G}'(f) = 1$ when $|f| \leq \frac{1-\alpha}{2B}$.

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

Property III: $\widehat{G}'(f) = 0$ when $|f| \geq \frac{1}{2B}$.

Property IV: $\|\widehat{G}' - \widehat{G}\|_\infty = \max_{f \in \mathbb{R}} |\widehat{G}'(f) - \widehat{G}(f)| \leq \frac{\delta}{\text{poly}(k, d)}$.

Proof. We define $\widehat{G}'(f)$ as follows; note that, similar to $\widehat{G}(f)$, this is also an even function:

$$\widehat{G}'(f) = \begin{cases} 1, & \forall |f| \leq \frac{1-\alpha}{2B}; \\ \widehat{G}(f), & \forall |f| \in (\frac{1-\alpha}{2B}, \frac{1}{2B}]; \\ 0, & \forall |f| > \frac{1}{2B}. \end{cases}$$

By construction, Properties I and III follows directly. Further, Property II follows from Property III of Lemma A.4, and Property IV follows from Properties II to IV of Lemma A.4.

This completes the proof of Lemma A.12. \square

A.5 Facts

The following facts are helpful in proving Claim A.5.

Claim A.13. $\int_0^{2/(\pi s_1)} (\text{sinc}_{s_1}(\xi))^\ell \cdot d\xi \approx \frac{1}{s_1} \cdot \ell^{-1/2}$.

Proof. Let $i^* = \lceil 2/\pi \cdot \sqrt{\ell/8} \rceil - 1$. We safely that assume $\ell = \Theta(\log(kd/\delta))$ is an integer larger than 1000 (see Definition A.3), which guarantees the following facts:

- (a): The integrand $(\text{sinc}_{s_1}(\xi))^\ell \geq 0$ for any $\xi \in \mathbb{R}$;
- (b): $i^* \geq 2/\pi \cdot \sqrt{\ell/8} - 1 \geq 2/\pi \cdot \sqrt{1000/8} - 1 \approx 6.118 \geq 6$.
- (c): $i^* \cdot \sqrt{8/\ell} \leq 2/\pi \cdot \sqrt{\ell/8} \cdot \sqrt{8/\ell} = 2/\pi$;
- (d): $(i^* + 1) \cdot \sqrt{8/\ell} \geq 2/\pi \cdot \sqrt{\ell/8} \cdot \sqrt{8/\ell} = 2/\pi$; and
- (e): $(i^* + 1) \cdot \sqrt{8/\ell} \leq (2/\pi \cdot \sqrt{\ell/8} + 1) \cdot \sqrt{8/\ell} \leq 2/\pi + \sqrt{8/1000} \approx \frac{2.281}{\pi} \leq \frac{2.3}{\pi}$.

These facts are useful in proving the current claim.

For the upper-bound part, we have

$$\int_0^{2/(\pi s_1)} (\text{sinc}_{s_1}(\xi))^\ell \cdot d\xi = \frac{1}{s_1} \cdot \int_0^{2/\pi} (\text{sinc}(\xi))^\ell \cdot d\xi$$

$$\begin{aligned}
&= \frac{1}{s_1} \cdot \int_0^{2/\pi} |\text{sinc}(\xi)|^\ell \cdot d\xi \\
&\leq \frac{1}{s_1} \cdot \int_0^{(i^*+1) \cdot \sqrt{8/\ell}} |\text{sinc}(\xi)|^\ell \cdot d\xi \\
&= \frac{1}{s_1} \cdot \sum_{i=0}^{i^*} \int_{i \cdot \sqrt{8/\ell}}^{(i+1) \cdot \sqrt{8/\ell}} |\text{sinc}(\xi)|^\ell \cdot d\xi,
\end{aligned} \tag{62}$$

where the first step is by substitution; the second step by because $\ell \in \mathbb{N}_{\geq 1}$ is an even integer (see Definition A.3); the third step follows from the above Fact (d); and the last step follows from the additivity of integration.

Given the above Fact (e), the whole interval of integral $\xi \in [0, (i^* + 1) \cdot \sqrt{8/\ell}]$ is a subset of $\xi \in [0, \frac{2.3}{\pi}]$, namely Parts (b) of Fact A.2 is applicable here. In particular, each i -th summand in Equation (62) equals

$$\begin{aligned}
\int_{i \cdot \sqrt{8/\ell}}^{(i+1) \cdot \sqrt{8/\ell}} |\text{sinc}(\xi)|^\ell \cdot d\xi &\leq \int_{i \cdot \sqrt{8/\ell}}^{(i+1) \cdot \sqrt{8/\ell}} (1 - \pi^2/8 \cdot \xi^2)^\ell \cdot d\xi \\
&\leq \int_{i \cdot \sqrt{8/\ell}}^{(i+1) \cdot \sqrt{8/\ell}} (1 - \pi^2 i^2/\ell)^\ell \cdot d\xi \\
&= \sqrt{8/\ell} \cdot (1 - \pi^2 i^2/\ell)^\ell \\
&\leq \sqrt{8/\ell} \cdot e^{-\pi^2 \cdot i^2},
\end{aligned} \tag{63}$$

where the first step is follows from Parts (b) of Fact A.2; the second step is because $1 - \pi^2/8 \cdot t^2 \leq (1 - \pi^2/8 \cdot t^2)|_{t=i \cdot \sqrt{8/\ell}} = 1 - \pi^2 i^2/\ell$; and the last step is by $0 \leq 1 - \frac{1}{x} \leq e^{-x}$ for any $x \in (0, 1)$.

Applying Equation (63) to Equation (62) over all $i \in [0 : i^*]$ results in

$$\begin{aligned}
\int_0^{2/(\pi s_1)} (\text{sinc}_{s_1}(\xi))^\ell \cdot d\xi &\leq \frac{1}{s_1} \cdot \sum_{i=0}^{i^*} \sqrt{8/\ell} \cdot e^{-\pi^2 \cdot i^2} \\
&\leq \frac{1}{s_1} \cdot \sqrt{8/\ell} \cdot \sum_{i=0}^{+\infty} e^{-\pi^2 \cdot i^2} \\
&\leq \frac{1}{s_1} \cdot \sqrt{8/\ell} \cdot \sum_{i=0}^{+\infty} \frac{1}{(1+i)^2} \\
&= \frac{1}{s_1} \cdot \sqrt{8/\ell} \cdot \frac{\pi^2}{6} \\
&\leq 5/s_1 \cdot \ell^{-1/2},
\end{aligned}$$

where the third step is because $e^{-\pi^2 \cdot i^2} \leq e^{-2i} \leq e^{-2 \ln(1+i)} = (1+i)^{-2}$ for each $i \in \mathbb{N}_{\geq 0}$; and the last step is because $\sqrt{8} \cdot \pi^2/6 \approx 4.6526 < 5$.

Further, we can infer the lower-bound part as follows:

$$\begin{aligned}
\int_0^{2/(\pi s_1)} (\text{sinc}_{s_1}(\xi))^\ell \cdot d\xi &= \frac{1}{s_1} \cdot \int_0^{2/\pi} (\text{sinc}(\xi))^\ell \cdot d\xi \\
&= \frac{1}{s_1} \cdot \int_0^{2/\pi} |\text{sinc}(\xi)|^\ell \cdot d\xi
\end{aligned}$$

$$\begin{aligned}
&\geq \frac{1}{s_1} \cdot \int_0^{i^* \cdot \sqrt{8/\ell}} |\text{sinc}(\xi)|^\ell \cdot d\xi \\
&\geq \frac{1}{s_1} \cdot \int_0^{\sqrt{8/\ell}} |\text{sinc}(\xi)|^\ell \cdot d\xi,
\end{aligned} \tag{64}$$

where the first step is by substitution; the second step by because $\ell \in \mathbb{N}_{\geq 1}$ is an even integer (see Definition A.3); the third step follows from the Fact (c) given in the beginning of this proof; and the last step is due to the above Fact (b) that $i^* \geq 6 > 1$.

Under the assumption $\ell \geq 1000$, we have $0 < \pi^2/6 \cdot \xi^2 \leq \pi^2/6 \cdot \frac{8}{\ell} \leq \pi^2/6 \cdot \frac{8}{1000} \approx 0.013 \leq 1$ for any $\xi \in [0, \sqrt{8/\ell}]$. Then for any $\xi \in [0, \sqrt{8/\ell}]$ we have

$$\begin{aligned}
|\text{sinc}(\xi)|^\ell &\geq (1 - \pi^2/6 \cdot \xi^2)^\ell \\
&\geq \left(1 - \pi^2/6 \cdot \frac{8}{\ell}\right)^\ell \\
&\geq \left(1 - \pi^2/6 \cdot \frac{8}{1000}\right)^{1000} \\
&= (1 - \pi^2/750)^{1000},
\end{aligned} \tag{65}$$

where the first step is by Part (a) of Fact A.2; and the third step is because $0 \leq \pi^2/6 \cdot \frac{8}{\ell} \leq 1$ and that $y = (1 - z)^{1/z}$ is a decreasing function for any $z \in (0, 1)$.

Plugging Equation (65) back into Equation (64) results in

$$\int_0^{2/(\pi s_1)} (\text{sinc}_{s_1}(\xi))^\ell \cdot d\xi \geq \frac{1}{s_1} \cdot \sqrt{8/\ell} \cdot (1 - \pi^2/750)^{1000} \geq 1/2^{18} \cdot \frac{1}{s_1} \cdot \ell^{-1/2},$$

where the last step follows from elementary calculation.

This completes the proof of Claim A.13. □

Claim A.14. $\int_{2/(\pi s_1)}^{+\infty} (\text{sinc}_{s_1}(\xi))^\ell \cdot d\xi = O\left(\frac{1}{s_1} \cdot 2^{-\ell}\right)$.

Proof. We check the claim as follows:

$$\begin{aligned}
\int_{2/(\pi s_1)}^{+\infty} (\text{sinc}_{s_1}(\xi))^\ell \cdot d\xi &= \int_{2/(\pi s_1)}^{+\infty} |\text{sinc}_{s_1}(\xi)|^\ell \cdot d\xi \\
&\leq \int_{2/(\pi s_1)}^{+\infty} \frac{1}{\pi^\ell \cdot |s_1 \xi|^\ell} \cdot d\xi \\
&= \frac{1}{\pi s_1} \cdot \int_2^{+\infty} \xi^{-\ell} \cdot d\xi \\
&= \frac{1}{\pi s_1} \cdot \frac{2}{\ell - 1} \cdot 2^{-\ell} \\
&= O\left(\frac{1}{s_1} \cdot 2^{-\ell}\right),
\end{aligned}$$

where the first step is because $\ell \in \mathbb{N}_{\geq 1}$ is an even integer (see Definition A.3); the second step is by Part (c) of Fact A.2; and the third step is by substitution.

This completes the proof of Claim A.14. □

B Filter, permutation and hashing in a single dimension

In this section, we first construct our single-dimensional filter function $(G(t), \widehat{G}(f))$ and investigate several properties of it, based on the building-block function $(G(t), \widehat{G}(f))$ introduced in Section A. In particular:

- Sections B.1 to B.3. We first leverage the function $(G(t), \widehat{G}(f))$ introduced in Definition A.3 to construct our ultimate single-dimensional filter $(G(t), \widehat{G}(f))$, and then prove the properties of this filter (by applying Lemma A.4 and extra arguments).
- Section B.4. We associate the filter $(G(t), \widehat{G}(f))$ with another standard window $(G'(t), \widehat{G}'(f))$ (in a manner similar to Lemma A.12), which is more convenient for our later use.

B.1 Construction of filter $(G(t), \widehat{G}(f))$

Definition B.1 (The single-dimensional filter). Recall the parameters defined in Definition A.3:

- The number of bins in a single dimension $B = \Theta(d \cdot k^{1/d})$ is a certain multiple of $d \in \mathbb{N}_{\geq 1}$.
- The noise level parameter $\delta \in (0, 1)$.
- $\alpha = \Theta(1/d)$ is chosen such that $\frac{1}{100 \cdot (d+1) \cdot \alpha} \in \mathbb{N}_{\geq 1}$ is an integer; clearly $\alpha \leq \frac{1}{100 \cdot (d+1)} \leq \frac{1}{200}$.
- $s_1 = \frac{2B}{\alpha}$ and $s_2 = \frac{1}{B+B/d}$.
- $\ell = \Theta(\log(kd/\delta))$ is an even integer. We safely assume $\ell \geq 1000$.

Further, the *width parameter* $W = \Theta(\frac{F}{B\eta})$ is chosen to be a sufficiently large integer. Based on the building-block function $(G(t), \widehat{G}(f))$ given in Definition A.3, for $i \in \mathbb{Z}$, define the shifted function

$$\begin{aligned}\widehat{G}_i(f) &:= \widehat{G}(f + i), \\ G_i(t) &:= \int_{-\infty}^{+\infty} \widehat{G}_i(\xi) \cdot e^{2\pi i t \cdot \xi} \cdot d\xi.\end{aligned}$$

Then for any $t, f \in \mathbb{R}$ the single-dimensional filter $(G(t), \widehat{G}(f))$ is given by

$$\begin{aligned}\widehat{G}(f) &= e^{-\frac{\delta}{\text{poly}(k, d)}} \cdot \sum_{i \in [-W:W]} \widehat{G}_i(f), \\ G(t) &= \int_{-\infty}^{+\infty} \widehat{G}(\xi) \cdot e^{2\pi i t \cdot \xi} \cdot d\xi.\end{aligned}$$

B.2 Properties of filter $(G(t), \widehat{G}(f))$

Later we will employ another slightly different filter $(G(t), \widehat{G}(f))$, just by shifting the one given in Definition B.1. For ease of presentation, the following Lemma B.2 is stated for the shifted filter, but we show in Section B.3 the counterpart claims for the unshifted filter.

Lemma B.2 (The single-dimensional filter). *The filter $(G(t), \widehat{G}(f))[B, \delta, \alpha, \ell, W]$ given in Definition B.1 satisfies the following (as Figure 20 illustrates):*

Property I: $e^{-\frac{\delta}{\text{poly}(k, d)}} \leq \widehat{G}(f) \leq 1$ when $|f - i| \leq \frac{1-\alpha}{2B}$ for some integer $|i| \leq W$.

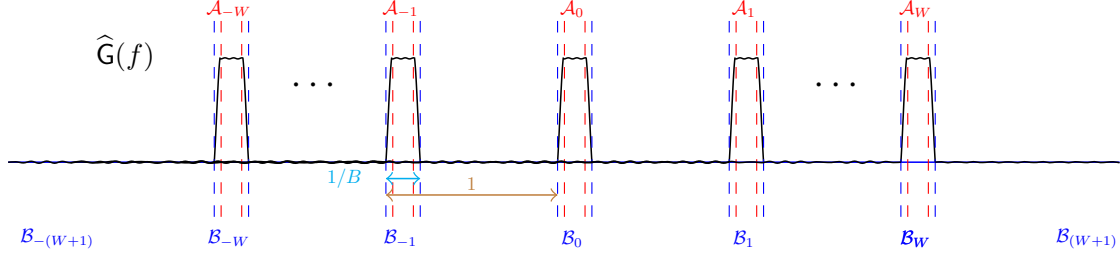


Figure 20: Demonstration for Lemma B.2. $\mathcal{B}_i = [i - 1/(2B), i + 1/(2B)]$ and $\mathcal{A}_i = [i - (1 - \alpha)/(2B), i + (1 - \alpha)/(2B)]$. Intuitively, $\widehat{G}(f)$ has $2W + 1$ “peaks” near the $2W + 1$ integers, where its value is very close 1. And its value drops very quickly and oscillates near 0 outside the “peaks”.

Property II: $0 \leq \widehat{G}(f) \leq 1$ when $\frac{1-\alpha}{2B} \leq |f - i| \leq \frac{1}{2B}$ for some integer $|i| \leq W$.

Property III: $0 \leq \widehat{G}(f) \leq \frac{\delta}{\text{poly}(k,d)}$ when $|f - i| \geq \frac{1}{2B}$ for any integer $|i| \leq W$.

Property IV: $G(t) = (2W + 1) \cdot e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot G(t) \cdot \frac{\text{sinc}_{(2W+1)}(t+1/2)}{\text{sinc}(t+1/2)}$ for any $t \in \mathbb{R}$.

Property V: $\text{supp}(G) \subseteq \text{supp}(G) \subseteq [-\ell \cdot \frac{B}{\alpha}, \ell \cdot \frac{B}{\alpha}]$.

Property VI: $\sum_{i \in \mathbb{Z}} G(i)^2 = e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot \sum_{i \in \mathbb{Z}} G(i)^2 \leq (1 + \frac{2}{d}) \cdot B^{-1}$.

Remark B.3. The function values $G(i)$ are the scaling coefficients for our samples in the time domain. In fact, we just need the values $G(i)$ at a few fixed points. Thus, we can calculate and store those effective coefficients before the sampling process.

B.3 Proof of properties

Recall that $e^{\delta/(4kd)} \cdot \widehat{G}(f) = \sum_{i \in [-W:W]} \widehat{G}_i(f)$ (see Definition B.1) and, that every summand function $\widehat{G}_i(f)$ has an axis of symmetry $f = -i$ (see Definition A.3). The next Claim B.4 suggests that, at any $f \in \mathbb{R}$, the value $e^{\delta/(4kd)} \cdot \widehat{G}(f)$ is dominated by one particular summand $\widehat{G}_{i^*}(f)$, and the other $2W$ summands are negligibly small. Given this observation, we can easily conclude Properties I to III of Lemma B.2 from Properties II to IV of Lemma A.4.

Claim B.4 (Auxiliary result for Lemma B.2). *Given any $f' \in \mathbb{R}$, let $i^* = \text{argmin}_{i \in [-W:W]} |f' + i|$ be the particular summand function $\widehat{G}_i(f)$ with the closest-to- f' axis of symmetry, then*

$$0 \leq e^{\frac{\delta}{\text{poly}(k,d)}} \cdot \widehat{G}(f') - \widehat{G}_{i^*}(f') = \sum_{i \in [-W:W] \setminus \{i^*\}} \widehat{G}_i(f') \leq \frac{\delta}{\text{poly}(k,d)}.$$

Proof. The first part $\sum_{i \in [-W:W] \setminus \{i^*\}} \widehat{G}_i(f') \geq 0$ follows because each summand function $\widehat{G}_i(f')$ is non-negative (see Properties II to IV of Lemma A.4).

We now show the second part $\sum_{i \in [-W:W] \setminus \{i^*\}} \widehat{G}_i(f) \leq \delta/(4kd)$. Clearly, the summand functions $\widehat{G}_i(f)$ have axes of symmetry $f = -i \in [-W : W]$ (see Definition B.1). Since $f = -i^*$ is the axis closet to f' , the distances between f' and either the left-hand-side axes (i.e. $f = -i \leq -i^* - 1$) or the right-hand-side axes (i.e. $f = -i \geq -i^* + 1$) are at least $(1 - \frac{1}{2})$, $(2 - \frac{1}{2})$, $(3 - \frac{1}{2})$, \dots . Further, when the distance between f' and an axis $f = -i$ is at least $j - \frac{1}{2} \geq \frac{j}{2} \geq \frac{1}{2B}$ (for some $j \in \mathbb{N}_{\geq 1}$;

recall Definition B.1 that $B = \Theta(d \cdot k^{1/d})$ is the number of bins in a single dimension), it follows from Property IV of Lemma A.4 that

$$\begin{aligned}\widehat{G}_i(f') &\leq (\pi B \cdot (j - 1/2))^{-\ell} \\ &\leq (\pi B \cdot j/2)^{-\ell}.\end{aligned}\tag{66}$$

Take all of the $2W$ remaining summands $i \in [-W : W] \setminus \{i^*\}$ into account:

$$\begin{aligned}\sum_{i \in [-W : W] \setminus \{i^*\}} \widehat{G}_i(f') &\leq \sum_{i \in \mathbb{Z} \setminus \{i^*\}} \widehat{G}_i(f') \\ &\leq 2 \cdot \sum_{j \in \mathbb{N}_{\geq 1}} (\pi B \cdot j/2)^{-\ell} \\ &\leq \frac{\delta}{\text{poly}(k, d)},\end{aligned}$$

where the first step follows because each summand function $\widehat{G}_i(f')$ is non-negative (see Properties II to IV of Lemma A.4); the second step follows from Inequality 66; and the last step holds whenever $B = \Theta(d \cdot k^{1/d})$ and $\ell = \Theta(\log(kd/\delta))$ are large enough.

This completes the proof of Claim B.4. \square

Claim B.5 (Property I of Lemma B.2). $e^{-\frac{\delta}{\text{poly}(k, d)}} \leq \widehat{\mathbf{G}}(f) \leq 1$ when $|f - i| \leq \frac{1-\alpha}{2B}$ for some integer $|i| \leq W$.

Proof. We let $i^* = \text{argmin}_{i \in [-W : W]} |f + i|$ index the summand function with the closest-to- f axis of symmetry. For the lower-bound part, we observe that

$$\begin{aligned}\widehat{\mathbf{G}}(f) &= e^{-\frac{\delta}{\text{poly}(k, d)}} \cdot \sum_{i \in [-W : W]} \widehat{G}_i(f) \\ &\geq e^{-\frac{\delta}{\text{poly}(k, d)}} \cdot \widehat{G}_{i^*}(f) \\ &\geq e^{-\frac{\delta}{\text{poly}(k, d)}} \cdot \left(1 - \frac{\delta}{\text{poly}(k, d)}\right) \\ &\geq e^{-\frac{\delta}{\text{poly}(k, d)}},\end{aligned}$$

where the first step follows from Claim B.4; the third step applies Property II of Lemma A.4; and the last step is because $1 - z/4 \geq e^{-3z/4}$ for any $z \in [0, 1]$ (recall Definition B.1 that the noise level parameter $0 < \delta < 1$).

In addition, for the upper-bound part we have

$$\begin{aligned}\widehat{\mathbf{G}}(f) &= e^{-\frac{\delta}{\text{poly}(k, d)}} \cdot \widehat{G}_{i^*}(f) + e^{-\frac{\delta}{\text{poly}(k, d)}} \cdot \sum_{i \in [-W : W] \setminus \{i^*\}} \widehat{G}_i(f) \\ &\leq e^{-\frac{\delta}{\text{poly}(k, d)}} \cdot 1 + e^{-\frac{\delta}{\text{poly}(k, d)}} \cdot \frac{\delta}{\text{poly}(k, d)} \\ &\leq 1,\end{aligned}$$

where the second step applies Property II of Lemma A.4 (to the first term) and Claim B.4 (to the second term); and the last step is because $1 + z \leq e^z$ for any $z \geq 0$.

This completes the proof of Claim B.5. \square

Claim B.6 (Property II of Lemma B.2). $0 \leq \widehat{G}(f) \leq 1$ when $\frac{1-\alpha}{2B} \leq |f-i| \leq \frac{1}{2B}$ for some integer $|i| \leq W$.

Proof. The first part $\widehat{G}(f) \geq 0$ follows because each summand function $\widehat{G}_i(f)$ is non-negative (see Properties II to IV of Lemma A.4). For the upper-bound part, by definition we have

$$\begin{aligned}\widehat{G}(f) &= e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot \widehat{G}_{i^*}(f) + e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot \sum_{i \in [-W:W] \setminus \{i^*\}} \widehat{G}_i(f) \\ &\leq e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot 1 + e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot \frac{\delta}{\text{poly}(k,d)} \\ &\leq 1,\end{aligned}$$

where the second step applies Property III of Lemma A.4 (to the first term) and Claim B.4 (to the second term); and the last step is because $1+z \leq e^z$ for any $z \geq 0$.

This completes the proof of Claim B.6. \square

Claim B.7 (Property III of Lemma B.2). $0 \leq \widehat{G}(f) \leq \frac{\delta}{\text{poly}(k,d)}$ when $|f-i| \geq \frac{1}{2B}$ for any integer $|i| \leq W$.

Proof. The first part $\widehat{G}(f) \geq 0$ has been justified in the proof of Claim B.6. For the upper-bound part, by definition we have

$$\begin{aligned}\widehat{G}(f) &= e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot \widehat{G}_{i^*}(f) + e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot \sum_{i \in [-W:W] \setminus \{i^*\}} \widehat{G}_i(f) \\ &\leq e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot \frac{\delta}{\text{poly}(k,d)} + e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot \frac{\delta}{\text{poly}(k,d)} \\ &\leq \frac{\delta}{\text{poly}(k,d)},\end{aligned}$$

where the second step applies Property IV of Lemma A.4 (to the first term) and Claim B.4 (to the second term); and the last step follows from elementary calculation.

This completes the proof of Claim B.7. \square

Claim B.8 (Property IV of Lemma B.2). $G(t) = (2W+1) \cdot e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot G(t) \cdot \frac{\text{sinc}(2W+1)(t)}{\text{sinc}(t)}$ for any $t \in \mathbb{R}$.

Proof. For convenient, in this proof we ignore the $\frac{0}{0}$ issue; this can be easily remedied by applying L'Hospital's rule. According to the definition of the inverse CFT,

$$\begin{aligned}G(t) &= \int_{-\infty}^{+\infty} \widehat{G}(\xi) \cdot e^{2\pi i t \cdot \xi} \cdot d\xi \\ &= e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot \sum_{i \in [-W:W]} \int_{-\infty}^{+\infty} \widehat{G}(\xi+i) \cdot e^{2\pi i t \cdot \xi} \cdot d\xi \\ &= e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot \sum_{i \in [-W:W]} e^{-2\pi i t \cdot i} \cdot \int_{-\infty}^{+\infty} \widehat{G}(\xi+i) \cdot e^{2\pi i t \cdot (\xi+i)} \cdot d\xi \\ &= e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot \sum_{i \in [-W:W]} e^{-2\pi i t \cdot i} \cdot \int_{-\infty}^{+\infty} \widehat{G}(\xi) \cdot e^{2\pi i t \cdot \xi} \cdot d\xi\end{aligned}$$

$$= e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot G(t) \cdot \sum_{i \in [-W:W]} e^{-2\pi i t \cdot i},$$

where the second step is by Definition B.1; the fourth step is by substitution; and the last step is by the definition of the CFT.

It remains to calculate the sum of the geometric sequence $\sum_{i \in [-W:W]} e^{-2\pi i \cdot (t+1/2) \cdot i}$. Concretely, for any $\tau \in \mathbb{R}$ we have

$$\begin{aligned} \sum_{i \in [-W:W]} e^{-2\pi i t \cdot i} &= e^{2\pi i t \cdot W} \cdot \left(\frac{1 - e^{-2\pi i t \cdot (2W+1)}}{1 - e^{-2\pi i t}} \right) \\ &= e^{2\pi i t \cdot W} \cdot \left(\frac{e^{-\pi i t \cdot (2W+1)}}{e^{-\pi i t}} \cdot \frac{e^{\pi i t \cdot (2W+1)} - e^{-\pi i t \cdot (2W+1)}}{e^{\pi i t} - e^{-\pi i t}} \right) \\ &= \frac{e^{\pi i t \cdot (2W+1)} - e^{-\pi i t \cdot (2W+1)}}{e^{\pi i t} - e^{-\pi i t}} \\ &= \frac{2i \cdot \sin(\pi t \cdot (2W+1))}{2i \cdot \sin(\pi t)} \\ &= \frac{\text{sinc}_{(2W+1)}(t)}{\text{sinc}(t)} \cdot (2W+1), \end{aligned}$$

where the fourth step is by Euler's formula (note that $\cos(z)$ is an even function while $\sin(z)$ is an odd function); and the last step follows from Definition A.1.

Combining everything together completes the proof of Claim B.8. \square

Claim B.9 (Property V of Lemma B.2). $\text{supp}(\mathbf{G}) \subseteq \text{supp}(G) \subseteq [-\ell \cdot \frac{B}{\alpha}, \ell \cdot \frac{B}{\alpha}]$.

Proof. By Claim B.8, the filter $\mathbf{G}(t) = (2W+1) \cdot e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot G(t) \cdot \frac{\text{sinc}_{(2W+1)}(t)}{\text{sinc}(t)}$ has the same support as the function $G(t)$. Thus we immediately infer this claim from Property V of Lemma A.4. \square

Claim B.10 (Property VI of Lemma B.2). *It follows that*

$$\sum_{i \in \mathbb{Z}} \mathbf{G}(i + 1/2)^2 = e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot \sum_{i \in \mathbb{Z}} G(i + 1/2)^2 \leq (1 + 2/d) \cdot B^{-1}.$$

Proof. The second part of the claim is a direct follow-up to Property VII of Lemma A.4. That is,

$$\begin{aligned} e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot \sum_{i \in \mathbb{Z}} G(i + 1/2)^2 &\leq e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot \left(1 + \frac{\delta}{4kd}\right)^2 \cdot \left(1 + \frac{1}{d}\right) \cdot B^{-1} \\ &\leq (1 + 2/d) \cdot B^{-1}, \end{aligned}$$

where the first step follows from Property VII of Lemma A.4; and the last step applies the fact that $e^{-z} \cdot (1 + z) \leq 1$ for any $z \in \mathbb{R}_{\geq 0}$.

To see the first part, it suffices to show that

$$|\mathbf{G}(i + 1/2)| = e^{-\frac{\delta}{\text{poly}(k,d)}} \cdot |G(i + 1/2)|, \forall i \in \mathbb{Z}.$$

Based on Claim B.8, this equation is equivalent to

$$\left| (2W+1) \cdot \frac{\text{sinc}_{(2W+1)}(i + 1/2)}{\text{sinc}(i + 1/2)} \right| = 1.$$

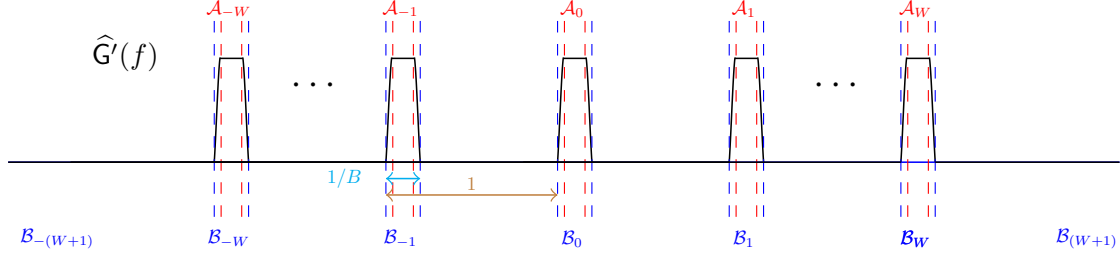


Figure 21: Demonstration for Lemma B.11. $\mathcal{B}_i = [i - 1/(2B), i + 1/(2B)]$ and $\mathcal{A}_i = [i - (1 - \alpha)/(2B), i + (1 - \alpha)/(2B)]$. Intuitively, $\hat{G}'(f)$ has $2W + 1$ “peaks”, where its value is exactly 1. And its value drops to 0 very quickly outside the “peaks”.

According to Definition A.1, we have

$$\begin{aligned}
 \left| (2W + 1) \cdot \frac{\text{sinc}_{(2W+1)}(i + 1/2)}{\text{sinc}(i + 1/2)} \right| &= \left| \frac{\sin(\pi \cdot (2W + 1) \cdot (i + 1/2))}{\sin(\pi \cdot (i + 1/2))} \right| \\
 &= \left| \frac{\sin(\pi \cdot (2i \cdot W + W + i) + \pi/2)}{\sin(\pi \cdot i + \pi/2)} \right| \\
 &= \left| \frac{(-1)^{2i \cdot W + i + W}}{(-1)^i} \right| \\
 &= 1,
 \end{aligned}$$

where the third step follows because both $(2i \cdot W + W + i)$ and i are integers; thus we can apply certain properties of the $\sin(z)$ function.

This completes the proof of Claim B.10. \square

B.4 Construction and properties of standard window $(G'(t), \hat{G}'(f))$

Now we associate the filter $(G(t), \hat{G}(f))$ introduced in Definition B.1 with another standard window $(G'(t), \hat{G}'(f))$ (in a manner similar to Lemma A.12), which is more convenient for our later use.

Lemma B.11 (The single-dimensional standard window). *For the filter $(G(t), \hat{G}(f))$ given in Definition A.3, there exists another function $(G'(t), \hat{G}'(f))$ such that(as Figure 21 illustrates):*

Property I: $\hat{G}'(f) = 1$ when $|f - i| \leq \frac{1-\alpha}{2B}$ for some integer $|i| \leq W$.

Property II: $\hat{G}'(f) \in [0, 1]$ when $\frac{1-\alpha}{2B} \leq |f - i| \leq \frac{1}{2B}$ for some integer $|i| \leq W$.

Property III: $\hat{G}'(f) = 0$ when $|f - i| \geq \frac{1}{2B}$ for any integer $|i| \leq W$.

Property IV: $\|\hat{G}' - \hat{G}\|_\infty = \max_{f \in \mathbb{R}} |\hat{G}'(f) - \hat{G}(f)| \leq \frac{\delta}{\text{poly}(k, d)}.$

Proof. Recall Definition B.1 for the parameters B, δ, α, ℓ and W . We define the single-dimensional standard window $\hat{G}'(f)$ as follows:

- $\hat{G}'(f) = 1$ when $|f - i| \leq \frac{1-\alpha}{2B}$ for some integer $|i| \leq W$;
- $\hat{G}'(f) = \hat{G}(f)$ when $\frac{1-\alpha}{2B} \leq |f - i| \leq \frac{1}{2B}$ for some integer $|i| \leq W$; and

- $\widehat{G}(f) = 0$ when $|f - i| \geq \frac{1}{2B}$ for any integer $|i| \leq W$.

By construction, Properties I and III follows directly. Further, Property II follows from Property II of Lemma B.2, and Property IV can be inferred from Properties I to III of Lemma B.2.

This completes the proof of Lemma B.11. □

References

- [AZGR16] Zeyuan Allen-Zhu, Rati Gelashvili, and Ilya Razenshteyn. Restricted isometry property for general p -norms. *IEEE Transactions on Information Theory*, 62(10):5839–5854, 2016.
- [BCG⁺12] Petros Boufounos, Volkan Cevher, Anna C Gilbert, Yi Li, and Martin J Strauss. What’s the frequency, Kenneth?: Sublinear Fourier sampling off the grid. In *Algorithmica (A preliminary version of this paper appeared in the Proceedings of RANDOM/APPROX 2012, LNCS 7408, pp.61–72)*, pages 1–28. Springer, 2012.
- [BD08] Thomas Blumensath and Mike E Davies. Iterative thresholding for sparse approximations. *Journal of Fourier analysis and Applications*, 14(5-6):629–654, 2008.
- [Ber24] Sergei Bernstein. On a modification of chebyshev’s inequality and of the error formula of laplace. *Ann. Sci. Inst. Sav. Ukraine, Sect. Math*, 1(4):38–49, 1924.
- [BIP⁺16] Arturs Backurs, Piotr Indyk, Eric Price, Ilya Razenshteyn, and David P Woodruff. Nearly-optimal bounds for sparse recovery in generic norms, with applications to k -median sketching. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 318–337. SIAM, <https://arxiv.org/pdf/1504.01076.pdf>, 2016.
- [BJP73] RP Boas Jr and H Pollard. Continuous analogues of series. *The American Mathematical Monthly*, 80(1):18–25, 1973.
- [BKM⁺21] Karl Bringmann, Michael Kapralov, Mikhail Makarov, Vasileios Nakos, Amir Yagudin, and Amir Zandieh. Sparse fourier transform by traversing cooley-tukey fft computation graphs. *arXiv preprint arXiv:2107.07347*, 2021.
- [BLL⁺19] Jaroslaw Blasiok, Patrick Lopatto, Kyle Luh, Jake Marcinek, and Shravas Rao. An improved lower bound for sparse reconstruction from subsampled hadamard matrices. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1564–1567. IEEE, <https://arxiv.org/pdf/1903.12135.pdf>, 2019.
- [BLSS20] Jan van den Brand, Yin Tat Lee, Aaron Sidford, and Zhao Song. Solving tall dense linear programs in nearly linear time. In *52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, 2020.
- [Bou14] Jean Bourgain. An improved estimate in the restricted isometry problem. In *Geometric Aspects of Functional Analysis*, pages 65–70. Springer, 2014.
- [CC13] Yuxin Chen and Yuejie Chi. Spectral compressed sensing via structured matrix completion. In *International Conference on Machine Learning*, pages 414–422. PMLR, 2013.
- [CF14] Emmanuel J Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.
- [CFG13] Emmanuel J Candès and Carlos Fernandez-Granda. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, 2013.

- [CFG14] Emmanuel J Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on pure and applied Mathematics*, 67(6):906–956, 2014.
- [CFJ13] Tony Cai, Jianqing Fan, and Tiefeng Jiang. Distributions of angles in random packing on spheres. *The Journal of Machine Learning Research*, 14(1):1837–1864, 2013.
- [CGV13] Mahdi Cheraghchi, Venkatesan Guruswami, and Ameya Velingker. Restricted isometry of Fourier matrices and list decodability of random linear codes. *SIAM Journal on Computing*, 42(5):1888–1914, 2013.
- [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 *57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 741–750. IEEE, <https://arxiv.org/pdf/1609.01361.pdf>, 2016.
- [KKSZ17] Volkan Cevher, Michael Kapralov, Jonathan Scarlett, and Amir Zandieh. An adaptive sublinear-time block sparse Fourier transform. In *Proceedings of the 49th Annual Symposium on the Theory of Computing (STOC)*. ACM, <https://arxiv.org/pdf/1702.01286>, 2017.
- [CLRS09] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2009.
- [CLS20] Sitan Chen, Jerry Li, and Zhao Song. Learning mixtures of linear regressions in subexponential time via Fourier moments. In *STOC*. <https://arxiv.org/pdf/1912.07629.pdf>, 2020.
- [CM21] Sitan Chen and Ankur Moitra. Algorithmic foundations for the diffraction limit. In *STOC*. <https://arxiv.org/pdf/2004.07659.pdf>, 2021.
- [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 *ICALP*, 2019.
- [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.
- [DBIPW10] Khanh Do Ba, Piotr Indyk, Eric Price, and David P Woodruff. Lower bounds for sparse recovery. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 1190–1197. SIAM, 2010.
- [Don06] David L. Donoho. Compressed sensing. *IEEE Trans. Information Theory*, 52(4):1289–1306, 2006.

- [Für09] Martin Fürer. Faster integer multiplication. *SIAM Journal on Computing*, 39(3):979–1005, 2009.
- [GLPS10] Anna C Gilbert, Yi Li, Ely Porat, and Martin J Strauss. Approximate sparse recovery: optimizing time and measurements. *SIAM Journal on Computing 2012 (A preliminary version of this paper appears in STOC 2010)*, 41(2):436–453, 2010.
- [GMS05] Anna C Gilbert, S Muthukrishnan, and Martin Strauss. Improved time bounds for near-optimal sparse Fourier representations. In *Optics & Photonics 2005*, pages 59141A–59141A. International Society for Optics and Photonics, 2005.
- [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. ACM, <https://arxiv.org/pdf/1201.2501.pdf>, 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, https://groups.csail.mit.edu/netmit/sFFT/soda_paper.pdf, 2012.
- [HK15] Qingqing Huang and Sham M Kakade. Super-resolution off the grid. In *Advances in Neural Information Processing Systems*, pages 2647–2655, 2015.
- [Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.
- [HR16] Ishay Haviv and Oded Regev. The restricted isometry property of subsampled Fourier matrices. In *SODA*, pages 288–297. <https://arxiv.org/pdf/1507.01768.pdf>, 2016.
- [HS90] Yingbo Hua and Tapan K Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 38(5):814–824, 1990.
- [IK14] Piotr Indyk and Michael Kapralov. Sample-optimal Fourier sampling in any constant dimension. In *IEEE 55th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 514–523. IEEE, <https://arxiv.org/pdf/1403.5804.pdf>, 2014.
- [IKP14] Piotr Indyk, Michael Kapralov, and Eric Price. (Nearly) Sample-optimal sparse Fourier transform. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 480–499. SIAM, 2014.
- [IP11] Piotr Indyk and Eric Price. K-median clustering, model-based compressive sensing, and sparse recovery for earth mover distance. In *Proceedings of the forty-third annual ACM symposium on Theory of computing (STOC)*, pages 627–636. ACM, <https://arxiv.org/pdf/1104.4674.pdf>, 2011.
- [IPW11] Piotr Indyk, Eric Price, and David P Woodruff. On the power of adaptivity in sparse recovery. In *52nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 285–294. IEEE, <https://arxiv.org/pdf/1110.3850.pdf>, 2011.

- [IR13] Piotr Indyk and Ilya Razenshteyn. On model-based rip-1 matrices. In *International Colloquium on Automata, Languages, and Programming*, pages 564–575. Springer, 2013.
- [JSWZ21] Shunhua Jiang, Zhao Song, Omri Weinstein, and Hengjie Zhang. Faster dynamic matrix inverse for faster lps. In *STOC*. <https://arxiv.org/pdf/2004.07470.pdf>, 2021.
- [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.
- [KP19] Akshay Kamath and Eric Price. Adaptive sparse recovery with limited adaptivity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2729–2744, 2019.
- [KPRvdO16] Stefan Kunis, Thomas Peter, Tim Römer, and Ulrich von der Ohe. A multivariate generalization of prony’s method. *Linear Algebra and its Applications*, 490:31–47, 2016.
- [KSBO00] Marc Van Kreveld, Otfried Schwarzkopf, Mark de Berg, and Mark Overmars. *Computational geometry algorithms and applications*. Springer, 2000.
- [KVZ19] Michael Kapralov, Ameya Velingker, and Amir Zandieh. Dimension-independent sparse Fourier transform. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2709–2728. SIAM, <https://arxiv.org/pdf/1902.10633.pdf>, 2019.
- [LDFU13] Yichao Lu, Paramveer Dhillon, Dean P Foster, and Lyle Ungar. Faster ridge regression via the subsampled randomized hadamard transform. In *Advances in neural information processing systems*, pages 369–377, 2013.
- [LN20] Yi Li and Vasileios Nakos. Deterministic sparse Fourier transform with an ℓ_∞ guarantee. In *ICALP*. <https://arxiv.org/pdf/1903.00995.pdf>, 2020.
- [LNNT16] Kasper Green Larsen, Jelani Nelson, Huy L Nguyễn, and Mikkel Thorup. Heavy hitters via cluster-preserving clustering. In *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*, pages 61–70. IEEE, <https://arxiv.org/pdf/1604.01357>, 2016.
- [LSZ19] Yin Tat Lee, Zhao Song, and Qiuyu Zhang. Solving empirical risk minimization in the current matrix multiplication time. In *COLT*. <https://arxiv.org/pdf/1905.04447.pdf>, 2019.
- [Moi15] Ankur Moitra. The threshold for super-resolution via extremal functions. In *STOC*. <https://arxiv.org/pdf/1408.1681.pdf>, 2015.

- [MZIC19] Sami Merhi, Ruochuan Zhang, Mark A Iwen, and Andrew Christlieb. A new class of fully discrete sparse fourier transforms: Faster stable implementations with guarantees. *Journal of Fourier Analysis and Applications*, 25(3):751–784, 2019.
- [NS19] Vasileios Nakos and Zhao Song. Stronger L2/L2 compressed sensing; without iterating. In *Proceedings of the 51st Annual ACM Symposium on Theory of Computing (STOC)*. <https://arxiv.org/pdf/1903.02742>, 2019.
- [NSW19] Vasileios Nakos, Zhao Song, and Zhengyu Wang. (Nearly) Sample-optimal sparse Fourier transform in any dimension; RIPless and Filterless. In *FOCS*. <https://arxiv.org/pdf/1909.11123.pdf>, 2019.
- [Pis73] Vladilen F Pisarenko. The retrieval of harmonics from a covariance function. *Geophysical Journal International*, 33(3):347–366, 1973.
- [Pri11] Eric Price. Efficient sketches for the set query problem. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 41–56. Society for Industrial and Applied Mathematics, 2011.
- [Pri13] Eric C Price. *Sparse recovery and Fourier sampling*. PhD thesis, Massachusetts Institute of Technology, 2013.
- [PS15] Eric Price and Zhao Song. A robust sparse Fourier transform in the continuous setting. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 583–600. IEEE, <https://arxiv.org/pdf/1609.00896.pdf>, 2015.
- [PW13] Eric Price and David P. Woodruff. Lower bounds for adaptive sparse recovery. In *SODA*. <https://arxiv.org/pdf/1205.3518.pdf>, 2013.
- [RV08] Mark Rudelson and Roman Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008.
- [Son19] Zhao Song. *Matrix Theory: Optimization, Concentration and Algorithms*. PhD thesis, The University of Texas at Austin, 2019.
- [SSBD14] Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.
- [SSWZ22] Zhao Song, Baocheng Sun, Omri Weinstein, and Ruizhe Zhang. Sparse fourier transform over lattices: A unified approach to signal reconstruction. <http://arxiv.org/abs/2205.00658>, 2022.
- [Sto93] Petre Stoica. List of references on spectral line analysis. *Signal Processing*, 31(3):329–340, 1993.
- [Tro15] Joel A Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.
- [Ver10] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.
- [Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.